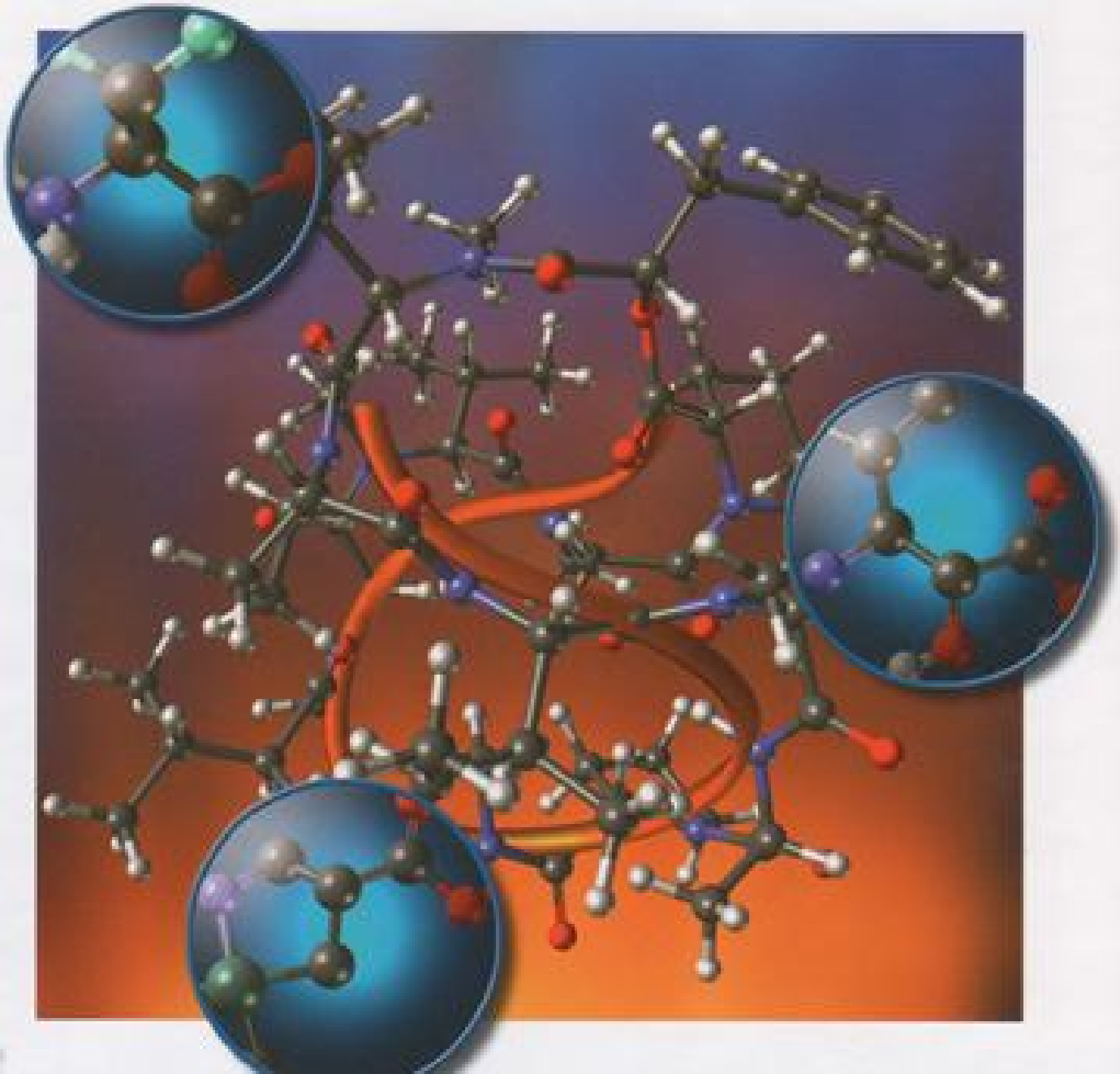
Edited by Andrew B. Hughes

WILEY-VCH

Amino Acids, Peptides and Proteins in Organic Chemistry

Volume 2
Modified Amino Acids, Organocatalysis and Enzymes



Amino Acids, Peptides and Proteins in Organic Chemistry

Edited by
Andrew B. Hughes

Further Reading

Fessner, W.-D., Anthonsen, T.

Modern Biocatalysis

Stereoselective and Environmentally Friendly Reactions

2009

ISBN: 978-3-527-32071-4

Sewald, N., Jakubke, H.-D.

Peptides: Chemistry and Biology

2009

ISBN: 978-3-527-31867-4

Lutz, S., Bornscheuer, U. T. (eds.)

Protein Engineering Handbook

2 Volume Set

2009

ISBN: 978-3-527-31850-6

Aehle, W. (ed.)

Enzymes in Industry

Production and Applications

2007

ISBN: 978-3-527-31689-2

Wiley-VCH (ed.)

Ullmann's Biotechnology and Biochemical Engineering

2 Volume Set

2007

ISBN: 978-3-527-31603-8

Budisa, N.

Engineering the Genetic Code

Expanding the Amino Acid Repertoire for the Design of Novel Proteins

2006

ISBN: 978-3-527-31243-6

Demchenko, A. V. (ed.)

Handbook of Chemical Glycosylation

Advances in Stereoselectivity and Therapeutic Relevance

2008

ISBN: 978-3-527-31780-6

Lindhorst, T. K.

Essentials of Carbohydrate Chemistry and Biochemistry

2007

ISBN: 978-3-527-31528-4

Amino Acids, Peptides and Proteins in Organic Chemistry

Volume 2 - Modified Amino Acids, Organocatalysis and Enzymes

Edited by
Andrew B. Hughes

WILEY-VCH Verlag GmbH & Co. KGaA

The Editor

Andrew B. Hughes
La Trobe University
Department of Chemistry
Victoria 3086
Australia

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Composition Thomson Digital, Noida, India
Printing Betz-Druck GmbH, Darmstadt
Bookbinding Litges & Dopf GmbH, Heppenheim
Cover Design Schulz Grafik Design, Fußgönheim

Printed in the Federal Republic of Germany
Printed on acid-free paper

ISBN: 978-3-527-32098-1

Contents

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

List of Contributors

Neil Audsley
The Food and Environment
Research Agency
Sand Hutton
York YO41 1LZ
UK

Sonia Barberis
Universidad Nacional de San Luis,
Chacabuco y Pedernera
Faculty of Chemistry, Biochemistry and
Pharmacy
San Luis
Argentina

Annette G. Beck-Sickinger
Leipzig University
Institute of Biochemistry
Brüderstraße 34
04103 Leipzig
Germany

Karen Brand
Leipzig University
Institute of Biochemistry
Brüderstraße 34
04103 Leipzig
Germany

Keith Brocklehurst
Queen Mary, University of London
School of Biological and Chemical
Sciences
Fogg Building, Mile End Road
London E1 4NS
UK

Arwen J. Cross
University of Sydney
School of Molecular and Microbial
Biosciences
G08 Biochemistry Building
NSW 2006
Sydney
Australia

Valery M. Dembitsky
The Hebrew University of Jerusalem
School of Pharmacy
Department of Medicinal Chemistry
and Natural Products
PO Box 12065
Jerusalem 91120
Israel

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

Marc Devocelle
Royal College of Surgeons in Ireland
Centre for Synthesis & Chemical Biology
Department of Pharmaceutical & Medicinal Chemistry
123 St. Stephens Green
Dublin 2
Ireland

Sean Doyle
National University of Ireland Maynooth
Department of Biology
Maynooth, Co. Kildare
Ireland

Nicholas Gathergood
Dublin City University
School of Chemical Sciences and National Institute for Cellular Biotechnology
Glasnevin, Dublin 9
Ireland

Giovanna Ghirlanda
Arizona State University
Department of Chemistry and Biochemistry
Tempe, AZ 85287-1604
USA

Darren Griffith
Royal College of Surgeons in Ireland
Centre for Synthesis & Chemical Biology
Department of Pharmaceutical & Medicinal Chemistry
123 St. Stephens Green
Dublin 2
Ireland

Sheraz Gul
European ScreeningPort GmbH
Schnackenburgallee 114
22525 Hamburg
Germany

Fanny Guzmán
Pontificia Universidad Católica de Valparaíso
Institute of Biology
Avenida Brasil 2950
Valparaíso
Chile

R. Elwyn Isaac
University of Leeds
Institute of Integrative and Comparative Biology
Faculty of Biological Sciences
Leeds LS2 9JT
UK

Thomas Hayes
Dublin City University
School of Chemical Sciences and National Institute for Cellular Biotechnology
Glasnevin
Dublin 9
Ireland

Usama M. Hegazy
Uppsala University
Biomedical Center
Department of Biochemistry and Organic Chemistry
Box 576
751 23 Uppsala
Sweden

Andrés Illanes
Pontificia Universidad Católica de Valparaíso
School of Biochemical Engineering
Avenida Brasil 2147
Valparaíso
Chile

Uli Kazmaier
Universität des Saarlandes
Institut für Organische Chemie
Im Stadtwald
66123 Saarbrücken
Germany

Valery P. Kukhar
National Academy of Sciences of Ukraine
Institute of Bioorganic Chemistry and Petrochemistry
Murmanskaya Street
Kiev 94
Ukraine

Joel P. Mackay
University of Sydney
School of Molecular and Microbial Biosciences
G08 Biochemistry Building
NSW 2006
Sydney
Australia

Robyn E. Mansfield
University of Sydney
School of Molecular and Microbial Biosciences
G08 Biochemistry Building
NSW 2006
Sydney
Australia

Bengt Mannervik
Uppsala University
Biomedical Center
Department of Biochemistry and Organic Chemistry
Box 576
751 23 Uppsala
Sweden

Celine J. Marmion
Royal College of Surgeons in Ireland
Centre for Synthesis and Chemical Biology
Department of Pharmaceutical and Medicinal Chemistry
123 St. Stephens Green
Dublin 2
Ireland

Jacqueline M. Matthews
University of Sydney
School of Molecular and Microbial Biosciences
G08 Biochemistry Building
NSW 2006
Sydney
Australia

Javier Narváez-Vásquez
University of California Riverside
Department of Botany and Plant Sciences
3401 Watkins Dr.
Riverside, CA 92521
USA

Martha L. Orozco-Cárdenas
University of California Riverside
Department of Botany and Plant Sciences
3401 Watkins Dr.
Riverside, CA 92521
USA

and

University of California Riverside
Plant Transformation Research Center
Riverside, CA 92521
USA

Gregory Pearce
Washington State University
Institute of Biological Chemistry
Pullman, WA 99164
USA

Richard W. Pickersgill
Queen Mary, University of London
School of Biological and Chemical Sciences
Joseph Priestley Building
Mile End Road
London E1 4NS
UK

Leonard J. Prins
University of Padova
Padova Section
Department of Chemical Sciences and ITM-CNR
Via Marzolo 1
35131 Padova
Italy

Yingmei Qi
Temple University
Department of Chemistry
1901 N. 13th Street
Philadelphia, PA 19122
USA

Vadim D. Romanenko
National Academy of Sciences of Ukraine
Institute of Bioorganic Chemistry and Petrochemistry
Murmanskaya Street
Kiev 94
Ukraine

Paolo Scrimin
University of Padova
Padova Section
Department of Chemical Sciences and ITM-CNR
Via Marzolo 1
35131 Padova
Italy

Scott McN. Sieburth
Temple University
Department of Chemistry
1901 N. 13th Street
Philadelphia, PA 19122
USA

Morris Srebnik
The Hebrew University of Jerusalem
School of Pharmacy
Department of Medicinal Chemistry and Natural Products
PO Box 12065
Jerusalem 91120
Israel

Joëlle Vidal
Université de Rennes 1
CNRS UMR 6510, Chimie et Photonique Moléculaires
Campus de Beaulieu, case 1012
35042 Rennes Cedex
France

Haibo Xie
Dublin City University
School of Chemical Sciences and National Institute for Cellular Biotechnology
Glasnevin, Dublin 9
Ireland

Part One
Synthesis and Chemistry of Modified Amino Acids

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

1
Synthesis and Chemistry of α,β-Didehydroamino Acids

Uli Kazmaier

1.1
Introduction

Although α,β-didehydroamino acids (DDAAs), where the term "didehydro-" is used to indicate the lack two hydrogen atoms, do not belong to the group of proteinogenic amino acids, they are commonly found in nature as building blocks of didehydropeptides (DDPs), mainly as secondary metabolites of bacteria and fungi or other lower organisms. Most of these compounds show interesting biological activities, such as the β-lactam antibiotics of the cephalosporin group [1], the herbicidal tetrapeptide tentoxin [2], or the antitumor agent azinomycin A (carzinophilin) [3] (Figure 1.1).

From a chemical point of view, DDAAs are interesting candidates for the synthesis of complex amino acids (e.g., via additions to the double bond). Therefore, it is not surprising that the research on this important class of amino acids has been reviewed frequently (e.g., by Schmidt [4], Chamberlin [5], and König *et al.* [6]). This chapter gives an overview of the different protocols for the synthesis of DDAAs and their typical reaction behavior.

1.2
Synthesis of DDAAs

1.2.1
DDAAs via Eliminations

1.2.1.1 DDAAs via β-Elimination

1.2.1.1.1 From β-Hydroxy Amino Acids The elimination of water from the corresponding β-hydroxy amino acids is a straightforward approach towards DDAAs, especially if the required hydroxy acids are readily available such as serine and threonine. On elimination didehydroalanine (ΔAla) and didehydroaminobutenoate

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

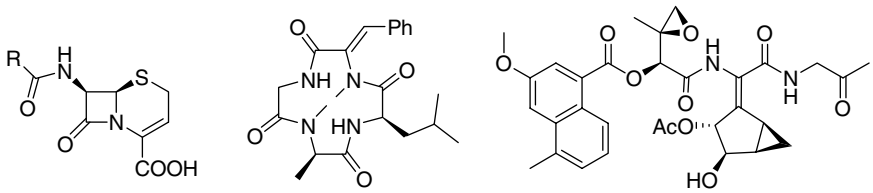


Figure 1.1 Naturally occurring DDPs.

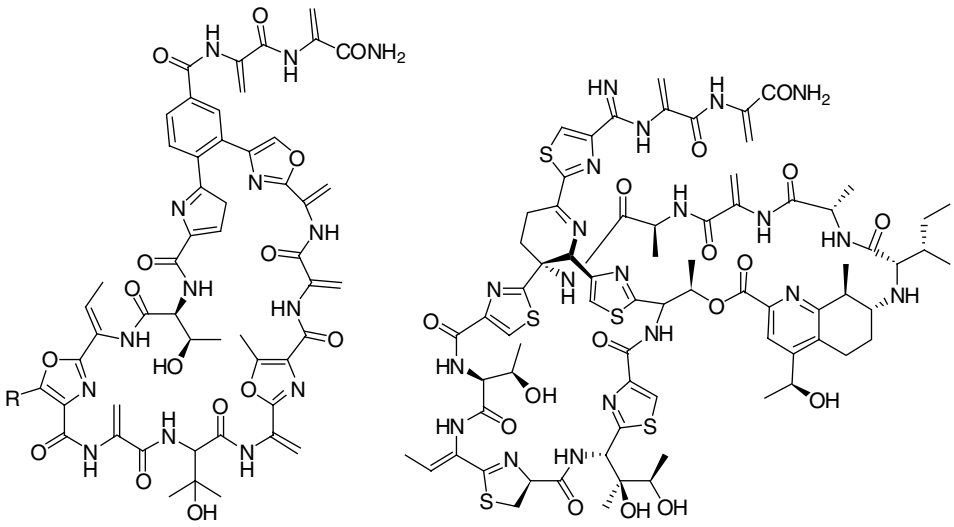
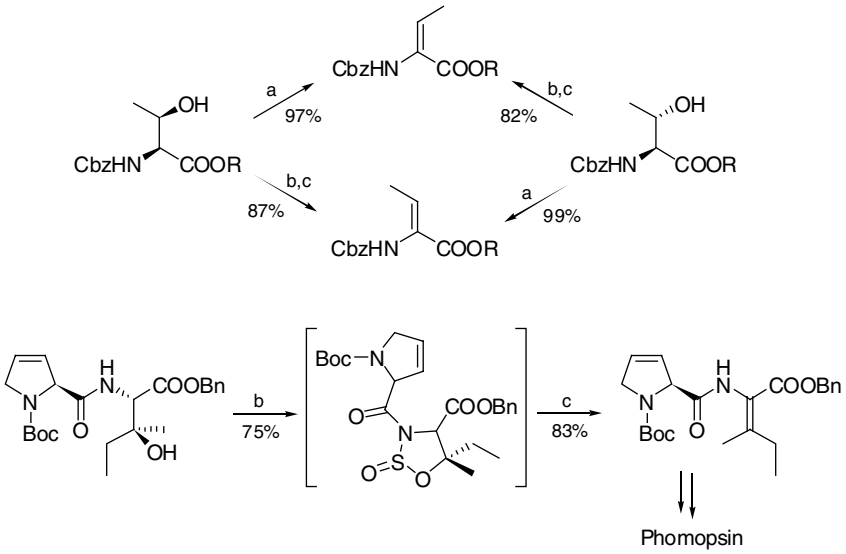


Figure 1.2 Naturally occurring didehydroalanine- and didehydroaminobutenoate-containing peptides.

(ΔAbu) are formed, two DDAAs also found widely in nature, such as in geninthiocin [7], berninamycin A [8], or thiostrepton (Figure 1.2) [9].

A wide range of reagents can be used for the activation of the OH group and elimination occurs in the presence of a suitable base. Useful combinations are oxalyl chloride [10], (diethylamino)sulfur trifluoride [11], dichloroacetyl chloride [12], tosyl chloride [13], and pyridine or NEt_3. PPh_3/diethyl azodicarboxylate [14] and carbodiimides in the presence of CuCl [15] can be used as well, and in general the thermodynamically more stable (*Z*) isomer is formed preferentially [16]. With respect to an application of this approach towards the synthesis of natural products, a stereoselective protocol is required, providing either the (*E*)- or (*Z*)-DDAA. Sai *et al.* reported a high selectivity for the (*E*)-ΔAbu from threonine by using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) in the presence of $CuCl_2$, while the (*Z*) isomer was obtained from *allo*-threonine (Scheme 1.1) [17]. Short reaction times

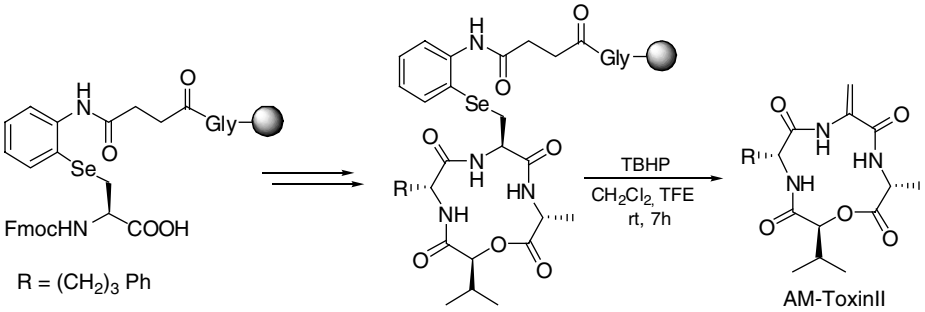


Scheme 1.1 Synthesis of DDAAs via β-elimination of water:
(a) 2 equiv. EDC, 0.1 equiv. $CuCl_2$, toluene, 80 °C, 0.5 h;
(b) 20 equiv. NEt_3, 10 equiv. $SOCl_2$, CH_2Cl_2, −78 °C, 30 min,
(c) 10 equiv. DBU, CH_2Cl_2, 0 °C, 30 min.

(0.5 h) are required for good (*E*) selectivity, because isomerization is observed under the reaction conditions. Therefore, longer reaction times strongly favor the thermodynamic (*Z*) product.

An alternative approach was reported by Wandless *et al.* [18]. They described a stereoselective elimination using $SOCl_2$/1,8-diazabicyclo[5.4.0]undec-7-ene (DBU). This reaction proceeds via a cyclic sulfamidite that can be isolated and purified, undergoing elimination on treatment with DBU. This protocol provides the opposite isomer compared to the Sai *et al.* procedure. Wandless *et al.* used their method for the stereoselective synthesis of disubstituted DDAAs in their synthesis of phomopsin and illustrated that this protocol is also suitable for the synthesis of DDPs.

1.2.1.1.2 **From β-Thio- and Selenoamino Acids** One of the best methods for the synthesis of didehydroalanine, and peptides containing this amino acid, starts from *S*-methylcysteine derivatives. *S*-Alkylation provides sulfonium salts that undergo elimination in a basic medium under relatively mild conditions [19]. Alternatively, the oxidation of thio- [20] and selenoamino acids [21] and subsequent thermolysis provides DDAAs in high yield. This approach was used successfully in natural product synthesis [22]. Nahamura *et al.* reported on a solid-phase synthesis of cyclic DDP AM-Toxin II using a selenated alanine as an anchoring residue (Scheme 1.2). After peptide synthesis and cyclization the seleno group was oxidized with *tert*-butylhydroperoxide (TBHP) and subsequent cleavage from the resin providing the natural product [23].



Scheme 1.2 Synthesis of AM-Toxin II according to Nakamura *et al*.

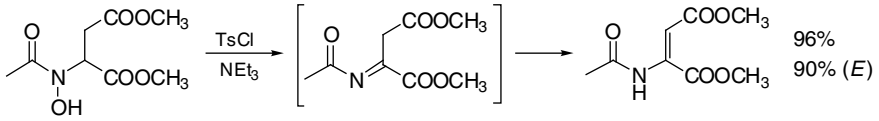
1.2.1.2 **Elimination from *N*-Hydroxylated and -Chlorinated Amino Acids and Peptides**

In 1944, Steiger reported on the elimination of H_2O from *N*-hydroxy amino acids in the presence of acetic anhydride/pyridine [24]. The *in situ* formed *O*-acetylated derivatives can be eliminated at room temperature in the presence of NEt_3 [25] or DBU [26]. The most convenient approach is the elimination using tosyl chloride/NEt_3, which gives the required DDAA in minutes with good (*E*) selectivity and nearly quantitative yields (Scheme 1.3) [27].

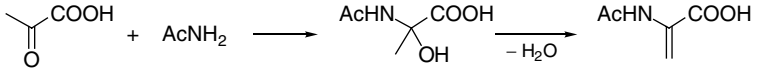
Similar good results are obtained in eliminations of *N*-chlorinated amino acid derivatives which can easily be obtained by oxidation of acylated amino acids with *t*BuOCl/NaOR [28] or NaOCl [29]. Primarily, an iminoester is formed, which undergoes isomerization to the enamide structure.

1.2.1.3 **DDAAs from α-Oxo Acids and Amides**

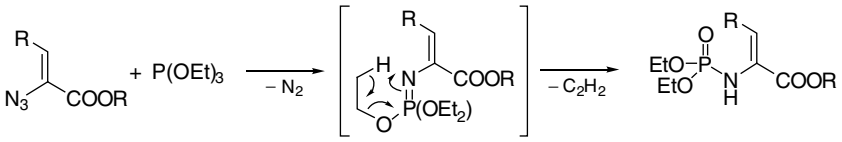
α-Oxo acids undergo addition of carboxamides on heating. α-Hydroxy-α-acylamino-carboxylic acids are formed primarily, which can undergo elimination of H_2O giving rise to DDAAs (Scheme 1.4) [30]. Using amino acid amides, this protocol can also be applied for the synthesis of DDPs. Best results are obtained with Cbz- or trifluoroacetic acid-protected amino acid amides, while side-products are observed with Boc-protected derivatives [31]. *N*-Alkylated DDAAs can be obtained in a similar manner by condensing primary amines with pyruvates, followed by acylation of the imine formed [32].



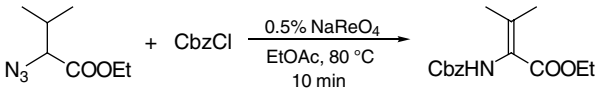
Scheme 1.3 Synthesis of DDAAs from *N*-hydroxyamino acids.



Scheme 1.4 Synthesis of DDAAs from α-oxo acids.



Scheme 1.5 Synthesis of DDAAs from azidoacrylates.



Scheme 1.6 Synthesis of DDAAs from azidoarboxylates.

1.2.1.4 DDAAs from Azides

α-Azidoacrylates are also suitable candidates for the synthesis of DDAAs. The azido group can be reduced electrolytically [33] or via Staudinger reaction [34]. In the latter case, with phosphines or phosphites the corresponding iminophosphoranes or phosphoric amides are obtained (Scheme 1.5).

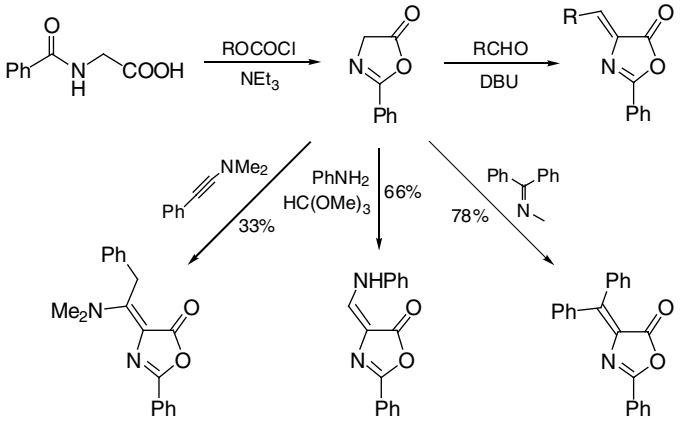
Saturated α-azidocarboxylates can be converted into DDAAs on treatment with strong bases such as BuLi or lithium diisopropylamide. The *in situ* formed iminoester can be directly acylated to the corresponding *N*-acylated DDAA [35]. By far the best method for the conversion of α-azidocarboxylates to DDAAs is their reaction with acyl halides or chloroformates in the presence of Re_2S_7 or $NaReO_4$ (Scheme 1.6) [36]. With phosgene as the acylating reagent, the corresponding Leuch's anhydrides are formed [37].

1.2.2 DDAAs via C=C Bond Formation

1.2.2.1 DDAAs via Azlactones [5(4H)-Oxazolones]

The Erlenmeyer azlactone synthesis [38] is a classical method for the synthesis of DDAAs, preferentially bearing aromatic or heteroaromatic substituents. Often this reaction is performed as a one-pot protocol by melting an aldehyde, acylglycine, acetic anhydride, and sodium acetate at about 140 °C [39]. For sensitive aldehydes a two-step procedure is more convenient, where the azlactone is prepared first and the subsequent aldol condensation is carried out in the presence of base under mild conditions [40]. The (*Z*) oxazolones are formed preferentially, but these can undergo isomerization to the corresponding (*E*) derivatives in the presence of phosphoric acid or HBr [39b]. Apart from aldehydes, a wide range of electrophiles can be reacted with the deprotonated azlactone (Scheme 1.7). While imines give the same products such as aldehydes and ketones [40, 41], amines in the presence of orthoesters give rise to β-aminoalkylidene oxazolidinones [42]. Similar structures are obtained in the reaction with ynamines [43].

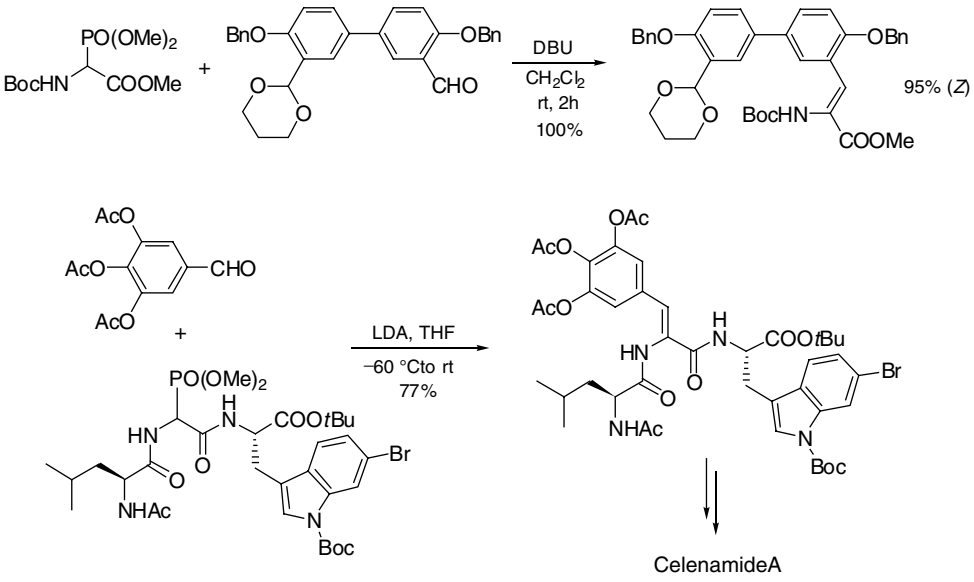
Ring opening of the oxazolinone is possible with a wide range of nucleophiles. While hydrolysis gives the *N*-benzoylated DDAAs, with alcohols the corresponding esters are obtained [39b]. Ring cleavage with amino acid esters gives direct access to DDPs containing a *N*-terminal DDAA [44].



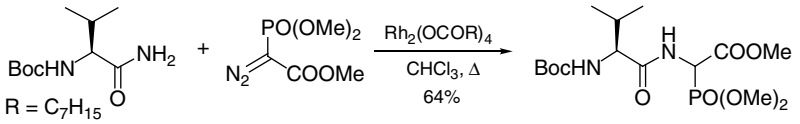
Scheme 1.7 Reactions of azlactones.

1.2.2.2 DDAAs via Horner–Emmons and Wittig Reactions

A quite popular approach towards DDAAs was developed by Schmidt *et al.* based on a phosphonate condensation of *N*-protected dimethoxyphosphoryl glycinates [45]. The Cbz- and Boc-protected derivatives are commercially available or can be prepared in large scale from glyoxylic acid [45, 46]. Condensations of these phosphonates with aldehydes, also highly functionalized ones, proceed well in the presence of KO*t*Bu or DBU as a base, and the (*Z*)-DDAAs are formed preferentially (Scheme 1.8) [47]. Subsequent asymmetric catalytic hydrogenation (see Section 1.3.1.4) provides



Scheme 1.8 Synthesis of DDAAs and DDPs via Horner–Emmons reaction.



Scheme 1.9 Phosphonopeptides via NH insertion.

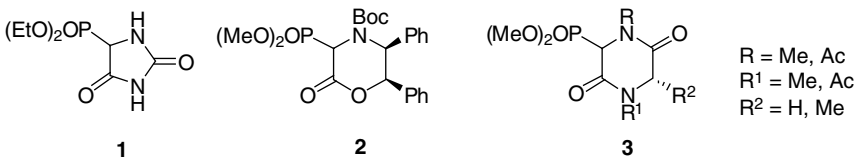
straightforward access to nonproteinogenic amino acids. Therefore, this approach has found many applications in amino acid [48] and natural product syntheses [49]. Incorporating dimethoxyphosphoryl glycine into peptides allows a direct synthesis of DDPs [47], as has been illustrated in the synthesis of celenamide A [47a, 50] and antrimycin D [51].

During their synthesis of stephanotic acid, Moody *et al.* [52] developed an independent approach towards phosphoryl glycine-containing peptides based on a rhodium-catalyzed NH insertion of the corresponding carbenes into amino acid amides (Scheme 1.9) [53].

By incorporating an alkoxyphosphoryl glycine into heterocycles such as hydantoins (**1**) this phosphonate condensation approach can be used for the synthesis of cyclic DDAA derivatives (Figure 1.3) [54]. Introduction of stereogenic centers into the heterocycle allows subsequent diastereoselective reactions of the DDAAs obtained. Williams *et al.* introduced the chiral phosponate **2** as a precursor for "chiral" didehydroalanine, which was subjected to modifications on the double bond, such as cycloadditions [55]. Chai *et al.* described phosphonates **3** as a precursor for methylene piperazine-2,5-diones, which were used as templates for amino acid syntheses [56].

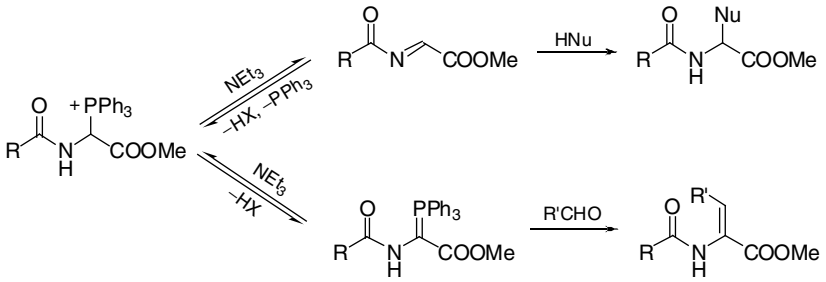
Comparable to these phosphonate condensations are the corresponding Wittig reactions using *N*-acyl-α-triphenylphosphonioglycinates. These Wittig reagents can be obtained either from the corresponding α-hydroxyglycinates via halogenation/PPh_3 substitution [46] or from the corresponding 4-triphenylphosphoranylidene azlactones [57]. Elimination occurs on treatment with base, giving an equilibrated mixture of the *N*-acylimino acetates and the phosphonium ylides. Addition of nucleophiles results in the formation of substitution products [58], while on addition of electrophiles, such as aldehydes, the formation of DDAAs (as a *E*/*Z* mixture) is observed (Scheme 1.10) [59]. This approach found several applications in the synthesis of β-lactams [60].

Steglich *et al.*, who were the first to describe the synthesis of these Wittig reagents, observed a dimerization of halogenated glycinates on treatment with PPh_3 [46]. This can easily be explained by a reaction of the *N*-acylimino acetates and the phosphonium ylides formed *in situ*. This dimerization process can also be transferred to peptides giving rise to cross-linked DDPs (Scheme 1.11) [61].

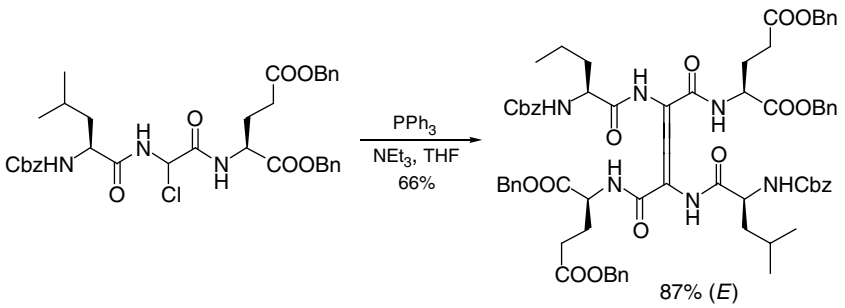


1 **2** **3**

Figure 1.3 Heterocyclic phosphonates.



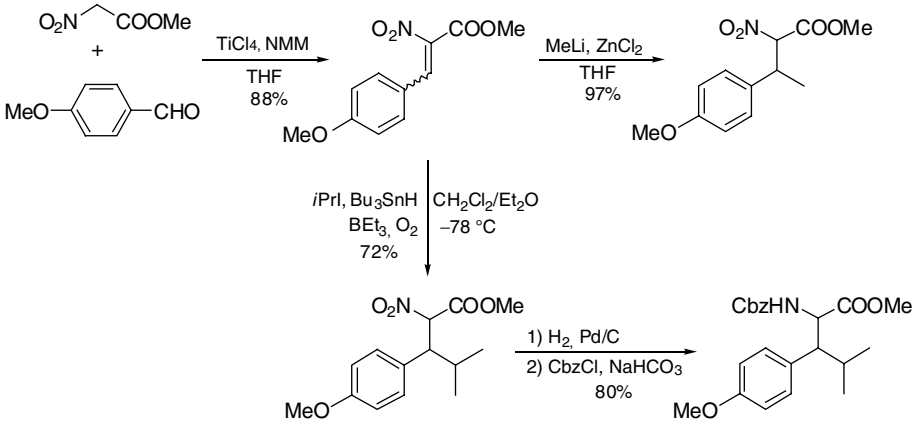
Scheme 1.10 Synthesis of DDAAs via Wittig reaction.



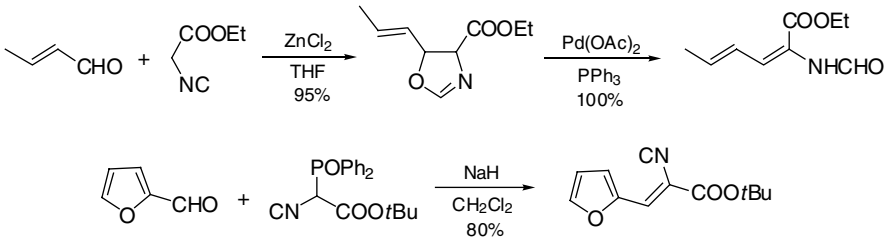
Scheme 1.11 Synthesis of cross-linked DDPs.

1.2.2.3 DDAAs via Enolates of Nitro- and Isocyano- and Iminoacetates

Nitroacetic esters can easily undergo Knoevenagel reactions with a wide range of aldehydes [62] or imines [63] giving rise to α,β-unsaturated α-nitro esters. β-Alkoxy- or β-amino-substituted derivatives are obtained from orthoformates [64] or dialkylformamide dialkylacetals [65]. Condensation in the presence of $TiCl_4$/base is an especially mild protocol and gives high yields of an isomeric mixture (Scheme 1.12) [66]. The unsaturated nitro esters obtained are excellent Michael



Scheme 1.12 Synthesis of DDAAs from nitroacetates and subsequent reactions.



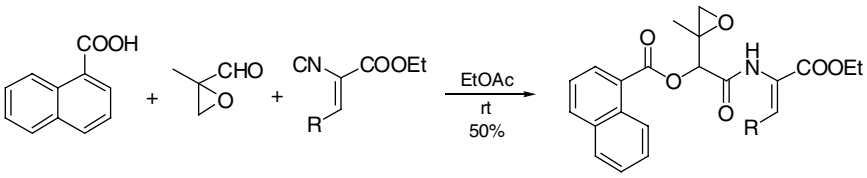
Scheme 1.13 Synthesis of DDAAs from isocyanoacetates.

acceptors. Nucleophilic or radical addition and subsequent reduction of the nitro group provides easy access to highly substituted amino acids [67b,67]. The nitro group of the unsaturated esters can be reduced easily without affecting the double bond using aluminum amalgam [68], zinc in glacial acetic acid [69], or by catalytic hydrogenation using Raney nickel [70] or Pt/C [71].

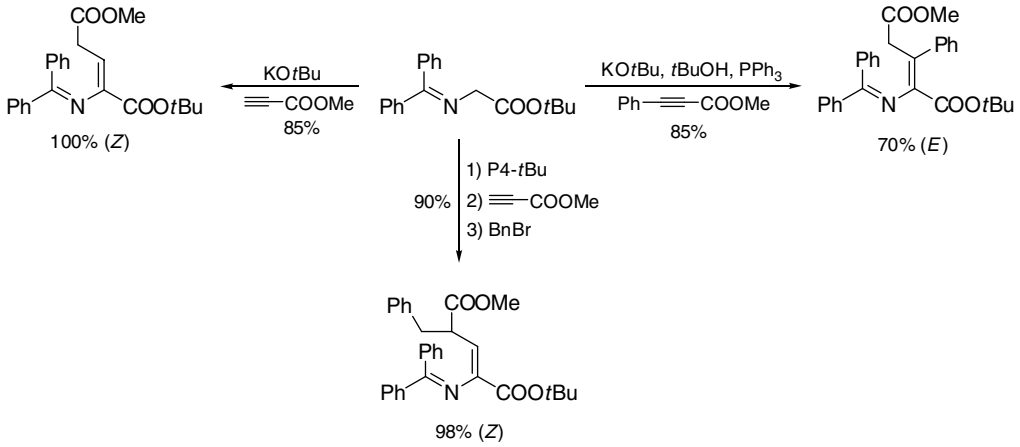
Schöllkopf *et al.* reported on similar condensation reactions using isocyano acetates. Condensation with aldehydes [72] and ketones [73] in aprotic solvents gives rise to *N*-formylated DDAAs, probably via oxazoline intermediates [74]. The reaction conditions are relatively mild and allow the condensation of sensitive carbonyl compounds. Best results with α,β-unsaturated carbonyls are obtained in the presence of Lewis acids such as $ZnCl_2$ or CuCl (Scheme 1.13) [75]. The unsaturated oxazoline obtained can be cleaved to the *N*-formylated DDAA using $Pd(OAc)_2/PPh_3$. According to comparable reactions described for the nitro acetates, orthoformates [76] and dimethylformamide acetals [77] give rise to the corresponding β-alkoxy or β-amino substituted DDAAs. α,β-Unsaturated isocyano acetates can be obtained via phosphonate condensation [78].

The isocyanides can not only be hydrolyzed to the corresponding *N*-formyl DDAAs [79], they can also be used in multicomponent couplings such as the Passerini [80] or Ugi [81] reactions. This allows direct incorporation of DDAAs into peptides. Armstrong *et al.* used such an approach during their synthesis of azinomycins (Scheme 1.14) [81a,b].

O'Donnell's imino glycinates are very useful nucleophiles and valuable precursors for the synthesis of complex amino acids [82]. Alvarez-Ibarra *et al.* reported on their applications in DDAA synthesis via nucleophilic addition to alkynoates (Scheme 1.15). Reaction with methyl propiolate gave rise to the (*Z*)-configured DDAA in a thermodynamically driven process. The *in situ* formed vinyl anion underwent 1,3-hydride shift and subsequent migration of the double bond to the



Scheme 1.14 DDPs via Passerini reaction.



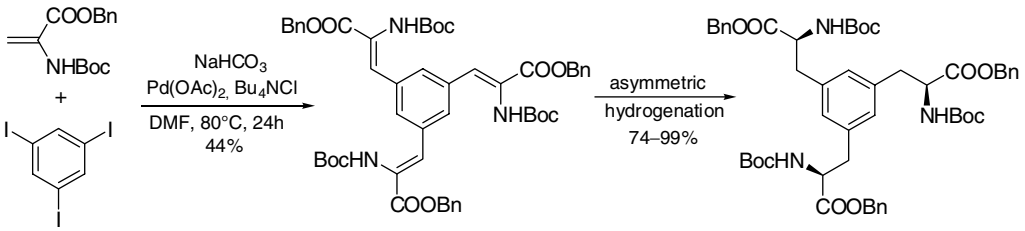
Scheme 1.15 Addition of imino glycinates towards alkynoates.

α,β-position [83]. On the other hand, β-substituted derivatives were best prepared by treatment of deprotonated imino glycinates with substituted alkynoates giving (*E*/*Z*) mixtures of products [84]. By using naked enolates, prepared in the presence of crown ethers or by using Schwesinger base [85], the enolate formed after isomerization could be trapped with electrophiles such as benzyl bromide [83].

1.2.3 DDAAs via C—C Bond Formation

1.2.3.1 DDAAs via Heck Reaction

The synthesis of a wide range of DDAAs from the most simple and easily available representative, didehydroalanine, is a straightforward and highly attractive approach [86]. Especially the reaction of aryl halides in combination with asymmetric catalytic hydrogenation of the DDAA formed gives easy access to libraries of substituted phenylalanines [87]. Frejd *et al.* applied this approach for the synthesis of dendrimers, containing a C-3-symmetric phenylalanine derivative as the center unit (Scheme 1.16) [88]. Gibson *et al.* used an intramolecular Heck reaction as the cyclization step in their synthesis of didehydrophenylalanine cyclophanes [89].



Scheme 1.16 Trifold Heck reaction of didehydroalanine.

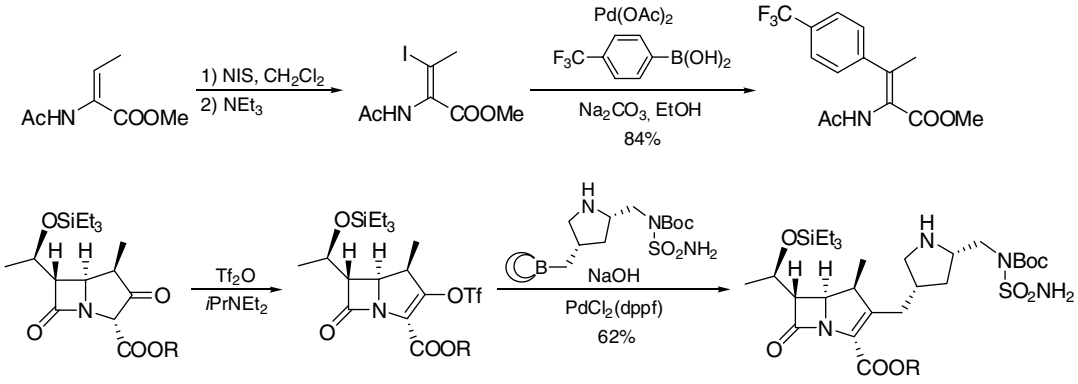
The Heck reaction can also be performed under solvent-free conditions in a ball mill [90] or on solid support [91].

1.2.3.2 DDAAs via Cross-Coupling Reactions

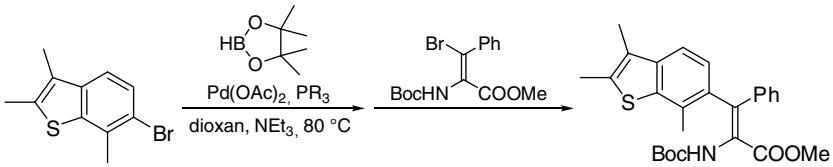
DDAAs containing a leaving group at the β-position can be subjected to a wide range of cross-coupling reactions such as Suzuki, Stille, or Sonogashira couplings, allowing the synthesis of highly functionalized and complex amino acids. β-Brominated or iodinated DDAAs, easily obtained by a halogenation/elimination approach (see Section 1.3.3) can be coupled with a wide range of boranes [92], borates [93], or boronic acids (Scheme 1.17) [94]. Cross-coupling occurs under retention of the olefin geometry. In general, the (*E*)-β-halogen DDAAs give higher yields of the substituted (*E*)-DDAAs, compared to the (*Z*) derivatives [95]. The corresponding triflates are easily obtained from the corresponding β-keto amino acids as nicely illustrated in the synthesis of functionalized carbapenems [96].

Queiroz *et al.* reported on an interesting one-pot reaction consisting of a palladium-catalyzed borylation of aryl halides and subsequent Suzuki coupling with β-brominated DDAAs (Scheme 1.18). The DDAAs obtained were subjected to a metal-assisted intramolecular cyclization giving indoles [97].

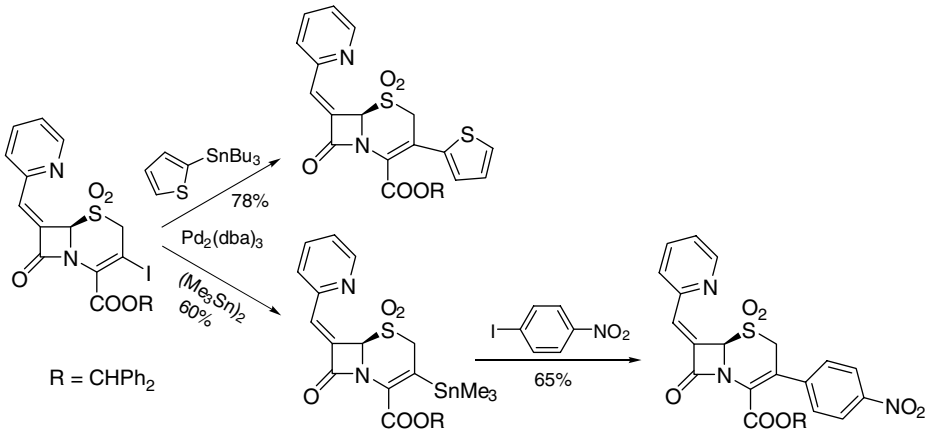
Stille couplings [98] have found several applications in the modification of β-lactams [99]. Iodinated bicyclic DDAAs were coupled with a wide range of nucleophiles such as vinyl and (het)aryl stannanes as well as stannyl acetate, thiolate, and acetylide (Scheme 1.19). The reaction with $(Me_3Sn)_2$ allowed the synthesis of stannylated β-lactams that could be coupled with electrophiles [100].



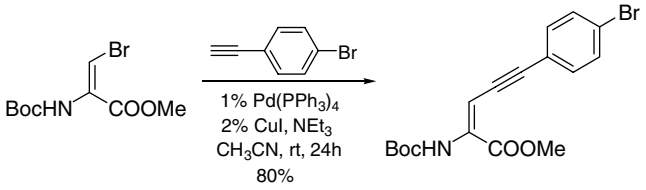
Scheme 1.17 Synthesis of DDAAs via Suzuki coupling.



Scheme 1.18 Domino borylation/Suzuki coupling.



Scheme 1.19 Synthesis of DDAAs via Stille coupling.



Scheme 1.20 Synthesis of DDAAs via Sonogashira coupling.

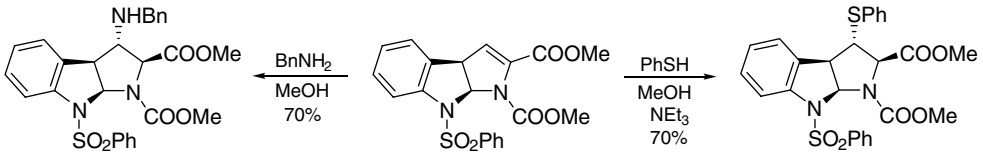
Sonogashira couplings of halogenated DDAA derivatives with terminal alkynes allows the synthesis of highly unsaturated amino acids [29]. Coupling with *p*-bromophenyl acetylene, for example, gives rise to a brominated DDAA, which can be further modified, for example, via Suzuki coupling (Scheme 1.20) [101].

1.3 Reactions of DDAAs

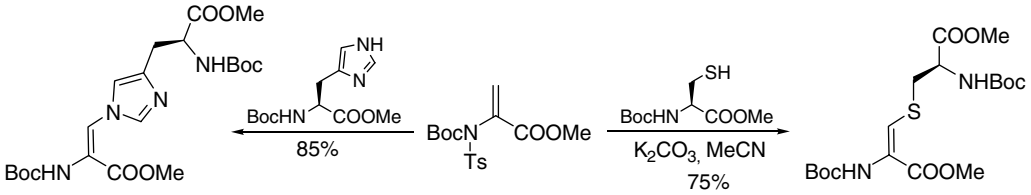
1.3.1 Additions to the C=C Bond

1.3.1.1 Nucleophilic Additions

In principle, suitably protected DDAAs react in the same way as acrylic acid derivatives, undergoing 1,4-addition of a wide range of nucleophiles, such as amines [102], thiols [103], or electron-rich *N*-heterocycles (Scheme 1.21) [104]. For example, addition to didehydroproline derivatives proceeds stereoselectively, giving the 1,4-addition products as single stereoisomers [102]. It should be mentioned that under acidic conditions the addition of the nucleophile occurs preferentially to the α-position of the DDAA [105].



Scheme 1.21 Nucleophilic attack on didehydroprolines.



Scheme 1.22 DDAAs via nucleophilic addition/elimination.

Best results in the 1,4-addition are obtained with substrates containing two electron-withdrawing groups on the nitrogen, as illustrated in the addition of several heterocycles [106]. If one of these electron-withdrawing groups is a tosyl group, the substituted DDAAs are obtained via an addition/elimination mechanism (Scheme 1.22) [107].

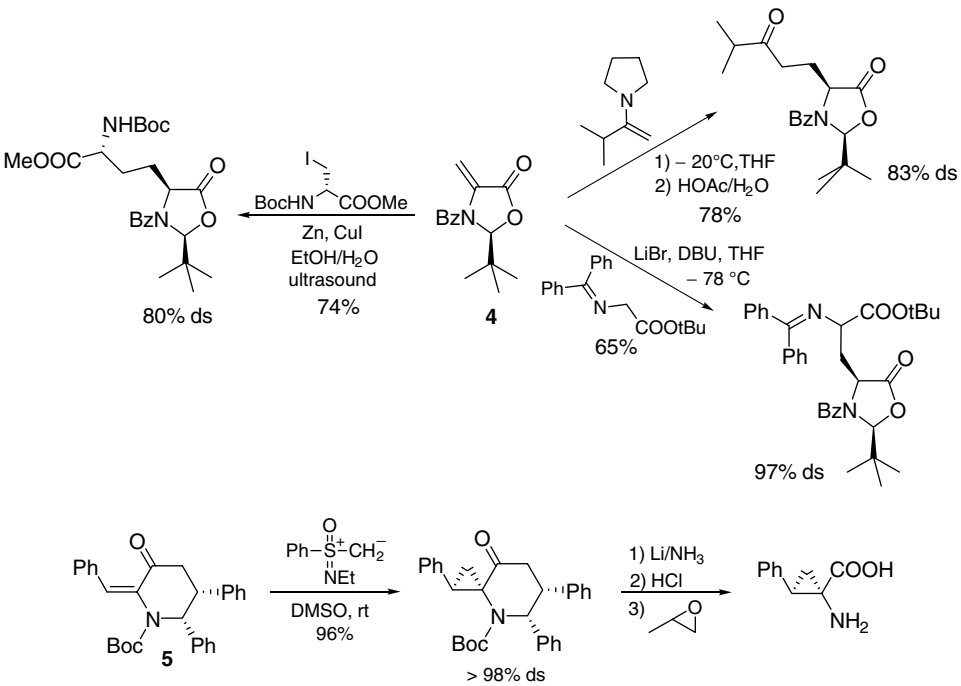
Additions of stabilized carbanions [108] and enols [109] or enamines [110] result in the elongation of the amino acid side-chain. The application of chirally modified DDAAs [111] allows the stereoselective synthesis of unnatural amino acid derivatives (Scheme 1.23) [112]. Furthermore, highly functionalized substituents can be introduced via cuprate addition [113] – a reaction that also gives good selectivities with cyclic chiral DDAAs such as **4** [114]. 1,4-Additions to acyclic chiral esters in general are less selective [115]. If the cuprate is generated *in situ* from a halide via halogen–zinc exchange (Luche conjugate addition) [116] the reaction can be carried out under aqueous conditions [117].

The addition of sulfur ylides to DDAAs is a straightforward approach to 1-aminocyclopropane carboxylic acids [118]. Williams *et al.* described the first asymmetric synthesis of a cyclopropane amino acid via addition of a sulfur ylide to a chirally modified DDAA **5** (Scheme 1.23). Excellent yields and diastereoselectivities were obtained, and the free amino acid was obtained via reduction under Birch conditions and subsequent cleavage of the Boc protecting group [55, 119].

Meanwhile, the additions of sulfur ylides to a range of other chiral DDAA derivatives, such as the pivane derivative **6** [120], the oxazinone **7** [121], the diketopiperazine **8** [122], or the oxazolone **9** [123], were described (Figure 1.4).

1.3.1.2 **Radical Additions**

DDAAs are good acceptors for radicals, generated for example, from alkyl or acyl halides using the Bu_3SnH/AIBN protocol [124]. This approach has found many applications, especially in cyclization reactions [125] (e.g., for the synthesis of pyroglutamates starting from α-halo amides [126]). In principle, the primarily



Scheme 1.23 Additions of *C*-nucleophiles towards chirally modified DDAA derivatives.

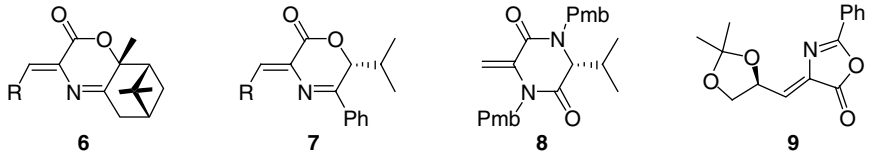
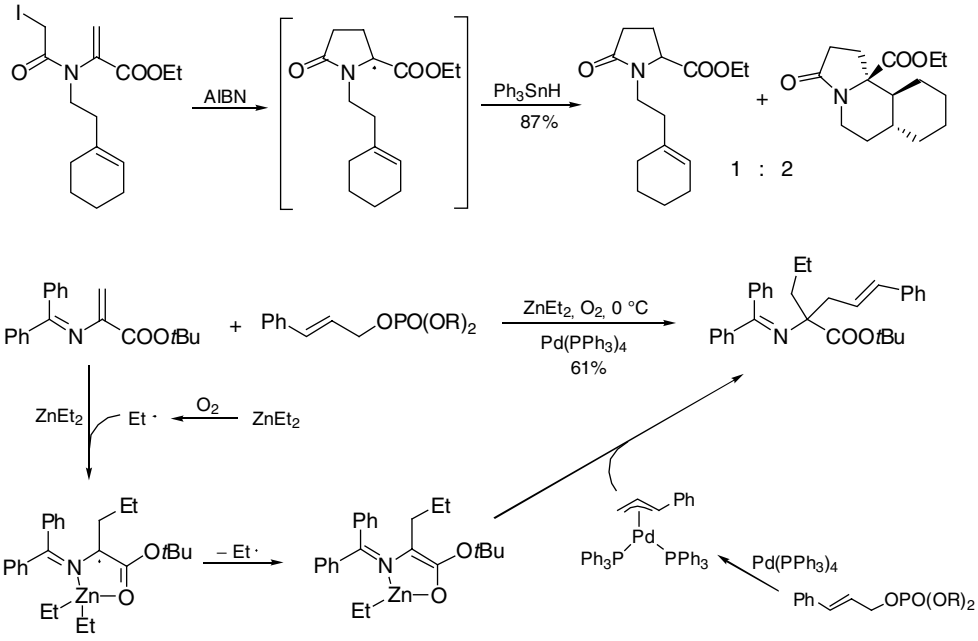


Figure 1.4 Chirally modified DDAAs used in cyclopropanation reactions.

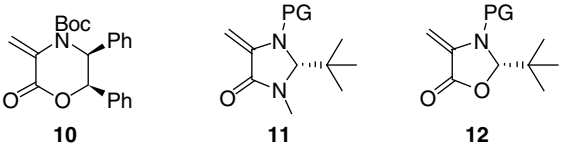
formed cyclic radical can be trapped with another radical acceptor [127] or can undergo a domino cyclization if a suitable double bond is present in the molecule (Scheme 1.24). Depending on the substitution pattern and the tin hydride used, mixtures of mono- and bicyclic products are obtained as single regio- and diastereomers, as a result of a 5-*endo*-6-*endo* cyclization [128]. Alternatively, the radicals can also be generated from epoxides using $TiCl_3$ – an approach which was used for the synthesis of glycosylated amino acids [129].

An interesting combination of radical addition and palladium-catalyzed allylic alkylation was reported by Takemoto *et al.* [130]. The initially formed chelated radical is converted into a chelated enolate, which then undergoes subsequent allylic alkylation (Scheme 1.24).

Vederas *et al.* reported on the generation of radicals from protected glutamates via the corresponding diacyloxoiodobenzene [131]. Additions to didehydroalanine derivatives gave rise to DDAAs, which were converted into diaminopimelic acids via



Scheme 1.24 Radical additions to DDAAs.



10 **11** **12**

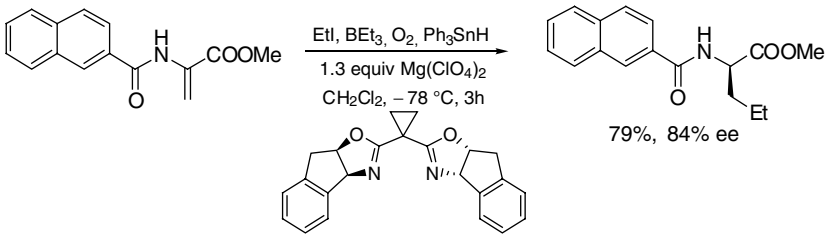
Figure 1.5 Chirally modified DDAAs used in radical additions.

catalytic hydrogenation. The reaction was also carried out with commonly used chiral didehydroalanine analogs, such as oxazinone **10** [132], imidazolidinone **11** [133], and oxazolidinone **12** (Figure 1.5) [134]. In the last case the expected dimerization products were also obtained. The chirally modified DDAAs allow the diastereoselective generation of amino acids. In principle, chiral auxiliaries can be used as well, such as chiral esters [135].

Sibi *et al.* reported another elegant protocol for enantioselective radical additions using a selective hydrogen atom transfer from tin hydride in the presence of a Lewis acid and a chiral ligand. The results strongly depend on the reaction conditions, especially the Lewis acid used, but under optimized conditions enantiomeric excesses up to 85% are possible (Scheme 1.25) [136].

1.3.1.3 **Cycloadditions**

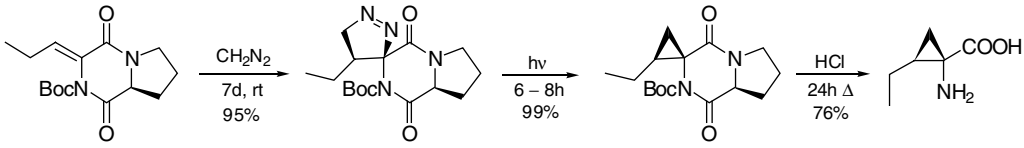
Cycloadditions of DDAAs give rise to quaternary amino acids. This area of reactions was covered by an excellent review by Cativiela and Diaz-de-Villegas [137]. Therefore, only the general principle and new developments will be discussed herein.



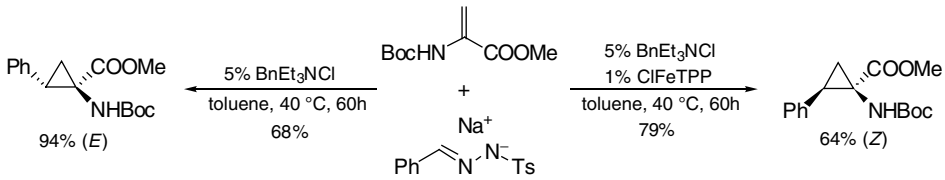
Scheme 1.25 Asymmetric radical addition towards DDAAs.

1.3.1.3.1 **[3 + 2] Cycloadditions** The addition of diazo compounds towards DDAAs is an interesting approach for the synthesis of aminocyclopropane carboxylic acids (see also *S*-ylide addition, in Section 1.3.1.1). The reaction occurs via a 1,3-dipolar cycloaddition providing a pyrazoline. Extrusion of N_2, either thermally or on photolysis, gives rise to cyclopropane derivatives [120b,138]. Various chirally modified DDAAs have been used to control the stereoselective outcome of the reaction [139]. Excellent results were obtained with proline-containing diketopiperazines, which gave the corresponding pyrazolines almost as single diastereomers (>95% d.s.). Photolysis produced the spirocyclopropanes, which could be cleaved under acidic conditions to the free amino acids (Scheme 1.26). Best results were obtained with the *N*-Boc-protected diketopiperazines [140].

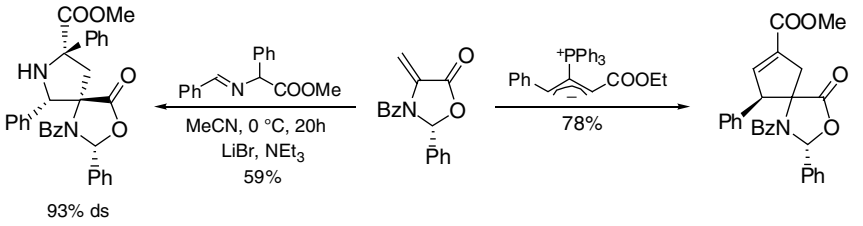
Due to the toxicity and lability of the diazo compounds, recent studies primarily focused on practical aspects and the handling of the diazo compounds. Aggarwal and Cox described asymmetric cyclopropanations with diazo compounds that were released *in situ* from tosylhydrazones. Interestingly, the (*E*)-configured product was formed preferentially by simple warming of the components to 40 °C. The (*Z*) isomer was the major one in the presence of an iron porphyrin (ClFeTPP) catalyst, although in this case the selectivity was moderate (Scheme 1.27) [141].



Scheme 1.26 [3 + 2] Cycloaddition of diazomethane.



Scheme 1.27 Cyclopropanation with *in situ* generated diazocompounds.



Scheme 1.28 [3 + 2] Cycloadditions of azomethine and phosphor ylides.

The 1,3-dipolar cycloaddition of azomethine ylides, easily obtained from *N*-alkylidene amino acid esters, to electron-deficient alkenes is a straightforward approach for the synthesis of functionalized prolines. Pyne *et al.* investigated the cycloaddition of chiral oxazolidinones (Scheme 1.28). The azomethine ylides were generated *in situ* in the presence of the DDAA-derivative by treating their tetrahydrofuran (THF) or MeCN solution with base (DBU or NEt_3). In nearly all cases investigated the reactions were completely regioselective with a high preference for the *exo* diastereomer. The auxiliary could be removed easily by saponification [142].

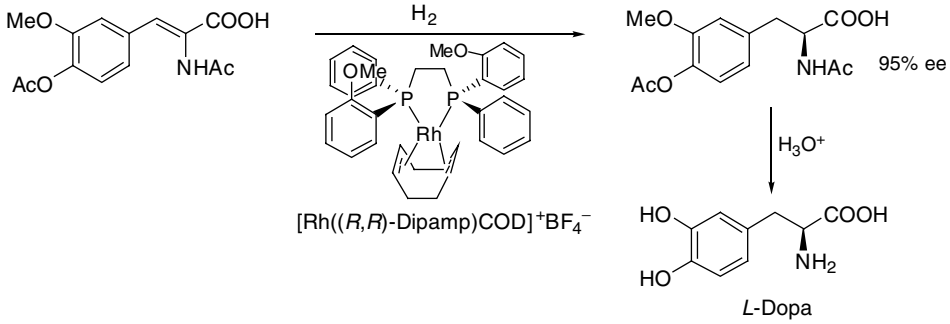
Similar results were also obtained with nitrones [143] and nitrile oxides [144]. Cyclopentenyl glutamates were obtained by a [3 + 2] cycloaddition of phosphorylides [145], obtained by nucleophilic attack of phosphines on allenic or alkynoic acid esters [146].

1.3.1.3.2 **[4 + 2] Cycloadditions** The Diels–Alder reaction is probably the most efficient method for the stereoselective synthesis of six-membered rings. This protocol has found widespread application in amino acid synthesis [137] and nearly all chirally modified DDAAs described so far have been used in this reaction [147]. Several functionalized cyclohexane α-amino acids have been synthesized as conformationally constrained amino acid analogs [148].

The reactions can be carried out thermally or in the presence of Lewis acids, while the rate of the Diels–Alder reaction as well as the *exo*/*endo* selectivity strongly depends on the Lewis acid used. Moderate selectivities were obtained with chirally modified aluminum and titanium complexes [149].

1.3.1.4 Catalytic Hydrogenations

The asymmetric catalytic hydrogenation of DDAAs is an important and straightforward approach to optically active amino acids. In principle, two major protocols are applied to introduce chirality: either hydrogenation under substrate control using chiral (modified) DDAAs or DDPs, or the application of chiral catalysts. DDAAs are standard substrates for the evolution of new chiral metal/ligand complexes. This chapter cannot go into detail, but the newest developments are covered in a series of recent reviews [150]. The development of the homogeneous asymmetric hydrogenation started with the discovery of Wilkinson's catalyst [151]. In the late 1960s, Horner [152] and Knowles [153] reported on the first asymmetric hydrogenations, albeit with moderate enantioselectivity. A breakthrough was the introduction of



Scheme 1.29 DOPA synthesis via asymmetric catalytic hydrogenation.

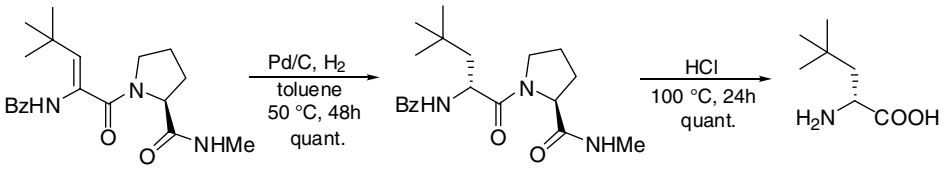
a bidentate chelating ligand 1,2-bis[(*o*-anisyl)-(phenyl)phosphino]ethane (DIPAMP) [154], a highly efficient ligand for rhodium-catalyzed hydrogenations. The process found application at Monsanto for the industrial production of 3,4-dihydroxyphenylalanine (DOPA) (Scheme 1.29) [155] – a development that won Knowles the Nobel Prize in 2001 [156]. This chiral catalyst system was applied by Schmidt *et al*. [49, 157] and others [158] to the synthesis of a wide range of unusual amino acids and peptides.

This was the starting point for the development of new ligands and hundreds of them are now in use or under investigation [150]. One of the advantages of the homogeneous hydrogenation is the possibility to use the catalyst not only for the catalytic hydrogenation, but also for other transition metal-catalyzed processes. For example, Robinson *et al*. reported on the synthesis of cyclic amino acids via a combination of a rhodium-catalyzed hydrogenation/hydroformylation and subsequent ring closure [159].

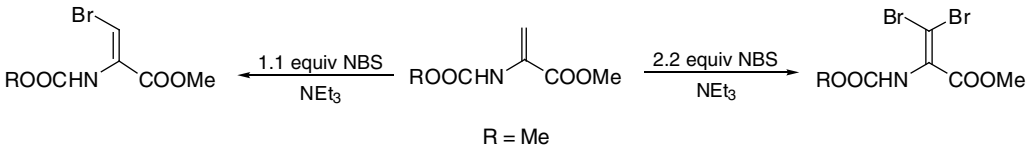
The major disadvantage of homogeneous catalysis is the problem of pollution of the product by the soluble catalysts. Therefore, attempts have been made to use chirally modified heterogeneous catalysts for asymmetric hydrogenations, but the selectivities are significantly worse compared to the homogeneous version [160]. The immobilization of homogeneous chiral catalysts on solid supports could solve some of the problems, such as difficult separation and recycling of the expensive chiral catalyst [152a]. However, the immobilized ligands or catalysts often display a lower selectivity and reactivity as compared to the corresponding homogeneous systems. Fan *et al*. applied a MeO-polyethylene glycol-supported ligand in the rhodium-catalyzed hydrogenation of α-acetamidocinnamic acid [161]. Enantiomeric excesses up to 96% were obtained and the polymer catalyst was recycled at least 3 times without loss of enantioselectivity. In contrast, the insoluble polymer-supported catalyst lost its selectivity in the second cycle [162].

Achiral (heterogeneous) catalysts can be used for asymmetric hydrogenations of chirally modified DDAAs. Best results are obtained with cyclic derivatives such as **6** [163], **8** [116a] (Figure 1.3), **10** [164], and **11** [165] (Figure 1.4).

In principle, the chiral information of a peptide chain can also be used to control the stereochemical outcome of a hydrogenation of an incorporated DDAA. In



Scheme 1.30 Substrate controlled hydrogenation of DDPs.



Scheme 1.31 Bromination of DDAAs.

general, the chiral induction is moderate, but can be increased by addition of metal salts [166]. Alternatively, chiral ligands can be applied in homogeneous hydrogenations of DDPs [167]. Schmidt *et al.* reported on highly diastereoselective heterogeneous hydrogenations of DDPs containing a C-terminal (*S*)-proline amide (Scheme 1.30) [168]. The (*R*)-configured amino acid could be obtained after hydrolysis of the dipeptide in quantitative yield.

1.3.2
Halogenations of DDAAs

β-Halogenated DDAAs are interesting building blocks and starting materials for cross-coupling reactions, which allow the synthesis of more complex DDAAs (see Section 1.2.3.2). The most commonly used β-brominated DDAAs are easily obtained using bromine [169] or *N*-bromosuccinimide (NBS) and a base such as NEt_3. In general, the (*Z*) isomer is obtained preferentially [29,94a,170]. Substituted DDAAs often give mixtures of isomers [171]. Application of more than 2 equiv. of NBS gives rise to β,β-dibrominated DDAAs (Scheme 1.31).

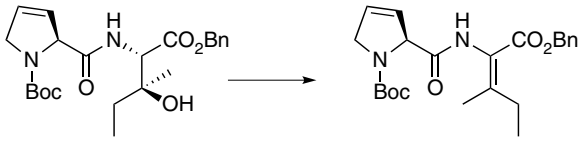
In an analogous manner the corresponding chlorinated and iodinated amino acids can be obtained by using either chorine [172] or *N*-iodosuccinimide [95b], while the fluorinated DDAAs requires a more complicated protocol [173].

1.4
Conclusions

DDAAs are not only interesting building blocks for the synthesis of natural products and drug-like molecules, but also important intermediates for the (stereoselective) synthesis of all kinds of α-amino acids. A wide range of protocols have been developed for their synthesis and especially, modern cross-coupling reactions have enlarged their synthetic potential dramatically.

1.5
Experimental Procedures

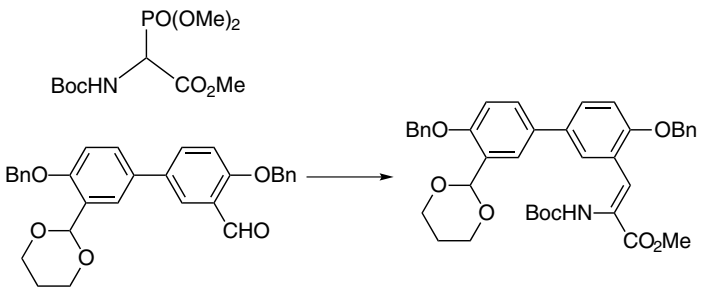
1.5.1
General Procedure for the Two-Step Synthesis of Dehydroisoleucine Derivatives [18]



Synthesis of sulfamidite from β-hydroxyisoleucine derivatives The β-hydroxyisoleucine derivative was dissolved in distilled CH_2Cl_2 (to give a concentration of 10 mM) in an oven-dried flask under an argon atmosphere and the solution was cooled to −78 °C. Distilled triethylamine (20 equiv.) was added via syringe and allowed to stir 5 min. Distilled thionyl chloride (10 equiv.) was added dropwise. The reaction was allowed to stir at −78 °C for 30 min before quenching with methanol (10 equiv.) at −78 °C. This solution was poured into a separatory funnel, and partitioned between water and CH_2Cl_2. The organic layer was dried with $MgSO_4$, concentrated, and the product was purified using silica gel chromatography.

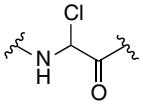
Elimination of sulfamidite to yield α,β-dehydroisoleucine derivative The purified sulfamidite (or mixture of diastereomers) was dissolved in distilled CH_2Cl_2 (to give a 100 mM solution) in an oven-dried flask under an argon atmosphere and the solution was cooled to 0 °C. DBU (10 equiv.) was added and the solution was stirred 30 min. This solution was poured into a separatory funnel, and partitioned between CH_2Cl_2 and saturated aqueous $NaHCO_3$. The organic layer was dried over $MgSO_4$, concentrated, and the product was purified using silica gel chromatography.

1.5.2
General Procedure for the Synthesis of α,β-Didehydroamino Acid Esters by the Phosphorylglycine Ester Method using DBU [45]



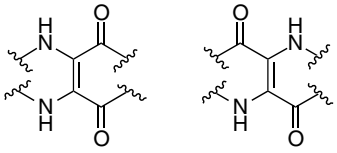
To a solution of methyl 2-acylamino-2-(dimethoxyphosphoryl)acetate (364 mg, 1.1 mmol) in the CH_2Cl_2 (2 ml) was added DBU (160 mg, 1.05 mmol). After 10 min, the respective carbonyl compound (1 mmol) was added. After 2 h the solution was diluted with EtOAc (20 ml), washed with 1 N H_2SO_4 (5 ml), dried ($MgSO_4$), and concentrated under vacuum. The residue was filtered through silica gel (hexane/EtOAc, 1: 1) to remove excess phosphorylglycine ester. The (*E*/*Z*) ratio of the product can be determined by high-performance liquid chromatography and ^{1}H-nuclear magnetic resonance (NMR) spectroscopy.

1.5.3
General Procedure for the Synthesis of α-Chloroglycine Derivatives: [61]



To a solution of the α-(ethylthio)glycyl peptide (1.0 mmol) in CH_2Cl_2 (30 ml) was added a 1 M solution of SO_2Cl_2 (1.1 ml, 1.1 mmol) in CH_2Cl_2 at 0 °C. After 30 min stirring, the solvent and all volatile byproducts were evaporated (cool trap), and the resulting residue was dried under high vacuum. The resulting α-chloropeptides were used for the next step without further purification.

1.5.4
General Procedure for the Synthesis of Homomeric Dimers [61]

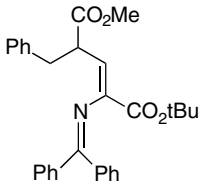


To a solution of the α-chloroglycyl peptide (1.0 mmol) and PPh_3 (0.2 mmol) in THF (50 ml) was added NEt_3 (0.15 ml, 1.1 mmol) in THF (30 ml) dropwise during 5 h. After continued stirring overnight, petroleum ether (200 ml) was added and the mixture filtered through Celite. Evaporation of the filtrate *in vacuo* yielded an oily residue which was purified by column chromatography on silica gel (petroleum ether/EtOAc).

1.5.5

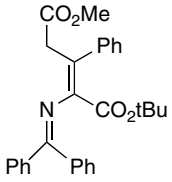
General Procedure for the Synthesis of (Z)-γ-Alkyl-α,β-Didehydroglutamates from Imino Glycinates [83]

Reactions with P4-*t*Bu base



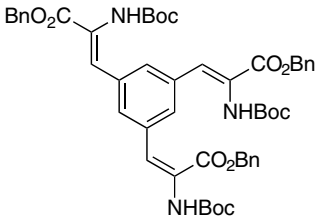
To a solution of the imino glycinate (0.48 mmol) in THF (2.0 ml) at −78 °C was added dropwise with vigorous stirring a 1.0 M solution of P4-*t*Bu base in hexane (0.48 ml, 0.48 mmol) prediluted in THF (1.0 ml) followed by a solution of the corresponding propiolate (0.48 mmol) in THF (0.5 ml). The mixture was stirred for 5 min, and the electrophile RX (5.0 mmol) was added. The temperature was slowly raised to 25 °C and stirring was maintained for 18 h. Et_2O was added, and the precipitate was filtered *in vacuo* and washed with Et_2O (3 × 2 ml). Evaporation of the solid afforded an oil which was purified by column chromatography with a hexane/Et_2O mixture (80: 20). The didehydroglutamates (colorless oils) were obtained as a mixture of epimers.

Reactions with 18-crown-6



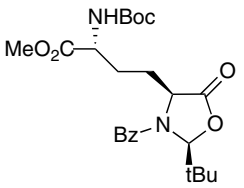
To a solution of KO*t*Bu (0.6 mmol) in THF (1.0 ml) at −78 °C was added a solution of the imino glycinate (0.48 mmol) in THF (1.0 ml) and the mixture was stirred for 30 min. A solution of the propiolate (0.6 mmol) in THF (0.5 ml) was added and the mixture was stirred for 10 min. A solution of 18-crown-6 (0.6 mmol) in THF (0.5 ml) was added and the mixture was stirred for 10 min. The corresponding electrophile RX (2.4 mmol) was added, the temperature was slowly raised to 25 °C, and stirring was continued for 18 h. Water (0.5 ml) was added, the organic layer was decanted, and the aqueous one was extracted with Et_2O (3 × 10 ml). The combined organic extracts were dried over $MgSO_4$. Evaporation of the solvent afforded an oil that was purified by column chromatography with a hexane/Et_2O mixture (80: 20). The didehydroglutamates (colorless oils) were obtained as a mixture of epimers.

1.5.6
Palladium-Catalyzed Trifold Heck Coupling [88a]



1,3,5-Triiodobenzene (0.46 g, 1.0 mmol), *N*-protected didehydroalanine (1.0 g, 3.6 mmol), $NaHCO_3$ (0.63 g, 7.5 mmol), Bu_4NCl (0.83 g, 3.0 mmol), and $Pd(OAc)_2$ (22 mg, 0.10 mmol) were mixed in DMF (5 ml) in a screwcap vial. A few crystals of hydroquinone were added to prevent polymerization of the acrylate and the mixture was freed from O_2 by N_2 bubbling for 5 min. The vial was sealed and heated at 80 °C for 24 h. The dark reaction mixture was allowed to cool and was diluted with EtOAc (50 ml). The resulting mixture was washed with water (2×50 ml) and brine (2×50 ml), and was then dried over Na_2SO_4. Evaporation of the solvent followed by flash chromatography using heptane/EtOAc (2: 1) as eluent gave the crude DDAA as a yellow semisolid (0.75 g), $R_f = 0.21$. Recrystallization twice from EtOAc/heptane yielded a pale yellow solid (0.40 g, 44%), melting point 123–131 °C, which contained about 20% (by ^{1}H-NMR) of an impurity of the same R_f. This material was used directly in the hydrogenation step.

1.5.7
General Experimental Procedure for Conjugate Addition of Alkyl iodides to Chiral α,β-Unsaturated Amino Acid Derivatives: [117c]

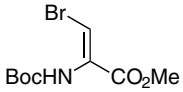


CuI (2 mmol) and zinc (6 mmol) were added to a solution of the chiral Michael acceptor **4** (1 mmol) and alkyl iodide (2–6 mmol) in aqueous EtOH (5 ml, 70%) under ultrasonic irradiation. After a few minutes, more aqueous EtOH (5 ml, 70%) was added and sonication was continued for 45 ± 90 min. In the cases where the α,β-unsaturated system was not completely consumed (thin-layer chromatography test), more CuI (1 mmol) and zinc (3 mmol) were added, and the sonication was continued for 3 h. The mixture was diluted with Et_2O (25 ml), sonicated a further

10 min, and filtered through a short pad of Celite. The solids were washed with Et_2O (3×30 ml). The organic phase was washed with brine (30 ml), dried (Na_2SO_4), filtered, and concentrated under reduced pressure (20–30 mmHg). The residue was purified by flash chromatography to afford, after concentration, the desired 1,4-addition product.

1.5.8

Bromination of *N-tert*-Butyloxycarbonyldehydroamino Acids: [171]



Boc-ΔAla-OMe (1.01 g, 5 mmol) was dissolved in dichloromethane (0.1 ml) and 1.2 equiv. of NBS was added with vigorous stirring. After reacting for 16 h, triethylamine (1.5 equiv.) was added and stirring was continued for an 1 h. The solvent was then evaporated at reduced pressure, and the residue was partitioned between dichloromethane (100 ml) and $KHSO_4$ solution (1 M, 50 ml). The organic phase was washed with $KHSO_4$ (1 M), $NaHCO_3$ (1 M), and brine (3×30 ml). After drying over $MgSO_4$ the extract was evaporated at reduced pressure to afford the (*E*)-configured brominated didehydroalanine (1.15 g, 82%) as a colorless oil.

References

1 Newton, G.G.F. and Abraham, E.P. (1956) *The Biochemical Journal*, **62**, 651–658.

2 Templeton, G.E. (1972) *Microbial Toxins*, **8**, 169–192.

3 Hata, T., Koga, F., Sano, Y., Kanamori, K., Matsumae, A., Sugawara, R., Hoshi, T., Shima, T., Ito, S. and Tomizawa, S. (1954) *Journal of Antibiotics*, **7**, 107–112.

4 Schmidt, U., Lieberknecht, A. and Wild, J. (1988) *Synthesis*, 159–172.

5 Humphrey, J.M. and Chamberlin, A.R. (1997) *Chemical Reviews*, **97**, 2243–2266.

6 Bonauer, C., Walenzyk, T. and König, B. (2006) *Synthesis*, 1–20.

7 Yun, B.-S., Hidaka, T., Furihata, K. and Seto, H. (1994) *Journal of Antibiotics*, **47**, 969–975.

8 Lau, R.C.M. and Rinehart, K.L. (1994) *Journal of Antibiotics*, **47**, 1466–1472.

9 Chen, G., Wang, G.-Y.-S., Li, X., Waters, B. and Davies, J. (2000) *Journal of Antibiotics*, **53**, 1145–1153.

10 Ranganathan, D., Shah, K. and Vaish, N. (1992) *Journal of the Chemical Society, Chemical Communications*, 1145–1147.

11 Somekh, L. and Shanzer, A. (1983) *The Journal of Organic Chemistry*, **48**, 907–908.

12 Goodall, K. and Parsons, A.F. (1995) *Tetrahedron Letters*, **36**, 3259–3260.

13 Photaki, I. (1963) *Journal of the American Chemical Society*, **85**, 1123–1126.

14 (a) Wojciechowska, H., Pawlowicz, R., Andruszkiewicz, R. and Grzybowska, J. (1978) *Tetrahedron Letters*, **19**, 4063–4064; (b) Cherney, R.J. and Wang, L. (1996) *The Journal of Organic Chemistry*, **61**, 2544–2546.

15 Miller, M.J. (1980) *The Journal of Organic Chemistry*, **45**, 3131–3132.

16 (a) Nugent, W.A. and Feaster, J.E. (1998) *Synthetic Communications*, **28**, 1617–1623; (b) Yokokawa, F. and Shioiri, T.

(2002) *Tetrahedron Letters*, **43**, 8679–8682.

17 Sai, H., Ogiku, T. and Ohmizu, H. (2003) *Synthesis*, 201–204.

18 Stohlmeyer, M.M., Tanaka, H. and Wandless, T.J. (1999) *Journal of the American Chemical Society*, **121**, 6100–6101.

19 (a) Sokolovsky, M., Sadeh, T. and Patchornik, A. (1964) *Journal of the American Chemical Society*, **86**, 1212–1217; (b) Rich, D.H. and Tam, J.P. (1975) *Tetrahedron Letters*, **16**, 211–212; (c) Rich, D.H. and Tam, J.P. (1977) *The Journal of Organic Chemistry*, **42**, 3815–3820; (d) Miao, Z. and Tam, J.P. (2000) *Organic Letters*, **2**, 3711–3713.

20 (a) Brown, R.F.C. and Meehan, G.V. (1968) *Australian Journal of Chemistry*, **21**, 1581–1589; (b) Wolfe, S., Bowers, R.J., Hasan, S.K. and Kazmaier, P.M. (1981) *Canadian Journal of Chemistry*, **59**, 406–421.

21 (a) Walter, R. and Roy, J. (1971) *The Journal of Organic Chemistry*, **36**, 2561–2563; (b) Reich, H.J., Jasperse, C.P. and Renga, J.M. (1986) *The Journal of Organic Chemistry*, **51**, 2981–2988; (c) Zhu, Y., Gieselman, M.D., Zhou, H., Averin, O. and van der Donk, W.A. (2003) *Organic and Biomolecular Chemistry*, **1**, 3304–3315; (d) Back, T.G. and Moussa, Z. (2003) *Journal of the American Chemical Society*, **125**, 13455–13460.

22 Nicolaou, K.C., Safina, B.S., Zak, M., Lee, S.H., Nevalainen, M., Bella, M., Estrada, A.A., Funke, C., Zecri, F.J. and Bulat, S. (2005) *Journal of the American Chemical Society*, **127**, 11159–11175.

23 Horikawa, E., Kodaka, M., Nakahara, Y., Okuno, H. and Nakamura, K. (2001) *Tetrahedron Letters*, **42**, 8337–8339.

24 Steiger, R.E. (1944) *The Journal of Biological Chemistry*, **153**, 691–692.

25 Shin, C., Nanjo, K., Ando, E. and Yoshimura, J. (1974) *Bulletin of the Chemical Society of Japan*, **47**, 3109–3113.

26 Herscheid, J.D.M., Scholten, H.P.H., Tijhuis, M.W. and Ottenheijm, H.C.J. (1981) *Recueil des Travaux Chimiques des Pays-Bas*, **100**, 73–78.

27 Kolasa, T. (1983) *Synthesis*, 539.

28 (a) Poisel, H. and Schmidt, U. (1975) *Chemische Berichte*, **108**, 2547–2553; (b) Kolar, A.J. and Olsen, R.K. (1977) *Synthesis*, 457–459; (c) Shimohiagashi, Y. and Stammer, C.H. (1983) *Journal of the Chemical Society, Perkin Transactions 1*, 803–808; (d) Zhu, Y.-F., Yamazaki, T., Tsang, J.W., Lok, S. and Goodman, M. (1992) *The Journal of Organic Chemistry*, **57**, 1074–1081.

29 Miossec, B., Danion-Bougot, R. and Danion, D. (1994) *Synthesis*, 1171–1174.

30 (a) Bergmann, M. and Grafe, K. (1930) *Hoppe-Seyler's Zeitschrift fur Physiologische Chemie*, **187**, 183–186; (b) Yonezawa, Y., Shin, C., Ono, Y. and Yoshimura, J. (1980) *Bulletin of the Chemical Society of Japan*, **53**, 2905–2909; (c) Labia, R. and Morin, C. (1986) *The Journal of Organic Chemistry*, **51**, 249–251.

31 Makowski, M., Rzeszotarska, B., Kubica, Z. and Pietrzynski, G. (1985) *Liebigs Annalen der Chemie*, 893–900.

32 Davies, D.T., Goodall, K., Kapur, N., O'Brien, M. and Parsons, A.F. (1997) *Synthetic Communications*, **27**, 3815–3822.

33 (a) Knittel, D. (1984) *Monatshefte fur Chemie*, **115**, 1335–1343; (b) Knittel, D. (1985) *Monatshefte fur Chemie*, **116**, 1133–1140.

34 (a) Yonezawa, Y., Shin, C., Kiyohara, M. and Yoshimura, J. (1979) *Tetrahedron Letters*, **20**, 3851–3854; (b) Fresneda, P.M., Molina, P., Delgado, S. and Bleda, J.A. (2000) *Tetrahedron Letters*, **41**, 4777–4780.

35 Manis, P.A. and Rathke, M.W. (1980) *The Journal of Organic Chemistry*, **45**, 4952–4954.

36 (a) Effenberger, F., Kühlwein, J. and Drauz, K. (1993) *Liebigs Annalen der Chemie*, 1295–1301; (b) Effenberger, F. and Beisswenger, T. (1984) *Chemische Berichte*, **117**, 1497–1512.

37 Effenberger, F., Niesert, C.-P., Kühlwein, J. and Ziegler, T. (1988) *Synthesis*, 218–219.

38 Erlenmeyer, E. (1900) *Ber*, **33**, 2036–2041.

39 (a) Gaset, A. and Gorrichon, J.P. (1982) *Synthetic Communications*, **12**, 71–79; (b) Cativiela, C., Diaz de Villegas, M.D., Mayoral, J.A. and Melendez, E. (1983) *Synthesis*, 899–902.

40 Tripathy, P.K. and Mukerjee, A.K. (1985) *Synthesis*, 285–288.

41 (a) Mukerjee, A.K. and Kumar, P. (1982) *Canadian Journal of Chemistry*, **60**, 317–322; (b) Ivanova, G.G. (1992) *Tetrahedron*, **48**, 177–186.

42 Wolfbeis, O.S. (1981) *Monatshefte fur Chemie*, **112**, 369–383.

43 Steglich, W., Höfle, G., König, W. and Weygand, F. (1968) *Chemische Berichte*, **101**, 308–322.

44 Nitz, T.J., Lindsey, J. and Stammer, C.H. (1982) *The Journal of Organic Chemistry*, **47**, 4029–4032.

45 Schmidt, U., Lieberknecht, A. and Wild, J. (1984) *Synthesis*, 53–60.

46 Kober, R. and Steglich, W. (1983) *Liebigs Annalen der Chemie*, 599–609.

47 (a) Schmidt, U. and Wild, J. (1984) *Angewandte Chemie (International Edition in English)*, **23**, 991–993; (b) Horenstein, B.A. and Nakanishi, K. (1989) *Journal of the American Chemical Society*, **111**, 6242–6246; (c) Kim, D., Li, Y., Horenstein, B.A. and Nakanishi, K. (1990) *Tetrahedron Letters*, **31**, 7119–7122; (d) Apitz, G. and Steglich, W. (1991) *Tetrahedron Letters*, **32**, 3163–3166; (e) Schmidt, U., Griesser, H., Leitenberger, V., Lieberknecht, A., Mangold, R., Meyer, R. and Riedl, B. (1992) *Synthesis*, 487–490.

48 (a) Ritzen, A., Basu, B., Wallberg, A. and Frejd, T. (1988) *Tetrahedron – Asymmetry*, **9**, 3491–3496; (b) Schmidt, U., Lieberknecht, A., Kazmaier, U., Griesser, H., Jung, G. and Metzger, J. (1991) *Synthesis*, 49–55; (c) Coleman, R.S. and Carpenter, A.J. (1992) *The Journal of Organic Chemistry*, **57**, 5813–5815; (d) Coleman, R.S. and Carpenter, A.J. (1993) *The Journal of Organic Chemistry*, **58**, 4452–4461; (e) Schmidt, U., Braun, C. and Sutoris, H. (1996) *Synthesis*, 223–229; (f) Jorgensen, K.B. and Gautun, O.R. (1999) *Tetrahedron*, **55**, 10527–10536; (g) Adamczyk, M., Akireddy, S.R. and Reddy, R.E. (2001) *Organic Letters*, **3**, 3157–3159; (h) Wang, W., Xiong, C., Yang, J. and Hruby, V.J. (2002) *Synthesis*, 94–98; (i) Krause, N., Hoffmann-Röder, A. and Canisius, J. (2002) *Synthesis*, 1759–1774.

49 (a) Schmidt, U., Weller, D., Holder, A. and Lieberknecht, A. (1988) *Tetrahedron Letters*, **29**, 3227–3230; (b) Schmidt, U., Stäbler, F. and Lieberknecht, A. (1992) *Synthesis*, 482–486; (c) Schmidt, U., Leitenberger, V., Griesser, G., Schmidt, J. and Meyer, R. (1992) *Synthesis*, 1248–1254; (d) Schmidt, U., Kleefeldt, A. and Mangold, R. (1992) *Journal of the Chemical Society, Chemical Communications*, 1687–1689; (e) Zhang, J., Xiong, C., Wang, W., Ying, J. and Hruby, V.J. (2002) *Organic Letters*, **4**, 4029–4032; (f) Yamada, K., Kurokawa, T., Tokuyama, H. and Fukuyama, T. (2003) *Journal of the American Chemical Society*, **125**, 6630–6631; (g) Okano, K., Tokuyama, H. and Fukuyama, T. (2006) *Journal of the American Chemical Society*, **128**, 7136–7137.

50 Schmidt, U. and Wild, J. (1985) *Liebigs Annalen der Chemie*, 1882–1894.

51 Schmidt, U. and Riedl, B. (1992) *Journal of the Chemical Society, Chemical Communications*, 1186–1187.

52 Bentley, D.J., Slawin, A.M.Z. and Moody, C.J. (2006) *Organic Letters*, **8**, 1975–1978.

53 (a) Buck, R.T., Clarke, P.A., Coe, D.M., Drysdale, M.J., Ferris, L., Haigh, D., Moody, C.J., Pearson, N.D. and Swann, E. (2000) *Chemistry – A European Journal*, **6**, 2160–2167; (b) Bentley, D.J. and Moody, C.J. (2004) *Organic and Biomolecular Chemistry*, **2**, 3545–3547.

54 (a) Molina, P., Tarraga, A., Curiel, D. and De Arellano, C.R. (1997) *Tetrahedron*, **53**, 15895–15902; (b) Renard, A., Kotera, M.,

Brochier, M.-C. and Lhomme, J. (2000) *European Journal of Organic Chemistry*, 1831–1840.

55 (a) Williams, R.M. and Fegley, G.J. (1991) *Journal of the American Chemical Society*, **113**, 8796–8806; (b) Williams, R.M. (1992) *Aldrichimica Acta*, **25**, 11–23; (c) Williams, R.M. and Fegley, G.J. (1992) *Tetrahedron Letters*, **33**, 6755–6758; (d) Williams, R.M., Fegley, G.J., Gallegos, R., Schaefer, F. and Pruess, D.L. (1996) *Tetrahedron*, **52**, 1149–1164.

56 (a) Chai, C.L.L. and King, A.R. (1995) *Tetrahedron Letters*, **36**, 4295–4298; (b) Chai, C.L.L. and King, A.R. (1997) *Journal of Chemical Research-S*, 382.

57 Mazurkiewicz, R., Grymel, M. and Kuźnik, A. (2004) *Monatshefte fur Chemie*, **135**, 799–806.

58 Mazurkiewicz, R. and Grymel, M. (1999) *Monatshefte fur Chemie*, **130**, 597–604.

59 Mazurkiewicz, R., Kuźnik, A., Grymel, M. and Kuźnik, N. (2004) *Monatshefte fur Chemie*, **135**, 807–815.

60 (a) Ernest, I. (1980) *Helvetica Chimica Acta*, **63**, 201–213; (b) Bateson, J.H., Robins, A.M. and Southgate, R. (1991) *Journal of the Chemical Society, Perkin Transactions 1*, 2399–2405; (c) Coulton, S. and Francois, I. (1991) *Journal of the Chemical Society, Perkin Transactions 1*, 2699–2705; (d) Dunlap, N.K., Dezube, M., Keith, D.D. and Weigele, M. (1992) *Tetrahedron Letters*, **33**, 6103–6106.

61 Schumann, S., Zeitler, K., Jäger, M., Polborn, K. and Steglich, W. (2000) *Tetrahedron*, **56**, 4187–4195.

62 Umezawa, S. and Zen, S. (1963) *Bulletin of the Chemical Society of Japan*, **36**, 1143–1145.

63 Dornow, A. and Menzel, H. (1954) *Liebigs Annalen der Chemie*, 40–44.

64 Wolfbeis, O. (1977) *Chemische Berichte*, **110**, 2480–2493.

65 Severin, T. and Böhme, H.-J. (1968) *Chemische Berichte*, **101**, 2925–2930.

66 (a) Lehnert, W. (1972) *Tetrahedron*, **28**, 663–666; (b) Fornicola, R.S., Oblinger, E. and Montgomery, J. (1998) *The Journal of Organic Chemistry*, **63**, 3528–3529.

67 (a) Versleijen, J.P.G., van Leusen, A.M. and Feringa, B.L. (1999) *Tetrahedron Letters*, **40**, 5803–5806; (b) Srikanth, G.S.C. and Castle, S.L. (2004) *Organic Letters*, **6**, 449–452; (c) He, L., Srikanth, G.S.C. and Castle, S.L. (2005) *The Journal of Organic Chemistry*, **70**, 8140–8147.

68 (a) Süs, O. (1948) *Liebigs Annalen der Chemie*, 92–101; (b) Shin, C.C., Masaki, M. and Ohta, M. (1967) *The Journal of Organic Chemistry*, **32**, 1860–1863.

69 Süs, O. (1948) *Liebigs Annalen der Chemie*, 31–46.

70 Süs, O., Schafer, W. and Grundkotter, M. (1951) *Liebigs Annalen der Chemie*, 201–225.

71 (a) Horwell, D.C., Nichols, P.D., Ratcliffe, G.S. and Roberts, E. (1994) *The Journal of Organic Chemistry*, **59**, 4418–4423; (b) Hammadi, A., Menez, A. and Genet, R. (1996) *Tetrahedron Letters*, **37**, 3309–3312.

72 Schöllkopf, U. and Meyer, R. (1981) *Liebigs Annalen der Chemie*, 1469–1475.

73 (a) Wüst, H.H., Bardenhagen, J. and Schöllkopf, U. (1985) *Liebigs Annalen der Chemie*, 1825–1837; (b) Stanfield, C.F. and Hruby, V.J. (1988) *Synthetic Communications*, **18**, 531–543; (c) Enders, D., Chen, Z.-X. and Raabe, G. (2005) *Synthesis*, 306–310.

74 Schöllkopf, U., Gerhart, F., Schröder, R. and Hoppe, D. (1972) *Liebigs Annalen der Chemie*, 116–129.

75 Ito, Y., Matsuura, T. and Saegusa, T. (1985) *Tetrahedron Letters*, **26**, 5781–5784.

76 Kantlehner, W., Wagner, F. and Bredereck, H. (1980) *Liebigs Annalen der Chemie*, 344–357.

77 Schöllkopf, U., Porsch, P.H. and Lau, H.H. (1979) *Liebigs Annalen der Chemie*, 1444–1446.

78 Rachon, J. and Schöllkopf, U. (1981) *Liebigs Annalen der Chemie*, 99–102.

79 Nunami, K., Hiramatsu, K., Hayashi, K. and Matsumoto, K. (1988) *Tetrahedron*, **44**, 5467–5478.

80 (a) Moran, E.J. and Armstrong, R.W. (1991) *Tetrahedron Letters*, **32**, 3807–3810; (b) Kim, S.W., Bauer, S.M. and Armstrong, R.W. (1998) *Tetrahedron Letters*, **39**, 7031–7034; (c) Bienayme, H. (1998) *Tetrahedron Letters*, **39**, 4255–4258.

81 Kolb, J., Beck, B. and Domling, A. (2002) *Tetrahedron Letters*, **43**, 6897–6901.

82 Reviews: (a) O'Donnell, M.J. (2001) *Aldrichimica Acta*, **34**, 3–15; (b) Soloshonok, V.A. (2002) *Current Organic Chemistry*, **6**, 341–364.

83 (a) Alvarez-Ibarra, C., Csáky, A.G., Ortega, E.M., De la Morena, M.J. and Quiroga, M.L. (1997) *Tetrahedron Letters*, **38**, 4501–4502; (b) Alvarez-Ibarra, C., Csáky, A.G., Martin, M.E. and Quiroga, M.L. (1999) *Tetrahedron*, **55**, 7319–7330.

84 Alvarez-Ibarra, C., Csáky, A.G. and de la Oliva, C.G. (2002) *The Journal of Organic Chemistry*, **67**, 2789–2797.

85 Schwesinger, R. and Schlemper, H. (1987) *Angewandte Chemie (International Edition in English)*, **26**, 1167–1169.

86 (a) Cutolo, M., Fiandanese, V., Naso, F. and Sciacovelli, O. (1983) *Tetrahedron Letters*, **24**, 4603–4606; (b) Carlström, A.S. and Frejd, T. (1989) *Synthesis*, 414–418; (c) Bozell, J.J., Vogt, C.E. and Gozum, J. (1991) *The Journal of Organic Chemistry*, **56**, 2584–2587; (d) Fairlamb, I.J.S., Kapdi, A.R., Lee, A.F., McGlacken, G.P., Weissburger, F., de Vries, A.H.M. and Schmieder-van de Vondervoort, L. (2006) *Chemistry – A European Journal*, **12**, 8750–8761.

87 (a) Gallou-Dagommer, I., Gastaud, P. and Rajanbabu, T.V. (2001) *Organic Letters*, **3**, 2053–2056; (b) Li, T., Tsuda, Y., Minoura, K., In, Y., Ishida, T., Lazarus, L.H. and Okada, Y. (2006) *Chemical & Pharmaceutical Bulletin*, **54**, 873–877.

88 (a) Ritzén, A., Basu, B., Wallberg, A. and Frejd, T. (1998) *Tetrahedron – Asymmetry*, **9**, 3491–3496; (b) Ritzén, A. and Frejd, T. (1999) *Journal of the Chemical Society, Chemical Communications*, 207–208; (c) Ritzén, A. and Frejd, T. (2000) *European Journal of Organic Chemistry*, 3771–3782.

89 Gibson, S.E., Jones, J.O., Kalindjian, S.B., Knight, J.D., Steed, J.W. and Tozer, M.J. (2002) *Chemical Communications*, 1938–1939.

90 Tullberg, E., Schacher, F., Peters, D. and Frejd, T. (2006) *Synthesis*, 1183–1189.

91 Yamazaki, K., Nakamura, Y. and Kondo, Y. (2003) *The Journal of Organic Chemistry*, **68**, 6011–6019.

92 (a) Collier, P.N., Patel, I. and Taylor, R.J.K. (2001) *Tetrahedron Letters*, **42**, 5953–5954; (b) Collier, P.N., Campbell, A.D., Patel, I., Raynham, T.M. and Taylor, R.J.K. (2002) *The Journal of Organic Chemistry*, **67**, 1802–1815.

93 Navarre, L., Darses, S. and Genet, J.-P. (2004) *European Journal of Organic Chemistry*, 69–73.

94 (a) Burk, M.-J., Allen, J.G., Kiesman, W.F. and Stoffan, K.M. (1997) *Tetrahedron Letters*, **38**, 1309–1312; (b) Hoerrner, R.S., Askin, D., Volante, R.P. and Reider, P.J. (1998) *Tetrahedron Letters*, **39**, 3455–3458; (c) Burk, M.J., Allen, J.G. and Kiesman, W.F. (1998) *Journal of the American Chemical Society*, **120**, 657–663; (d) Zhang, J., Xiong, C., Wang, W., Ying, J. and Hruby, V.J. (2002) *Organic Letters*, **4**, 4029–4032; (e) Chapman, C.J. and Frost, C.G. (2003) *Advanced Synthesis and Catalysis*, **345**, 353–355.

95 Roff, G.J., Lloyd, R.C. and Turner, N.J. (2004) *Journal of the American Chemical Society*, **126**, 4098–4099.

96 Narukawa, Y., Nishi, K. and Onoue, H. (1997) *Tetrahedron*, **53**, 539–556.

97 Abreu, A.S., Ferreira, P.M.T., Queiroz, M.-J.R.P., Ferreira, I.C.F.R., Calhelha, R.C. and Estevinho, L.M. (2005) *European Journal of Organic Chemistry*, 2951–2957.

98 Stille, J.K. (1986) *Angewandte Chemie (International Edition in English)*, **25**, 508–523.

99 Cook, G.K., Hornback, W.J., Jordan, C.L., McDonald, J.H. III and Munroe, J.E.

(1989) *The Journal of Organic Chemistry*, **54**, 5828–5830.

100 (a) Buynak, J.D., Vogeti, L. and Chen, H. (2001) *Organic Letters*, **3**, 2953–2956; (b) Buynak, J.D., Vogeti, L., Doppalapudi, V.R., Solomon, G.M. and Chen, H. (2002) *Bioorganic & Medicinal Chemistry Letters*, **12**, 1663–1666.

101 Abreu, A.S., Ferreira, P.M.T., Queiroz, M.-J.R.P., Gatto, E. and Venanzi, M. (2004) *European Journal of Organic Chemistry*, 3985–3991.

102 Bruncko, M. and Crich, D. (1994) *The Journal of Organic Chemistry*, **59**, 4239–4249.

103 Probert, J.M., Rennex, D. and Bradley, M. (1996) *Tetrahedron Letters*, **37**, 1101–1104.

104 Yokoyama, Y., Hikawa, H., Mitsuhashi, M., Uyama, A., Hiroki, Y. and Murakami, Y. (2004) *European Journal of Organic Chemistry*, 1244–1253.

105 (a) Jin, S. and Liebscher, J. (1999) *Synlett*, 459–461; (b) Jin, S., Wessig, P. and Liebscher, J. (2000) *European Journal of Organic Chemistry*, 1993–1999; (c) de la Hoz, A., Diaz-Ortiz, A., Gomez, M.V., Mayoral, J.A., Moreno, A., Sanchez-Migallon, A.M. and Vazquez, E. (2001) *Tetrahedron*, **57**, 5421–5428.

106 Ferreira, P.M.T., Maia, H.L.S., Monteiro, L.S., Sacramento, J. and Sebastião, J. (2000) *Journal of the Chemical Society, Perkin Transactions 1*, 3317–3324.

107 (a) Ferreira, P.M.T., Maia, H.L.S. and Monteiro, L.S. (2003) *European Journal of Organic Chemistry*, 2635–2644; (b) Abreu, A.S., Silva, N.O., Ferreira, P.M.T., Queiroz, M.-J.R.P. (2003) *European Journal of Organic Chemistry*, 1537–1544.

108 (a) Ezquerra, J., Escribano, A., Rubio, A., Remuinan, M.J. and Vaquero, J.J. (1996) *Tetrahedron – Asymmetry*, **7**, 2613–2626; (b) Dugave, C., Cluzeau, J., Menez, A., Gaudry, M. and Marquet, A. (1998) *Tetrahedron Letters*, **39**, 5775–5778; (c) Crossley, M.J., Fung, Y.M., Potter, J.J. and Stamford, A.W. (1998) *Journal of the Chemical Society, Perkin Transactions 1*, 1113–1122; (d) Ballini, R., Balsamini, C., Diamantini, G. and Savoretti, N. (2005) *Synthesis*, 1055–1057.

109 Guevel, A.-C. and Hart, D.J. (1996) *The Journal of Organic Chemistry*, **61**, 473–479.

110 Pyne, S.G., Javidan, A., Skelton, B.W. and White, A.H. (1995) *Tetrahedron*, **51**, 5157–5168.

111 Zimmermann, J. and Seebach, D. (1987) *Helvetica Chimica Acta*, **70**, 1104–1114.

112 Javidan, A., Schafer, K. and Pyne, S.G. (1997) *Synlett*, 100–102.

113 Cardellicchio, C., Fiandanese, V., Marchese, G., Naso, F. and Ronzini, L. (1985) *Tetrahedron Letters*, **26**, 4387–4390.

114 (a) Seebach, D., Bürger, H.M. and Schickli, C.P. (1991) *Liebigs Annalen der Chemie*, 669–684; (b) Toyooka, N., Tanaka, K., Momose, T., Daly, J.W. and Garraffo, H.M. (1997) *Tetrahedron*, **53**, 9553–9574; (c) Bull, S.D., Davies, S.G. and O'Shea, M.D. (1998) *Journal of the Chemical Society, Perkin Transactions 1*, 3657–3658; (d) Toyooka, N., Okumura, M. and Nemoto, H. (2002) *The Journal of Organic Chemistry*, **67**, 6078–6081.

115 Cardillo, G., Gentilucci, L., Tolomelli, A. and Tomasini, C. (1999) *Tetrahedron*, **55**, 6231–6242.

116 (a) Petrier, C., Dupuy, C. and Luche, J.L. (1986) *Tetrahedron Letters*, **27**, 3149–3152; (b) Luche, J.L. and Allavena, C. (1988) *Tetrahedron Letters*, **29**, 5369–5372.

117 (a) Sestelo, J.P., de Una, O., Mourino, A. and Sarandeses, L.A. (2002) *Synlett*, 719–722; (b) Huang, T., Keh, C.C.K., Li, C.-J. (2002) *Chemical Communications*, 2440–2441; (c) Suarez, R.M., Sestelo, J.P. and Sarandeses, L.A. (2003) *Chemistry – A European Journal*, **9**, 4179–4187.

118 (a) Ahmad, S., Phillips, R.S. and Stammer, C.H. (1992) *Journal of Medicinal Chemistry*, **35**, 1410–1417; (b) Cativiela, C., Diaz-de-Villegas, M.D. and Jimenez, A.I. (1994) *Tetrahedron*, **50**, 9157–9166; (c) Matsumura, Y., Inoue, M., Nakamura, Y., Talib, I.L., Maki, T. and Onomura, O. (2000) *Tetrahedron Letters*, **41**, 4619–4622.

119 Williams, R.M. and Fegley, G.J. (1993) *The Journal of Organic Chemistry*, **58**, 6933–6935.

120 Calmes, M., Daunis, J. and Escale, F. (1996) *Tetrahedron – Asymmetry*, **7**, 395–396.

121 Chinchilla, R., Falvello, L.R., Galindo, N. and Najera, C. (1998) *Tetrahedron – Asymmetry*, **9**, 2223–2227.

122 Bunuel, E., Bull, S.D., Davies, S.G., Garner, A.C., Savory, E.D., Smith, A.D., Vickers, R.J. and Watkin, D.J. (2003) *Organic and Biomolecular Chemistry*, **1**, 2531–2542.

123 Cativiela, C., Diaz-de-Villegas, M.D. and Jimenez, A.I. (1995) *Tetrahedron*, **51**, 3025–3032.

124 Yim, A.M., Vidal, Y., Viallefont, P. and Martinez, J. (1999) *Tetrahedron Letters*, **40**, 4535–4538.

125 (a) Gibson, S.E., Guillo, N. and Tozer, M.J. (1997) *Chemical Communications*, 637–638; (b) Horneman, A.M. and Lundt, I. (1998) *The Journal of Organic Chemistry*, **63**, 1919–1928.

126 (a) Goodall, K. and Parsons, A.F. (1996) *Tetrahedron*, **52**, 6739–6758; (b) Tamura, O., Matsukida, H., Toyao, A., Takeda, Y. and Ishibashi, H. (2002) *The Journal of Organic Chemistry*, **67**, 5537–5545.

127 Baker, S.R., Parsons, A.F. and Wilson, M. (1998) *Tetrahedron Letters*, **39**, 2815–2818.

128 (a) Baker, S.R., Parsons, A.F., Pons, J.-F. and Wilson, M. (1998) *Tetrahedron Letters*, **39**, 7197–7200; (b) Baker, S.R., Burton, K.I., Parsons, A.F., Pons, J.-F. and Wilson, M. (1999) *Journal of the Chemical Society, Perkin Transactions 1*, 427–436.

129 Parrish, J.D. and Little, R.D. (2002) *Organic Letters*, **4**, 1439–1442.

130 Miyabe, H., Asada, R., Yoshida, K. and Takemoto, Y. (2004) *Synlett*, 540–542.

131 Sutherland, A. and Vederas, J.C. (2002) *Chemical Communications*, 224–225.

132 (a) Axon, J.R. and Beckwith, A.L.J. (1995) *Journal of the Chemical Society, Chemical Communications*, 549–550; (b) Pyne, S.G. and Schafer, K. (1998) *Tetrahedron*, **54**, 5709–5720; (c) Jones, R.C.F., Berthelot, D.J.C. and Iley, J.N. (2001) *Tetrahedron*, **57**, 6539–6555.

133 Beckwith, A.L.J. and Chai, C.L.L. (1990) *Journal of the Chemical Society, Chemical Communications*, 1087–1088.

134 Kabat, M.M. (2001) *Tetrahedron Letters*, **42**, 7521–7524.

135 Yim, A.-M., Vidal, Y., Viallefont, P. and Martinez, J. (2002) *Tetrahedron – Asymmetry*, **13**, 503–510.

136 Sibi, M.P., Asano, Y. and Sausker, J.B. (2001) *Angewandte Chemie International Edition*, **40**, 1293–1296.

137 Cativiela, C. and Diaz-de-Villegas, M.D. (2000) *Tetrahedron – Asymmetry*, **11**, 645–732.

138 (a) Zhu, Y.F., Yamazaki, T., Tsang, J.W., Lok, S. and Goodman, M. (1992) *The Journal of Organic Chemistry*, **57**, 1074–1081; (b) Adams, L.A., Charmant, J.P.H., Cox, R.J., Walter, M. and Whittingham, W.G. (2004) *Organic and Biomolecular Chemistry*, **2**, 542–553.

139 (a) Alami, A., Calmes, M., Daunis, J., Escale, F., Jacquier, R., Roumestant, M.L. and Viallefont, P. (1991) *Tetrahedron – Asymmetry*, **2**, 175–178; (b) Cativiela, C., Diaz-de-Villegas, M.D., Garcia, J.I. and Jimenez, A.I. (1997) *Tetrahedron*, **53**, 4479–4486; (c) Rife, J. and Ortuno, R.M. (1999) *Tetrahedron – Asymmetry*, **10**, 4245–4260.

140 Alcaraz, C., Fernandez, M.D., de Frutos, M.P., Marco, J.L., Bernabe, M., Foces-Foces, C. and Cano, F.H. (1994) *Tetrahedron*, **50**, 12443–12456.

141 (a) Aggarwal, V.K., Alonso, E., Fang, G.Y., Ferrara, M., Hynd, G. and Porcelloni, M. (2001) *Angewandte Chemie International Edition*, **40**, 1433–1436; (b) Adams, L.A., Aggarwal, V.K., Bonnert, R.V., Bressel, B., Cox, R.J., Shepherd, J., de Vicente, J., Walter, M., Whittingham, W.G. and Winn, C.L. (2003) *The Journal of Organic Chemistry*, **68**, 9433–9440.

142 (a) Pyne, S.G., Safaei-G., J. and Koller, F. (1995) *Tetrahedron Letters*, **36**, 2511–2514;

(b) Pyne, S.G., Safaei-G., J., Schafer, A.K., Javidan, A., Skelton, B.W. and White, A.H. (1998) *Australian Journal of Chemistry*, **51**, 137–158.

143 Pyne, S.G., Safaei-G., J., Skelton, B.W. and White, A.H. (1995) *Australian Journal of Chemistry*, **48**, 1511–1533.

144 Chai, C.L.L., Edwards, A.J., Wilkes, B.A. and Woodgate, R.C.J. (2003) *Tetrahedron*, **59**, 8731–8739.

145 (a) Pyne, S.G., Schafer, K., Skelton, B.W. and White, A.H. (1997) *Chemical Communications*, 2267–2268; (b) Ung, A.T., Schafer, K., Lindsay, K.B., Pyne, S.G., Amornraksa, K., Wouters, R., van der Linden, I., Biesmans, I., Lesage, A.S.J., Skelton, B.W. and White, A.H. (2002) *The Journal of Organic Chemistry*, **67**, 227–233; (c) Pham, T.Q., Pyne, S.G., Skelton, B.W. and White, A.H. (2005) *The Journal of Organic Chemistry*, **70**, 6369–6377.

146 (a) Trost, B.M. and Kazmaier, U. (1992) *Journal of the American Chemical Society*, **114**, 7933–7935; (b) Methot, J.L. and Roush, W.R. (2004) *Advanced Synthesis and Catalysis*, **346**, 1035–1050.

147 (a) Cativiela, C., Lopez, P. and Mayoral, J.A. (1991) *Tetrahedron – Asymmetry*, **2**, 449–456; (b) Pyne, S.G., Dikic, B., Gordon, P.A., Skelton, B.W. and White, A.H. (1991) *Journal of the Chemical Society, Chemical Communications*, 1505–1506; (c) Bunuel, E., Cativiela, C. and Diaz-de-Villegas, M.D. (1995) *Tetrahedron*, **51**, 8923–8934; (d) Chinchilla, R., Falvello, L.R., Galindo, N. and Najera, C. (1999) *Tetrahedron – Asymmetry*, **10**, 821–825; (e) Chai, C.L.L., Johnson, R.C. and Koh, J. (2002) *Tetrahedron*, **58**, 975–982.

148 Avenoza, A., Cativiela, C., Busto, J.H., Fernandez-Recio, M.A., Peregrina, J.M. and Rodriguez, F. (2001) *Tetrahedron*, **57**, 545–548.

149 Cativiela, C., Lopez, P. and Mayoral, J.A. (1991) *Tetrahedron – Asymmetry*, **2**, 1295–1304.

150 (a) Fan, Q.-H., Li, Y.-M. and Chan, A.S.C. (2002) *Chemical Reviews*, **102**, 3385–3465; (b) Noyori, R. (2002) *Angewandte Chemie International Edition*, **41**, 2008–2022; (c) Chelucci, G., Orru, G. and Pinna, G.A. (2003) *Tetrahedron*, **59**, 9471–9515; (d) Genet, J.-P. (2003) *Accounts of Chemical Research*, **36**, 908–918; (e) Tang, W. and Zhang, X. (2003) *Chemical Reviews*, **103**, 3029–3069; (f) Jäkel, C. and Paciello, R. (2006) *Chemical Reviews*, **106**, 2912–2942.

151 Osborn, J.A., Jardine, F.H., Young, J.F. and Wilkinson, G. (1966) *Journal of the Chemical Society, A: Inorganic, Physical, Theoretical*, 1711–1732.

152 Horner, L., Siegel, H. and Büthe, H. (1968) *Angewandte Chemie (International Edition in English)*, **7**, 942–943.

153 Knowles, W.S. and Sabacky, M.J. (1968) *Chemical Communications*, 1445–1446.

154 (a) Vineyard, B.D., Knowles, W.S., Sabacky, M.J., Bachman, G.L. and Weinkauff, D.J. (1977) *Journal of the American Chemical Society*, **99**, 5946–5952; (b) Knowles, W.S. (1983) *Accounts of Chemical Research*, **16**, 106–112.

155 Knowles, W.S. (1986) *Journal of Chemical Education*, **63**, 222–225.

156 Knowles, W.S. (2002) *Angewandte Chemie International Edition*, **41**, 1998–2007.

157 (a) Schmidt, U., Lieberknecht, A., Griesser, H. and Bartkowiak, F. (1984) *Angewandte Chemie (International Edition in English)*, **23**, 318–320; (b) Schmidt, U., Leitenberger, V., Meyer, R. and Griesser, H. (1992) *Journal of the Chemical Society, Chemical Communications*, 951–953.

158 (a) Poulin, J.C. and Kagan, H.B. (1982) *Journal of the Chemical Society, Chemical Communications*, 1261–1262; (b) Carlström, A.S. and Frejd, T. (1991) *Journal of the Chemical Society, Chemical Communications*, 1216–1217; (c) Nagel, U. and Krink, T. (1993) *Angewandte Chemie (International Edition in English)*, **32**, 1052–1054; (d) Ghosh, A.K. and Wang, Y. (1999) *Journal of the Chemical Society, Perkin Transactions 1*, 3597–3601.

159 (a) Teoh, E., Campi, E.M., Jackson, W.R. and Robinson, A.J. (2002) *Chemical*

Communications, 978–979; (b) Teoh, E., Campi, E.M., Jackson, W.R. and Robinson, A.J. (2003) *New Journal of Chemistry*, **27**, 387–394.

160 (a) Borszeky, K., Mallat, T. and Baiker, A. (1997) *Tetrahedron – Asymmetry*, **8**, 3745–3753; (b) Blaser, H.U., Hönig, H., Studer, M., Wedemeyer-Exl, C. (1999) *Journal of Molecular Catalysis A – Chemical*, **139**, 253–257; (c) Sugimura, T., Watanabe, J., Okuyama, T. and Nitta, Y. (2005) *Tetrahedron – Asymmetry*, **16**, 1573–1575; (d) Szöllösi, G., Szabo, E. and Bartok, M. (2007) *Advanced Synthesis and Catalysis*, **349**, 405–410.

161 (a) Fan, Q.-H., Deng, G.-J., Lin, C.-C. and Chan, A.S.C. (2001) *Tetrahedron – Asymmetry*, **12**, 1241–1247; (b) Hu, X.-P., Huang, J.-D., Zeng, Q.-H. and Zheng, Z. (2006) *Chemical Communications*, 293–295.

162 Nagel, U. and Leipold, J. (1996) *Chemische Berichte*, **129**, 815–821.

163 (a) Cativiela, C., Diaz-de-Villegas, M.D. and Galvez, J.A. (1992) *Tetrahedron – Asymmetry*, **3**, 567–572; (b) Horti, A., Redmond, D.E. Jr and Soufer, R. (1995) *Journal of Labelled Compounds & Radiopharmaceuticals*, **36**, 409–423; (c) Kassem, T., Wehbe, J., Rolland-Fulcrand, V., Rolland, M., Roumestant, M.L. and Martinez, J. (2001) *Tetrahedron – Asymmetry*, **12**, 2657–2661.

164 Vigneron, J.P., Kagan, H. and Horeau, A. (1968) *Tetrahedron Letters*, **9**, 5681–5683.

165 (a) Poisel, H. and Schmidt, U. (1973) *Chemische Berichte*, **106**, 3408–3420; (b) Herscheid, J.D.M., Nivard, R.J.F., Tijhuis, M.W. and Ottenheijm, H.C.J. (1980) *The Journal of Organic Chemistry*, **45**, 1880–1885; (c) Davies, S.G., Rodriguez-Solla, H., Tamayo, J.A., Cowley, A.R., Concellon, C., Garner, A.C., Parkes, A.L. and Smith, A.D. (2005) *Organic and Biomolecular Chemistry*, **3**, 1435–1447.

166 Lisichkina, I.N., Ushakova, O.M., Alekseeva, M.O., Peregudov, A.S. and Belikov, V.M. (1999) *Russian Chemical Bulletin*, **48**, 1682–1684.

167 (a) Onuma, K., Ito, T. and Nakamura, A. (1980) *Chemistry Letters*, 481–482; (b) Yamagishi, T., Yatagai, M., Hatakeyama, H. and Hida, M. (1984) *Bulletin of the Chemical Society of Japan*, **57**, 1897–1901; (c) El-Baba, S., Nuzillard, J.M., Poulin, J.C. and Kagan, H.B. (1986) *Tetrahedron*, **42**, 3851–3861.

168 Schmidt, U., Kumpf, S. and Neumann, K. (1994) *Journal of the Chemical Society, Chemical Communications*, 1915–1916.

169 Nakatsuka, S., Tanino, H. and Kishi, Y. (1975) *Journal of the American Chemical Society*, **97**, 5008–5010.

170 Danion-Bougot, R., Danion, D. and Francis, G. (1990) *Tetrahedron Letters*, **31**, 3739–3742.

171 Silva, N.O., Abreu, A.S., Ferreira, P.M.T., Monteiro, L.S. and Queiroz, M.-J.R.P. (2002) *European Journal of Organic Chemistry*, 2524–2528.

172 Richards, K.D., Kolar, A.J., Srinivasan, A., Stephenson, R.W. and Olsen, R.K. (1976) *The Journal of Organic Chemistry*, **41**, 3674–3677.

173 (a) Zhou, H. and van der Donk, W.A. (2001) *Organic Letters*, **3**, 593–596; (b) Zhou, H. and van der Donk, W.A. (2002) *Organic Letters*, **4**, 1335–1338.

2
Synthesis and Chemistry of α-Hydrazino Acids

Joëlle Vidal

2.1
Introduction

α-Hydrazino acids are described by the general formula **1** and have a close structural relationship to the naturally occurring α-amino acids, when $R^a = R^b = R^c = R^d = H$. Thus, the usual names of α-amino acids are used for the trivial nomenclature of **1**. For instance, 2-hydrazino propanoic acid **2** is named *N*-amino alanine and abbreviated NH_2-Ala-OH. Sometimes, this compound **2** is also called hydrazinoalanine in the literature. According to the International Union of Pure and Applied Chemistry nomenclature of hydrazines and derivatives, the nitrogen atom attached to the α-carbon is numbered 1 and the other 2 (Figure 2.1).

The first synthesis of an α-hydrazino acid, 2-hydrazino-2-methylpropanoic acid, was reported in 1894 [1]. The real basis of the synthesis and chemistry of α-hydrazino acids was established by Darapsky between 1916 and 1936 [2–6]. During that time, the racemic amino homologs of α-amino acids NH_2-Ala, NH_2-Val, NH_2-Leu, and NH_2-Phe, and the two enantiomers of *N*-aminophenylglycine were described.

The chemistry and synthesis of α-hydrazino acids have developed significantly because the N−N−C−C=O motif is found in a lot of potent biologically active compounds and it also induces a turn conformation in pseudopeptides. For instance, carbidopa **3** is a drug used in the treatment of Parkinson's disease. Furthermore, α-hydrazino acid derivatives are important intermediates in the asymmetric synthesis of α-amino acids. The synthesis and chemistry of α-hydrazino acids were partially reviewed by Marraud and Vanderesse [7].

2.1.1
α-Hydrazino Acids are Potent Inhibitors of Pyridoxal Phosphate Enzymes

At the beginning of the 1960s, α-hydrazino acids were shown to be inhibitors of pyridoxal phosphate enzymes. Merck laboratories discovered that (±)-carbidopa [8], L-carbidopa **3** (Figure 2.1) [9], and (±)-β–(5-hydroxy-3-indoyl)-α-hydrazinopropanoic

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

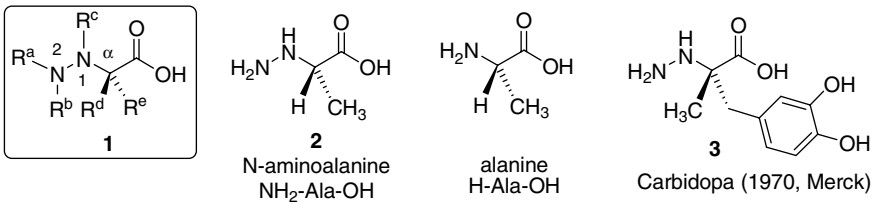


Figure 2.1 General formula for α-hydrazino acids and notation.

acid [10] were very potent inhibitors of 3,4-dihydroxyphenylalanine (DOPA) decarboxylase. Nowadays, L-carbidopa **3** is still an important drug for the treatment of Parkinson's disease, in association with L-DOPA. More than 2300 references are associated with it in the MEDLINE database. As it cannot cross the blood–brain barrier, it inhibits only peripheral DOPA decarboxylase, thus preventing the conversion of L-DOPA to dopamine peripherally and increasing the plasma half-life of L-DOPA targeted to the brain. The mechanism of inhibition is the formation of a hydrazone linkage between the hydrazine function and the aldehyde function of the coenzyme [11]. Other pyridoxal phosphate enzymes such as ornithine decarboxylase ((±)-N^{α}-aminoornithine [12–16]), histidine decarboxylase (L-N^{α}-aminohistidine [17, 18]), mesodiaminopimalate decarboxylase (2-hydrazino-6-amino-pimelic acid [19, 20]), aspartate aminotransferase (D- and L-2-hydrazinosuccinic acid [21, 22]), or γ-aminobutyric acid (GABA) transaminases (*N*-aminoglycine [23]) are also inhibited by α-hydrazino acids.

Modest anticancer activity of α-hydrazino acids was also reported [24].

2.1.2
Natural Products Containing the N–N–C–C=O Fragment

To the best of our knowledge, free α-hydrazino acids **1** are not found in nature, but several natural compounds contain the N–N–C–C=O fragment. Linatine is an N^2-acylated D-aminoproline derivative isolated from flax seeds (Figure 2.2). It behaves as a vitamin B_6 antagonist and is responsible for the poor growth of chickens or turkeys reared on linseed meal [25]. Negamycine is a dipeptide composed of the *N*-aminosarcosine residue and an unusual ε,β-diamino acid. It was isolated from *Streptomyces purpeofuscus* [26] and displays interesting biological activities. It is active against multidrug-resistant Gram-negative bacteria [26], it exhibits very low acute toxicity, and it also inhibits protein synthesis [27].

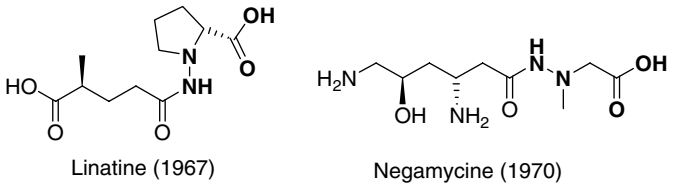


Figure 2.2 Natural peptide derivatives containing the N–N–C–C=O fragment.

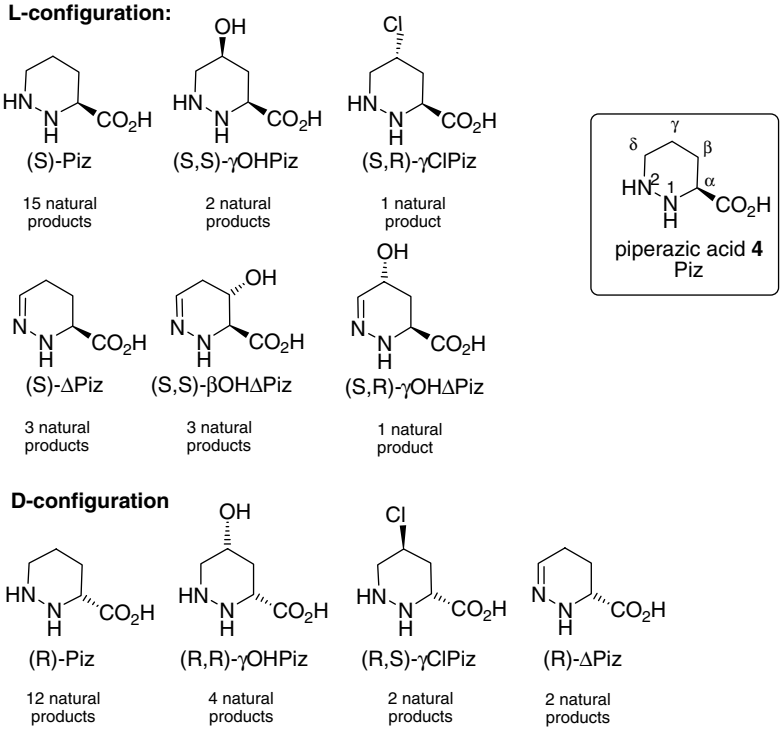


Figure 2.3 Piperazic acid derivatives found in natural products.

Since 1971, more than 30 linear or macrocyclic natural peptide derivatives containing one to three modified piperazic acid **4** residues and displaying remarkable biological properties have been isolated from microorganisms (*Streptomyces*, *Actinomycetes*, etc.) (Figure 2.3, recently discovered products [28–30]). Both L- and D-piperazic acid residues are found in these natural products. The γ-carbon is often substituted by a chlorine atom or a hydroxyl group and sometimes the δ-carbon is unsaturated [dehydropiperazic acid (ΔPiz)]. Synthesis, chemistry, and conformational properties of piperazic acids have been reviewed by Ciufolini and Xi [31]. Some structures of important macrocyclic or linear peptides containing piperazic acid residues are shown in Figure 2.4. In all these compounds, the piperazic acid derivative is linked at the C=O by a peptide bond and N^1 is acylated, except in the case of Sanglifehrin which is N^2-acylated (Figure 2.4).

In many cases (e.g., azinothricin [32] or piperazimycin [30]; Figure 2.4) both the L- and D-piperazic acid units are present in the same molecule. Monamycins [33] were the first members of this family to be described in the literature (Figure 2.4). Their cyclohexadepsipeptide structure is representative of these *Streptomyces* metabolites: it is present in 19 compounds out of 34 containing piperazic acid residues. Antrimycin [34] is a linear hexapeptide containing the ΔPiz unit. Luzopeptins [35] and himastatin [36] have dimeric structures and impressive

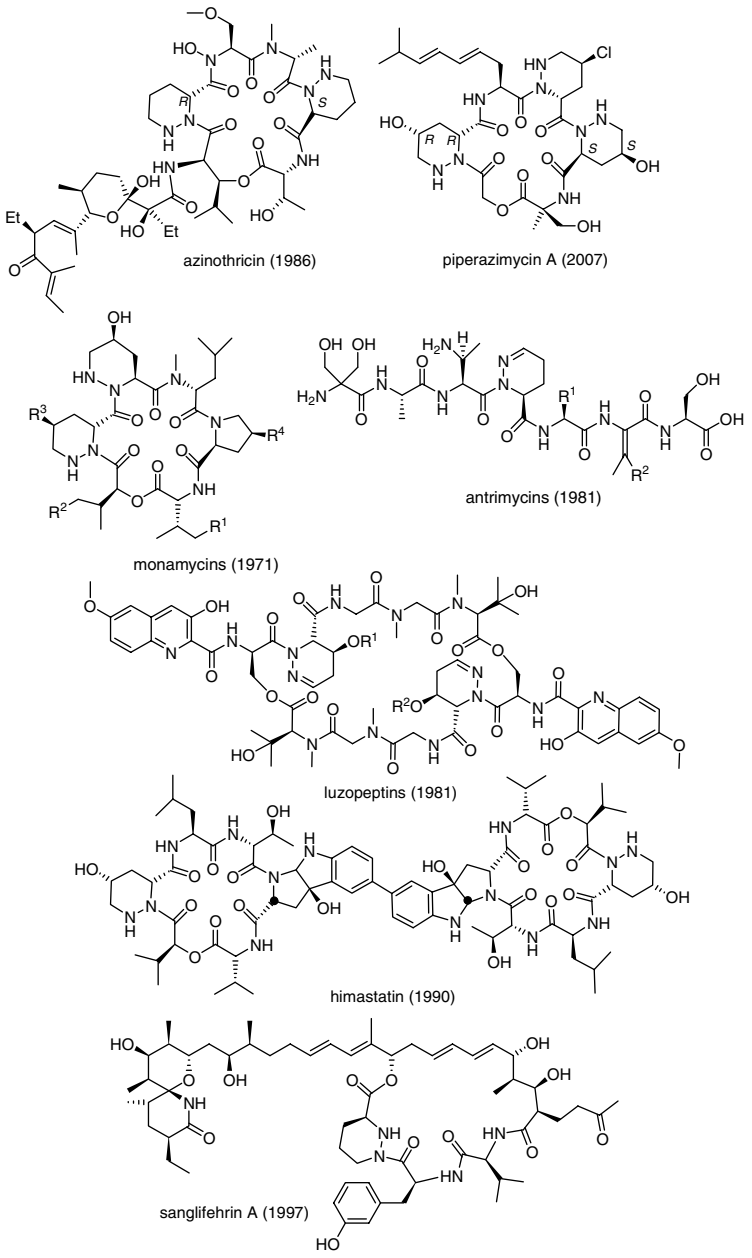


Figure 2.4 Some natural products containing the piperazic acid residue.

biological properties. Sanglifehrin [37] was discovered at Novartis and is a novel type of immunosuppressant, whose mode of action is different from that of other known compounds [38–40]. Its 22-membered macrocycle is unique among these derivatives, with a tripeptide unit and an unsaturated carbon chain.

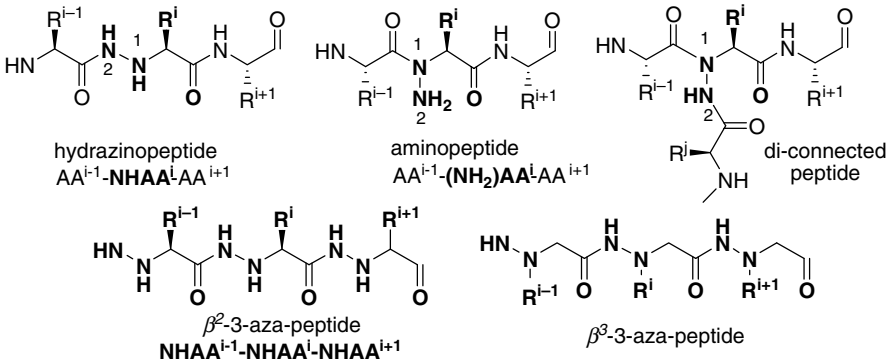


Figure 2.5 Known pseudopeptidic structures containing one or several α-hydrazino acids.

The potent biological activities and the complex structures of these derivatives generated broad interest leading to several syntheses of piperazic acid derivatives [31, 41, 42].

2.1.3
Synthetic Bioactive Products Containing the N–N–C–C=O Fragment

The replacement in a peptide of an α-amino acid residue by an α-hydrazino acid was first investigated by Niedrich *et al*. in the 1960s. The presence of two nitrogen atoms in α-hydrazino acids led Niedrich to connect the peptidic chain either on the N^2 atom to give a pseudopeptide that he named a "hydrazinopeptide" or on the N^1 atom to give a pseudopeptide that he named an "aminopeptide" (Figure 2.5) [43–55]. α-Hydrazino acids were introduced in eledoisin, which has vasodilatory, hypotensive, and extravascular smooth muscle stimulant properties [56, 57]. In some eledoisin analogs, the biological activity of the parent compound was retained and the resistance to the hydrolysis by proteases was enhanced.

More recently, hydrazinopeptides were shown to behave as reversible inhibitors of human leukocyte elastase [58, 59]. Pseudopeptides bearing an N-terminal achiral hydrazinoacetic acid residue were shown to display anticancer properties [60], or to be very useful intermediates in the preparation of lipopeptides [61], clustered glycoside–antigen conjugates [62], or synthetic mannose receptor ligands grafted on vesicles [63].

Diconnected peptides via an α-hydrazino acid were successfully used in the design of low-molecular-weight reverse-turn mimetics acting in a similar way to a native protein [64–67]. Unlike analogous peptides that are readily hydrolyzed, trimers and hexamers of α-hydrazino acids (named β^2-3-aza-peptides) proved to be very stable in the presence of several proteases [68]. β^3-3-Aza-dipeptides, which are N^1-substituted hydrazinoacetic oligomers, were found to be fair inhibitors of proteasome chymotrypsin-like activity and to inhibit cancer cell growth [69]. Twelve dodecapeptides derived from histone H4 sequence, in which each residue has been successively

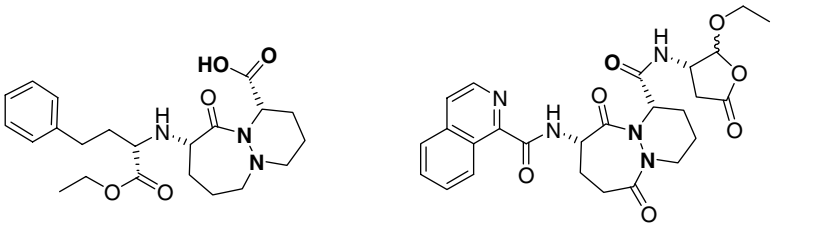


Figure 2.6 Drugs containing the piperazic acid residue.

replaced by a 3-aza β^3-amino acid in order to gain metabolic stability, have been prepared [70].

The L-enantiomer of piperazic acid also resides within the bicyclic ring system of many bioactive synthetic products such as cilazapril (Vascace) [71], a drug widely used in the treatment of hypertension or pralnacasan [72, 73], which went in phase IIb clinical trials as an anti-inflammatory agent (Figure 2.6).

2.1.4 The CO–N–N–C–CO–NH Fragment is a Turn Inducer in Pseudopeptides

Several crystalline structures of hydrazinopeptides and aminopeptides have been solved [68, 74–84]. The CO–N–N–C–CO–NH fragment is capable of stabilizing a turn via a bifurcated hydrogen bond closing an eight-membered cycle, which was named by Marraud as a "hydrazino turn" (Figure 2.7). Quantum chemical calculations carried out on β-3-aza-peptides reveal the presence of a wide variety of secondary structures [85]. Experimental studies with linear or macrocyclic oligomers of N^1-substituted hydrazinoacetic acid (β^3-3-aza peptides) showed remarkable conformations, the presence of hydrazino turns, and foldamer behavior [86–89].

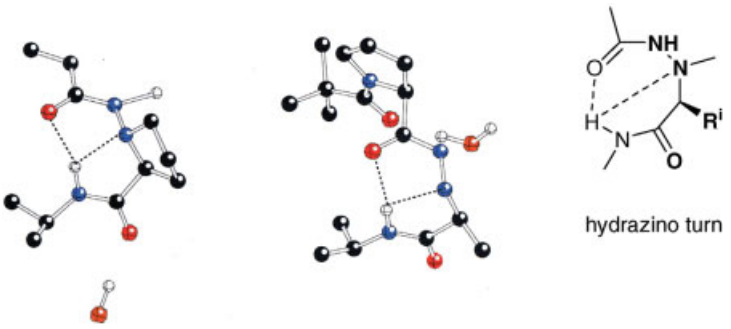
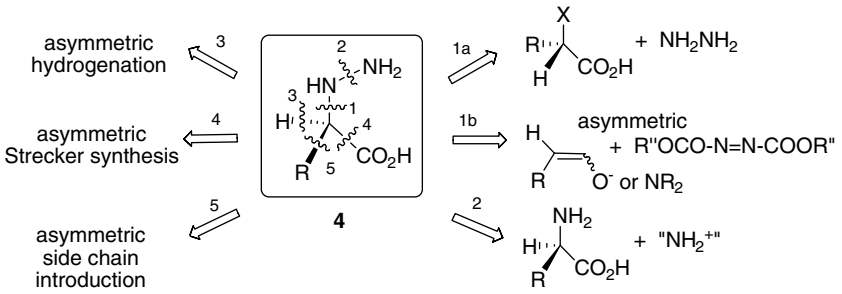


Figure 2.7 Hydrazino turn and X-ray structures of Moc-NHPro-NH*i*Pr and Piv-Pro-NHAla-NH*i*Pr.



Scheme 2.1 Disconnections used in the synthesis of α-hydrazino acids and derivatives.

2.2 Synthesis

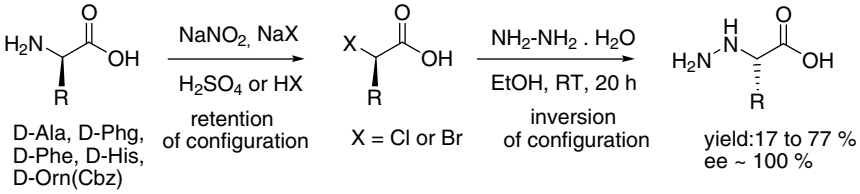
As mentioned, racemic α-hydrazino acids have been known for a long time. Owing to their use in the synthesis of biologically active compounds, we focus here only on the preparation of nonracemic α-hydrazino acids. Only three resolutions of racemic α-hydrazino acids (NH_2-Phe [4], NH_2-Val [90], and carbidopa **3** [91]) have been described and yields are generally low.

Examination of the general formula **4** leads to the identification of five main disconnections (Scheme 2.1). All of them have been used in the synthesis of α-hydrazino acids and derivatives.

Synthetic methods corresponding to disconnection *1a* (Scheme 2.1) mainly use nucleophilic amination of α-halogeno or α-hydroxy acids prepared from the chiral pool. They have been developed in the early synthesis of α-hydrazino acids. Modern techniques of asymmetric synthesis have been applied to electrophilic amination of carbonyl compounds using azodicarboxylates (disconnection *1b*). Enantioselective catalysis of this reaction is currently the object of much study. Electrophilic amination of α-amino acids (disconnection *2*) has also been developed. Disconnections *3*–*5* (hydrogenation, introduction of the carboxyl group, or introduction of the side-chain) have been explored in the last 15 years and mainly use catalytic asymmetric synthesis, which has also been applied to the preparation of α-amino acids [92]. Synthesis of α-hydrazino acids is reviewed according to these disconnections. Specific methods for the construction of the piperazic acid cycle are also reviewed.

2.2.1 Disconnection *1a*: Reaction of Hydrazine Derivatives with Carbon Electrophiles

Since nonracemic α-hydroxyacids ([93] and references cited therein) or α-halogenoacids ([94] and references cited therein) are easily available and have been known for a long time, they have been widely used as starting enantiopure materials. A reliable and inexpensive method for their preparation is the diazotization of α-amino acids, which proceeds with retention of configuration. Their reaction with hydrazine or several substituted hydrazines has been developed using diverse conditions, producing diversely substituted α-hydrazino acids with inversion of configuration,



Scheme 2.2 Reaction of hydrazine hydrate with α-halogeno acids.

in good yields and high enantiomeric excesses [2, 15, 17, 22, 46, 67, 95–98, 100–103]. Stereoselective synthesis from electrophilic unsaturated carbon compounds is less commonly used [104].

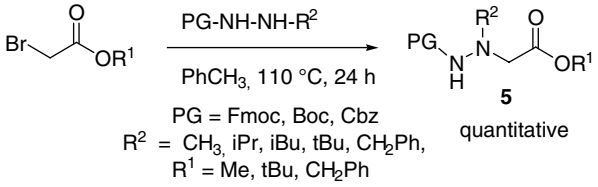
2.2.1.1 Reaction of Hydrazine Derivatives with Enantiopure α-Halogeno Acids

The first reported enantiopure α-hydrazino acid NH_2Phg was prepared by reacting hydrazine hydrate with the corresponding α-bromo acid [2]. This reaction is general (R = Me [46], CH_2Ph [46], Ph [2], imidazoylmethyl [17], or $(CH_2)_3$NHCbz [15]) and proceeds at room temperature (RT) or within a few hours in refluxing ethanol (Scheme 2.2). Yields are low to good (17–77%). Inversion of configuration is generally observed, except in the case of the valine derivative (R = $CHMe_2$ [46]), where total racemization occurs because of the steric hindrance. Thus, starting from a D-amino acid leads to an L-hydrazino acid by this approach.

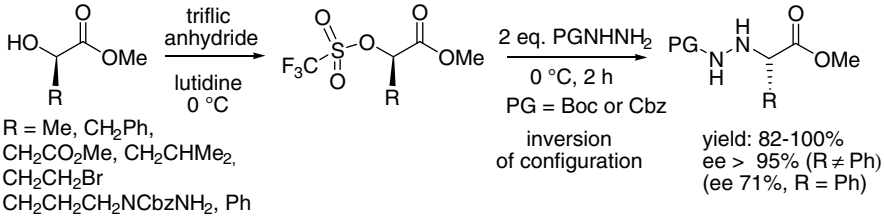
Carbazates ($ROCONHNH_2$) are more stable in air than hydrazines. Their reaction with α-bromo acetates needs harsh conditions (overnight reflux in toluene) to yield quantitatively achiral N^2-protected, N^1-substituted hydrazino acetates **5** [95] (Scheme 2.3). These derivatives are 3-aza-analogs of β^3-amino acids.

2.2.1.2 Reaction of Hydrazine Derivatives with Enantiopure Activated α-Hydroxy Esters

Both *tert*-butylcarbazate (Boc-$NHNH_2$) and benzyl carbazate (Cbz-$NHNH_2$) are less reactive than hydrazine hydrate, but they are stable and possess very useful protecting groups. Their reactions occur in mild conditions (0 °C) with sulfonates prepared from α-hydroxy esters and with inversion of configuration (Scheme 2.4) [22, 67, 96, 97]. In the reaction with *tert*-butylcarbazate, triflate ($CF_3SO_3^-$) proved to be a better leaving group than nosylate ($4\text{-}NO_2C_6H_4SO_3^-$) and achieved very good enantiomeric excesses [96]. N^2-Protected α-hydrazino esters were afforded in a short reaction time (2 h), with very good yields (82–100%) and in greater than 95% e.e. In the case of the easily racemizable



Scheme 2.3 Reaction of substituted hydrazine derivatives with α-bromoacetates.



Scheme 2.4 Synthesis of L-N^2-protected α-hydrazino esters from D-α-hydroxy esters.

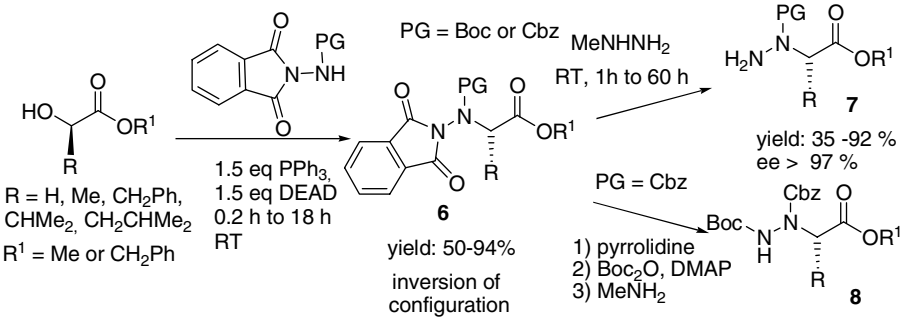
phenylglycine derivative (R = Ph), a lower enantiomeric excess (71%) was obtained. A similar intramolecular transformation of α-hydroxy-δ-hydrazino esters [98] yields piperazic acid derivatives.

2.2.1.3 Mitsunobu Reaction of Aminophthalimide Derivatives with Enantiopure α-Hydroxy Esters

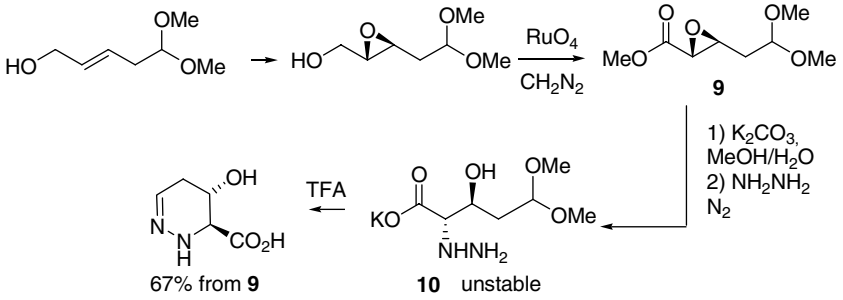
The Mitsunobu reagent [99] [Ph_3P and diethylazodicarboxylate (DEAD)] has been used to activate the reaction between enantiopure α-hydroxy esters and *N*-protected aminophthalimides in mild conditions (Scheme 2.5) [100, 101]. Treatment of **6** with methylhydrazine allowed the smooth deprotection of the phthalimide moiety yielding N^1-protected α-hydrazino esters **7** in fair to very good overall yield, excellent enantiomeric excess, and with inverted configuration [100]. Replacement of the phthalimide moiety by the more valuable Boc protecting group was also done via a three-step process leading to **8** [101]. Solid-phase synthesis of orthogonally N^1,N^2-diprotected α-hydrazino esters was also developed using this strategy [102].

2.2.1.4 Reaction of Hydrazine Derivatives with Nonracemic Epoxides

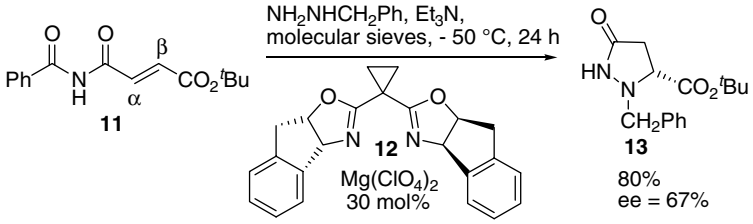
Nonracemic glycidic ester **9** [103], obtained via enantioselective Sharpless allylic epoxidation, is a very good electrophile in the regioselective and stereoselective reaction with a hydrazine nucleophile (Scheme 2.6). Only the α-carbon of **9** is attacked, with inversion of configuration to give the *anti*-α-hydrazino-β-hydroxy acid derivative **10**, which is unstable and a very good intermediate in the synthesis of β-hydroxy dehydropiperazic acid [103].



Scheme 2.5 Synthesis of L-N^1-protected α-hydrazino esters from D-α-hydroxy esters.



Scheme 2.6 Reaction of glycidic ester with hydrazine.



Scheme 2.7 Enantioselective synthesis of D-pyrazolidinone from α,β-unsaturated imide and benzyl hydrazine.

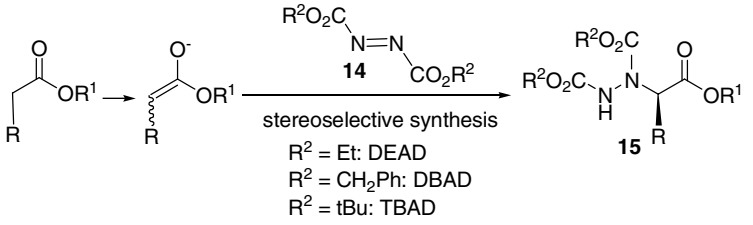
2.2.1.5 **Enantioselective Conjugate Addition of Hydrazines to α,β-Unsaturated Imides**

Complexation of Mg(II) by diimide **11** and the chiral ligand **12** catalyzed the regioselective conjugate addition of benzylhydrazine to the β-carbon in **11** (Scheme 2.7) [104]. Hydrazine attack was regioselective at N^1 and stereoselective. Cyclization occurred to form pyrazolidinone **13** and benzamide in very good yield and fair enantiomeric excess (67%). Only trace amounts of the isomeric pyrazolidinone (N^2 attack of hydrazine) were detected.

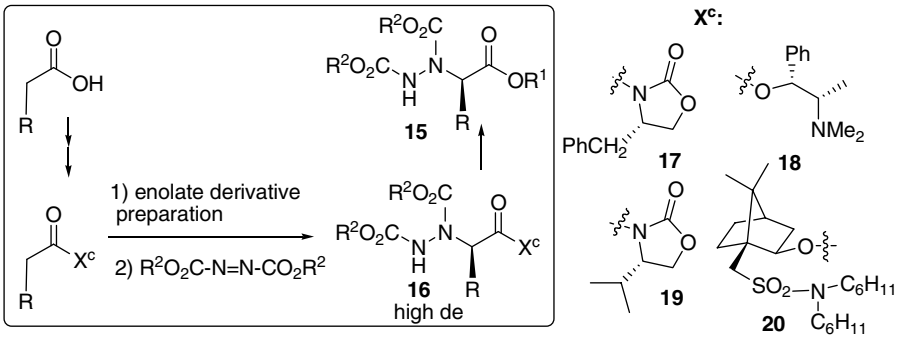
2.2.2

Disconnection *1b*: Stereoselective Synthesis using Azodicarboxylates

Asymmetric electrophilic α-amination of carbonyl compounds by azodicarboxylates **14** efficiently produces nonracemic N^1,N^2-diprotected α-hydrazino esters **15** (Scheme 2.8). The reactive enolate intermediate is mainly generated from α-saturated



Scheme 2.8 General scheme for α-hydrazination of esters using azodicarboxylates.



Scheme 2.9 General scheme for stereoselective amination of enolates bearing chiral auxiliaries with azodicarboxylates.

esters. As this reaction was reviewed by Greck *et al.* [105] and Erdik [106] in 2004, only its important features and current developments are depicted here.

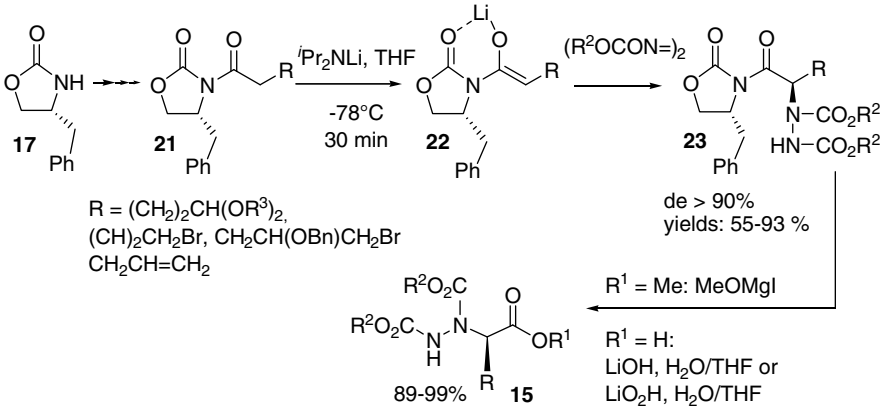
Different techniques of stereoselective synthesis have been applied to this reaction. Twenty years ago, Evans *et al.* [107], Gennari *et al.* [108], Vederas and Trimble [109], and Oppolzer and Moretti [110] used enolates carrying chiral auxiliaries X^c (X^c = **17**, **18**, **19** and **20**, respectively) and obtained **16** in high diastereomeric excess (Scheme 2.9). They further cleaved the auxiliary [LiOH, tetrahydrofuran (THF)/H_2O, 0 °C or MeOMgBr, MeO, 0 °C or $LiOCH_2Ph$, 0 °C or LiOOH, THF/H_2O, 0 °C] [111] and the N−N bond of **15** in order to prepare enantiopure α-amino acids.

Vederas also proposed the use of chiral azodicarboxylates (R^2 = mentyl, bornyl, or isobornyl), but the diastereomeric excesses of **15** were not good [112, 113]. Highly stereoselective amination of optically active β-hydroxy esters or β-amino esters has also been developed. Since 2002, direct catalyzed enantioselective α-amination of aldehydes, β-keto esters or α-cyano esters with azodicarboxylates is the object of many developments [114–116].

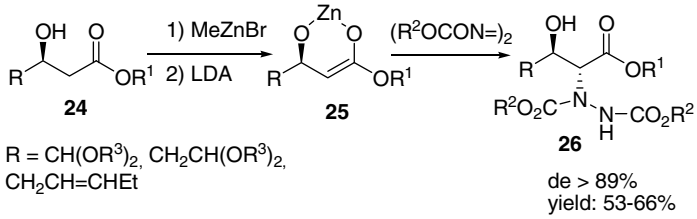
2.2.2.1 Stereoselective α-Hydrazination of Chiral Carbonyl Compounds using Azodicarboxylates

Stereoselective α-hydrazination of chiral carbonyl compounds using azodicarboxylates has been reviewed [105, 106]. Only examples related to the synthesis of piperazic acid derivatives will be covered here.

The commercially available 4-(*R*)- or 4-(*S*)-benzyloxazolidinones **17** are by far the most used chiral auxiliaries in the stereoselective α-hydrazination of carbonyl compounds and especially in the synthesis of piperazic acid derivatives (Scheme 2.10; R = $(CH_2)_2CH(OR^3)_2$ [117, 118], $(CH)_2CH_2Br$ [42], $CH_2CH(OBn)CH_2Br$ [119], or $CH_2CH{=}CH_2$ [120]). The formation of the (*Z*)-enolate **22** from amides **21** is usually highly stereoselective. The trapping of (*R*)-**22** by azodicarboxylate is also highly stereoselective and gives (*R,R*)-**23** with high diastereomeric excess. Nuclear magnetic resonance analysis of **23** is generally difficult because of the signal broadening due to the restricted rotation around the N−CO_2R^2 bond. In some cases, a resolved spectrum has been obtained at 125 °C in d_6-dimethylsulfoxide [42]. Removal of the



Scheme 2.10 Synthesis of (*R*)-piperazic acid precursors using Evans (*R*)-chiral auxiliary and azodicarboxylates.



Scheme 2.11 Synthesis of dehydropiperazic derivatives using stereoselective α-hydrazination of β-hydroxy esters with azodicarboxylates.

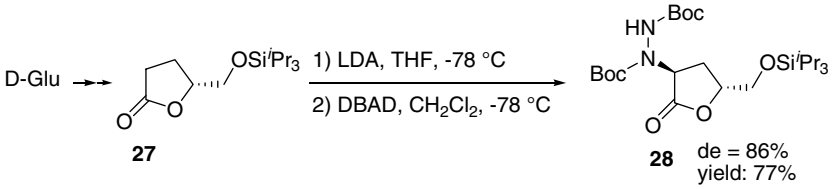
auxiliary gives **15** in very good yields. Further functional group interconversions allow the preparation of piperazic acid derivatives.

Nonracemic β-hydroxy esters **24** proved to be useful synthons in the synthesis of hydroxy-substituted piperazic derivatives **26** (Scheme 2.11; $CH_2CH(OR^3)_2$ [121, 122], $CH_2CH{=}CHEt$ [123], and $CH(OR^3)_2$ [124]). Chelation of enolate **25** by zinc resulted in a highly stereoselective attack on the azodicarboylate [125], yielding *anti* α-hydrazino, β-hydroxy esters **26**. Further functional group interconversions of the side-chain R allowed the preparation of piperazic acid derivatives.

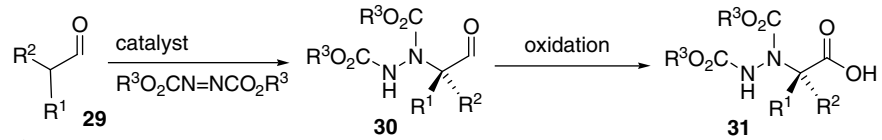
The D-glutamic acid skeleton was used in the stereoselective α-hydrazination of **27** (Scheme 2.12) [126]. The triisopropylsilyl protection induces better selectivity than the *tert*-butyldimethylsilyl ether. α-Hydrazino lactone **28** is a good precursor to γ-chloro- or γ-hydroxy-substituted piperazic acid.

2.2.2.2 Catalytic Enantioselective α-Hydrazination of Carbonyl Compounds using Azodicarboxylates

Jørgensen *et al.* [127] and List [128] reported the catalytic enantioselective amination of achiral aldehydes **29** ($R^2 = H$, $R^1 = Me$, Et, *i*Pr, *t*Bu, allyl, CH_2Ph, $(CH_2)_2CH_2Br$ [41], or $CH_2CH_2OCH_2Ar$ [129]) using azodicarboxylates as the nitrogen source and L-proline (2–50 mol%) as catalyst in good to excellent yields (57–92%)



Scheme 2.12 Stereoselective α-hydrazination of D-glutamic acid derivative.



Scheme 2.13 Enantioselective two-step synthesis of diprotected α-hydrazino acids from aldehydes.

and excellent enantioselectivities (89–95%) (Scheme 2.13). Reaction times varied from 0.5 to 4 h in CH_2Cl_2 at RT. Very good enantioselectivities and yields were obtained in multigram-scale experiments [41, 129]. The (*R*) configuration was assigned to the easily epimerizable α-hydrazinoaldehydes **30** ($R^2 = H$) when L-proline was used as catalyst. As D-proline is easily available, this versatile reaction can produce the (*S*) enantiomer. Intermediate **30** ($R^1 = (CH_2)_2CH_2Br$ and $R^2 = H$) was used in a short and highly efficient preparation of D-piperazic acid [41].

This reaction using proline [130–132] or **34a** [133] has been successfully applied to racemic α-disubstituted aldehydes **29** ($R^1 \neq H$ and $R^2 \neq H$) to yield the nonepimerizable hydrazino aldehyde **30** in good to excellent enantiomeric excesses and yields.

Further oxidation of the aldehyde group in **30** with buffered $KMnO_4$ [127] or $NaClO_2$ [131–133] leads to the diprotected α-hydrazino acids **31a** [127], **31b** [132], **31c** [131], or **31d** [133] in good yields and high enantiomeric excesses (Scheme 2.13 and Figure 2.8).

Numerous studies about the catalyzed amination of β-keto esters or α-cyano esters **32** have been developed, using basic catalysts **34** or catalytic organometallic complexes **35** (Table 2.1 and Figure 2.9) [134–146]. Optimization of reaction

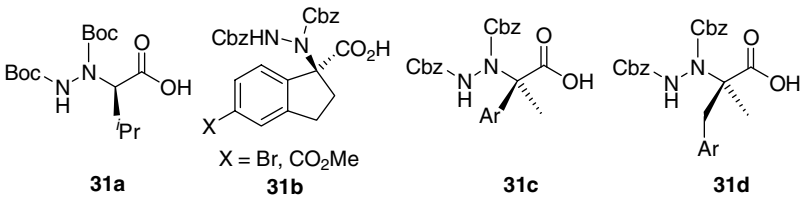
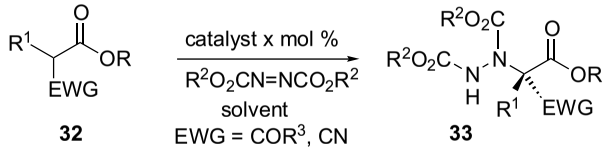


Figure 2.8 α-Hydrazino acids prepared according to Scheme 2.13.

Table 2.1 Catalytic enantioselective α-hydrazination of β-keto esters or α-cyano esters **32**.



32	Catalyst	R^2	Solvent	Temperature (°C)	Time (h)	Yield (%)	Enantiomeric excess (%)	Reference
Ten examples EWG = COR^3	**35a** (S)-PhBOX $Cu(OTf)_2$ 10 mol%	CH_2Ph	CH_2Cl_2	RT	16	76–98	95–99	[134]
Four examples EWG = COR^3	**34b** 5 mol%	*t*Bu	toluene	−50 or RT	16–143	86–99	83–90	[135]
Nine examples EWG = CN R^1 = Ar	**34b** 5 mol%	*t*Bu	toluene	−78; −50	16–20	84–99	89–98	[135]
Seven examples EWG = COR^3	**34e** or **34f** 20 mol%	CH_2Ph	CH_2Cl_2	−25	0.03–160	51–99	50–90	[136]
One example EWG = COR^3	**35a** Ph-BOX $Cu(OTf)_2$ 10 mol%	CH_2Ph	CH_2Cl_2	0	3	not isolated; used in further tandem reaction		[137]

Ten examples EWG = CN R^1 = Ar	**34c** **34d** 5–10 mol %	*t*Bu CH_2Ph	toluene	−78	0.5–12	72–98	87–99	[138]
One example EWG = COR^3	**35b** $Cu(OTf)_2$ 1 mol%	CH_2Ph	CH_2Cl_2	0	16	91	99	[139]
Seven examples EWG = COR^3	**34h** 10 mol %	*t*Bu	toluene	−78 −40	3–96	52–99	87–91	[140]
Ten examples EWG = COR^3	$Pd(PF_6)_2$, H_2O, CH_3CN, **35d** 5 mol%	Et *i*Pr CH_2Ph *t*Bu	MeOH	RT	0.5–170	56–99	91–99	[141]
Ten examples EWG = COR^3	**34j** 2 mol%	*t*Bu	THF	−60	0.5–24	54–99	15–98	[142]
Four examples R^1 = F EWG = COR^3	**35a** Ph-BOX $Cu(OTf)_2$ 0.5 mol%	Et CH_2Ph	CH_2Cl_2	RT	48	78–94	81–94	[143]
Three examples EWG = COR^3	**35c** $Eu(OTf)_3$ 10 mol%	*t*Bu	CH_2Cl_2	−41 to >0	3 + 2	81	95	[144]
One example EWG = CN R^1 = Ph	**34g** (19 screened catalysts) 50 mol%	*t*Bu	toluene/hexane	−78	0.5	91	84	[145]
One example EWG = COR^3	**34i** $La(OiPr)_3$ 4 mol%	*t*Bu	$CHCl_3$	0	1	99	92	[146]

EWG = electron-withdrawing group.

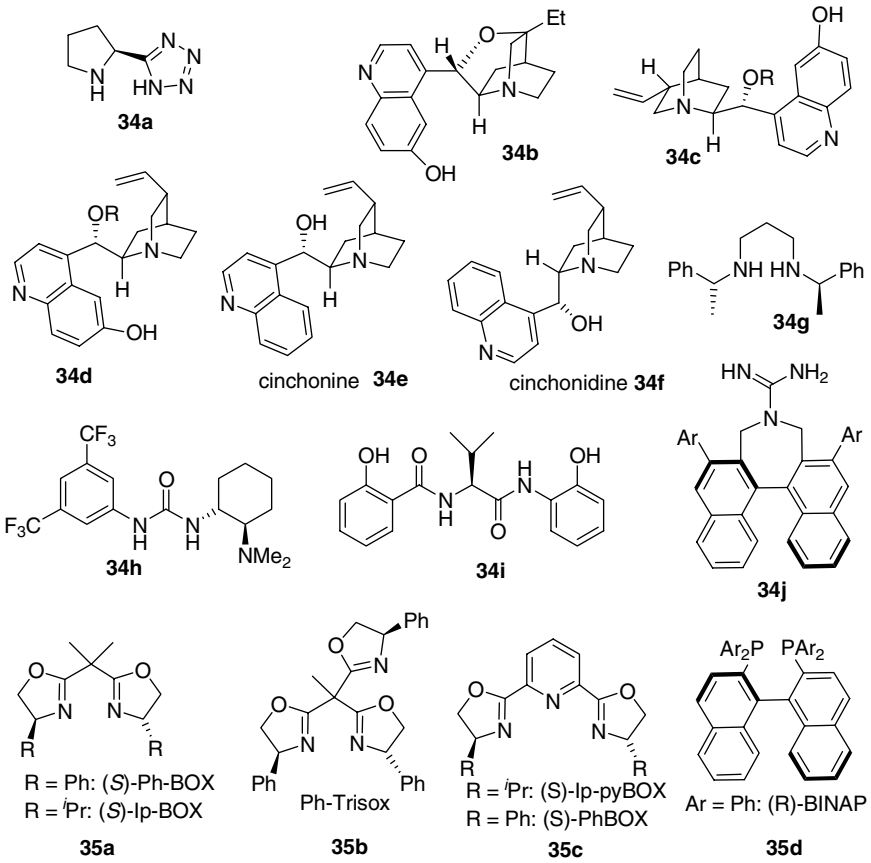


Figure 2.9 Enantiopure catalysts and ligands.

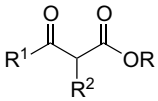
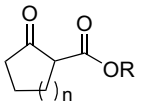
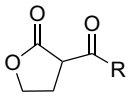
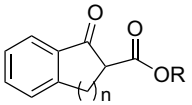
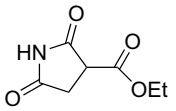
conditions (azodicarboxylate, temperature, solvent, and catalyst) resulted in very good to excellent yields and enantiomeric excesses for a wide variety of substrate. This reaction is the shortest and most efficient way to make nonracemic α-N^1,N^2-diprotected hydrazino β-keto esters **33** (EWG = COR^3) or α-N^1,N^2 diprotected hydrazino α-cyano esters **33** (EWG = CN). Moreover, it was run efficiently on a multigram scale (23 g) in the total synthesis of AS 3201 [146].

β-Keto esters or lactones used in catalytic enantioselective α-hydrazination with azodicarboxylates are listed in Table 2.2. In many cases, the absolute configuration of product **33** was not assigned.

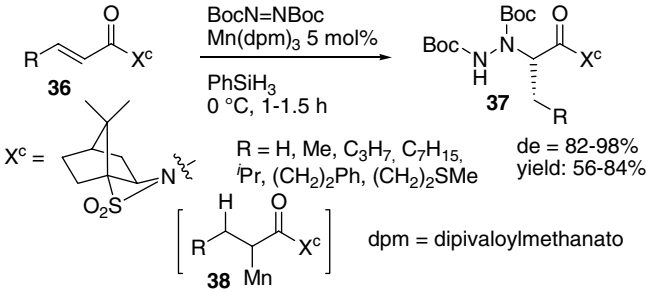
2.2.2.3 Stereoselective α-Hydrazination of Chiral α,β-Unsaturated Carboxylates using Azodicarboxylates

Organometallic catalyzed hydrohydrazination of olefins is the object of renewed interest [147, 148]. Chiral manganese enolates **38** can be generated *in situ* from α,β-unsaturated carbonyl compounds **36** using phenylsilane and catalytic Mn(III) tris

Table 2.2 β-keto esters or lactones **32** used in catalytic enantioselective α-hydrazination with azodicarboxylates.

β-keto ester or lactone 32	Substituent	Reference
	$R^1 = H$, $R^2 = Me$	[142]
	$R^1 = Me$, $R^2 = Me$	[134, 139, 141, 142]
	$R^1 = Me$, $R^2 = Et$	[136, 142]
	$R^1 = Me$, $R^2 = allyl$	[134]
	$R^1 = Et$, $R^2 = Me$	[134, 135, 142]
	$R^1 = iPr$, $R^2 = Me$	[134]
	$R^1 = PhCH_2$, $R^2 = Me$	[134]
	$R^1 = Ph$, $R^2 = Me$	[134]
	$R^1 = Me$, tBu, Ph, $R^2 = F$	[143]
	$n = 1$	[134–136, 140–142, 144]
	$n = 2$	[134–136, 140, 142]
	$n = 3$	[134, 140]
	R = Me	[136, 141, 142]
	R = *i*Pr	[136]
	R = *t*Bu	[136]
	$n = 1$	[140, 141]
	$n = 2$	[140–142]
		[146]

(dipivaloylmetanato) (Scheme 2.14) [149]. Trapping of this intermediate with azodicarboxylate is highly diastereoselective when the Oppolzer sultam is used as the chiral auxiliary. Diprotected α-hydrazino acid derivatives **37**, similar to those described in the previous section, are obtained in good yields.



Scheme 2.14 Stereoselective α-hydrazination of α,β-unsaturated carbonyl compounds.

2.2.3 Disconnection 2: Synthesis from Chiral Nonracemic α-Amino Acids

Electrophilic amination of proteinogenic α-amino acids is an attractive method since it avoids the creation of the stereogenic center. The rearrangement of hydantoic acids prepared from α-amino acids or the use of several electrophilic aminating reagents (nitrosonium cation, hydroxylamine derivatives, or oxaziridines) have been developed. These methods have received great interest in the preparation of carbidopa **3** from L-α-methyl-DOPA.

2.2.3.1 Schestakow Rearrangement of Hydantoic Acids Prepared from α-Amino Acids

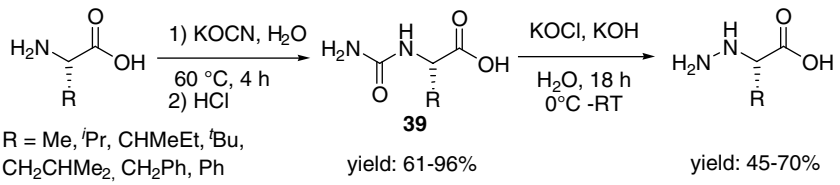
Schestakow prepared hydrazines from urea, using a reaction similar to the well-known Hofmann rearrangement of amides [150]. Hydantoic acids **39**, which are easily prepared by reacting α-amino acids and potassium cyanate, undergo rearrangement to form α-hydrazino acids without racemization in the presence of hypochlorite in strongly basic aqueous solution (Scheme 2.15). The use of potassium hypochlorite [151] instead of sodium hypochlorite [152–154] greatly improves the yields. α-Hydrazino acids with aliphatic or aromatic side-chains that are stable in basic and oxidizing conditions are thus obtained in fair to good overall yields. As reagents are cheap, the preparation of carbidopa **3** using this two-step sequence has been patented several times by different inventors.

2.2.3.2 Reduction of *N*-Nitroso-α-Amino Esters

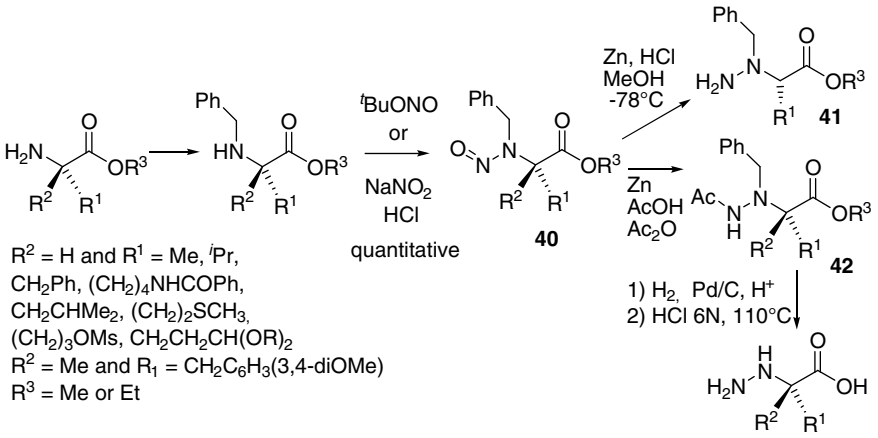
Sodium nitrite in hydrochloric acid or *tert*-butyl nitrite are inexpensive aminating reagents for secondary amines. *N*-Nitroso proline [25] or *N*-benzyl,*N*-nitroso-α-amino esters **40** were quantitatively obtained using these reagents [98, 155, 156] (Scheme 2.16). Controlled reduction of **40** by zinc dust at low temperature gave the unstable α-hydrazino esters **41** [156]. The stable protected α-hydrazino esters **42** were obtained in good yields by zinc reduction in the presence of acetic anhydride [98, 155]. Hydrogenolysis of the benzyl group proceeded without cleavage of the N−N bond and subsequent hydrolysis furnished α-hydrazino acids in good yields [155]. This sequence has been applied to the preparation of carbidopa **3** [155], NH_2-Pro [25], and piperazic acid derivatives [98].

2.2.3.3 Amination of α-Amino Acids by Hydroxylamine Derivatives

In order to prepare carbidopa **3**, Merck laboratories also explored direct amination of L-α-methyl-DOPA with sodium hydroxylamine sulfonate. The reaction was not complete and carbidopa isolation was difficult [152].



Scheme 2.15 Schestakow rearrangement of hydantoic acids prepared from α-amino acids.



Scheme 2.16 Synthesis of protected α-hydrazino esters by reduction of *N*-nitroso-α-amino esters.

2.2.3.4 **Amination of α-Amino Acids by Oxaziridines**

Unsubstituted oxaziridines **43** [157] and *N*-alkoxycarbonyl oxaziridines **44** [158] are excellent reagents for electrophilic amination of various nucleophiles (Figure 2.10).

The Schmitz oxaziridine **43** is an unstable reagent, which can be stored in toluene solution. It is easily prepared *in situ* from cyclohexanone and hydroxylamine *O*-sulfonic acid. It reacts with α-amino esters at 90 °C to give hydrazones **45**, which are hydrolyzed to yield enantiopure α-hydrazino acids in fair to good overall yields (Scheme 2.17) [157]. Conditions were optimized for the industrial preparation of carbidopa **3** (yield 77%) in Eastern Germany [159]. The terminal *N*-amino group in peptides can be similarly aminated by oxaziridine **43** [160].

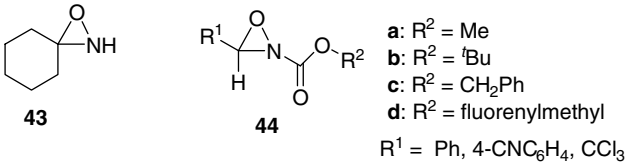
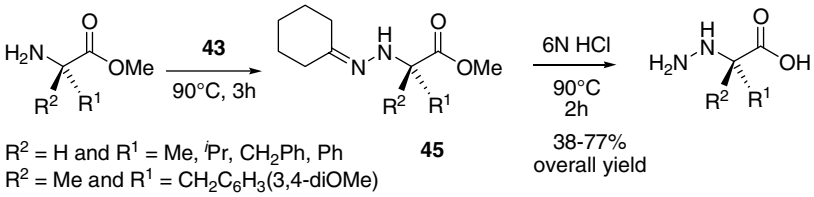
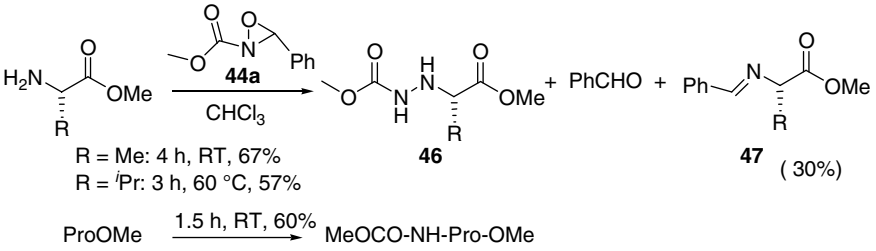


Figure 2.10 Oxaziridines for electrophilic amination of α-amino acid derivatives.



Scheme 2.17 Synthesis of α-hydrazino acids from α-amino esters using oxaziridine **43**.

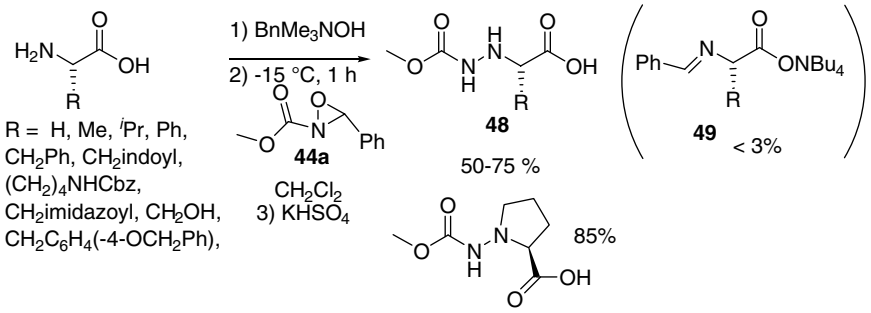


Scheme 2.18 Synthesis of N^2-protected α-hydrazino esters using oxaziridine **44a**.

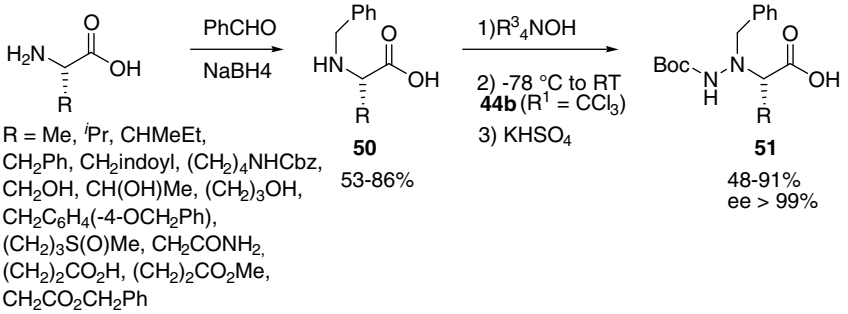
Oxaziridines **44a–d**, which are *N*-substituted by protecting groups, methoxycarbonyl (Moc) [161], Boc [158, 162–164], Cbz [158, 165], and Fmoc [158], are stable at RT and more reactive than **43**. These oxaziridines transfer their *N*-protected group to secondary and primary amines in mild conditions (Scheme 2.18). N^2-Moc-α-hydrazino esters **46** are obtained in one step from amino esters but yields are fair, since a side-reaction leads to imine **47** from the starting primary amine and the released benzaldehyde [161].

α-Amino acids, in the form of their benzyltrimethylammonium salts in order to obtain dissolution in CH_2Cl_2, react faster than the corresponding esters with **44a** (Scheme 2.19) [158]. Formation of the side-product **49** is avoided at −15 °C. α-Amino acids bearing functionalized side-chain can be aminated by **44a**. Protection of the weakly basic side-chain (imidazole in histidine or indole in tryptophan) is not needed. This one-pot sequence furnishes N^2-protected α-hydrazino acids **48** in good yields. However, deprotection of the Moc group requires rather harsh conditions (BBr_3 or Me_3SiI, 60 °C).

Orthogonal protection of both N^1 and N^2 nitrogens is recommended in order to use α-hydrazino acids in conventional peptide synthesis. A *N*-Boc group can be delivered by oxaziridine **44b** to an orthogonally protected α-amino acid **50** (Scheme 2.20) [166]. The benzyl group was chosen since nucleophilicity of the amino group is kept. A wide variety of *N*-benzyl amino acids can be obtained by reductive alkylation of the corresponding amino acid. Electrophilic amination of the *N*-benzyl α-amino acid, tetralkyl ammonium salt uses *N*-Boc oxaziridine **44b** ($R^1 = CCl_3$), which is easily prepared on a large scale (75 g) and in two steps from chloral [163]. Good yields of



Scheme 2.19 Synthesis of N^2-protected α-hydrazino acids using oxaziridine **44a**.



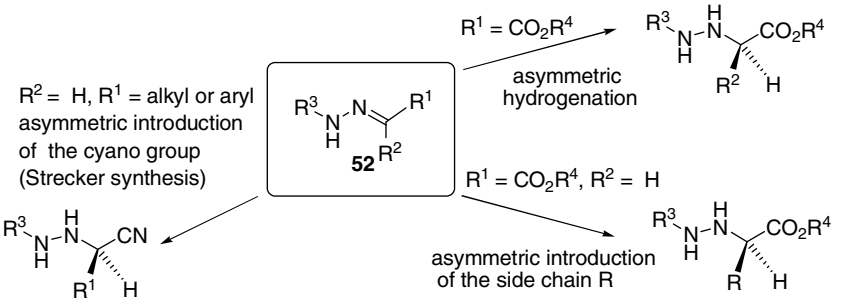
Scheme 2.20 Synthesis of orthogonally N^1,N^2-diprotected α-hydrazino acids using oxaziridine **44b** ($R^1 = CCl_3$).

N^1-benzyl, N^2-Boc-α-hydrazino acids **51** bearing the representative side-chains of proteinogenic amino acids are obtained on a multigram scale (Scheme 2.20) [166].

Several D- or L-N^1-benzyl, N^2-Boc-hydrazino acids **51** [Boc-NH(Bn)-Tyr, Boc-NH(Bn)-Ala, Boc-NH(Bn)-Ile, Boc-NH(Bn)-Met, Boc-NH(Bn)-Ser, Boc-NH(Bn)-Trp, Boc-NH(Bn)-Val, and Boc-NH-Pro] are commercially available.

2.2.4
Disconnections *3*, *4*, and *5*: Syntheses from Hydrazones or α-Diazoesters

The prochiral hydrazone function is an excellent precursor to α-hydrazino acid derivatives, but it was used only recently in catalytic or stereoselective reactions (Scheme 2.21). Preparation of hydrazones from carbonyl compounds and hydrazine derivatives is easier than that of the corresponding imines and often gives crystalline compounds. The first catalytic enantioselective hydrogenation of **52** ($R^1 = CO_2R^4$ and $R^3 = COPh$) was realized in 1992 [167] and this opened the door to the wide use of N^2-benzoyl hydrazone **52** ($R^3 = COPh$) in catalytic asymmetric synthesis since it can coordinate metals or Lewis acids by its N^1 atom and Bz carbonyl (Figure 2.11) [168, 169]. Hydrazones from glyoxylic acid derivatives **52** ($R^1 = H$, $R^2 = CO_2R^4$) are quite stable, even in water. They have proved to be useful precursors in the asymmetric introduction of the side-chains to yield α-hydrazino acid



Scheme 2.21 General synthetic pathways to α-hydrazino acid derivatives from hydrazones **52**.

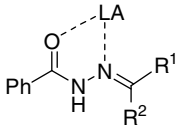


Figure 2.11 Chelation of *N*-acyl hydrazones with Lewis acids (LA) and electrophilic activation of the carbon.

derivatives [173–181]. Essentially weak nucleophiles (allylsilane [174, 175], silyl enol ethers [176–179], electron-rich arenes [180], and radicals [181]) reacted with these hydrazones in stereoselective reactions. Asymmetric hydrocyanation of hydrazones was also realized, as well as the cyano group hydrolysis to an acid group [171, 172].

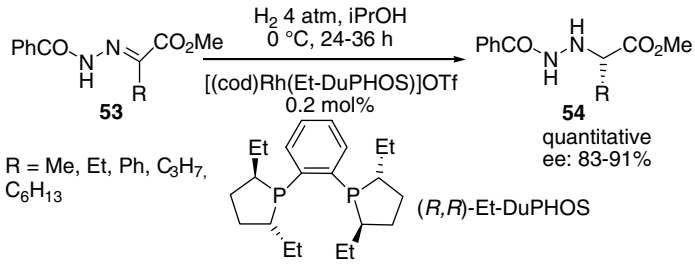
2.2.4.1 Catalytic Enantioselective Hydrogenation of Hydrazones

Efficient rhodium-catalyzed enantioselective hydrogenation of *N*-benzoylhydrazone **54** was performed using (*R,R*)-Et-DuPhos as the chiral ligand (Scheme 2.22) [167, 170]. The hydrazido carbonyl function appears to be the crucial structural feature required for this hydrogenation to proceed, suggesting that substrate chelation on the rhodium center occurs via the Bz carbonyl oxygen and the N^1 nitrogen moiety of the hydrazido moiety (Figure 2.11). Although reaction time was rather long, N^2-protected α-hydrazino esters **54** are obtained in excellent yields and very good enantiomeric excess.

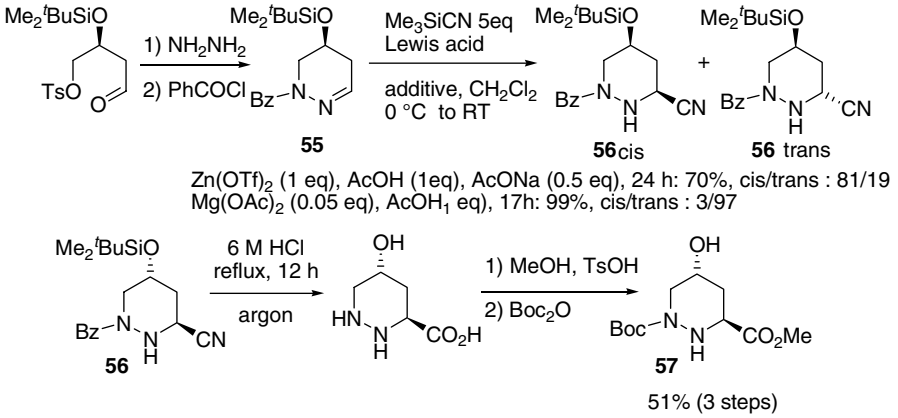
2.2.4.2 Stereoselective and Catalytic Enantioselective Strecker Reaction

Lewis acid promoted diastereoselective addition of HCN to nonracemic cyclic hydrazone **55** (Strecker synthesis) was the key step in the synthesis of 5-hydroxypiperazic acid derivative **57** (Scheme 2.23) [171]. An interesting reversal of diastereoselectivity by the choice of Lewis acid catalyst was observed. Further hydrolysis under inert atmosphere of *cis*-**56** or *trans*-**56** led respectively to *cis*- or *trans*-5-hydroxypiperazic acid derivatives **57**.

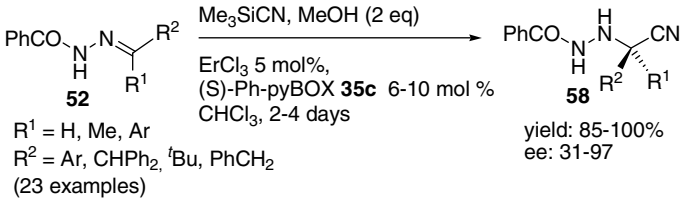
The enantioselective addition of HCN to hydrazones **52** was catalyzed in high yields by the easily prepared (Ph-pyBOX)$ErCl_3$ complex (Scheme 2.24) [172]. When $R^1 = H$, greater enantiomeric excess were obtained for aromatic R^2 (76–97%) than for aliphatic R^2 (31–69%). The reaction rate is very slow and electron-rich aromatic R^2 are



Scheme 2.22 N^2-protected α-hydrazino esters from catalytic enantioselective hydrogenation of hydrazones.



Scheme 2.23 Diastereoselective hydrocyanation of hydrazone **55**, as a key step to 5-hydroxy piperazic acid **57**.



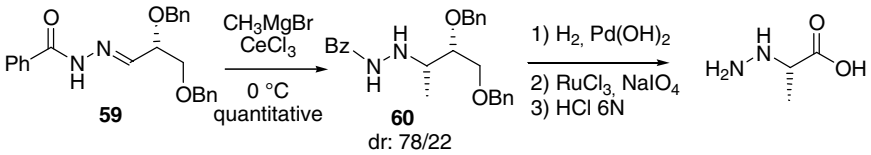
Scheme 2.24 N^2-protected α-hydrazino nitrile **58** from catalytic enantioselective hydrocyanation of hydrazones.

needed. As products **58** were found to be solids, their enantiopurity can be raised through recrystallization.

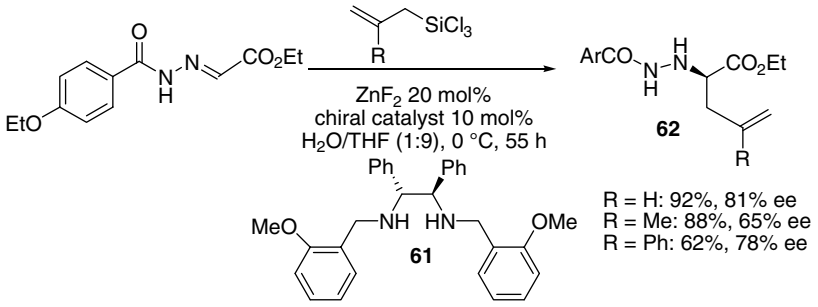
2.2.4.3 Stereoselective Addition of Organometallic Reagents to Hydrazones

Addition of Grignard reagents to the chiral hydrazone **59** derived from D-glyceraldehyde stereoselectively gives **60** in the presence of $CeCl_3$ (Scheme 2.25) [173]. Further deprotection and oxidation furnishes L-NH_2-Ala.

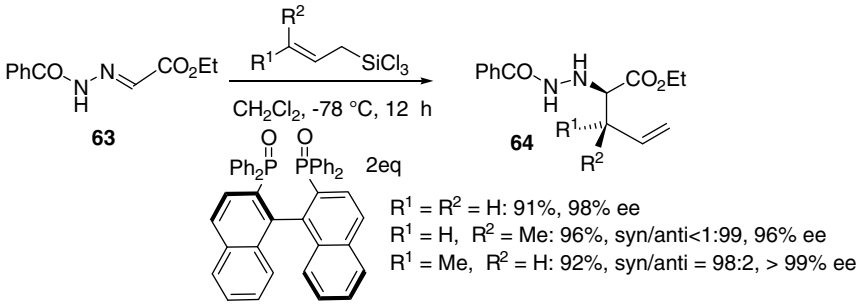
The asymmetric allylation of benzoyl-type hydrazones like **52** ($R^3 = COPh$) with trichloro allylsilanes has been achieved by Kobayashi *et al.* in two ways (Schemes 2.26 and 2.27). The use of catalytic amounts of zinc fluoride and catalyst **61** in aqueous media yields N^2-aroyl α-hydrazino α-allylic esters **62** in modest enantiomeric excess at 0 °C (Scheme 2.26) [174]. Water plays a key role since the reaction does not occur in dry THF.



Scheme 2.25 Diastereoselective addition of organometallics to hydrazone **60**.



Scheme 2.26 Preparation of α N^2-aroylhydrazino α-allylic esters **62** using allyl silanes.



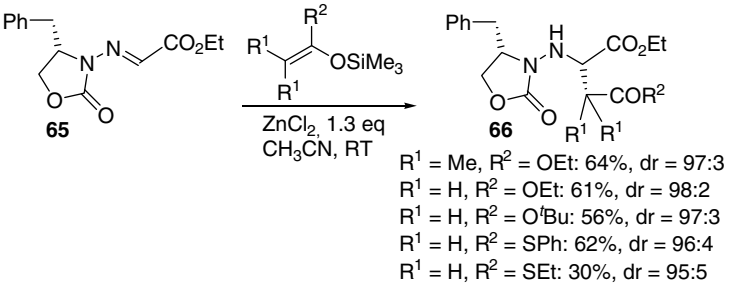
Scheme 2.27 Enantioselective and stereospecific allylation of hydrazone using BINAP dioxide as additive.

The reaction can also be run in organic solvent, at low temperature using 2,2′-bis (diphenylphosphino)1,1′-binaphthyl (BINAP) oxide as additive (Scheme 2.27) [175]. BINAP oxide coordinates to the allyl silane to form a chiral hypervalent silicon compound, which reacts very efficiently with hydrazone **63**. The enantiomeric excess can be improved up to 98%. Furthermore, the reaction proceeds stereospecifically. (*E*)-Allylic silanes give the *syn* adduct **64**, whereas the *anti* adduct **64** is obtained from the (*Z*)-allylic silane.

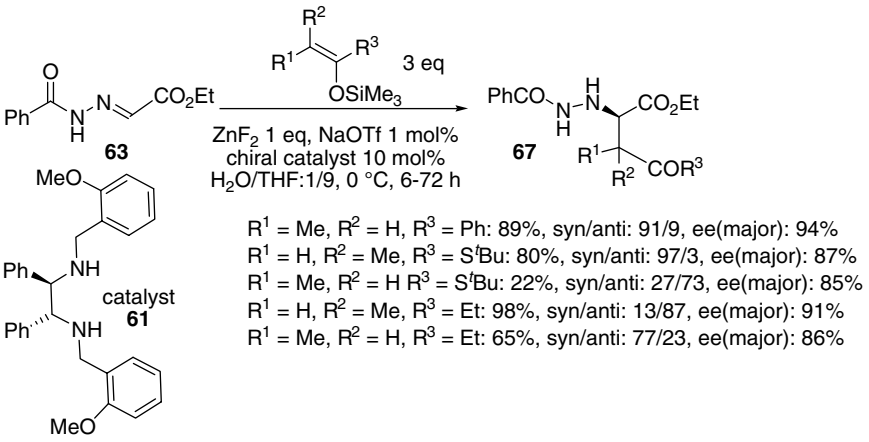
2.2.4.4 Stereoselective or Catalytic Enantioselective Mannich-Type Reaction with Hydrazones

The Lewis acid-catalyzed activation of glyoxylic hydrazones allows their Mannich-type reaction with silyl enol ethers. A highly stereoselective reaction occurs with non-racemic hydrazone **65** derived from Evans oxazolidinone **17** and silyl ketene acetals or thiocetals (Scheme 2.28) [176]. Simple activation by excess zinc chloride is necessary and yields of **66** are good.

Diastereo- and enantioselective Mannich-type reactions of achiral glyoxylic hydrazone **63** and silyl enol ethers in aqueous medium are achieved using ZnF_2 and the chiral catalyst **61** (Scheme 2.29) [177–179]. This reaction proceeds by double activation, in which Zn(II) acts as a Lewis acid to activate the hydrazone and fluoride acts as a Lewis base to activate the silyl enol ether. Fair to good yields of **67** are obtained, in fair to good diastereomeric ratio.



Scheme 2.28 Diastereoselective Mannich-type reaction between silyl enol ethers and chiral glyoxylic hydrazone **65**.



Scheme 2.29 Catalytic enantioselective Mannich-type reaction between silyl enol ethers and glyoxylic hydrazone **63**.

2.2.4.5 **Enantioselective Friedel–Crafts Alkylations with Hydrazones**

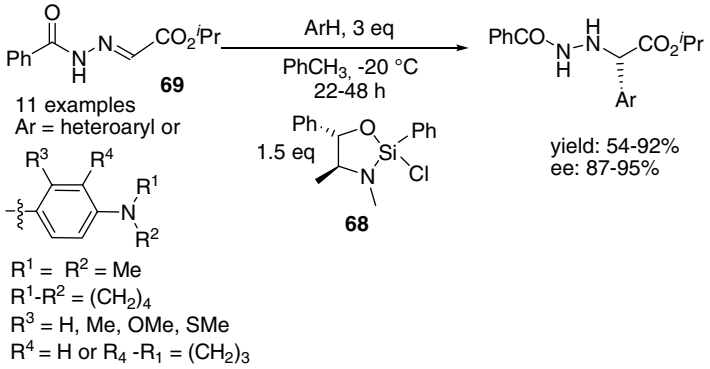
Chlorosilane **68** mediates the enantioselective Friedel–Crafts reaction between electron rich aromatic compounds ArH and isopropyl glyoxylate hydrazone **69** (Scheme 2.30) [180]. Although chlorosilane **68** is obtained in a 2 : 1 diastereomeric ratio from pseudoephedrine and trichlorophenylsilane, the reaction proceeds with high enantiomeric excess.

2.2.4.6 **Diastereoselective Zinc-Mediated Carbon Radical Addition to Hydrazones**

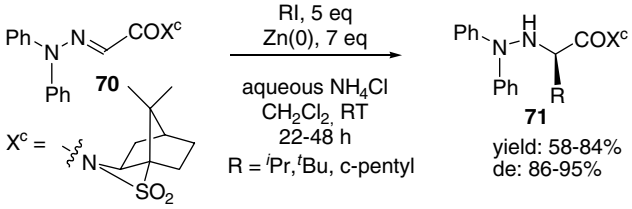
Radicals generated from alkyl iodides and Zn(0) in aqueous media add, in a highly stereoselective way, to N^2-diphenyl hydrazones **70** carrying the Oppolzer chiral auxiliary (Scheme 2.31) [181]. N^2-Diphenyl α-hydrazino acids **71** are obtained in fair to good yields and with good diastereomeric excess.

2.2.4.7 **Catalytic Enantioselective Reaction of α-Diazoesters with Aldehydes and Subsequent Stereoselective Reduction**

Asymmetric phase-transfer catalysis by **72** allows the synthesis of α-diazo-β-hydroxy esters **73** in good to excellent yields and modest to good enantiomeric excesses from

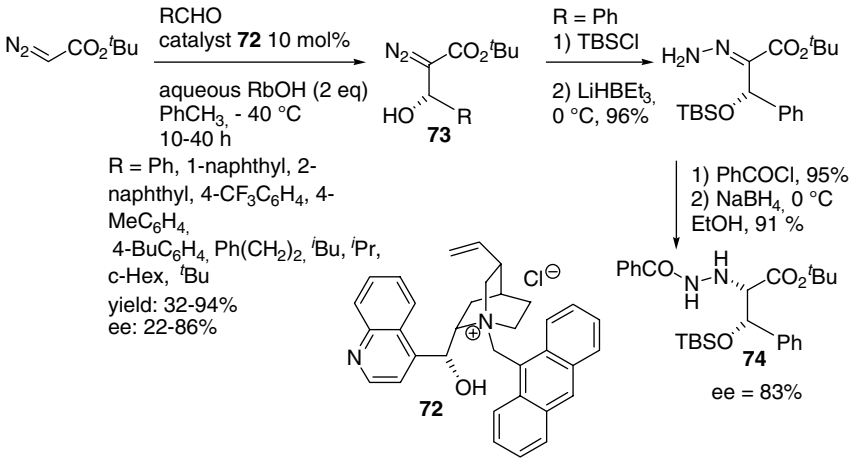


Scheme 2.30 Enantioselective Friedel–Crafts reaction between electron-rich aromatic compounds and isopropyl glyoxylate hydrazone **69** mediated by chiral chlorosilane **68**.

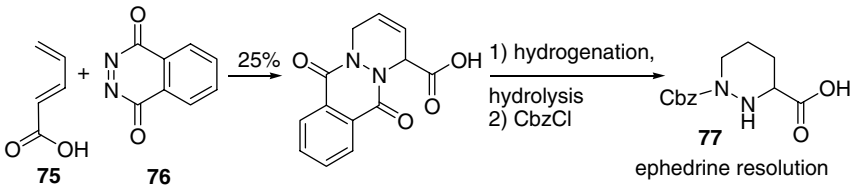


Scheme 2.31 Diastereoselective zinc-mediated carbon radical addition to hydrazone **70**.

α-diazo *tert*-butyl esters and aldehydes (Scheme 2.32) [182]. Further reactions of the diazo group, including 100% stereoselective reduction, furnish, for instance, fully protected **74**.



Scheme 2.32 Catalytic enantioselective synthesis of α-diazo-β-hydroxy esters and diastereoselective synthesis of protected α-hydrazino-β-hydroxy ester **74**.



75 76 77

Scheme 2.33 Resolution of protected piperazic acid prepared using a Diels–Alder reaction.

2.2.5
Piperazic Acid and Derivatives by Cycloaddition Reactions

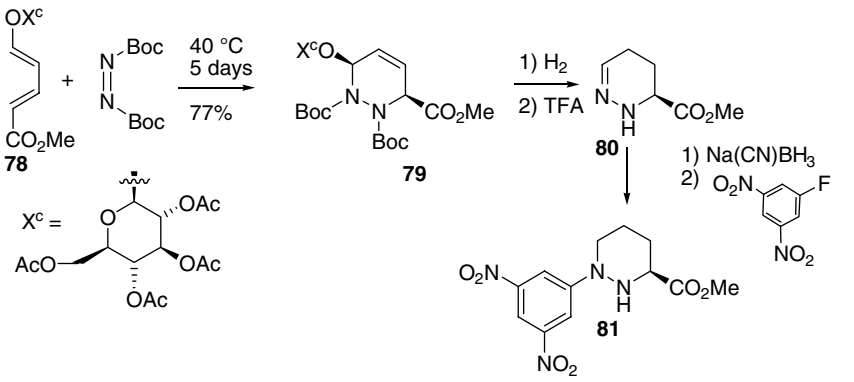
Hetero-cycloaddition reactions are pertinent tools to build the piperazic acid framework. The earliest synthesis used azodienophiles and dienes. These reactions have been conducted in stereoselective modes.

2.2.5.1 **Diels–Alder Cycloaddition**

The first reported synthesis of piperazic acid used diene **75** and phthalazinedione **76**. Hydrogenolysis, hydrolysis, and protection furnished racemic protected piperazic acid **77**, which was resolved using ephedrine (Scheme 2.33) [183].

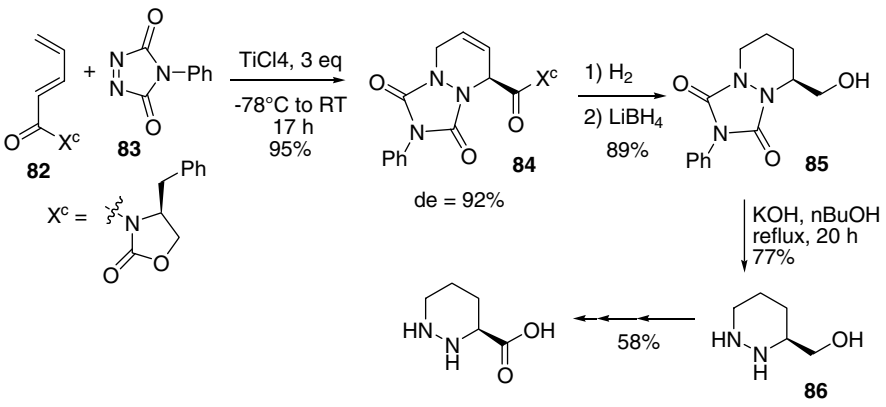
Two chiral auxiliaries have been introduced on the diene. Diene **78** carrying a glucopyranose auxiliary reacts slowly with di(*tert*-butyl)azodicarboxylate (TBAD) (Scheme 2.34) [184, 185]. Cycloadduct **79** is obtained in good yield and excellent stereoselectivity. Hydrogenation and acidolysis of the auxiliary easily furnishes hydrazone **80**. Reduction to the unstable piperazic ester and derivatization with Sanger's reagent gives the stable **81**.

Evans chiral auxiliary has also been introduced into diene **82** (Scheme 2.35) [186]. Reaction with azadienophile **83** is efficiently catalyzed by $TiCl_4$, and provides cycloadduct **84** in very good yield and diastereomeric excess (Scheme 2.35). However, further transformation to piperazic acid derivative proves to be long. Reductive



78 79 80 81

Scheme 2.34 Preparation of piperazic acid using a Diels–Alder reaction between di-*tert*-butyl azodicarboxylate and a chiral diene.

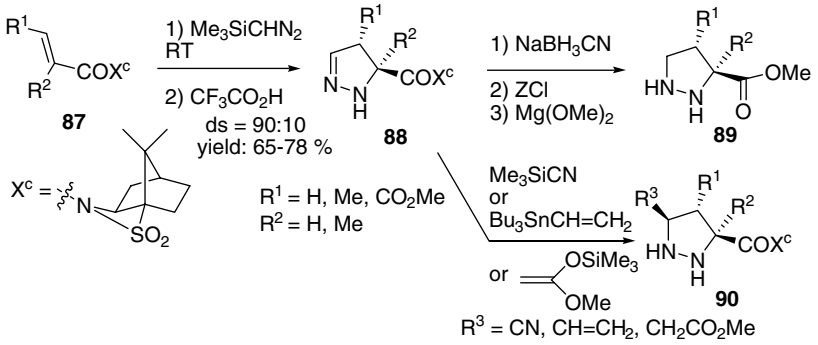


Scheme 2.35 Preparation of piperazic acid using a Diels–Alder reaction between 4-phenyltriazoline-3,5-dione and a chiral diene.

removal of the chiral auxiliary yields **85**, which is hydrolyzed to **86** in harsh conditions. Oxidation to piperazic acid needs nitrogen protection.

2.2.5.2 **1,3-Dipolar Cycloaddition**

Highly diastereoselective 1,3-dipolar cycloaddition between trimethylsilyl diazomethane and camphor sultam-derived dipolarophile **87** followed by acidic treatment gives the five-membered heterocycle **88** in good yields and good diastereoselectivity (Scheme 2.36) [187]. Further reduction, protection, and chiral auxiliary removal provides aza analogs of proline **89**. The electrophilic unsaturated carbon in hydrazones **88** can also be functionalized using $Bu_3SnCH_2CH{=}H_2$, Me_3SiCN or silylketene acetals (see reactions developed in Section 2.2.4) to give, stereoselectively, compound **90** [188].



Scheme 2.36 1,3-Dipolar cycloaddition between trimethylsilyl diazomethane and camphor sultam-derived dipolarophile.

2.3 Chemistry

The two reactive functions in α-hydrazino acid derivatives give rise to a very rich chemistry. In comparison with analogous α-amino acids [189], reactivity is due to the presence of the two nucleophilic nitrogen atoms and the possible cleavage of the rather weak N−N bond. Reduction of this bond is the last step in the asymmetric synthesis of α-amino acids using hydrazones or azodicarboxylates. Owing to their close structural relationship to the naturally occurring α-amino acids, α-hydrazino acids have been incorporated in peptides. Methods of peptide synthesis have been adapted to that goal and methods for the regioselective N^1- or N^2-acylation have been developed since the mid-1960s. Only specific reactions of the carboxyl group due to the presence of the hydrazino function are covered herein. The formation of heterocycles from α-hydrazino acids is also reviewed.

Obviously, general properties of compounds having N−N bonds apply to α-hydrazino acids derivatives [190]. Air oxidation of free hydrazines may occur, especially in a basic medium. However, solid zwitterionic α-hydrazino acids and also hydrazides (R−CO−N−N) or carbazates (R−O−CO−N−N) are quite stable at RT and in air. Hydrazines are less basic than the corresponding amines (Table 2.3) and the extra nitrogen in hydrazides $RCONHNH_2$ or in N^2-protected α-hydrazino acids is weakly basic. In N^2-protected α-hydrazino acids, the acidity of the carboxyl function is enhanced and similar to that of unprotected amino acids.

2.3.1 Cleavage of the N−N Bond

Cleavage of the N−N bond in nonracemic α-hydrazino acid derivatives **91** is a very convenient method to prepare α-amino acids **92** substituted by a wide diversity of side-chains R^4 or R^5 (Table 2.4). Methods to reduce hydrazines to amines have been known for a long time [190]. In the 1960s, racemic N^1-methyl α-hydrazino acids had been converted into *N*-methyl α-amino acids by two methods: quantitative cleavage by nitrous acid or hydrogenolysis using Raney nickel in ethanol [192]. Since the first efficient asymmetric synthesis of α-hydrazino acid derivatives in 1986, basically two methods (hydrogenolysis and reductive cleavage using samarium iodide) have been applied for their conversion into α-amino acids. In most cases, the conditions required to cleave that bond depend markedly on substituents R^1, R^2, or R^3 (Table 2.4). Avoiding racemization or reduction of the side-chains R^4 or R^5 are the main difficulties encountered.

Hydrogenolysis of N^1,N^2-unsubstituted α-hydrazino acid derivatives readily occurs using Raney nickel [109, 111, 127, 128, 146]. This inexpensive reagent was used in a large-scale synthesis of AS-3201 – a potent aldose reductase inhibitor [146]. When N^1 and N^2 are substituted by alkoxycarbonyl groups, no reaction occurs [111] or low yields are obtained [132]. Epimerization was observed in one case [193]. Hydrogenolysis of α-hydrazino acid derivatives that are N^1,N^2-unsubstituted or substituted by aryl or alkyl substituents also occurs using palladium hydroxide [181] or platinum

Table 2.3 pK_as of some amine, hydrazine, and acid derivatives.

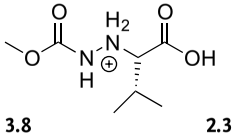
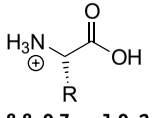
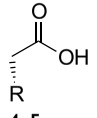
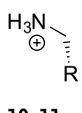
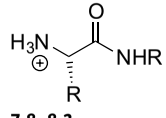
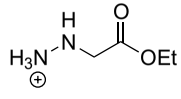
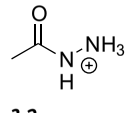
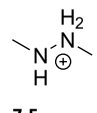
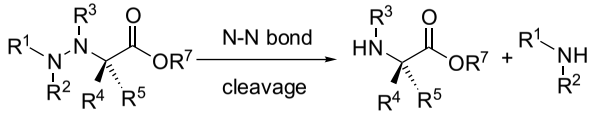
										
pK_a (H_2O)	**3.8**	**2.3**	**8.8–9.7**	**1.9–2.3**	**4–5**	**10–11**	**7.8–8.3**	**5.9**	**3.2**	**7.5**
Reference	[191]		[189]				[189]	[61]	[190]	[190]

Table 2.4 Reagents and conditions for N–N bond cleavage.



91 **92**

R^4 = H, Me, CN, CO_2Et

R^5 = Me, Et, ^{i}Pr, ^{t}Bu, C_6H_{11}, CH_2Ph, Ph, allyl, aryl, heteroaryl, CO_2Et, CHOHMe...

Reagent	R^1	R^2	R^3	Conditions	Remark	Reference
H_2 Raney nickel	H^+, H	H	H	Up to 550 psi, EtOH, RT		[109, 111, 127, 128, 146]
	H	Boc	Boc	550 psi, EtOH	no reaction	[111]
	Me	Me	H		epimerization	[193]
	H	Cbz	Cbz	high pressure, $R^4 \neq H$ and $R^5 \neq H$	yield <10%	[132]
H_2 $Pd(OH)_2$	Ph	Ph	H	sulfonic acid, MeOH, RT		[181]
H_2 PtO_2	H^+, H	H	H	Up to 80 psi, EtOH or water, RT	reduction of C_6H_5 in C_6H_{11}, racemization on pH increase	[108, 110]
	H^+, H	alkyl	Ac			[188]
	Me	Me	H			[193]
H_2 Pd/C	Cbz	H	Cbz		Cbz or CH_2Ph cleavage, no N–N cleavage	[109, 111]
	H	H or COR^6	CH_2Ph	10% Pd/C, up to 50 psi		[59, 68, 156, 173]
	H	CO_2R^6	COPh			

(*Continued*)

Table 2.4 (*Continued*)

Reagent	R^1	R^2	R^3	Conditions	Remark	Reference
	H	Cbz	CH_2Ph		Cbz, CH_2Ph and N–N cleavage	[165]
SmI_2	H	PhCO	H	THF/MeOH 0 °C to RT		[170, 175]
	H	PhCO	COR^6			[177]
	H	PhCO	CO_2R^6			[174, 180]
	H	CO_2R^6	COPh			[176]
	H	$COCF_3$	CO_2R^6			[131–133, 135]
	Ac	Boc	CO_2R^6			[140]
Zn	H	H	CO_2R^6	acetone		[127, 134]
HNO_2	H	H	COR^6	$NaNO_2$, AcOH, HCl, reflux		[131]
Electrochemistry	alkyl	COPh	COPh	reduction potential around −1.75 V (saturated calomel electrode)	LiOAc/AcOH mixture at pH 7–8	[195]

oxide [108, 110, 188, 193]. However, these conditions are not suitable with aromatic side-chains since they are reduced to cyclohexyl. Hydrogen and palladium on charcoal are selective for benzyl or benzyloxycarbonyl groups and are generally unable to cleave N−N bonds [59, 68, 109, 111, 156, 173], except in the case of aminoserine, which has been converted to serine [165]. Very efficient reductive cleavage of N^2-benzoyl hydrazines using SmI_2 was described in 1992 [170] and has gained prominence in many recent studies [140, 174–177, 180]. The reaction proceeds at RT within 30 min in MeOH/THF mixtures. Reduction of N−N is selective in the presence of unsaturated side-chains [174]. It has also been shown that N^1-trifluoroacetylation also activates the N−N bond cleavage by SmI_2 [194]. In the case of N^1,N^2-dialkoxycarbonyl substitution, cleavage using SmI_2 does not work. Treatment with trifluoroacetic anhydride in the presence of pyridine allows N^2-trifluoroacetylation and activation of the reductive cleavage [131–133, 135]. N−N bond cleavage using zinc [127, 134], nitrous acid [131], or electrochemistry [195] has been also described.

2.3.2
Reactivity of the Hydrazino Function

To the best of our knowledge, intermolecular alkylation of α-hydrazino acid derivatives has not been described. However, reactions of α-hydrazino acid derivatives with numerous acylating reagents used in peptide synthesis have been studied since the mid-1960s. Results are reviewed according to the substitution pattern of the hydrazine moiety (Figure 2.12). Nonsubstituted hydrazines **93** are more reactive than the monosubstituted **94** or **95**, but difficulties in the regioselective acylation may occur. Guidelines for the efficient and regioselective incorporation of α-hydrazino acid residues in peptides are provided. The easy hydrazone formation from **93** or **94** is also described.

2.3.2.1 Reaction of Unprotected α-Hydrazino Acid Derivatives with Acylating Reagents

Owing to the presence of two reactive nitrogen atoms, acylation of **93** may produce a mixture of monoacylated **96** and **97** or diacylated **98** (Table 2.5). It is well known that the most substituted nitrogen in hydrazines is more nucleophilic [190]. However, the site of acylation is markedly sensitive to steric effects as well as to the nature of the acylating agent [190]. The influence of the reagent and the side-chain R^1 on the regioselective protection of **93** by Boc or Cbz has been studied (Table 2.5). The major product is **97** when ethyl hydrazino acetate is acylated with benzylchloroformate (entry 2) or pyrocarbonates (entries 3 and 4). Using the activated ester, CbzOSu (entry 5), the regioselectivity is reversed and **96** is obtained in good yield. In the case of

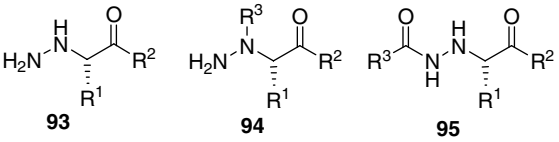
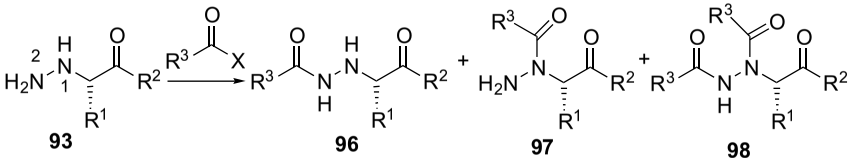


Figure 2.12 Starting α-hydrazino acid derivatives used in acylation reactions.

Table 2.5 Regioselective acylation of unprotected α-hydrazino acid derivatives **93**.



	R^1	R^2	R^3COX	Yield **96** (%)	Yield **97** (%)	Yield **98** (%)	Reference
1	H	OH	CbzCl		41		[44]
2	H	OEt	CbzCl/DMAP	10	65	15	[78]
3	H	OEt	Cbz_2O	10	80	5	[78]
4	H	OEt	Boc_2O	10	80	5	[78]
5	H	OEt	CbzOSu	85	10	0	[78]
6	CH_3	OH	CbzCl	1[a]	1[a]	1[a]	[55]
7	CH_2Ph	OH	CbzCl	55	—	—	[48]
8	CH_3	OMe	Boc_2O	85	15	1–2	[196]
9	H	OEt	DCC, Boc-Ile	78			[52]
10	H	OEt	Fmoc-Pro-OSu	80	—	—	[74, 78]
11	H	peptide	$CH_3(CH_2)_{14}CO_2Su$, pH: 5.1	69			[61]
12	CH_3	OEt	Cbz-AA-ONp or DCC Cbz-AA (AA = Asn, Phe, Ala)	46–65			[55]
13	CH_3	Ala-NH*i*Pr	Cbz-Val-OSu	40			[59]
14	CH_3	Gly-OEt	Boc-Phe-OCO_2*i*Bu or DCC/HOBt, Boc-Phe	58			[196]
15	CH_2Ph	OEt	DCC, Cbz-Gly or Cbz-Ala	54–60			[48]
16	*i*Pr	Ala-Ala-NH*i*Pr	Cbz-Ala–Ala–Pro-OSu	25			[59]
17	*i*Pr	Ala-Ala-NH*i*Pr	Boc-Ala–Ala–Pro-OCO_2*i*Bu	55			[59]
18	(*R*)-CHMeEt	Leu-OMe	Cbz-Val-OCO_2*i*Bu	60			[59]

DCC = *N*,*N*′-dicyclohexylcarbodiimide; DMAP = 4-dimethylaminopyridine; HOBt = 1-hydroxybenzotriazole.
[a] 1/1/1 mixture was obtained.

the more hindered aminoalanine (entry 6) or aminophenyl alanine (entry 7) acylation at N^2 by CbzCl is increased and becomes the major reaction with Boc_2O (entry 8).

N^2-Acylated products **96** are the only products in reactions of hydrazino acetate derivatives with activated protected amino acids (entries 9 and 10) or succinimidyl fatty esters (entry 11). More hindered hydrazino acid derivatives ($R^1 = CH_3$, CH_2Ph, *i*Pr, or CHMeEt) are selectively acylated at N^2 by active esters or mixed anhydrides in fair to good yields (entries 9–18). In two cases (entries 13 and 16) yields obtained with the corresponding α-amino acid derivatives are similar (32 and 44%, respectively). Reaction times are generally longer than that needed for the corresponding α-amino acids. In conclusion, the N^1-acylated product **97** is produced only in the case of aminoglycine derivatives reacting with chloroformates or pyrocarbonates. In all the other cases, steric hindrance of the side-chain or the use of activated amino acids favors the formation of the N^2-acylated derivative **96**.

2.3.2.2 Reaction of N^1-Substituted α-Hydrazino Acid Derivatives with Acylating Reagents

Steric hindrance of the side-chain in **94** and of the N^1 substituent greatly reduces the N^2 nucleophilicity towards acylating reagents (Table 2.6). Activation of acids using carbodiimides allows the acylation of hydrazino acetic derivatives **94** ($R^1 = H$) in fair to very good yields, in solution-phase synthesis (entries 1–3) [43, 49, 84, 87] and in solid-phase synthesis (entry 4) [70]. Reaction times are longer than those used with α-amino acid derivatives. In the case of amino (Cbz)-alanine (entries 5–9), coupling conditions to Fmoc-alanine have been optimized and the reaction is rather difficult [101]. Acyl fluoride or acyl chloride prove to be the best reagents. Yields are fair when steric hindrance is important (entry 10). Acylation of N^1-benzyl substituted α-hydrazino acid derivatives is also difficult (entries 11–16) and strong acylating reagents are needed for the reaction to proceed [59, 68].

2.3.2.3 Reaction of N^2-Protected α-Hydrazino Acid Derivatives with Acylating Reagents

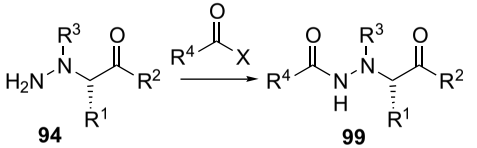
The N^1 atom in **95** is weakly basic and sterically hindered. Reaction with acylating reagents is generally difficult (Table 2.7). Acyl chlorides, *N*-carboxy anhydrides (NCAs), or acid activation using O-(7-azabenzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluroniumhexafluorophosphate (HATU) are the best reagents, but yields of **100** are fair.

Piperazic acids occur in nature as components of peptide natural products and they are acylated at N^1 except in the case of Sanglifehrin. In a review, Ciufolini and Xi pointed out that N^1 acylation of piperazic acid derivatives is also problematic, and have proposed reasons for this unusually poor nucleophilicity [31]. Efficient acylation reagents such as acyl chlorides were generally used for the N^1-acylation to proceed.

2.3.2.4 Reaction with Aldehydes and Ketones

In the 1960s, smooth formation of hydrazones from N^2-unsubstituted α-hydrazino acid derivatives was used for characterization or purification. Acetone hydrazone as an N^2-protecting group allowed Niedrich to prepare *N*-aminopeptides [47, 48]. Intramolecular hydrazone formation from α-hydrazino and δ-oxo functions is generally one of the last steps in dehydropiperazic acid synthesis (see Section 2.3.4.1 and Table 2.10).

Table 2.6 Acylation of N^1-substituted α-hydrazino acid derivatives.

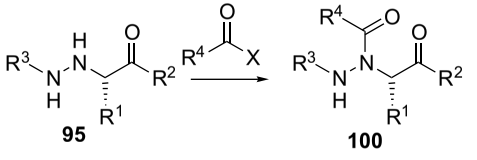


	R^1	R^3	R^2	R^4COX	Yield **99** (%)	Reference
1	H	Boc	OEt	DCC, Cbz-Ile	78	[49]
2	H	Cbz	OEt, OMe	DCC Cbz-Gly or Cbz-Val or Cbz-Pro or Cbz-Asn or Cbz-Pro–Leu	40–92	[43]
3	H	Me, *i*Pr, *i*Bu, CH_2Ph	NH-NR^3-CH_2COR^2	DCC/DMAP BocNH-NR^3-CH_2CO_2H	55–95	[84, 87]
4	H	Me, *i*Bu, CH_2Ph, $(CH_2)_4NHBoc$, $(CH_2)_3NH(NBoc)NHBoc$, $CH_2C_6H_4OCH_2OEt$	peptide chain	DIC, HOBt, Fmoc-AA (AA Tyr (CH_2OEt), Ala, Leu, Lys (Boc), Arg (Boc_2), Gln, Gly), solid-phase synthesis	—	[70]

5	CH_3	Cbz	OMe	HATU, Fmoc-Ala	30	[101]
6	CH_3	Cbz	OMe	DCC/HOBt, Fmoc-Ala	73	[101]
7	CH_3	Cbz	OMe	Fmoc-Ala- OCO_2*i*Bu	86	[101]
8	CH_3	Cbz	OMe	Fmoc-Ala-Cl	90	[101]
9	CH_3	Cbz	OMe	Fmoc-Ala-F	98	[101]
10	(*R*)-CHMeEt	Cbz	OMe	Fmoc-Val-F	68	[101]
11	CH_3	CH_2Ph	Ala-NH*i*Pr	Boc-Val-OSu	no reaction	[59]
12	CH_3	CH_2Ph	Ala-NH*i*Pr	Boc-Val-OCO_2*i*Bu	22	[59]
13	CH_3	CH_2Ph	Ala-NH*i*Pr	Boc-Val/PyBop	28	[59]
14	CH_3	CH_2Ph	Ala-NH*i*Pr	Boc-Val-NCA	60	[59]
15	CH_3	CH_2Ph	NH-($PhCH_2$)Leu-OMe	Boc-NH-($PhCH_2$)Val, HATU	88	[68]
16	*i*Bu	CH_2Ph	OMe	Boc-NH-($PhCH_2$)Ala, EDC/HOBt or HATU	76–87	[68]
17		$(CH_2)_3$	Val-Ala-Ala-NH*i*Pr	Boc-Ala-OSu	42	[59]

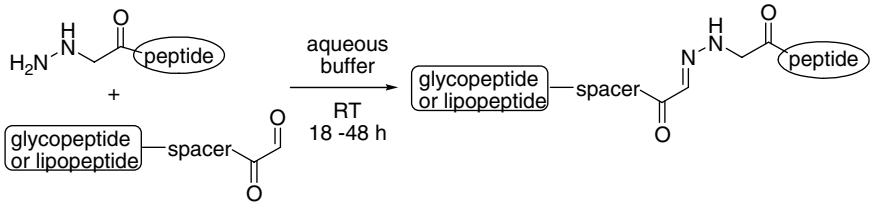
EDC = *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide; DIC = *N*,*N*′-diisopropylcarbodiimide.

Table 2.7 Acylation of N^2-protected α-hydrazino acid derivatives **95**.



	R^1	R^3	R^2	R^4COX	Yield 100 (%)	Reference
1	H	Cbz-Gly	OEt	Cbz-Gly-CO_2iBu	70	[47]
2	CH_2Ph	Boc	OEt	Cbz-Gly-CO_2iBu	24	[48]
3	*i*Bu	Alloc	OMe	HATU, DIEA, Fmoc-Phe	30	[197]
				(*S*)-$PhCH_2CH(N_3)COF$	46	
4	CH_2OCH_2Ph	Cbz	OMe	Fmoc-Ser(CH_2Ph)-NCA	32	[65]
5	$(CH_2)_2(OR^5)_2$	Boc	OMe	Fmoc-AA-Cl	52–65	[117, 198]
6	$CH(OAc)CH_2(OR^5)_2$	Boc	O*i*Bu	Ac-AA-Cl	60	[199]

DIEA = diisopropylethylamine.

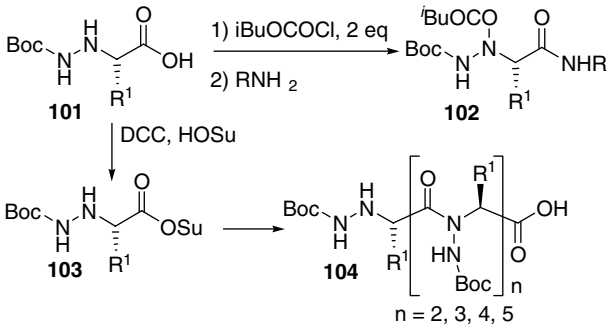


Scheme 2.37 Chemical ligation using α-hydrazino acetyl peptides.

More recently, α-hydrazino acetyl peptides, prepared by solid-phase functionalization of peptides [200], have proved to be valuable partners in the hydrazone chemical ligation with oxo-aldehydes derivatives giving a straightforward access to synthetic peptide conjugates and high-molecular-weight constructs (Scheme 2.37) [62, 63, 201].

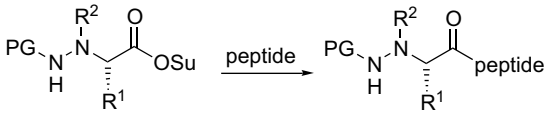
2.3.3
Reactivity of the Carboxyl Function

Generally, reactivities of the carboxyl or ester functions in α-hydrazino acid derivatives are similar to those of analogous α-amino acids. Activation of the carboxyl group and subsequent peptidic coupling have been realized using several reagents. In the case of N^1,N^2-diprotected α-hydrazino acids, activation and coupling proceed without any problems using liquid-phase [43, 59, 68, 84, 87] or solid-phase synthesis techniques [70, 202]. No detectable racemization occurs when Boc-NH-(CH_2Ph) Ala-OH is coupled to Ala-NH*i*Pr via isobutyl chloroformate activation [59]. Owing to the very weak nucleophilicity of the N^1 atom, the use of N^2-monoprotected α-hydrazino acids **101** in amide formation may be pertinent. Activation methods must be carefully chosen in order to avoid N^1-acylation. Isobutyl chloroformate reacts selectively with the N^1 nitrogen and a second equivalent is needed in order to activate the carboxyl group, so leading to N^1,N^2-diprotected α-hydrazino amides **102** (Scheme 2.38) [59].



Scheme 2.38 Side-reactions occurring during carboxyl activation of N^2-monoprotected α-hydrazino acids.

Table 2.8 Use of N^1-protected α-hydrazino activated esters in peptide coupling.



105

	PG	R^1	R^2	Peptide	Yield **105** (%)	Reference
1	Cbz	CH_2Ph	H	Gly–Gly-OEt	70	[50]
2	Boc	CH_2Ph	H	Ile–Gly–Leu–Met-NH_2	92	[50]
3	Boc	CH_3	H	Ala-NH*i*Pr	44	[59]
4	Boc	CH_3	CH_2Ph	Ala-NH*i*Pr	46	[59]
5	Boc	(*R*)-CHMeEt	H	Leu-OMe	67	[59]
6	Boc	(*R*)-CHMeEt	CH_2Ph	Leu-OMe	72	[59]

PG = protecting group.

Succinimidyl **103** or 1,3,4-trichlorophenyl activated esters ($R^1 = CH_2Ph$ [48, 50, 55] or CH_3 [55]) can be isolated in fair to good yields (40–80%) by crystallization at low temperature (−15 °C). However, partial to total decomposition may occur after recrystallization of **103** from refluxing methanolic solution. Oligomers **104** resulting of the self N^1-acylation were observed by mass spectrometry of the recrystallization mother liquors [59].

Nevertheless, peptidic coupling of isolated active esters **103** to free N-terminal peptides gives very good yields [50], even in the case of hindered residues (Ile, Table 2.8, entry 2), because N-free peptides are far more nucleophilic than the unprotected N^1 of the hydrazino moiety (Table 2.8, entries 1 and 2). Similar results are obtained when in situ generated activated esters from N^1-subtituted or N^1-unsubstituted hydrazino acids are coupled to amino acids (Table 2.8, entries 3–6).

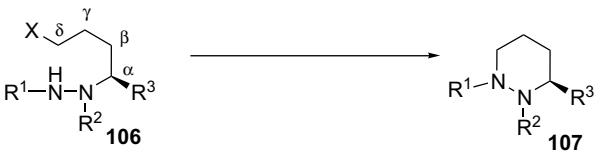
2.3.4 Synthesis of Heterocycles

The presence of at least two reactive functions in α-hydrazino acids has been used in cyclization reactions leading to heterocycles. Intramolecular reaction between the hydrazine group and an electrophilic function in the side-chain is the main access to piperazic acid derivatives [41, 42, 98, 103, 117, 119, 121, 124, 129, 166, 198, 199, 203, 204]. Other heterocycles (1-aminohydantoins [205], 5,5-fused heterocyclic ring systems [206, 207], or pyrazoles [208]) with interesting biological properties have been prepared from α-hydrazino acids.

2.3.4.1 Cyclization Leading to Piperazic Acid Derivatives

N,N-Diacylated α-hydrazino compounds **106** which are substituted at C^δ by a good leaving group X are excellent precursors to piperazic acid derivatives **107**

Table 2.9 Cyclization reaction leading to piperazic acid derivatives.



	R^1	R^2	R^3	Chain Substitution	Reagent	X	Yield 107 (%)	Reference
1	Boc	Boc	COX[c]	—	LDA, DMPU	Br	60	[42]
2	Boc	Boc	COX[c]	(*S*)γ-OCH_2Ph	LDA, DMPU	Br	60	[119]
3	Ac	Cbz	CO_2iPr	—	NaH	OMs	84	[98]
4	Cbz	Cbz	CH_2OR	—	NaH	Br	97	[41]
5	CO_2iPr	CO_2iPr	CH_2OR	(*S*)β, (*R*)γ-$OCMe_2O$	NaH	OTs	81	[129]

LDA = lithium diisopropylamide; DMPU = *N,N′*-dimethypropyleneurea.

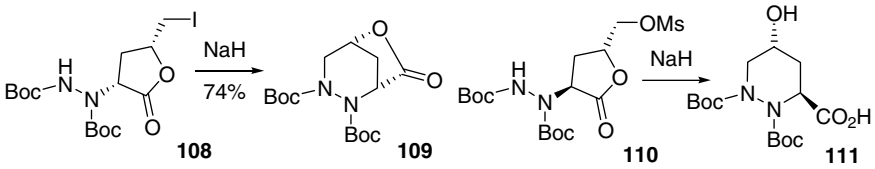
(Table 2.9). Deprotonation of the hydrazide function results in a smooth cyclization reaction, which is the key step in the most used method for piperazic acid preparation (entry 1) [42]. A promising precursor to piperazic acid (entry 4) is prepared in high yields and the R^3 hydroxymethyl group was further oxidized to the acid [41].

Diastereomeric γ-hydroxy piperazic acid derivatives **109** [120] and **111** [126] have been prepared from the corresponding lactones **108** and **110** via an efficient cyclization reaction (Scheme 2.39).

The hydrazone function in dehydropiperazic derivatives **114** is generally created from hydrazino aldehyde **113** (Table 2.10). The aldehyde function is generated in situ from acidic hydrolysis of acetal **112** (entries 1–8) or by oxidation of the corresponding alcohol (entry 9). Subsequent cyclization occurs nearly quantitatively in acidic media, which is compatible with β-hydroxy substitution of the side-chain.

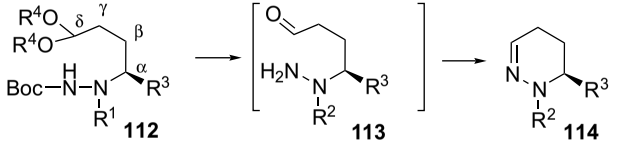
2.3.4.2 Other Heterocycles

In order to discover new bioactive compounds, scientists at Procter and Gamble developed two general solid-phase syntheses of heterocyclic scaffolds, 1-aminohydantoins **115** (Scheme 2.40) [205] and 5,5-fused heterocyclic ring systems **116**

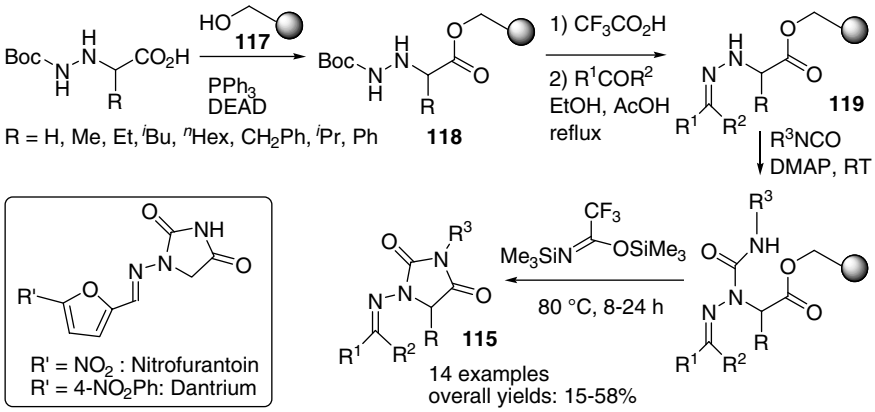


Scheme 2.39 Intramolecular substitution reaction leading to γ-hydroxy piperazic acid derivatives.

Table 2.10 Cyclization reaction leading to dehydropiperazic acid derivatives **114**.



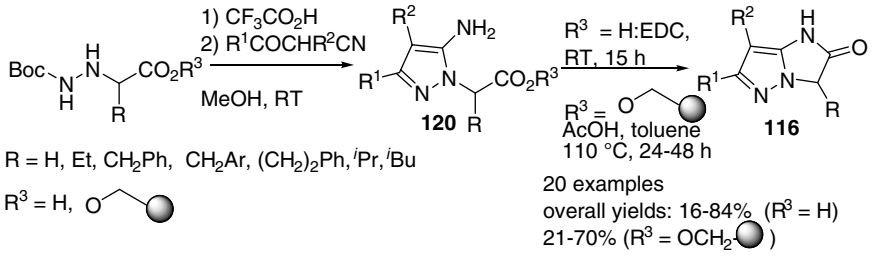
	R[1]	R[2]	R[3]	R[4]	Chain substitution	Reagent	Yield **114** (%)	Reference
1	Boc	H	CO_2Me	-CH_2CH_2-	—	TFA	90	[117]
2	Boc	H	CO_2Me	CH_3	—	TFA	95	[203]
3	Cbz	Cbz	CO_2Me	-$CH_2CMe_2CH_2$-	—	TFA	87	[98]
4	RCO	RCO	CO_2Me	-$CH_2CH_2CH_2$-	—	TFA	100	[198]
5	Boc	H	CO_2Et	CH_3	(*S*)-β-OH	TFA	70–100	[103, 199]
6	RCO	RCO	CO_2iBu	-CH_2CH_2-	(*S*)-β-OAc	TFA	97	[199]
7	RCO	RCO	CO_2Bn	-$CH_2CH_2CH_2$-	(*S*)-β-OH	TFA	55	[204]
8	H	H	CO_2Me	CH_3	(*S*)-β-OH	TFA	100	[121, 124]
9	Bn	Bn	CO_2Me	**113** generated by corresponding alcohol oxidation	—	TFA	76	[166]



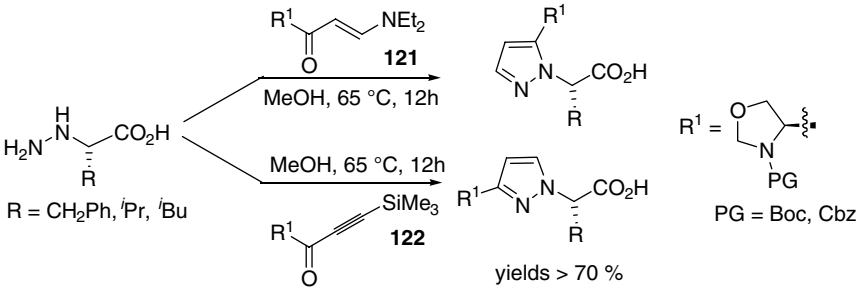
Scheme 2.40 Synthesis of 1-aminohydantoin library from α-hydrazino acids.

(Scheme 2.41) [206, 207], using α-hydrazino acids. 1-Aminohydantoins **115** have important pharmacological properties (Scheme 2.40). For instance, nitrofurantoin and dantrium are, respectively, anti-infective and skeletal muscle relaxant drugs. N^2-Boc α-hydrazino acids were coupled to hydroxymethylene polystyrene resin **117**. After deprotection of **118**, hydrazone **119** was readily prepared and then functionalized at N^1 using isocyanates or equivalent reagents. Traceless cyclizative cleavage from the resin afforded a library of 1-aminohydantoins **115** bearing four points of diversity in good overall yields (average yield per step 73–92%).

5,5-Fused heterocyclic ring systems **116** were shown to inhibit both interleukin-1 and tumor necrosis system ($R^1 = pF\text{-}C_6H_4$, $R^2 = 4$-pyrimidyl) or to inhibit mitogen-activated protein kinase with anti-inflammatory activity (Scheme 2.41) [206, 207]. Libraries containing this scaffold were prepared using solution-phase synthesis ($R^3 = H$) [206] or solid-phase synthesis ($R^3 =$ hydroxymethylene polystyrene) [207]. The hydrazine function of α-hydrazino acid derivatives was reacted with malononitriles to give the expected aminopyrazoles **120**. The carboxyl function allowed cyclization and, thus, access to a library of 5,5-fused heterocyclic ring systems **116**.



Scheme 2.41 Solution-phase and solid-phase synthesis of fused pyrazoles from α-hydrazino acids.



Scheme 2.42 Regioselective synthesis of pyrazole from α-hydrazino acids.

Other regioselective pyrazole syntheses from α-hydrazino acids were achieved using enamino ketones **121** or silyl acetylenic ketone **122** (Scheme 2.42) [208]. The isomeric pyrazoles obtained were precursors to mimics of the cis amide bond in peptides.

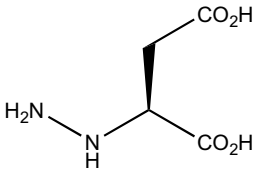
2.4 Conclusions

This chapter covers a century of chemistry and synthesis of nonracemic α-hydrazino acids and piperazic acid derivatives. The early preparations used resolution or starting materials from the chiral pool, such as α-amino, hydroxy, or halogeno acids. Then, chiral auxiliaries, mainly those of Evans and Oppolzer, were applied to highly stereoselective preparations. Currently, intensive efforts are made in the search for enantioselective catalytic reactions using essentially azodicarboxylates or hydrazones. Diversely substituted α-hydrazino acids have been obtained. Derivatives bearing on the α-carbon not only the side-chains present in proteinogenic α-amino acids, but also a lot of functionalized substituents have been prepared. Methods producing quaternary α-carbons are also available. Substitution of the two nitrogens has also been controlled and substituents are mainly protecting groups. Preparation of α-hydrazino acid derivatives has greatly developed because the N−N−C−C=O fragment is found in drugs that have been marketed for a long time (carbidopa or cilazapril) or in several bioactive products. They are also very important intermediates in the asymmetric synthesis of α-amino acids. Chemoselective cleavage of the N−N bond by mild methods avoiding racemization has been proposed. Owing to their close structural relationship to the naturally occurring α-amino acids, α-hydrazino acids have been included in peptidic structures using conventional techniques of peptide synthesis. The question of the mono or the orthogonal bis protection of the two nitrogens in synthons useful for peptide synthesis is still open. Except in the case of hydrazino acetic derivatives, the hydrazino moiety is generally weakly nucleophilic, thus requiring strong acylating reagents. The presence of at least two reactive

functions in α-hydrazino acid derivatives has been used in the synthesis of several bioactive heterocycles.

2.5
Experimental Procedures

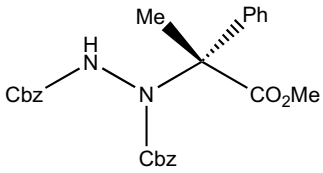
2.5.1
(S)-2-hydrazinosuccinic Acid Monohydrate [22]



A solution of tetra-*N*-methylammonium hydroxide (9.10 ml, 20% in hexanes, 19 mmol) was added to a solution of (*R*)-malic acid (1.34 g, 10.0 mmol) in methanol (10 ml). The mixture was concentrated *in vacuo* and the white solid was suspended in 3 : 1 dry DMF/acetonitrile (15 ml). Benzyl bromide (2.40 ml, 20 mmol) was added and the mixture was stirred for 21 h. Ether (150 ml) was added and the mixture was filtered. The filtrate was washed with water (3×50 ml), dried (Na_2SO_4), and concentrated *in vacuo*. The residual yellow oil was purified using flash chromatography (35% EtOAc/hexanes) to afford dibenzyl (*R*)-2-hydroxysuccinate (2.73 g, 86% yield) as a colorless, homogeneous oil; $[\alpha]_D = +19.3$ (*c* 1.9, $CHCl_3$). A solution of dibenzyl (*R*)-2-hydroxysuccinate (1.57 g, 5.0 mmol) and pyridine (0.5 ml, 6.2 mmol) in dichloromethane (5 ml) was added to a solution of trifluoromethanesulfonic anhydride (0.9 ml, 5.3 mmol) in dichloromethane (10 ml) at $-20\,^{\circ}C$. The mixture was stirred at RT for 1 h. After three successive additions of CCl_4 (3×10 ml) and concentration *in vacuo* to about 5 ml each time, the mixture was filtered through a plug of anhydrous Na_2SO_4. The filtrate was concentrated *in vacuo* to give the triflate as an oil (2.21 g, 99%). Benzyl carbazate (1.66 g, 10 mmol) was added to a solution of the triflate (2.21 g, 5.0 mmol) in dichloromethane (10 ml) and the mixture was stirred for 24 h. Water (5 ml) and saturated aqueous $NaHCO_3$ (5 ml) were added and the mixture was extracted with ether (3×35 ml). The ether extracts were dried (Na_2SO_4) and concentrated *in vacuo*. The residual oil was purified by flash chromatography (45% EtOAc/hexanes) to give dibenzyl (*S*)-2-(N^2-benzyloxycarbonylhydrazino)succinate as a colorless oil (2.07 g, 89% yield); $[\alpha]_D = +7.7$ (*c* 2, $CHCl_3$). A solution of dibenzyl (*S*)-2-(N^2-benzyloxycarbonylhydrazino)succinate (1.0 g, 2.2 mmol) in 10% AcOH/EtOAc (10 ml) was hydrogenated for 8 h in the presence of 5% Pd/C catalyst (0.1 g). The mixture was filtered through a Celite pad and the filtrate was concentrated *in vacuo*. The white solid was recrystallized from THF/water to give (*S*)-2-hydrazinosuccinic acid monohydrate (0.191 g, 52% yield);

enantiomeric excess greater than 99%; melting point (m.p.) 116–118 °C; $[\alpha]_D = -14.2$ (c 1, H_2O).

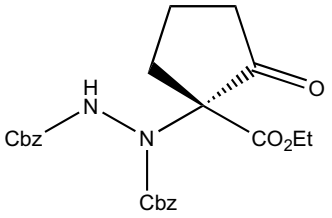
2.5.2
(−)-(*R*)-N^1,N^2-dibenzyloxycarbonyl-2-hydrazino-2-phenyl Propionic Acid, Methyl Ester [130, 131]



A suspension of L-proline (102.6 mg, 0.9 mmol) in dichloromethane (14 ml) was stirred at RT for 30 min, followed by addition of 2-phenylpropionaldehyde (362 mg, 2.7 mmol) at 0 °C. After 1 h of stirring at RT dibenzyloxycarbonyl azodicarboxylate (537 mg, 1.8 mmol) was added and the mixture was stirred at RT under argon until the color of the azodicarboxylate disappeared (3 days). The reaction was then quenched by addition of water and the aqueous phase was extracted 3 times with diethyl ether. The combined organic phases were dried with magnesium sulfate and the solvent was removed by evaporation at reduced pressure. Subsequent flash chromatography on silica delivered N^1,N^2-bis (benzyloxocarbonyl)-2-hydrazino-2-phenyl propionaldehyde (0.65 g, 1.5 mmol, 83% yield). N^1,N^2-Bis(benzyloxocarbonyl)-2-hydrazino-2-phenyl propionaldehyde (0.65 g, 1.5 mmol) was dissolved in acetonitrile (5 ml) along with NaH_2PO_4 (0.050 g, 0.4 mmol) in water (2 ml) and 35% H_2O_2 (0.14 ml, 1.5 mmol). To the resulting mixture was added a solution of $NaClO_2$ (0.23 g, 2.6 mmol) in water (5 ml) dropwise, keeping the temperature of the mixture below 10 °C. After the mixture was stirred for 12 h, Na_2SO_3 (0.015 g, 0.1 mmol) was added and the resulting mixture was acidified (pH 2–3) with 10% aqueous HCl. The resulting mixture was partitioned between brine and dichloromethane, the layers were separated. The aqueous layer was extracted with dichloromethane. The combined organic extracts were washed with brine, dried, and concentrated under reduced pressure to give a residue that was taken up in toluene/methanol 1 : 2 and treated with an ethereal solution of trimethylsilyldiazomethane at 0 °C for 5 min until bubbling subsided. The excess trimethylsilyldiazomethane was quenched with a few drops of acetic acid and the solvent was then removed *in vacuo*. Column chromatography on silica (diethyl ether/pentane, 1 : 1) delivered (−)-N^1,N^2-dibenzyloxycarbonyl-2-hydrazino-2-phenyl propionic acid, methyl ester (0.6 g, 1.3 mmol, 87% over two steps) of a colorless solid with 82% e.e. High-performance liquid chromatography (HPLC; Chiralcel OD, *n*-heptane/isopropanol 90 : 10, 0.7 ml/min): $R_{t(maj)} = 30.1$ min, $R_{t(min)} = 46.4$ min; $[\alpha]_D = -38.55$ (c 1.00, $CHCl_3$). The (*R*) configuration was assigned by chemical transformation into known compounds.

2.5.3

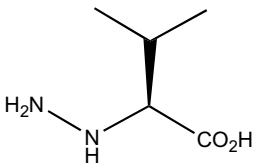
(+)-(*R*)-N^1,N^2-Bis(benzyloxycarbonyl)-1-hydrazino-2-oxocyclopentane Carboxylic Acid, Ethyl Ester [136]



To a solution of dibenzylazodicarboxylate (32.0 mg, 0.1 mmol) in CH_2Cl_2 (1 ml) at −25 °C was added ethyl 2-oxocyclopentane carboxylate (15.6 mg, 0.1 mmol) and cinchonine (5.9 mg, 0.02 mmol). The resulting yellow solution was stirred for 5 min at −25 °C. The reaction mixture was partitioned between CH_2Cl_2 (3 ml) and 0.5 M HCl (2 ml). The aqueous layer was extracted with CH_2Cl_2 (3 × 1 ml). The combined organic extracts were dried (Na_2SO_4) and concentrated *in vacuo*. Purification by flash chromatography afforded ethyl (*R*)-*N*,*N*′-bis(benzyloxycarbonyl)-1-hydrazino-2-oxocyclopentane carboxylate as a colorless glass (95% yield, 88% e.e.). HPLC (Chiralcel AS, hexane/isopropanol 80 : 20, 1 ml/min): $R_{t(maj)} = 17.2$ min, $R_{t(min)} = 36.1$ min; $[\alpha]_D = +2$ ($c = 1.1$, CH_2Cl_2).

2.5.4

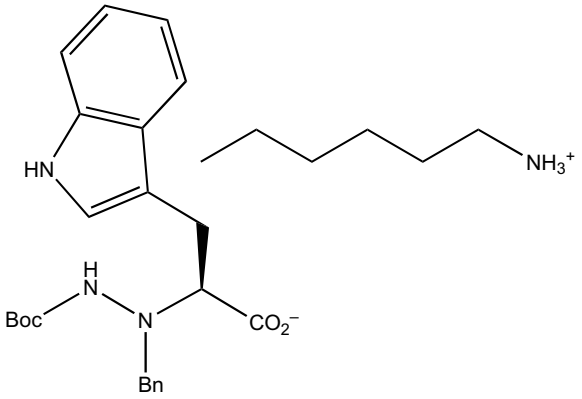
(−)-L-*N*-Aminovaline [151]



To a suspension of L-valine (29.25 g, 0.25 mol) in water (500 ml) was added KOCN (142 g, 1.75 mol). After 4 h at 60 °C under stirring, the mixture was cooled to 0 °C. The ureo derivative was precipitated by slow addition of 10 M HCl (270 ml) and collected (36.0 g, 90%). $[\alpha]_D^{23} = +15.7$ (*c* 1, 95% EtOH). To this ureido derivative (0.500 g, 3.12 mmol) in 5 M KOH (1.9 ml) at 0 °C and in the dark was added, dropwise, aqueous KOCl (2.01 M, 1.9 ml). This mixture was then stirred at RT for 19 h in the dark. On acidification (0 °C, 10 M HCl) a gummy material separated and was removed by ether extraction. The aqueous phase was percolated through a Dowex 1X2-100 ion-exchange resin (OH^- form). The resin was eluted with water until the eluent was neutral, then (−)-L-*N*-aminovaline hemi-hydrate (0.25 g, 61% yield) was recovered as a white solid after elution with 1 M aqueous acetic acid, lyophilization and recrystallization from boiling 95% ethanol. $[\alpha]_D^{25} = -16.8$ (*c* 0.8, 5 M HCl).

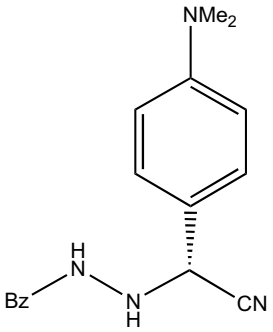
2.5.5

(+)-L-*N*-benzyl-*N*-(tert-butoxycarbonylamino)tryptophan, Hexylamine Salt [166]



To a mixture of L-tryptophan (4.08 g, 20 mmol) and NaOH (0.8 g, 20 mmol) in water (25 ml) was added freshly distilled benzaldehyde (2 ml, 20 mmol) at RT. The mixture was stirred for 1 h and then cooled to 0 °C before addition of $NaBH_4$ (0.23 g, 6 mmol) in small portions. After 2 h at RT the treatment with benzaldehyde and $NaBH_4$ was repeated. The resulting mixture was cooled to 0 °C and acidified with 6 M HCl to pH 6. The solid was collected, and washed with iced water and diethyl ether. Recrystallization from a boiling mixture of water (20 ml) and AcOH (40 ml) afforded L-*N*-benzyltryptophan as a beige solid (4.15 g, 72%); m.p. 256 °C (dec.); $[\alpha]_D^{25} + 10.7$ (*c* 1, HCl 6 M). A suspension of *N*-benzyltryptophan (3.52 g, 12 mmol) in MeOH (24 ml) was treated at 0 °C with a 2.2 M solution of Me_4NOH in MeOH (5.46 ml, 12 mmol). After stirring for 5 min the solution was concentrated *in vacuo*. To the residue dissolved in CH_2Cl_2 (60 ml) was added dropwise at −78 °C a solution of *N*-*tert*-butoxycarbonyl-3-trichloromethyl oxaziridine [163] (3.3 g, 12.6 mmol) in CH_2Cl_2 (60 ml). The cooling bath was allowed to warm slowly overnight. The reaction mixture was extracted with water (3 × 200 ml). The aqueous phase was washed with CH_2Cl_2 (30 ml), acidified with $KHSO_4$ (1.63 g) until pH 3–4, and then extracted with CH_2Cl_2 (3 × 150 ml). The CH_2Cl_2 phase was dried over Na_2SO_4 and concentrated *in vacuo* to afford crude L-*N*-benzyl-*N*-(*tert*-butoxycarbonylamino)tryptophan (4.60 g, 93%) as a glassy solid, pure enough for most purposes. A portion of this solid (0.331 g) was dissolved in ether (10 ml) and treated by hexylamine (0.132 ml, 1 mmol) to yield pure L-*N*-benzyl-*N*-(*tert*-butoxycarbonylamino)tryptophan hexylamine salt as a colorless solid (0.322 g, 63%); m.p. 142 °C; $[\alpha]_D^{25} + 28.6$ (*c* 1.8, MeOH).

2.5.6
(*R*)-2-(*N*^{2}-benzoylhydrazino)-2-(4-dimethylaminophenyl) Acetonitrile [172]

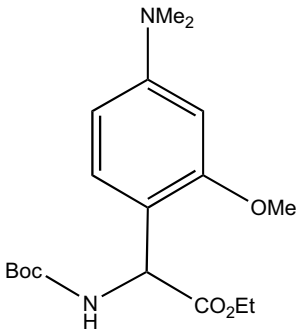


Catalyst preparation To a flame-dried flask under argon was added $ErCl_3$ (188 mg, 0.67 mmol, hygroscopic!), (*S*)-Ph-PyBOX (517 mg, 1.35 mmol, 2.0 equiv.), and dry acetonitrile (135.0 ml, 0.1 M in $ErCl_3$). The mixture was heated to reflux under argon for 8 h. The resulting cloudy mixture was concentrated to dryness by rotary evaporation, and the pinkish-white residue was redissolved in CH_2Cl_2 and filtered. The clear filtrate was concentrated to dryness under high vacuum to afford the $(Ph\text{-}PyBOX)_2$-$ErCl_3$ complex as a pale-pink foam (701 mg, 99% yield). The material was used without further purification.

Hydrazinonitrile preparation To the solid, preformed catalyst prepared as described above (5 mol%) was added under argon 4-dimethylaminobenzaldehyde, *N*-benzoylhydrazone, and freshly purified chloroform in order to obtain 0.25–0.1 M final concentration. The heterogeneous mixture was cooled to 0 °C and then charged with trimethylsilyl cyanide (2.0 equiv.) followed immediately by methanol (2.0 equiv.). The flask was sealed under an argon atmosphere and kept at 0–4 °C for 3 days with stirring. The product was isolated by adding the cold reaction mixture directly to a silica gel column and eluting with a CH_2Cl_2/EtOAc solvent mixture. (*R*)-2-(*N*^{2}-Benzoylhydrazino)-2-(4-dimethylaminophenyl) acetonitrile was obtained as a yellow foam (yield 85%): m.p. 79–81 °C, 97% e.e.; HPLC (Chiralcel OD, hexane/isopropanol 80: 20): $R_{t(maj)} = 24.3$ min, $R_{t(min)} = 13.5$ min; $[\alpha]_D^{22} = -6$ (*c* 0.47, CH_2Cl_2).

2.5.7

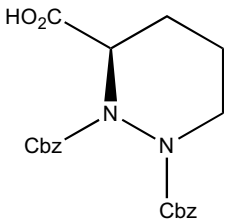
tert-Butoxycarbonylamino-(4-dimethylamino-2-methoxy-phenyl)-acetic Acid Ethyl Ester by Reduction using SmI_2 [180]



To a cooled (−78 °C) degassed solution of (N^2-benzoyl-N^1-*tert*-butoxycarbonyl-hydrazino)-(4-dimethylamino-phenyl)-acetic acid isopropyl ester (0.388 g, 0.8 mmol, 90% e.e.) in MeOH (8 ml) was added SmI_2 (17.6 ml, 1.76 mmol, 0.1 M in THF). Upon addition the blue color of the SmI_2 solution bleached to pale yellow. After 30 min, the reaction mixture was concentrated at reduced pressure. The residue was purified by flash chromatography on silica gel (hexane/EtOAc 5 : 1 as eluent) to give *tert*-butoxycarbonylamino-(4-dimethylamino-2-methoxy-phenyl)-acetic acid ethyl ester (0.284 g, 0.78 mmol, 97% yield); 90% e.e.; HPLC (DAICEL Chiralpak AD-H, hexane/*i*PrOH = 90 : 10, 1.0 ml/min): $R_{t(min)} = 15.0$ min and $R_{t(maj)} = 21.7$ min; $[\alpha]_D^{28} = +142.8$ (c 0.51, $CHCl_3$).

2.5.8

(*R*)-1,2-bis(benzyloxycarbonyl)piperazine-3-carboxylic Acid [41]



To a stirred solution of 5-bromopentanal (13.8 g, 83.3 mmol) and dibenzyl azodicarboxylate (16.1 g, 56 mmol) in CH_3CN (400 ml) at 0 °C was added (*S*)-proline (638 mg, 5.5 mmol, 10 mol%). After stirring the mixture at 0 °C for 15 h, ethanol (160 ml) and $NaBH_4$ (1.70 g, 45.0 mmol) were added, and the mixture was stirred at 0 °C for 40 min. The reaction was quenched by slow addition of 10% aqueous citric acid and the whole solution was concentrated *in vacuo*. This residue was diluted with ethyl acetate, washed with saturated brine, dried over sodium sulfate, filtered, and concentrated *in vacuo* to give (*R*)-5-bromo-2-*N*,*N*′-dibenzyloxycarbonyhydrazino-1-pentanol as a colorless solid (>99% e.e. as judged by the chiral HPLC analysis). The crude material

was purified by recrystallization from ethyl acetate/hexane to give pure (*R*)-5-bromo-2-*N*,*N*′-dibenzyloxycarbonyhydrazino-1-pentanol (23.5 g, 50 mmol, 91%) as a colorless solid: m.p. 94.0–95.5 °C; $[\alpha]_D^{22} = -10.7$ (*c* 1.04, $CHCl_3$). To a stirred solution of (*R*)-5-bromo-2-*N*,*N*′-dibenzyloxycarbonyhydrazino-1-pentanol (121 mg, 0.3 mmol) and imidazole (91 mg, 1.3 mmol) in DMF (3 ml) at 23 °C was added *tert*-butyldimethylsilyl chloride (45 mg, 0.3 mmol) and the mixture was stirred for 3.5 h. The reaction mixture was diluted with hexane/ethyl acetate, washed with saturated brine, dried over sodium sulfate, filtered, and concentrated *in vacuo*. The residue was purified by silica gel column chromatography (hexane/ethyl acetate 10 : 1) to give (*R*)-5-bromo-2-*N*, *N*′-dibenzyloxycarbonyl hydrazino-1-(*tert*-butyldimethylsilyloxy)pentane (150 mg, 0.26 mmol, 100%) as a colorless solid; m.p. 87 °C; $[\alpha]_D^{22} = +12.3$ (*c* 0.95, MeOH). To a stirred solution of (*R*)-5-bromo-2-*N*,*N*′-dibenzyloxycarbonylhydrazino-1-(*tert*-butyldimethylsilyloxy)pentane (12.0 g, 21 mmol) in DMF (100 ml) at 0 °C was added a suspension of sodium hydride (60%, 1.78 g, 37.0 mmol) in DMF (30 ml). After stirring the mixture at 0 °C for 40 min, the reaction mixture was quenched with 10% citric acid solution, and extracted with hexane/ethyl acetate. The combined extracts were washed with saturated brine, dried over sodium sulfate, filtered, and concentrated *in vacuo*. The residue was purified by silica gel column chromatography (hexane/ethyl acetate 10 : 1) to give (*R*)-dibenzyl 3-(*tert*-butyldimethylsilyloxy)methyl)piperazine-1,2-dicarboxylate (10.0 g, 20 mmol, 97%) as a colorless oil; $[\alpha]_D^{23} = +17.5$ (*c* 1.10, $CHCl_3$). To a stirred solution of (*R*)-dibenzyl 3-(*tert*-butyldimethylsilyloxy)methyl)piperazine-1,2-dicarboxylate (1.65 g, 3.3 mmol) in THF (20 ml) at 0 °C was added 0.5 M tetrabutylammonium fluoride (8.0 ml, 4.00 mmol). After stirring the mixture at 0 °C for 40 min, the reaction was quenched with saturated brine, and extracted with ethyl acetate. The combined extracts were washed with brine, dried over sodium sulfate, filtered, and concentrated *in vacuo*. The residue was purified by silica gel column chromatography (hexane/ethyl acetate 2 : 1 to 1 : 1) to give the alcohol (*R*)-dibenzyl 3-(hydroxymethyl) piperazine-1,2-dicarboxylate (1.27 g, 100%) as a colorless oil; $[\alpha]_D^{23} = -11.0$ (*c* 1.48, $CHCl_3$). To a stirred solution of (*R*)-dibenzyl 3-(hydroxymethyl)piperazine-1,2-dicarboxylate (1.53 g, 4.0 mmol), 2,2,6,6-tetramethylpiperidinooxy (TEMPO) (98 mg, 0.63 mmol), and $NaClO_2$ (726 mg, 8.0 mmol) in CH_3CN (20 ml) and phosphate buffer (pH 6.8, 20 ml) at 23 °C was added 1.6 M NaClO (500 µl). After stirring the mixture at RT for 4 h, 1 M NaOH (10 ml) was added, and the mixture added to 1 M aqueous sodium sulfite (30 ml). The mixture was acidified with 1 M aqueous potassium hydrogen sulfate and extracted with ethyl acetate. The combined organic extracts were washed with saturated brine, dried over sodium sulfate, filtered, and concentrated *in vacuo*. The residue was purified by silica gel column chromatography (hexane/ethyl acetate 1 : 1) to give (*R*)-1,2-bis(benzyloxycarbonyl)piperazine-3-carboxylic acid (1.43 g, 3.6 mmol, 90%) as a colorless oil; $[\alpha]_D^{23} = +20.0$ (*c* 0.975, $CHCl_3$).

Acknowledgments

This chapter is dedicated to the memory of Professor André Collet. Drs Joëlle Viret-Peyrard, Catherine Dubar, Laure Guy, and Jean-Christophe Hannachi are

acknowledged for their contribution to the synthesis and chemistry of α-hydrazino acids. Drs Michel Marraud and André Aubry are also sincerely thanked for their contribution to our research.

References

1 Thiele, J. and Stange, O. (1894) *Justus Liebigs Annalen der Chemie*, 1–46.

2 Darapsky, A., Prabrakar, M. and Bergur, H. (1917) *Journal für Praktische Chemie*, **96**, 251–327.

3 Darapsky, A. and Adamczewski, B. (1918) *Journal für Praktische Chemie*, **97**, 182–217.

4 Darapsky, A. (1919) *Journal für Praktische Chemie*, **99**, 179–231.

5 Darapsky, A., Germscheid, J., Kreuter, C., Engelmann, E., Engels, W. and Trinius, W. (1936) *Journal für Praktische Chemie*, **146**, 219–249.

6 Darapsky, A., Loevenich, J., Creifelds, O., Bellingen, W., Koster, E., Binet, V., Wasserfuhr, H. and Beck, H. (1936) *Journal für Praktische Chemie*, **146**, 268–306.

7 Marraud, M. and Vanderesse, R. (2002) Peptides containing C/N/O amide bond replacement, in *Houben–Weyl Methods of Organic Chemistry: Synthesis of Peptides and Peptidomimetics*, vol. **E22c** (eds M. Goodman, A. Felix, L. Moroder and C. Toniolo), Thieme, Stuttgart.

8 Porter, C.C., Watson, L.S., Titus, D.C., Totaro, J.A. and Byer, S.S. (1962) *Biochemical Pharmacology*, **11**, 1067–1077.

9 Lotti, V.J. and Porter, C.C. (1970) *Journal of Pharmacology and Experimental Therapeutics*, **172**, 406–415.

10 Glamkowski, E.J., Gal, G., Sletzinger, M., Porter, C.C. and Watson, L.S. (1967) *Journal of Medicinal Chemistry*, **10**, 852–855.

11 Burkhard, P., Dominici, P., Borri-Voltattorni, C., Jansonius, J.N. and Malashkevich, V.N. (2001) *Nature Structural Biology*, **8**, 963–967.

12 Harik, S.I. and Snyder, S.H. (1973) *Biochimica et Biophysica Acta, Enzymology*, **327**, 501–509.

13 Harik, S.I., Hollenberg, M.D. and Snyder, S.H. (1974) *Nature*, **249**, 250–251.

14 Inoue, H., Kato, Y., Takigawa, M., Adachi, K. and Takeda, Y. (1975) *Journal of Biochemistry*, **77**, 879–893.

15 Sawayama, T., Kinugasa, H. and Nishimura, H. (1976) *Chemical & Pharmaceutical Bulletin*, **24**, 326–329.

16 Takano, T., Takigawa, M. and Suzuki, F. (1983) *Journal of Biochemistry*, **93**, 591–598.

17 Sletzinger, M., Firestone, R.A., Reinhold, D.F., Rooney, C.S. and Nicholson, W.H. (1968) *Journal of Medicinal Chemistry*, **11**, 261–263.

18 Tanase, S., Guirard, B.M. and Snell, E.E. (1985) *Journal of Biological Chemistry*, **260**, 6738–6746.

19 Kelland, J.G., Arnold, L.D., Palcic, M.M., Pickard, M.A. and Vederas, J.C. (1986) *Journal of Biological Chemistry*, **261**, 13216–13223.

20 Lam, L.K.P., Arnold, L.D., Kalantar, T.H., Kelland, J.G., Lane-Bell, P.M., Palcic, M.M., Pickard, M.P. and Vederas, J.C. (1988) *Journal of Biological Chemistry*, **263**, 11814–11819.

21 Yamada, R., Wakabayashi, Y., Iwashima, A. and Hasegawa, T. (1987) *Biochimica et Biophysica Acta*, **911**, 372–375.

22 Scaman, C.H., Palcic, M.M., McPhalen, C., Gore, M.P., Lam, L.K.P. and Vederas, J.C. (1991) *Journal of Biological Chemistry*, **266**, 5525–5533.

23 Oehme, P., Glaesmer, R., Kalusa, W. and Heine, U. (1970) *Acta Biologica et Medica Germanica*, **24**, 141–153.

24 Pollak, G., Yellin, H. and Carmi, A. (1964) *Journal of Medicinal Chemistry*, **7**, 220–224.

25 Klosterman, H.J., Lamoureux, G.L. and Parsons, J.L. (1967) *Biochemistry*, **6**, 170–177.

26 Hamada, M., Takeuchi, T., Kondo, S., Ikeda, Y., Naganawa, H., Maeda, K., Okami, Y. and Umezawa, H. (1970) *Journal of Antibiotics*, **23**, 170–171.

27 Uehara, Y., Hori, M. and Umezawa, H. (1976) *Biochimica et Biophysica Acta*, **442**, 251–262.

28 Broberg, A., Menkis, A. and Vasiliauskas, R. (2006) *Journal of Natural Products*, **69**, 97–102.

29 Huang, X., Roemer, E., Sattler, I., Moellmann, U., Christner, A. and Grabley, S. (2006) *Angewandte Chemie International Edition*, **45**, 3067–3072.

30 Miller, E.D., Kauffman, C.A., Jensen, P.R. and Fenical, W. (2007) *The Journal of Organic Chemistry*, **72**, 323–330.

31 Ciufolini, M.A. and Xi, N. (1998) *Chemical Society Reviews*, **27**, 437–445.

32 Maehr, H., Liu, C.M., Palleroni, N.J., Smallheer, J., Todaro, L., Williams, T.H. and Blount, J.F. (1986) *Journal of Antibiotics*, **39**, 17–25.

33 Hassall, C.H., Bevan, K., Davies, J.S., Morton, R.B. and Phillips, D.A.S. (1971) *Journal of the Chemical Society C*, 514–522.

34 Shimada, N., Morimoto, K., Naganawa, H., Takita, T., Hamada, M., Maeda, K., Takeuchi, T. and Umezawa, H. (1981) *Journal of Antibiotics*, **34**, 1613–1614.

35 Boger, D.L., Ledeboer, M.W. and Kume, M. (1999) *Journal of the American Chemical Society*, **121**, 1098–1099.

36 Kamenecka, T.M. and Danishefsky, S.J. (2001) *Chemistry – A European Journal*, **7**, 41–63.

37 Fehr, T., Kallen, J., Oberer, L., Sanglier, J.-J. and Schilling, W. (1999) *Journal of Antibiotics*, **52**, 474–479.

38 Zhang, L.-H., Youn, H.-D. and Liu, J.O. (2001) *Journal of Biological Chemistry*, **276**, 43534–43540.

39 Zhang, L.-H. and Liu, J.O. (2001) *Journal of Immunology*, **166**, 5611–5618.

40 Sedrani, R., Kallen, J., Cabrejas, L.M.M., Papageorgiou, C.D., Senia, F., Rohrbach, S., Wagner, D., Thai, B., Eme, A.-M.J., France, J., Oberer, L., Rihs, G., Zenke, G. and Wagner, J. (2003) *Journal of the American Chemical Society*, **125**, 3849–3859.

41 Henmi, Y., Makino, K., Yoshitomi, Y., Hara, O. and Hamada, Y. (2004) *Tetrahedron – Asymmetry*, **15**, 3477–3481.

42 Hale, K.J., Cai, J., Delisser, V., Manaviazar, S., Peak, S.A., Bhatia, G.S., Collins, T.C. and Jogiya, N. (1996) *Tetrahedron*, **52**, 1047–1068.

43 Knobloch, W. and Niedrich, H. (1962) *Journal für Praktische Chemie*, **17**, 273–281.

44 Niedrich, H. and Knobloch, W. (1962) *Journal für Praktische Chemie*, **17**, 263–272.

45 Niedrich, H. (1963) *Chemische Berichte*, **96**, 2774–2779.

46 Niedrich, H. and Grupe, R. (1965) *Journal für Praktische Chemie*, **27**, 108–112.

47 Niedrich, H. (1965) *Chemische Berichte*, **98**, 3451–3461.

48 Grupe, R. and Niedrich, H. (1966) *Chemische Berichte*, **99**, 3914–3924.

49 Niedrich, H. (1967) *Chemische Berichte*, **100**, 3273–3282.

50 Grupe, R. and Niedrich, H. (1967) *Chemische Berichte*, **100**, 3283–3288.

51 Grupe, R. and Niedrich, H. (1970) *Journal für Praktische Chemie*, **312**, 1087–1096.

52 Niedrich, H. and Pirrwitz, J. (1972) *Journal für Praktische Chemie*, **314**, 735–750.

53 Grupe, R., Baeck, B. and Niedrich, H. (1972) *Journal für Praktische Chemie*, **314**, 751–758.

54 Niedrich, H. (1972) *Journal für Praktische Chemie*, **314**, 769–779.

55 Niedrich, H. and Koeller, G. (1974) *Journal für Praktische Chemie*, **316**, 729–740.

56 Oehme, P., Rex, H. and Ackermann, E. (1964) *Acta Biologica et Medica Germanica*, **12**, 234–252.

57 Oehme, P., Vogt, W.E., Niedrich, H., Katzwinkel, S., Bergmann, J. and Jung, F.

(1973) *Acta Biologica et Medica Germanica*, **30**, 415–422.

58 Amour, A., Collet, A., Dubar, C. and Reboud-Ravaux, M. (1994) *International Journal of Peptide & Protein Research*, **43**, 297–304.

59 Guy, L., Vidal, J., Collet, A., Amour, A. and Reboud-Ravaux, M. (1998) *Journal of Medicinal Chemistry*, **41**, 4833–4843.

60 Bouget, K., Aubin, S., Delcros, J.-G., Arlot-Bonnemains, Y. and Baudy-Floc'h, M. (2003) *Biorganic & Medicinal Chemistry*, **11**, 4881–4889.

61 Bonnet, D., Ollivier, N., Gras-Masse, H. and Melnyk, O. (2001) *The Journal of Organic Chemistry*, **66**, 443–449.

62 Grandjean, C., Gras-Masse, H. and Melnyk, O. (2001) *Chemistry – A European Journal*, **7**, 230–239.

63 Chenevier, P., Grandjean, C., Loing, E., Malingue, F., Angyalosi, G., Gras-Masse, H., Roux, D., Melnyk, O. and Bourel-Bonnet, L. (2002) *Chemical Communications*, 2446–2447.

64 Gardner, B., Nakanishi, H. and Kahn, M. (1993) *Tetrahedron*, **49**, 3433–3448.

65 Saragovi, H.U., Fitzpatrick, D., Raktabutr, A., Nakanishi, H., Kahn, M. and Greene, M.I. (1991) *Science*, **253**, 792–795.

66 Kahn, M. (1993) *Synlett*, 821–826.

67 Chen, S., Chrusciel, R.A., Nakanishi, H., Raktabutr, A., Johnson, M.E., Sato, A., Weiner, D., Hoxie, J., Saragovi, H.U., Greene, M.I. and Kahn, M. (1992) *Proceedings of the National Academy of Sciences of the, United States of America*, **89**, 5872–5876.

68 Lelais, G. and Seebach, D. (2003) *Helvetica Chimica Acta*, **86**, 4152–4168.

69 Aubin, S., Martin, B., Delcros, J.-G., Arlot-Bonnemains, Y. and Baudy-Floc'h, M. (2005) *Journal of Medicinal Chemistry*, **48**, 330–334.

70 Busnel, O., Bi, L., Dali, H., Cheguillaume, A., Chevance, S., Bondon, A., Muller, S. and Baudy-Floc'h, M. (2005) *The Journal of Organic Chemistry*, **70**, 10701–10708.

71 Attwood, M.R., Hassall, C.H., Krohn, A., Lawton, G. and Redshaw, S. (1986) *Journal of the Chemical Society, Perkin Transactions 1*, 1011–1019.

72 Batchelor, M.J., Bebbington, D., Bemis, G.W., Fridman, W.H., Gillespie, R.J., Golec, J.M.C., Gu, Y., Lauffer, D.J., Livingston, D.J., Matharu, S.S., Mullican, M.D., Murcko, M.A., Murdoch, R., Nyce, P.L., Robidoux, A.L.C., Su, M., Wannamaker, M.W., Wilson, K.P. and Zelle, R.E. (1997) *Chemical Abstracts*, **127**, 122000.

73 Siegmund, B. and Zeitz, M. (2003) *IDrugs*, **6**, 154–158.

74 Viret, J., Collet, A., Pichon-Pesme, V. and Aubry, A. (1988) *New Journal of Chemistry*, **12**, 253–256.

75 Aubry, A. and Marraud, M. (1989) *Biopolymers*, **28**, 109–122.

76 Viret, J., Collet, A., Lecoq, A., Marraud, M. and Aubry, A. (1989) *New Journal of Chemistry*, **13**, 849–852.

77 Aubry, A., Bayeul, D., Mangeot, J.P., Vidal, J., Sterin, S., Collet, A., Lecoq, A. and Marraud, M. (1991) *Biopolymers*, **31**, 793–801.

78 Lecoq, A., Marraud, M. and Aubry, A. (1991) *Tetrahedron Letters*, **32**, 2765–2768.

79 Marraud, M., Dupont, V., Grand, V., Zerkout, S., Lecoq, A., Boussard, G., Vidal, J., Collet, A. and Aubry, A. (1993) *Biopolymers*, **33**, 1135–1148.

80 Aubry, A., Mangeot, J.P., Vidal, J., Collet, A., Zerkout, S. and Marraud, M. (1994) *International Journal of Peptide & Protein Research*, **43**, 305–311.

81 Zerkout, S., Dupont, V., Aubry, A., Vidal, J., Collet, A., Vicherat, A. and Marraud, M. (1994) *International Journal of Peptide & Protein Research*, **44**, 378–387.

82 Marraud, M., Boussard, G., Vanderesse, R., Collet, A., Vidal, J. and Aubry, A. (1996) *Actualités de Chimie Thérapeutique*, **22**, 67–82.

83 Aubry, A., Del Duca, V., Pedone, C., Zerkolt, S. and Marraud, M. (1999) *Acta Crystallographica, Section C: Crystal Structure, Communications*, **C55**, 439–441.

84 Chéguillaume, A., Salaün, A., Sinbandhit, S., Potel, M., Gall, P., Baudy-Floc'h, M.

and Le Grel, P. (2001) *The Journal of Organic Chemistry*, **66**, 4923–4929.
85 Gunther, R. and Hofmann, H.-J. (2001) *Journal of the American Chemical Society*, **123**, 247–255.
86 Salaün, A., Potel, M., Roisnel, T., Gall, P. and Le Grel, P. (2005) *The Journal of Organic Chemistry*, **70**, 6499–6502.
87 Le Grel, P., Salaün, A., Potel, M., Le Grel, B. and Lassagne, F. (2006) *The Journal of Organic Chemistry*, **71**, 5638–5645.
88 Salaün, A., Favre, A., Le Grel, B., Potel, M. and Le Grel, P. (2006) *The Journal of Organic Chemistry*, **71**, 150–158.
89 Le Grel, P. and Guichard, G. (2007) In *Foldamers: Structure, Properties and Applications* (eds F. Diederich, S. Hecht and I. Huc), Wiley-VCH Verlag GmbH, Weinheim, pp. 35–74.
90 Stephenson, L.M. and Mattern, D.L. (1976) *The Journal of Organic Chemistry*, **41**, 3614–3619.
91 Karady, S., Ly, M.G., Pines, S.H. and Sletzinger, M. (1971) *The Journal of Organic Chemistry*, **36**, 1946–1948.
92 Najera, C. and Sansano, J.M. (2007) *Chemical Reviews*, **107**, 4584–4671.
93 Bauer, T. and Gajewiak, J. (2004) *Tetrahedron*, **60**, 9163–9170.
94 Moumne, R., Lavielle, S. and Karoyan, P. (2006) *The Journal of Organic Chemistry*, **71**, 3332–3334.
95 Chéguillaume, A., Doubli-Bounoua, I., Baudy-Floc'h, M. and Le Grel, P. (2000) *Synlett*, 331–334.
96 Hoffman, R.V. and Kim, H.O. (1990) *Tetrahedron Letters*, **31**, 2953–2956.
97 Kim, H.-O., Lum, C. and Lee, M.S. (1997) *Tetrahedron Letters*, **38**, 4935–4938.
98 Schmidt, U., Braun, C. and Sutoris, H. (1996) *Synthesis*, 223–229.
99 Mitsunobu, O. (1981) *Synthesis*, 1–28.
100 Brosse, N., Pinto, M.-F., Bodiguel, J. and Jamart-Grégoire, B. (2001) *The Journal of Organic Chemistry*, **66**, 2869–2873.
101 Bouillon, I., Brosse, N., Vanderesse, R. and Jamart-Grégoire, B. (2007) *Tetrahedron*, **63**, 2223–2234.
102 Bouillon, I., Brosse, N., Vanderesse, R. and Jamart-Grégoire, B. (2004) *Tetrahedron Letters*, **45**, 3569–3572.
103 Hughes, P. and Clardy, J. (1989) *The Journal of Organic Chemistry*, **54**, 3260–3264.
104 Sibi, M.P. and Soeta, T. (2007) *Journal of the American Chemical Society*, **129**, 4522–4523.
105 Greck, C., Drouillat, B. and Thomassigny, C. (2004) *European Journal of Organic Chemistry*, 1377–1385.
106 Erdik, E. (2004) *Tetrahedron*, **60**, 8747–8782.
107 Evans, D.A., Britton, T.C., Dorow, R.L. and Dellaria, J.F. (1986) *Journal of the American Chemical Society*, **108**, 6395–6397.
108 Gennari, C., Colombo, L. and Bertolini, G. (1986) *Journal of the American Chemical Society*, **108**, 6394–6395.
109 Trimble, L.A. and Vederas, J.C. (1986) *Journal of the American Chemical Society*, **108**, 6397–6399.
110 Oppolzer, W. and Moretti, R. (1986) *Helvetica Chimica Acta*, **69**, 1923–1926.
111 Evans, D.A., Britton, T.C. and Dellaria, J.F. Jr (1988) *Tetrahedron*, **44**, 5525–5540.
112 Harris, J.M., Bolessa, E.A., Mendonca, A.J., Feng, S.-C. and Vederas, J.C. (1995) *Journal of the Chemical Society, Perkin Transactions 1*, 1945–1950.
113 Harris, J.M., McDonald, R. and Vederas, J.C. (1996) *Journal of the Chemical Society, Perkin Transactions 1*, 2669–2674.
114 List, B. (2006) *Chemical Communications*, 819–824.
115 Marigo, M. and Jørgensen, K.A. (2006) *Chemical Communications*, 2001–2011.
116 Janey, J.M. (2005) *Angewandte Chemie International Edition*, **44**, 4292–4300.
117 Schmidt, U. and Riedl, B. (1993) *Synthesis*, 809–814.
118 Nakamura, Y., Ito, A. and Shin, C.-G. (1994) *Bulletin of the Chemical Society of Japan*, **67**, 2151–2161.
119 Hale, K.J., Jogiya, N. and Manaviazar, S. (1998) *Tetrahedron Letters*, **39**, 7163–7166.

120 Kamenecka, T.M., and Danishefsky, S.J. (1998) *Angewandte Chemie International Edition*, **37**, 2993–2995.

121 Poupardin, O., Greck, C. and Genêt, J.-P. (1998) *Synlett*, 1279–1281.

122 Ciufolini, M.A. and Xi, N. (1994) *Journal of the Chemical Society, Chemical Communications*, 1867–1868.

123 Greck, C., Bischoff, L. and Genêt, J.P. (1995) *Tetrahedron – Asymmetry*, **6**, 1989–1994.

124 Poupardin, O., Greck, C. and Genêt, J.P. (2000) *Tetrahedron Letters*, **41**, 8795–8797.

125 Greck, C., Bischoff, L., Ferreira, F., Pinel, C., Piveteau, E. and Genêt, J.P. (1993) *Synlett*, 475–477.

126 Ushiyama, R., Yonezawa, Y. and Shin, C.-G. (2001) *Chemistry Letters*, 1172–1173.

127 Bogevig, A., Juhl, K., Kumaragurubaran, N., Zhuang, W. and Jørgensen, K.A. (2002) *Angewandte Chemie International Edition*, **41**, 1790–1793.

128 List, B. (2002) *Journal of the American Chemical Society*, **124**, 5656–5657.

129 Chandrasekhar, S., Parimala, G., Tiwari, B., Narsihmulu, C. and Sarma, G.D. (2007) *Synthesis*, 1677–1682.

130 Vogt, H., Vanderheiden, S. and Bräse, S. (2003) *Chemical Communications*, 2448–2449.

131 Baumann, T., Vogt, H. and Bräse, S. (2007) *European Journal of Organic Chemistry*, 266–282.

132 Suri, J.T., Steiner, D.D. and Barbas, C.F. III (2005) *Organic Letters*, **7**, 3885–3888.

133 Chowdari, N.S. and Barbas, C.F. III (2005) *Organic Letters*, **7**, 867–870.

134 Marigo, M., Juhl, K. and Jørgensen, K.A. (2003) *Angewandte Chemie International Edition*, **42**, 1367–1369.

135 Saaby, S., Bella, M. and Jørgensen, K.A. (2004) *Journal of the American Chemical Society*, **126**, 8120–8121.

136 Pihko, P.M. and Pohjakallio, A. (2004) *Synlett*, 2115–2118.

137 Ma, S., Jiao, N., Zheng, Z., Ma, Z., Lu, Z., Ye, L., Deng, Y. and Chen, G. (2004) *Organic Letters*, **6**, 2193–2196.

138 Liu, X., Li, H. and Deng, L. (2005) *Organic Letters*, **7**, 167–169.

139 Foltz, C., Stecker, B., Marconi, G., Bellemin-Laponnaz, S., Wadepohl, H. and Gade, L.H. (2005) *Chemical Communications*, 5115–5117.

140 Xu, X., Yabuta, T., Yuan, P. and Takemoto, Y. (2006) *Synlett*, 137–140.

141 Kang, Y.K. and Kim, D.Y. (2006) *Tetrahedron Letters*, **47**, 4565–4568.

142 Terada, M., Nakano, M. and Ube, H. (2006) *Journal of the American Chemical Society*, **128**, 16044–16045.

143 Huber, D.P., Stanek, K. and Togni, A. (2006) *Tetrahedron – Asymmetry*, **17**, 658–664.

144 Comelles, J., Pericas, A., Moreno-Manas, M., Vallribera, A., Drudis-Sole, G., Lledos, A., Parella, T., Roglans, A., Garcia-Granda, S. and Roces-Fernandez, L. (2007) *The Journal of Organic Chemistry*, **72**, 2077–2087.

145 Liu, Y., Melgar-Fernandez, R. and Juaristi, E. (2007) *The Journal of Organic Chemistry*, **72**, 1522–1525.

146 Mashiko, T., Hara, K., Tanaka, D., Fujiwara, Y., Kumagai, N. and Shibasaki, M. (2007) *Journal of the American Chemical Society*, **129**, 11342–11343.

147 Waser, J. and Carreira, E.M. (2004) *Journal of the American Chemical Society*, **126**, 5676–5677.

148 Waser, J., Gaspar, B., Nambu, H. and Carreira, E.M. (2006) *Journal of the American Chemical Society*, **128**, 11693–11712.

149 Sato, M., Gunji, Y., Ikeno, T. and Yamada, T. (2005) *Chemistry Letters*, **34**, 316–317.

150 Schestakow, P. (1906) *Chemical Abstracts*, 150102.

151 Viret, J., Gabard, J. and Collet, A. (1987) *Tetrahedron*, **43**, 891–894.

152 Karady, S., Ly, M.G., Pines, S.H. and Sletzinger, M. (1971) *The Journal of Organic Chemistry*, **36**, 1949–1951.

153 Gustafsson, H. and Ragnarsson, U. (1974) *Acta Pharmaceutica Suecica*, **11**, 493–494.

154 Gustafsson, H. (1975) *Acta Chemica Scandinavica, Series B: Organic Chemistry, and Biochemistry*, **B29**, 93–98.

155 Achiwa, K. and Yamada, S. (1975) *Tetrahedron Letters*, 2701–2704.
156 Oguz, U., Guilbeau, G.G. and McLaughlin, M.L. (2002) *Tetrahedron Letters*, **43**, 2873–2875.
157 Andreae, S. and Schmitz, E. (1991) *Synthesis*, 327–341.
158 Vidal, J., Damestoy, S., Guy, L., Hannachi, J.-C., Aubry, A. and Collet, A. (1997) *Chemistry – A European Journal*, **3**, 1691–1709.
159 Schmitz, E., Andreae, S., Schramm, S., Albert, F.M. and Lohmann, D. (1985) *Chemical Abstracts*, **105**, 227301.
160 Szurdoki, F., Andreae, S., Baitz-Gacs, E., Tamas, J., Valko, K., Schmitz, E. and Szantay, C. (1988) *Synthesis*, 529–533.
161 Vidal, J., Drouin, J. and Collet, A. (1991) *Journal of the Chemical Society, Chemical Communications*, 435–437.
162 Vidal, J., Guy, L., Sterin, S. and Collet, A. (1993) *The Journal of Organic Chemistry*, **58**, 4791–4793.
163 Vidal, J., Hannachi, J.-C., Hourdin, G., Mulatier, J.-C. and Collet, A. (1998) *Tetrahedron Letters*, **39**, 8845–8848.
164 Vidal, J., Damestoy, S. and Collet, A. (1995) *Tetrahedron Letters*, **36**, 1429–1442.
165 Niederer, D.A., Kapron, J.T. and Vederas, J.C. (1993) *Tetrahedron Letters*, **34**, 6859–6862.
166 Hannachi, J.-C., Vidal, J., Mulatier, J.-C. and Collet, A. (2004) *The Journal of Organic Chemistry*, **69**, 2367–2373.
167 Burk, M.J. and Feaster, J.E. (1992) *Journal of the American Chemical Society*, **114**, 6266–6267.
168 Friestad, G.K. (2005) *European Journal of Organic Chemistry*, 3157–3172.
169 Sugiura, M. and Kobayashi, S. (2005) *Angewandte Chemie International Edition*, **44**, 5176–5186.
170 Burk, M.J., Martinez, J.P., Feaster, J.E. and Cosford, N. (1994) *Tetrahedron*, **50**, 4399–4428.
171 Makino, K., Jiang, H., Suzuki, T. and Hamada, Y. (2006) *Tetrahedron – Asymmetry*, **17**, 1644–1649.
172 Keith, J.M. and Jacobsen, E.N. (2004) *Organic Letters*, **6**, 153–155.
173 Cativiela, C., Diaz-de-Villegas, M.D. and Galvez, J.A. (1997) *Tetrahedron – Asymmetry*, **8**, 1605–1610.
174 Hamada, T., Manabe, K. and Kobayashi, S. (2003) *Angewandte Chemie International Edition*, **42**, 3927–3930.
175 Ogawa, C., Sugiura, M. and Kobayashi, S. (2004) *Angewandte Chemie International Edition*, **43**, 6491–6493.
176 Jacobsen, M.F., Ionita, L. and Skrydstrup, T. (2004) *The Journal of Organic Chemistry*, **69**, 4792–4796.
177 Kobayashi, S., Hamada, T. and Manabe, K. (2002) *Journal of the American Chemical Society*, **124**, 5640–5641.
178 Hamada, T., Manabe, K. and Kobayashi, S. (2004) *Journal of the American Chemical Society*, **126**, 7768–7769.
179 Hamada, T., Manabe, K. and Kobayashi, S. (2006) *Chemistry – A European Journal*, **12**, 1205–1215.
180 Shirakawa, S., Berger, R. and Leighton, J.L. (2005) *Journal of the American Chemical Society*, **127**, 2858–2859.
181 Ueda, M., Miyabe, H., Sugino, H. and Naito, T. (2005) *Organic & Biomolecular Chemistry*, **3**, 1124–1128.
182 Hasegawa, K., Arai, S. and Nishida, A. (2006) *Tetrahedron*, **62**, 1390–1401.
183 Hassall, C.H., Johnson, W.H. and Theobald, C.J. (1979) *Journal of the Chemical Society, Perkin Transactions 1*, 1451–1454.
184 Aspinall, I.H., Cowley, P.M., Mitchell, G. and Stoodley, R.J. (1993) *Journal of the Chemical Society, Chemical Communications*, 1179–1180.
185 Aspinall, I.H., Cowley, P.M., Mitchell, G., Raynor, C.M. and Stoodley, R.J. (1999) *Journal of the Chemical Society, Perkin Transactions 1*, 2591–2599.
186 Makino, K., Henmi, Y., Terasawa, M., Hara, O. and Hamada, Y. (2005) *Tetrahedron Letters*, **46**, 555–558.
187 Mish, M.R., Guerra, F.M. and Carreira, E.M. (1997) *Journal of the American Chemical Society*, **119**, 8379–8380.

188 Guerra, F.M., Mish, M.R. and Carreira, E.M. (2000) *Organic Letters*, **2**, 4265–4267.
189 Benoiton, N.L. (2006) *Chemistry of Peptide Synthesis*, Taylor & Francis, Boca Raton, FL.
190 Smith, P.A.S. (1983) *Derivatives of Hydrazine and Other Hydronitrogens Having N–N Bonds*, Benjamin/Cummings, Reading, MA.
191 Guy, L. (1995) PhD Thesis, École normale supérieure de Lyon.
192 Carmi, A., Pollak, G. and Yellin, H. (1960) *The Journal of Organic Chemistry*, **25**, 44–46.
193 Claremon, D.A., Lumma, P.K. and Phillips, B.T. (1986) *Journal of the American Chemical Society*, **108**, 8265–8266.
194 Ding, H. and Friestad, G.K. (2004) *Organic Letters*, **6**, 637–640.
195 Chauveau, A., Martens, T., Bonin, M., Micouin, L. and Husson, H.-P. (2002) *Synthesis*, 1885–1890.
196 Dubar, C. (1992) PhD Thesis, Université de Lyon 1.
197 Kim, H.-O., Gardner, B. and Kahn, M. (1995) *Tetrahedron Letters*, **36**, 6013–6016.
198 Schmidt, U. and Riedl, B. (1992) *Journal of the Chemical Society, Chemical Communications*, 1186–1187.
199 Ciufolini, M.A. and Xi, N. (1997) *The Journal of Organic Chemistry*, **62**, 2320–2321.
200 Bonnet, D., Grandjean, C., Rousselot-Pailley, P., Joly, P., Bourel-Bonnet, L., Santraine, V., Gras-Masse, H. and Melnyk, O. (2003) *The Journal of Organic Chemistry*, **68**, 7033–7040.
201 Grandjean, C., Rommens, C., Gras-Masse, H. and Melnyk, O. (2000) *Angewandte Chemie International Edition*, **39**, 1068–1072.
202 Bouillon, I., Vanderesse, R., Brosse, N., Fabre, O. and Jamart-Gregoire, B. (2007) *Tetrahedron*, **63**, 9635–9641.
203 Nakamura, Y. and Shin, C.G. (1991) *Chemistry Letters*, 1953–1956.
204 Boger, D.L. and Schule, G. (1998) *The Journal of Organic Chemistry*, **63**, 6421–6424.
205 Wilson, L.J., Li, M. and Portlock, D.E. (1998) *Tetrahedron Letters*, **39**, 5135–5138.
206 Blass, B.E., Srivastava, A., Coburn, K.R., Faulkner, A.L., Janusz, J.J., Ridgeway, J.M. and Seibel, W.L. (2004) *Tetrahedron Letters*, **45**, 619–621.
207 Blass, B.E., Srivastava, A., Coburn, K.R., Faulkner, A.L., Janusz, J.J., Ridgeway, J.M. and Seibel, W.L. (2004) *Tetrahedron Letters*, **45**, 1275–1277.
208 De Luca, L., Falorni, M., Giacomelli, G. and Porcheddu, A. (1999) *Tetrahedron Letters*, **40**, 8701–8704.

3
Hydroxamic Acids: Chemistry, Bioactivity, and Solution- and Solid-Phase Synthesis

Darren Griffith, Marc Devocelle, and Celine J. Marmion

3.1
Introduction

Despite their discovery by Lossen in 1869 [1], the chemistry and bioactivity of hydroxamic acids remained elusive until the early 1980s, after which time much information emerged with respect to their bioactivity, their synthesis, and the synthesis and structural diversity of their metal complexes. Hydroxamic acids, whether naturally occurring or synthetic, now constitute a leading class of organic acids with a diverse range of pharmacological, pathological, and toxicological properties. Their rich bioactivity has undoubtedly been the impetus behind new developments into efficient methods for their synthesis and the synthesis of hydroxamic acid derivatives. This chapter highlights recent advances in relation to the chemistry and versatile bioactivity of hydroxamic acids, which is followed by an overview of well-established in addition to more recent novel and innovative approaches to their solution- and solid-phase synthesis and that of their derivatives.

3.2
Chemistry, Bioactivity, and Clinical Utility

3.2.1
Chemistry

Hydroxamic acids are a group of weak organic acids of general formula RCONHOH, which can exist in two tautomeric forms (Figure 3.1): the keto tautomer, which is predominant under acidic conditions and behaves as a monobasic acid, and the enol (iminol) form, which is more stable in alkaline media [2]. Nuclear magnetic resonance (NMR) studies [3] have shown that each tautomer can exist as (*E*) and (*Z*) isomers (Figure 3.1). From *ab initio* molecular orbital calculations [4] as well as crystallographic data, the (*E*) isomer has been shown to be more stable in the solid

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

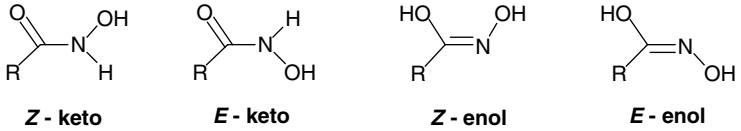


Figure 3.1 The keto and enol tautomeric forms of hydroxamic acids and their corresponding (*E*) and (*Z*) isomers.

state, while the (*Z*) isomer is more stable under aqueous conditions [5], the latter being stabilized by hydrogen bonding.

Typically, hydroxamic acids have pK_a values in aqueous solution of approximately 9, but can range, depending on the electronic properties of substituents, from 7.05 for *ortho*-nitrobenzohydroxamic acid, for example, to 11.33 for *N*-phenyl-*n*-butyrohydroxamic acid [6]. The acidity of hydroxamic acids has been somewhat contentious though as, depending on structure, phase, or solvent, hydroxamic acids can behave as *N*- or *O*-acids (Figure 3.2). It was thought initially that hydroxamic acids were predominantly *O*-acids [7], giving rise, upon dissociation, to structures (*c*) and (*d*); however, Bauer and Exner concluded on the basis of infrared and ultraviolet spectroscopic studies, and taking into account their pK_a values, that hydroxamic acids with common substituents were *N*-acids with structure (*e*) being their conjugate base [8]. An X-ray photoelectron spectroscopy study in the solid state [9] and an ^{17}O-NMR investigation of the benzohydroxamate ion in methanol [10] further support these findings. Ventura *et al.* [11] subsequently showed from *ab initio* methods that formo- and acetohydroxamic acids behave as *N*-acids in the gas phase. Their findings also indicated that in dimethylsulfoxide acetohydroxamic acid is an *N*-acid, whereas in contrast benzohydroxamic acid may act as an *O*-acid. It was further surmised that both (*d*) and (*e*)/(*f*) could coexist in equilibrium under aqueous conditions, with OH dissociation having a larger stabilization effect with respect to water. More recently, a soft X-ray absorption spectroscopy study indicated that acetohydroxamic acid in aqueous solution could behave as a mixed acid, present primarily as an *O*-acid at high pH with the *N*-acid being the minor species [12].

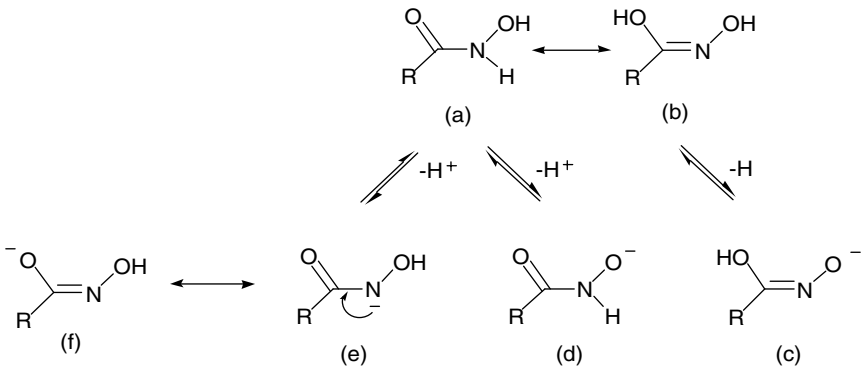


Figure 3.2 Possible hydroxamic acid dissociation pathways.

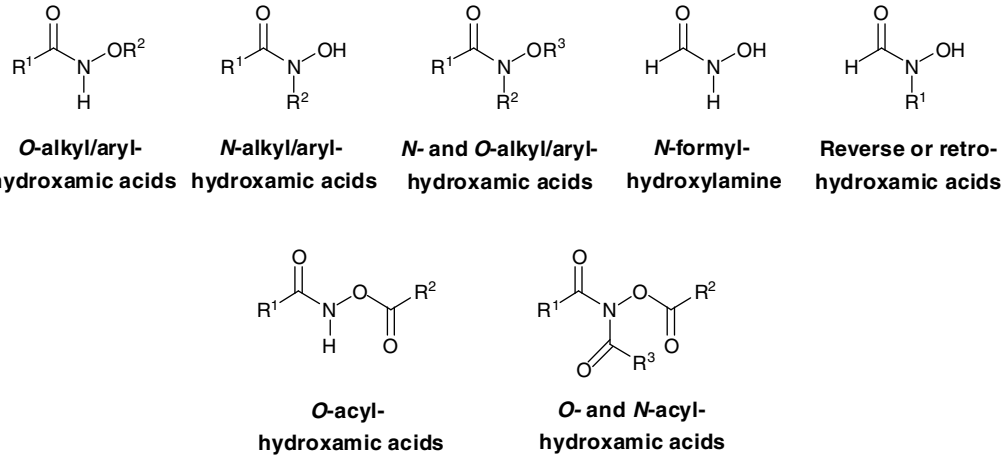


Figure 3.3 Structures of *N*- and *O*-alkyl, aryl or acyl derivatives of hydroxamic acids and reverse or retro-hydroxamic acids.

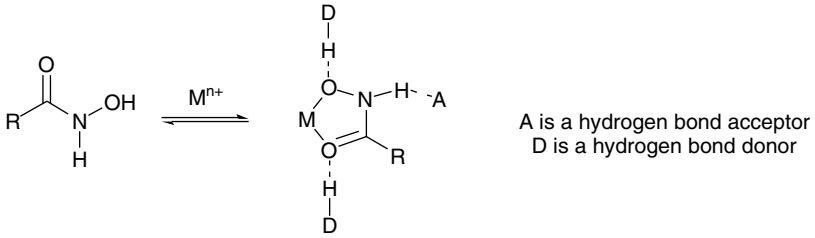
Other important members of the hydroxamic acid family include *N*- and *O*-alkyl or aryl or acylhydroxamic acids and reverse or retro-hydroxamic acids, which are *N*-alkyl or aryl derivatives of *N*-formylhydroxylamine (formohydroxamic acid) (Figure 3.3).

3.2.2 Bioactivity and Clinical Utility

There is no doubt that many of the biomedical applications of hydroxamic acids are due to their ability to bind strongly to metal ions at the catalytic active sites of metalloenzymes, although some may be due to their ability to release nitric oxide [13] – a view strengthened by the now established importance of nitric oxide in many physiological and pathophysiological processes [14, 15]. Typically, hydroxamic acids chelate metal ions via the carbonyl oxygen and deprotonated hydroxyl oxygen to form stable singly deprotonated hydroxamato complexes, although coordination via the deprotonated nitrogen may also occur under certain reaction conditions to give doubly deprotonated hydroximato derivatives [16, 17]. Their strong affinity for transition metal ions is reflected in the magnitudes of their overall stability constants (log β values) [18] which for trisbenzohydroxamato Fe(III), for example, is 27.20 [19]. The hydroxamic acid moiety also contains multiple sites for hydrogen bonding interactions (Scheme 3.1), often crucial structural elements that can confer selectivity to a drug for its biological target or indeed increase the binding of a drug to its receptor [16, 20].

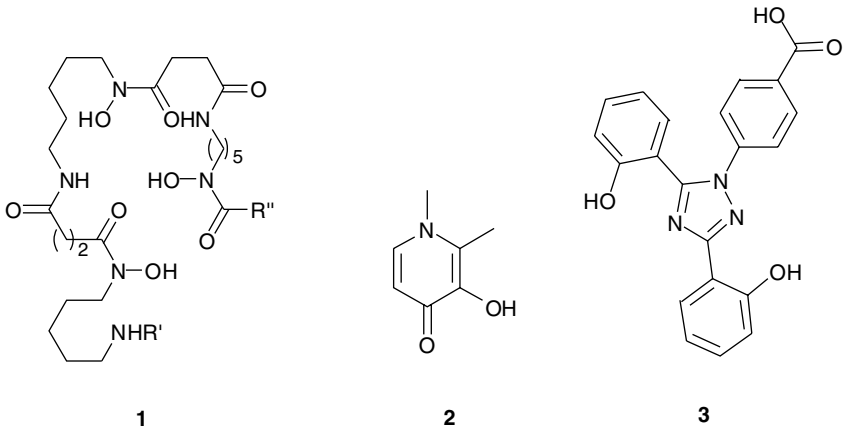
3.2.2.1 Hydroxamic Acids as Siderophores

One of the first physiological roles associated with hydroxamic acids was their use as siderophores – a class of low-molecular-weight Fe(III)-sequestering agents synthesized by bacteria for iron uptake and transport. Iron is one of the most abundant elements on Earth and crucial to many life-sustaining processes, yet is in poor supply



Scheme 3.1 Multiple sites for potential hydrogen bonding interactions.

due to poor aqueous solubility [21]. Fe(III), its most common oxidation state, for example, has a concentration as low as 10^{-9} M under physiological conditions, thus rendering its acquisition by any organism a major challenge [22]. Siderophores, produced by bacteria to assimilate iron into a soluble form, possess hydroxamic acid, catechol, or α-hydroxy carboxylic acids as key functional groups capable of binding Fe(III) giving stable, water-soluble iron complexes, which are recognized by cell surface proteins, facilitating the transport of the iron-containing siderophore to the cytoplasm where the essential nutrient iron is released [17, 21, 22]. A remarkable feature of siderophores is their ability to selectively bind Fe(III) ($\log \beta > 30$) over other, potentially poisonous, environmentally prevalent metal ions [23]. Despite their enormous therapeutic potential, few practical applications of siderophores have come to fruition. One success story is the use of desferrioxamine B (DFOB, **1**, Figure 3.4), a naturally occurring trihydroxamic acid siderophore derived from cultures of *Streptomyces pilosus* to treat β-thalassemia – an inherited blood disorder characterized by a person's inability to synthesize hemoglobin properly. To supplement their hemoglobin levels, thalassemic patients require regular blood transfusions leading to an accumulation of Fe(III) in the body, which, if untreated, can result in organ failure and death. Chelation therapy, utilizing the hexadentate chelator DFOB that complexes with iron in a 1 : 1 molar ratio to form ferrioxamine, is an



1 **2** **3**

Figure 3.4 Chemical structures of DFOB (**1**), deferiprone (**2**), and deferasirox (**3**).

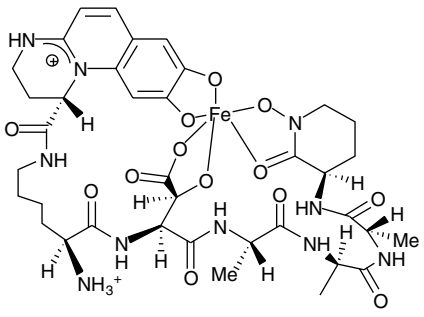
essential component of the treatment regime for these patients, ultimately prolonging their life expectancy [24, 25]. Despite its widespread use as an iron-sequestering agent in hematology, its lack of oral bioavailability has led to problems with patient compliance [24, 26] and this fuelled research into finding orally active DFOB alternatives, either naturally occurring or synthetic. Two non-hydroxamate-based DFOB alternatives are now in clinical use as orally active iron chelators: deferiprone (**2**, 1,2-dimethyl-3-hydroxy-4-pyridone), which gained limited approval over a decade ago [24, 25, 27–29], and deferasirox (**3**, 4-[3,5-bis(2-hydroxyphenyl)-1,2,4-triazol-1-yl] benzoic acid, ICL 670), which received US Food and Drug Administration (FDA) approval in 2005 [24, 25, 28, 30]. DFOB and deferiprone are also being utilized in combination therapy [27, 31].

Enormous progress has been achieved in recent years with respect to the synthesis of naturally occurring and synthetic siderophores, including pseudobactin (**4**) [21], nannochelin (**5**) [21], acinetonferrin (**6**) [21], desferrisalmycin B (**7**) [32], and exochelin MN (**8**) [33], to name but a few (Figure 3.5) [21].

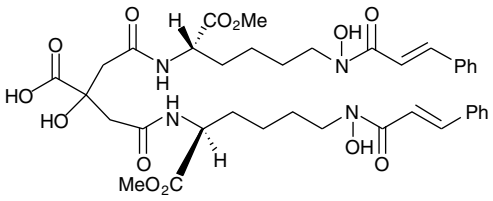
One potentially powerful application of siderophores is the development of bioconjugates in which siderophores are covalently linked to, for example, antimicrobial agents. Microbes have the ability to distinguish and employ only certain siderophores. A siderophore linked to an antimicrobial agent could therefore act as a "Trojan horse" to selectively deliver antimicrobial prodrugs to their site of action, thereby targeting pathogenic microbes [21]. This is now of considerable importance in this time of increasing resistance to conventional chemotherapeutics. Numerous hydroxamic acid-based siderophores have thus been synthesized and linked to not only antimicrobial agents, but to antiviral, antifungal, anticancer, and nontoxic organ-selective magnetic resonance imaging agents [21]. It is noteworthy that many old drugs that relied on passive diffusion and to which resistance had developed may now be revitalized by the formation of such conjugates.

3.2.2.2 **Hydroxamic Acids as Enzyme Inhibitors**

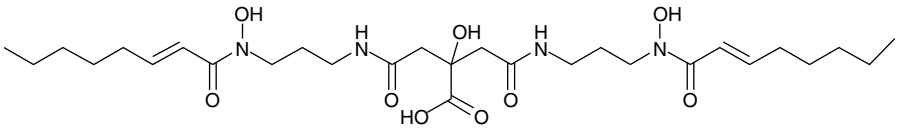
The strong metal-chelating ability of hydroxamic acids has been a critical feature in transforming small molecules into active metallo-enzyme inhibitors with chemotherapeutic properties [34, 35]. A large number of potent inhibitors containing the hydroxamic acid moiety have been disclosed in published and patent literature. While hydroxamic acids have been shown to inhibit a diverse range of enzymes including ureases [36], tumor necrosis factor (convertases (TACEs) [34, 37], angiotensin-converting enzyme [35], matrix metalloproteases (MMPs), histone deacetylases (HDACs) and prostaglandin H_2 synthases (PGHSs), only the last three will be discussed as representative examples to illustrate the biodiversity of hydroxamic acids as enzyme inhibitors. MMPs were chosen as the majority of MMP inhibitors studied to date contain a hydroxamic acid as the metal-binding group. A hydroxamic acid recently entered the clinic as a therapeutic agent targeting HDACs, hence the reason for including this class of enzymes. PGHSs were selected as they represent a family of enzymes that contain two active sites; a heme-containing peroxidase site and a non-metal-containing cyclooxygenase site, both sites of which are inhibited by hydroxamic acids via entirely different mechanisms.



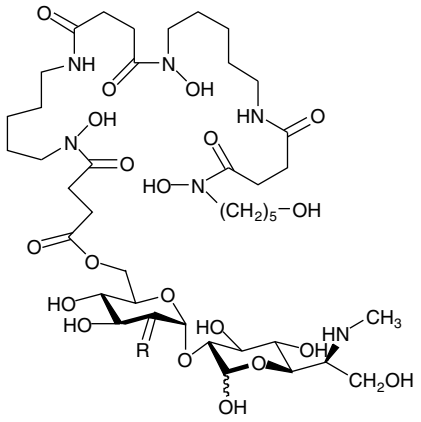
4



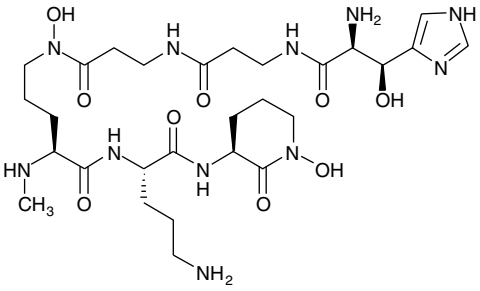
5



6



7 R = O or $(OH)_2$



8

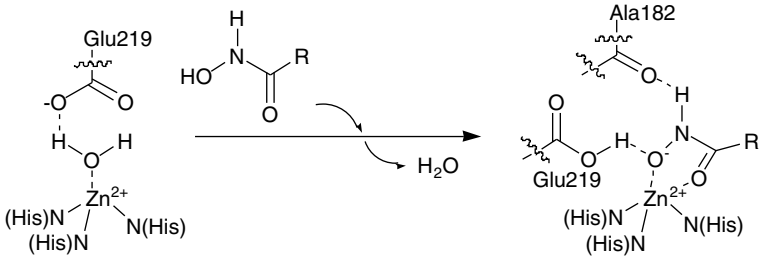
Figure 3.5 Chemical structures of pseudobactin (**4**), nannochelin (**5**), acinetonferrin (**6**), desferrisalmycin B (**7**), and exochelin MN (**8**).

3.2.2.2.1 **MMP Inhibitors** The MMPs are a family of more than 20 zinc-containing proteolytic enzymes that play a key role in healthy tissue remodeling and degradation of proteinaceous components of the extracellular matrix. Their abnormal expression has been implicated in a variety of disease states, including rheumatoid and osteoarthritis [38], tumor metastases [39–41], and cardiovascular disease [42]. The search for modulators of MMPs as therapeutic agents for diseases in which excess MMP activity has been implicated has long been the subject of intense investigation

and has resulted in a plethora of small-molecule, orally active, MMP inhibitors [43–48]. Both macromolecular (natural tissue inhibitors of metalloproteinases and monoclonal antibodies) and small-molecule inhibitors (synthetic and naturally occurring) have been investigated. Most small-molecule synthetic inhibitors follow a basic design strategy – a peptidomimetic backbone coupled to a zinc-binding group (ZBG) containing three bound histidine residues [45–47]. While the former has varied considerably with increasing knowledge of MMP structure and function, hydroxamic acids have been utilized as the ZBG in most inhibitors to date [43–48]. The hydroxamic acid has been shown by X-ray crystallography to chelate the catalytic Zn^{2+} ion via *O,O′*-bidentate coordination to form a slightly distorted trigonal bipyramidal coordination geometry [46] with average Zn−O bond lengths of around 2.0 Å [47]. The potential for one or more noncovalent interactions between the inhibitor and enzyme backbone is also a requirement for inhibitory activity. Hydrogen bonding between the hydroxamato oxyanion and the carboxylate oxygen of the catalytic Glu219 in human MMP-1, for example, and between the hydroxamato NH and the carbonyl oxygen of Ala182, also contribute to the binding (Scheme 3.2) [46].

Such binding prevents the natural substrate gaining access to the catalytic active site, thus rendering the metal incapable of catalyzing peptide hydrolysis. Hydroxamate coordination to the zinc and additional hydrogen bonding interactions between all the heteroatoms of the hydroxamate moiety and the enzyme active site account for the greater potency of hydroxamate-based MMP inhibitors over other classes of compounds [45].

However, the clinical performance of early hydroxamate-based MMP inhibitors, of which the three best known examples are batimastat, BB-1101, and marimastat (**9**, Figure 3.6), proved disappointing for three reasons. (i) The hydroxamic acid ZBG was found to be metabolically labile, resulting in loss of the hydroxamate ZBG by its reduction to the amide, its hydrolysis to the corresponding carboxylate, and its conjugation as a glucuronide [48]. (ii) The inhibitors were nonselective and, in many cases, were found to inhibit other classes of metallo-enzymes such as the TACE protease family [53]. (ii) Only upon clinical patient exposure did it become evident that these compounds induced dose-limiting musculoskeletal toxicity (MST) or musculoskeletal syndrome in a large cohort of patients [45, 48].



Scheme 3.2 Hydrogen bonding interactions between the hydroxamato oxyanion and the carboxylate oxygen of the catalytic Glu219 in human MMP-1 and between the hydroxamato NH and the carbonyl oxygen of Ala182 [46].

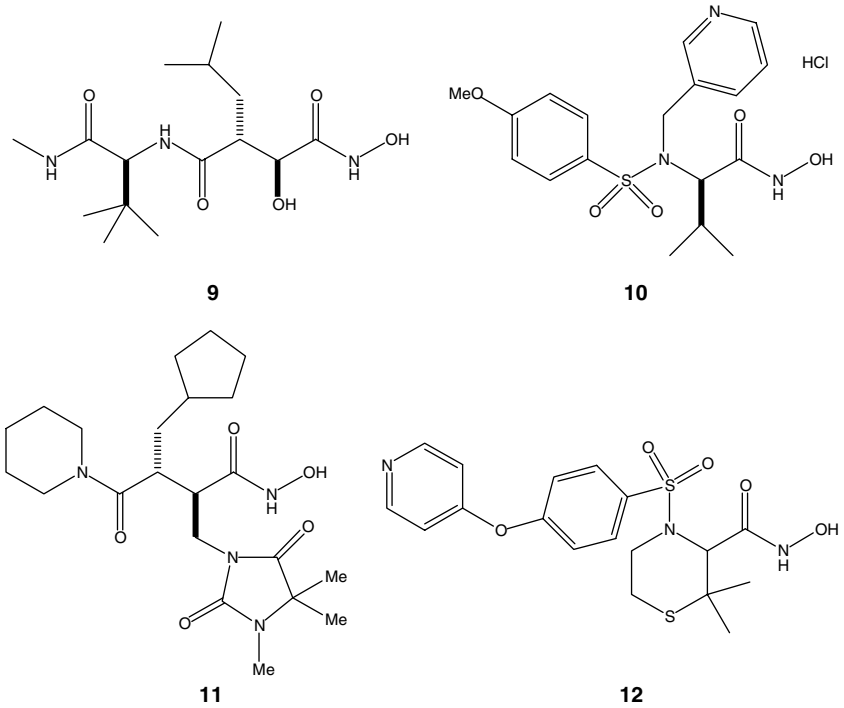


Figure 3.6 Chemical structures of the hydroxamate-based MMP inhibitors marimastat (**9**, BB-2516, GI-5712, KB-R-8898), MMI-270 (**10**, CGS-27023A), cipemastat (**11**, Trocade, Ro 32-3555), and prinomastat (**12**, AG 3340).

MMI-270 (**10**) is representative of the transition from early initiatives to the next generation of MMP inhibitors. Although a broad-spectrum inhibitor, it had the advantage of being orally active. It was found, however, to have limited efficacy due to metabolic liability and MST side-effects [45, 48]. These limitations were addressed to some degree with the development of cipemastat (**11**), amongst others [48]. Cipemastat was shown to inhibit MMP-1, -3, and -9, and was developed as a treatment for rheumatoid and osteoarthritis; however, despite having animal efficacy, its clinical trial was terminated prematurely [48]. Of the hydroxamate-based inhibitors developed to date, prinomastat (**12**) is one of the most extensively studied and has also been utilized to highlight the complex relationship between MMP inhibition and metastases [48]. While it demonstrated excellent anticancer efficacy in preclinical [48] and clinical trials [49], and was also shown to be efficacious in a type I diabetes animal model [50], a recurring limitation of prinomastat and hydroxamate-based inhibitors, in general, is the susceptibility of the hydroxamic acid moiety to metabolic degradation as mentioned earlier in this section. Suppression of metabolism and increased selectivity formed the focus of subsequent studies. Some orally active diaryl ether hydroxamic acids, a few of which advanced to clinical trials, were shown to suppress metabolism while having

greater selectivity and were investigated as anticancer, antiarthritic, and antiangiogenic agents [48]. The trend is now to focus on disease types in which excessive activity of a single MMP has been implicated, such as MMP-13 in arthritis. Despite their aforementioned limitations, new-generation peptidomimetic hydroxamate-based MMP inhibitors, with increased suppression of metabolism and increased selectivity, are still being disclosed and while none have progressed thus far to the clinic, research on this class of compounds has been instrumental in advancing our understanding of the diverse roles played by these intriguing enzymes. Hydroxamate-based MMP inhibitors thus far reported have been categorized as substrate-analog peptides, succinyl, sulfonamide, phosphinamide hydroxamic acids, and derivatives [46].

A schematic representation of the interaction between a hydroxamate-based peptide inhibitor with the active site of MMP-1 is given in Figure 3.7, and serves to illustrate the structural requirements necessary to maximize selectivity and potency [46].

It was recently reported by Ganea *et al.* [51] that, of the MMP inhibitors (hydroxamate and non-hydroxamate-based) investigated as therapeutic candidates especially for the treatment of cancer, arthritis, and cardiovascular disease, 82% have been discontinued and of the seven compounds remaining under clinical evaluation, only one, periostat (**13**, Figure 3.8), a non-hydroxamate-based inhibitor, has received US FDA approval to date [51].

The lack of success of these compounds can be attributed to their low specificity, toxicity, the ability of MMPs to adapt their binding sites to the shape of the inhibitor, and the underestimated level of complexity of MMP cell biology [51].

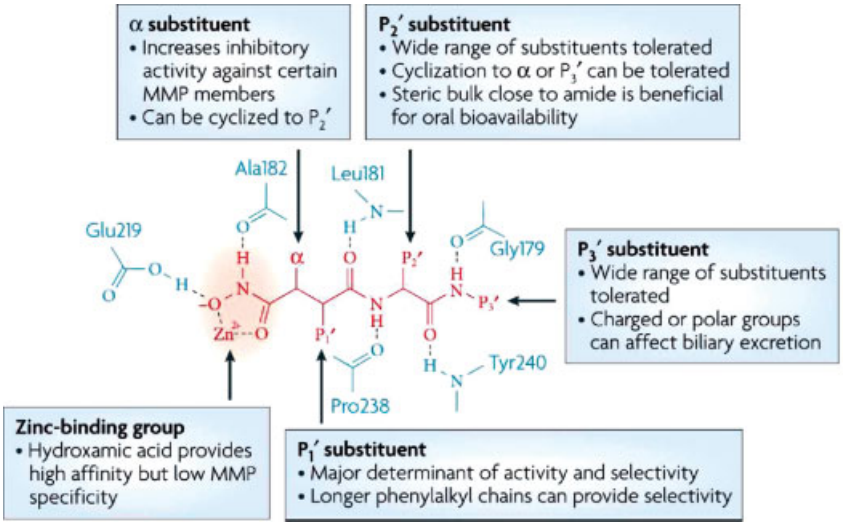
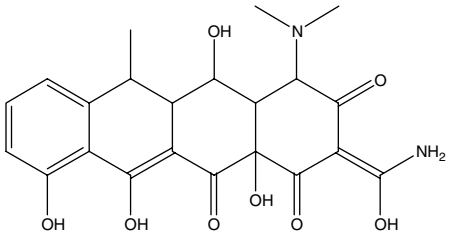


Figure 3.7 A schematic representation of the interaction between a hydroxamate-based peptide inhibitor with the active site of MMP-1. (Reprinted with permission from [46]).



13

Figure 3.8 Structure of periostat (**13**).

Fisher and Mobashery provide an excellent general review of recent developments in MMP inhibitor design [48]. Rao summarizes recent advances in the design of MMP inhibitors based on structural and computational studies [45]. Excellent reviews on the role of MMP inhibitors as anticancer therapeutics are provided by Mannello *et al*. [44] and Vihinen *et al*. [41], while Hu *et al*. compare the major classes of MMP inhibitors as therapeutics for inflammatory and vascular diseases [46]. A bioinorganic perspective on the inhibition of MMPs is provided by Puerta and Cohen [47].

3.2.2.2.2 **HDAC Inhibitors** Histone acetyltransferases (HATs) and HDACs are classes of two enzymes that have been identified as therapeutic targets in eukaryotic cells [52, 53]. This is due to their ability to regulate chromatin structure and function through the dynamic acetylation and deacetylation, respectively, of lysine residues of core histones – proteins around which DNA coils [52, 53]. Such post-translational covalent modification can impact on a diverse range of biological processes such as gene regulation, DNA repair, and mitosis. Abnormal expression or a disruption in the activity of these enzymes has been found in malignant tissues and, as such, there has been an explosion of interest into the possible use of HAT or HDAC inhibitors as potential anticancer agents. Inhibitors of HDACs (of which 18 enzyme types have been identified in humans) lead to hyperacetylation of histone lysine residues, and have been shown to promote cell cycle arrest and apoptosis and inhibit angiogenesis [54–58]. Naturally derived and synthetic HDAC inhibitors can be divided into several distinct structural classes, including hydroxamic acids, short-chain fatty acids, benzamides, cyclic tetrapeptides, electrophilic ketones, and some miscellaneous, and at least 11 have progressed to clinical trials [56, 57]. Each class has limitations associated with their use, including hydroxamate-based inhibitors which have been found to have an unfavorable pharmacokinetic profile as a result of glucuronidation and sulfation, together with the fact that the hydroxamic acid functional group is subject to metabolic hydrolysis (see Section 3.2.2.2.1) and all of which result in short *in vivo* half-lives of the hydroxamic acid moiety [59]. Hydrolysis of the hydroxamic acid group, for example, may also result in the formation of hydroxylamine, which has potent undesirable mutagenic properties [60]. Despite these limitations, a large number of hydroxamate-based

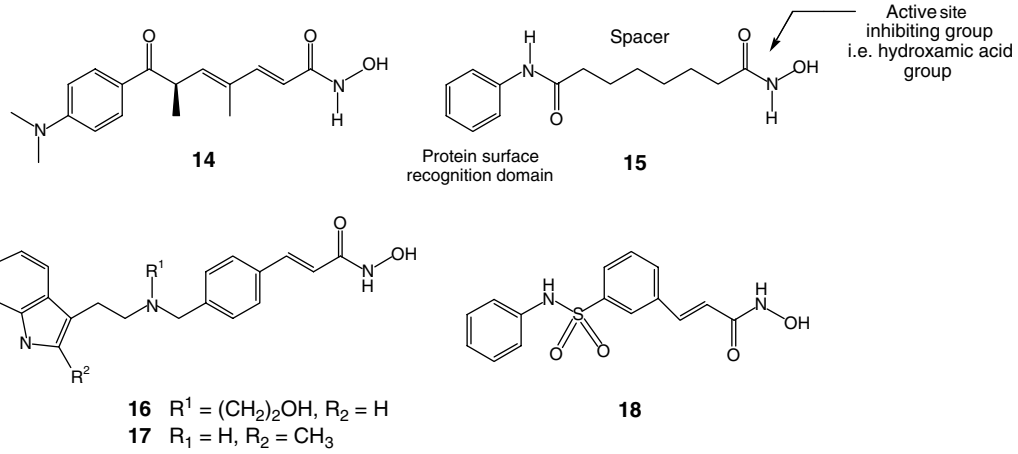
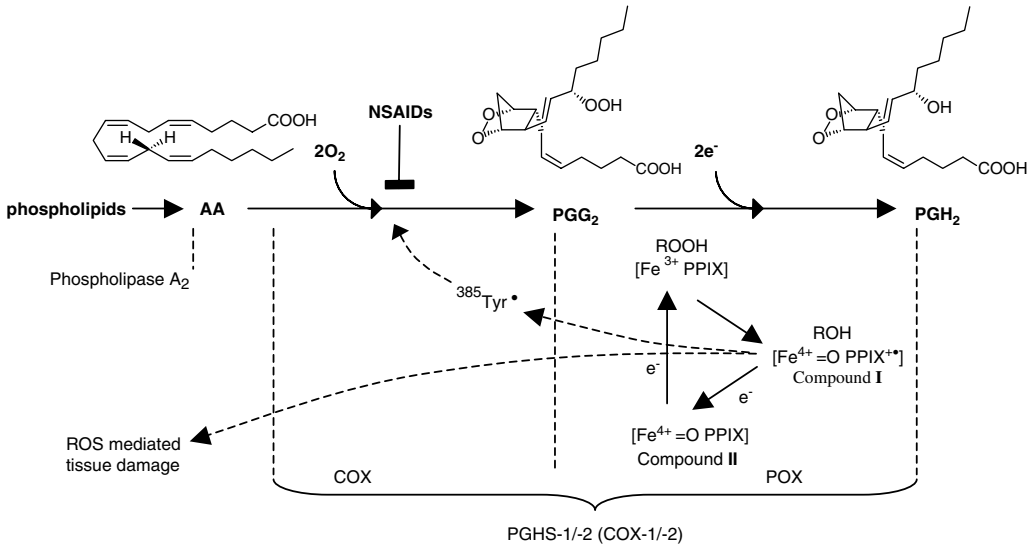


Figure 3.9 Structure of trichostatin A (**14**), SAHA (**15**), LAQ824 (**16**), LBH589 (**17**), and belinostat (**18**).

HDAC inhibitors have been investigated. Trichostatin A (**14**, Figure 3.9) [61] was the first natural product hydroxamate to be discovered with HDAC-inhibitory properties and since then, already one, suberoylanilide hydroxamic acid (SAHA, Vorinostat, Zolinza, **15**), has been approved by the US FDA as a once-daily oral treatment of cutaneous manifestations of advanced cutaneous T cell lymphoma [62–64].

SAHA is well tolerated by patients at doses which elicit a potent anticancer effect [62, 63]. Crystal structures of a human HDAC8 [65] and a bacterial HDAC homolog [66], HDAC-like protein, with SAHA bound shows that the catalytic domain consists of a narrow tube-like channel in which a Zn^{2+} ion is deeply buried. Elucidation of this and other HDAC enzyme structures provided the platform upon which new HDAC inhibitors could be designed with three distinct structural requirements identified as necessary for enzyme inhibition: a pharmacophore containing a ZBG that can interact directly with the catalytic Zn^{2+} ion, a spacer or linker that occupies the tube-like channel, and a surface domain that interacts with residues on the surface of the active site [65, 66]. The pharmacophore in many of the HDAC inhibitors developed to date consists of a hydroxamic acid moiety and in fact many hydroxamate-based HDAC inhibitors as anticancer agents are currently under clinical evaluation, including LAQ824 (**16**), [67] LBH589 (**17**), and belinostat (**18**, PXD101) [57, 68, 69].

Bolden *et al.* [57] provide a comprehensive summary of recent advances in the understanding of the molecular events that underlie the anticancer effects associated with HDAC inhibitors, including those containing a hydroxamic acid moiety as their ZBG. The review by Bolden *et al.* also includes a discussion as to how such information could be utilized in progressing these agents to the clinic [57, 63]. Marks and Breslow provide an excellent perspective on the significant milestones in the discovery and development of SAHA [64].

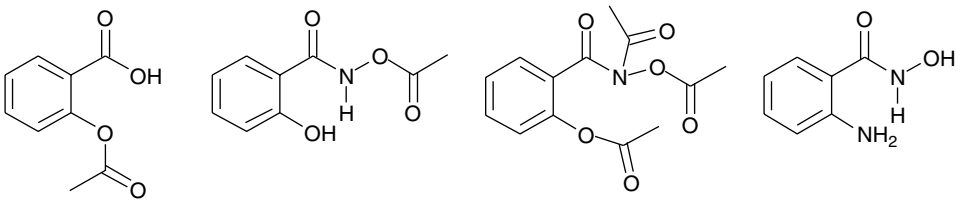


Scheme 3.3 Schematic representation of PGHS activity.

3.2.2.2.3 **PGHS Inhibitors** PGHS is an enzyme responsible for the generation of prostaglandins – bioactive lipids that mediate a variety of physiological processes, including platelet aggregation, modification of vascular tone, inflammation, nociception, and temperature control [70, 71]. Pharmacological control of PGHS activity is therefore central to the treatment of fever, inflammation, and heart disease, and has also been implicated in the prevention of colon cancer and Alzheimer's disease [70, 72]. The enzyme, of which three isoforms exist, consists of two spatially separate and functionally distinct active sites: a cyclooxygenase (COX) site and a heme-containing peroxidase (POX) site [70]. The synthesis of prostaglandin H_2 (PGH_2) by PGHS takes place in two steps; the catalytic oxidation of its natural substrate arachidonic acid at the COX active site to form hydroperoxide prostaglandin G_2 (PGG_2), which is then reduced at the POX active site to the alcohol PGH_2 (Scheme 3.3) [70].

This enzyme, regardless of isoform, is inhibited by a class of drugs known as nonsteroidal anti-inflammatory drugss (NSAIDs) such as aspirin (**19**, Figure 3.10) that target the COX active site. Aspirin, for example, covalently modifies the COX active site through the irreversible acetylation of a serine residue (Ser-529), thereby preventing the entry of arachidonic acid, Figure 3.11.

A drawback associated with the use of NSAIDs, including aspirin, is the fact that they do not exert any influence on the POX site. Thus, despite the beneficial effects of NSAID treatment, POX activity can continue to produce damaging free radical species. Hydroxamic acid analogs of aspirin as COX inhibitors have been investigated because the hydroxamic acid moiety (i) is less acidic than the carboxylic acid group of aspirin and should therefore cause less topical irritation, (ii) when acetylated, should, like aspirin, inhibit the COX active site via the transfer of its acetyl group to



19 **20** **21** **22**

Figure 3.10 Structures of aspirin (**19**), *O*-acetylsalicylhydroxamic acid (**20**) [73] triacetylsalicylhydroxamic acid (**21**) [74], and anthranilic hydroxamic acid (**22**).

the Ser-529 residue, and (iii) may also inhibit the POX active site through hydroxamato coordination to the heme iron group, thereby preventing the formation of free radical species. Acetylated hydroxamic acid derivatives such as *O*-acetylsalicylhydroxamic acid (**20**) have been shown to be effective inhibitors of COX through the acetylation of the active-site serine residue in much the same way as that done by aspirin [73, 74]. Triacetylsalicylhydroxamic acid (**21**), in particular, was found to be more potent an inhibitor of COX than aspirin [74]. It can potentially acetylate two molecules of PGHS via the formation first of *O*-acetylsalicylhydroxamic acid (**22**). Loss of the acetyl groups would yield salicylhydroxamic acid – a known inhibitor of myeloperoxidase [75]. A number of aromatic hydroxamic acids have been shown to have POX inhibitory activity through spectroscopic as well as crystallographic studies [76–78]. The *ortho*-substituted aromatic hydroxamic acids with a group capable of hydrogen bonding were shown to represent the minimum structural determinants for COX and POX inhibitors [75]. However, an investigation into the POX inhibitory activity of *O*-acetylsalicylhydroxamic acid (**20**), a potential metabolite of the parent COX inhibitor, found it not to be an effective POX inhibitor. Anthranilic hydroxamic acid (**22**), in contrast, was found to significantly inhibit POX and some

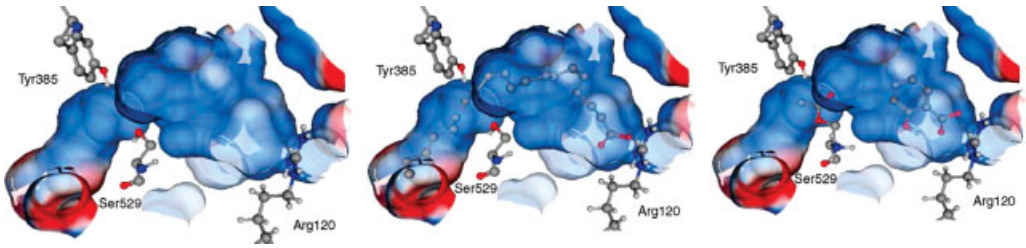


Figure 3.11 Inhibition of PGHS-1 by aspirin. (Left) MOE screen shot of the COX active site of PGHS-1, showing the active-site residues Tyr385, Ser529, and Arg120. (Middle) MOE screen shot of the COX active site of PGHS-1 containing its natural substrate, arachidonic acid. (Right) MOE screen shot showing interaction of the carboxylate of salicylic acid with Arg120 and acetylation of the active site residue Ser529, thus preventing entry of arachidonic acid into the COX active site. The Protein Data Bank structures used to build this model are 1PTH, 1DIY, and 1EBV. MOE (Molecular Operating Environment) from the Chemical Computing Group (www.chemcomp.com) was used to draw these figures.

29 anthranilic hydroxamic acid derivatives have been investigated [79]. Of those studied, 3,5-diiodoanthranilic hydroxamic acid exhibited a 10-fold improvement in inhibition [79].

Given the now established importance of hydroxamic acids as potential therapeutics in medicine, in particular as metallo-enzyme inhibitors, it is not surprising that there has been an increasing number of published methods with regard to their synthesis and that of their derivatives.

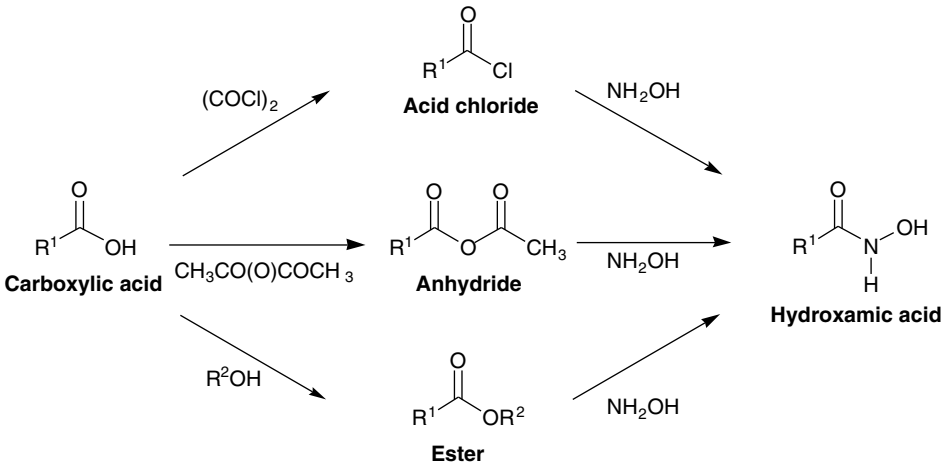
3.3 Solution-Phase Synthesis of Hydroxamic Acids

Up until 1996, solution phase was the only method available for the synthesis of hydroxamic acids and although solid-phase synthesis has become increasingly important, solution-phase synthesis continues to play an important role for good reasons: (i) more solution-phase relative to solid-phase reactions have been reported and their reaction conditions optimized; (ii) solution-phase reactions generally can be followed more conveniently than solid-phase reactions by standard analytical techniques such as thin-layer chromatography or NMR spectroscopy; and, very importantly, (iii) solution-phase reactions can be carried out routinely on a larger scale relative to solid-phase synthesis. The solution-phase synthesis of hydroxamic acids was last reviewed in 2003 by Yang and Lou [20]. Herein, progress over the last 35 years in the solution-phase synthesis of hydroxamic acids will be summarized based on the type of precursor from which hydroxamic acids can be derived. In addition, the synthesis of *N*-formylhydroxylamine, reverse or retro-hydroxamic acids, and acylhydroxamic acids will be discussed. Exemplary syntheses of hydroxamic acids are given in Section 3.6.

3.3.1 Synthesis of Hydroxamic Acids Derived from Carboxylic Acid Derivatives

Hydroxamic acids can be considered to be *N*-acylated derivatives of hydroxylamine. In fact, many are synthesized by the reaction of carboxylic acid derivatives with hydroxylamine or *O*/*N*-protected hydroxylamines (generally generated *in situ* by the addition of base to the hydrochloride salts). Direct reaction of carboxylic acids with hydroxylamine is difficult as the OH is a poor leaving group and the use of an acid catalyst to effect a direct transformation of carboxylic acids to hydroxamic acids is not feasible as hydroxylamines are basic in nature [80]. Therefore, *N*-acylation with activated carboxylic acids (readily generated from the corresponding carboxylic acid) such as esters, acid chlorides, and anhydrides typically results in formation of the analogous hydroxamic acid (Scheme 3.4) [16, 20, 81, 82]. Activation of the carboxy component of carboxylic acids involves attachment of a good leaving group to the acyl carbon, thereby enabling successful attack of hydroxylamine.

There are three different approaches to generating hydroxamic acids from carboxylic acid derivatives as outlined in Scheme 3.4. (i) The activated carboxylic



Scheme 3.4 Synthesis of carboxylic acid derivatives and corresponding hydroxamic acid derivatives.

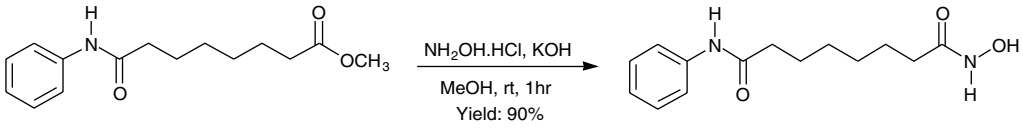
acid derivative, RCOX, can be generated from a carboxylic acid, RCOOH, followed by direct treatment with hydroxylamine. (ii) RCOX can be generated from a carboxylic acid and isolated or purified prior to treatment with hydroxylamine. (iii) RCOX can be generated from a carboxylic acid in the presence of hydroxylamine in a one-pot synthesis. The mechanism for the formation of hydroxamic acids involves nucleophilic attack of hydroxylamine on the carbonyl carbon of the carboxylic acid derivative to give a tetrahedral intermediate, which subsequently rearranges, resulting in loss of the leaving group and formation of the corresponding hydroxamic acid.

In certain instances it is necessary to protect the hydroxamic acid functional groups with a view to eliminating undesirable side reactions in subsequent reactions. Common *O*-protecting groups include butyl, benzyl, allyl, trityl, tetrahydropyranyl (THP), and silyl groups, whereas *N*-protecting groups comprise Boc, phenoxycarbonyl, and silyl groups, amongst others [83, 84].

3.3.1.1 From Esters

Heinrich Lossen isolated oxalohydroxamic acid after treating hydroxylamine with diethyloxalate [1], thus pioneering the synthesis of hydroxamic acids via the reaction of esters with hydroxylamine. This historic and extremely effective method was subsequently used to synthesize a vast array of different hydroxamic acids [85–90]. Hydroxylamine is generally generated by adding KOH or NaOH to an aqueous or methanolic solution of hydroxylamine hydrochloride. If commercially unavailable, esters are readily generated by standard reactions such as the S_N2 reaction of a carboxylate anion and a primary alkyl halide or via Fischer esterification of a carboxylic acid with an alcohol (Scheme 3.4) [91].

The synthesis of SAHA, a potent HDAC inhibitor (see Section 3.2.2.2.1), from methyl suberanilate is representative of how easily hydroxamic acids can be synthesized from their corresponding esters, Scheme 3.5 [92].



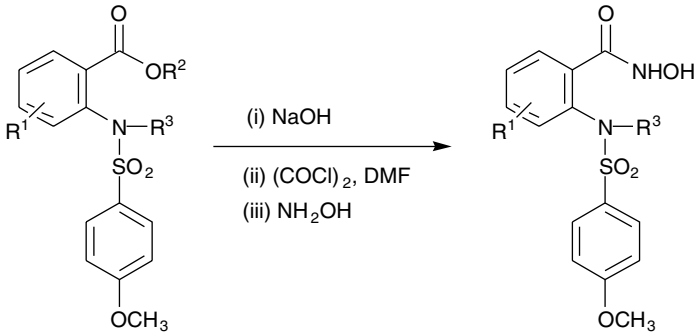
Methyl suberanilate **15**

Scheme 3.5 Synthesis of SAHA (**15**).

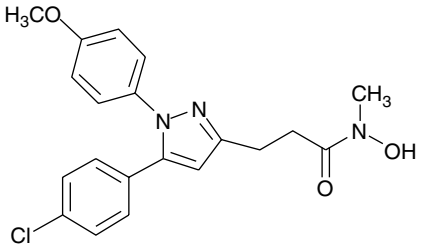
One potential problem is that efficient transformation of esters to hydroxamic acids does not proceed under neutral conditions rather it occurs at high pH (>10). Therefore this method is not suitable if the ester substrates contain base sensitive groups [83]. In addition, esters can hydrolyze to the corresponding carboxylic acids under alkaline conditions [93]. Ho *et al.* found that the addition of small quantities of KCN can accelerate the conversion of esters to hydroxamic acids upon reaction with hydroxylamine [94]. This method is similar to that of the well established use of cyanide as a catalyst for the aminolysis of nonactivated esters [95]. In addition Massaro *et al.* demonstrated that microwave irradiation dramatically accelerates the transformation of esters to hydroxamic acids upon reaction with hydroxylamine. Certain reactions were completed within 6 min, for example [93]. Two exemplary reactions for the synthesis of hydroxamic acids from esters are given in Sections 3.6.1 and 3.6.2.

3.3.1.2 From Acid Halides

Carboxylic acids readily give acid chlorides upon treatment with reagents such as $SOCl_2$, PCl_5, and $(COCl)_2$ [96] thereby replacing the OH of the acid with a much better Cl leaving group (Scheme 3.4). These reagents may not be compatible with carboxylic acid substrates containing acid-sensitive groups though, due to the production of acid as the reaction progresses [96]. Hydroxamic acids can be synthesized from acid chlorides upon treatment with hydroxylamine in a suitable unreactive dry solvent [e.g., tetrahydrofuran (THF) or *N,N*-dimethylformamide (DMF)]. For example, Levin *et al.* successfully synthesized a series of hydroxamic acid inhibitors of MMPs and TACE [42, 97–100]. A general synthesis for a series of sulfonamide hydroxamic acids is shown in Scheme 3.6, where an ester is hydrolyzed, followed by formation of the acid chloride and subsequent conversion to the hydroxamic acid [97].



Scheme 3.6 Synthesis of sulfonamide hydroxamic acids.



23

Figure 3.12 Chemical structure of tepoxalin (**23**).

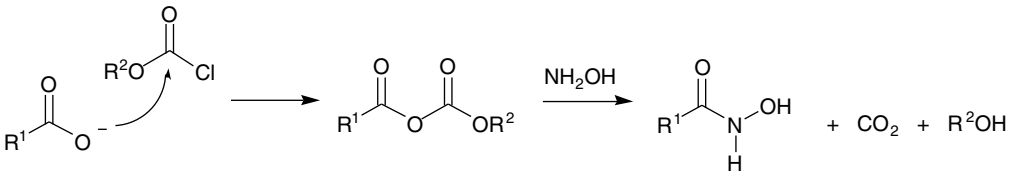
The preparation of hydroxamic acids from acid chlorides can be tedious due to the high reactivity of acid chlorides and the difficulty in preventing further acylation of hydroxylamine [83]. Synthesis of *N*-alkylhydroxamic acids can be achieved with less difficulty by reacting *N*-alkylhydroxylamines with acid chlorides [5, 101]. For example, Murray *et al*. synthesized the potent anti-inflammatory agent, tepoxalin (**23**, Zubrin) (Figure 3.12), and analogs by first generating acid chlorides from diarylpyrazolylpropanoic acids using oxalylchloride and subsequently converting the acid chlorides to *N*-methylhydroxamic acids upon reaction with *N*-methylhydroxylamine [102].

More recently, Flipo *et al*. synthesized *O*-*tert*-butyl or *O*-tritylhydroxamic acids based on a malonic template. Malonic acid derivatives can be treated with oxalylchloride to give the corresponding acid chlorides, followed by treatment with *O*-*tert*-butylhydroxylamine or *O*-tritylhydroxylamine to give the corresponding *O*-alkylhydroxamic acids. The *tert*-butyl and trityl groups are easily removed to give the parent hydroxamic acids using boron tris(trifluoroacetate) in trifluoroacetic acid (TFA) and TFA alone, respectively [103].

As acid chlorides react vigorously with water, the acid chloride substrates need to be soluble in anhydrous organic solvents. In general, acid fluorides are found to be more stable than their chloride analogs to tertiary bases and even neutral oxygen nucleophiles such as water [80]. Therefore, one might consider in certain instances using acid fluorides as alternative precursors to hydroxamic acids. An exemplary reaction for the synthesis of hydroxamic acids from an acid chloride is given in Section 3.6.3.

3.3.1.3 From Anhydrides

An alternative approach for the activation of carboxylic acids is generation of the corresponding anhydrides. Anhydrides are generally prepared by nucleophilic acyl substitution of an acid chloride with a carboxylate anion. Carboxylic acids also readily give mixed anhydrides upon treatment with acetic anhydride (Scheme 3.4) [91]. Perhaps the best approach in this regard is to generate the mixed anhydride, such as a mixed carbonic acid anhydride, from the reaction of a carboxylic acid, ethylchloroformate, and a slight excess of tertiary amine base in a suitable unreactive dry solvent (Scheme 3.7) [80, 83]. It is noteworthy that in addition to being moisture sensitive ethylchloroformate is also a strong irritant [104]. With regard to the mechanism in Scheme 3.7, the chloroformate-derived carbonyl group (R^2CO) in the mixed



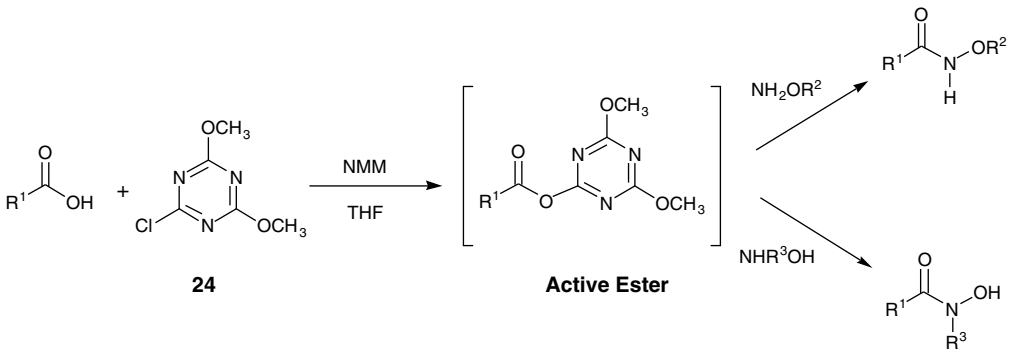
Scheme 3.7 Synthesis of hydroxamic acids via mixed carbonic acid anhydrides.

anhydride is less susceptible to nucleophilic attack than the carbonyl (R^1CO) of the original carboxylic acid as it is flanked by two oxygen atoms [80]. Therefore, nucleophilic attack of hydroxylamine will give the hydroxamic acid analog of the original carboxylic acid in good yield.

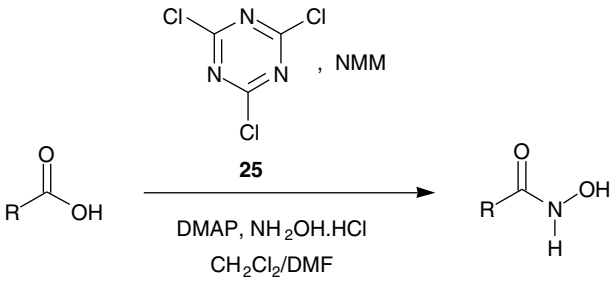
Reddy *et al*. successfully employed ethylchloroformate utilizing the method described in Scheme 3.7 under neutral pH conditions to synthesize aliphatic/aromatic hydroxamic acids that contain hydroxyl, halo, ester, and other base-sensitive substituents [83]. Similarly ethylchloroformate can be used to generate trichostatin A (**14**) [105] and SAHA (**15**) [106] – well-known HDAC inhibitors (see Section 3.2.2.2.2). An exemplary reaction for the synthesis of hydroxamic acids from an anhydride is given in Section 3.6.4.

3.3.1.4 From [1.3.5]Triazine-Coupled Carboxylic Acids

Carboxylic acids can be treated with triazine derivatives such as 2-chloro-4,6-dimethoxyl[1.3.5]triazine (CDMT, **24**) to give active esters, which can be reacted *in situ* with hydroxylamine derivatives (Scheme 3.8) to give the corresponding hydroxamic acids. For example, a one-pot reaction of 2-phenylpropionic acid with CDMT in the presence of *N*-methylmorpholine (NMM) in THF at room temperature for 1 h, followed by treatment with *O*-benzylhydroxylamine gives *O*-benzyl-2-phenylpropionhydroxamic acid in 85% yield [107]. Any unreacted carboxylic acid can be separated using a straightforward work-up procedure so no chromatography is necessary [107]. As CDMT is an irritant, 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride can be used as an alternative triazine, but reaction



Scheme 3.8 Synthesis of hydroxamic acids from carboxylic acids using CDMT (**24**).



Scheme 3.9 Synthesis of hydroxamic acids from carboxylic acids using TCT (**25**).

times are longer. This method can also be successfully employed to synthesize Weinreb amides (*N*-methoxy-*N*-methyl amides) [107].

In addition, a one-pot synthesis of hydroxamic acids from carboxylic acids and *N*-protected amino acids using 2,4,6-trichloro[1.3.5]triazine (TCT, cyanuric chloride, **25**) as a coupling agent (Scheme 3.9) has also been reported [104].

An exemplary reaction for the synthesis of hydroxamic acids from a CDMT-coupled carboxylic acid is given in Section 3.6.5.

3.3.1.5 From Carbodiimide-Coupled Carboxylic Acids

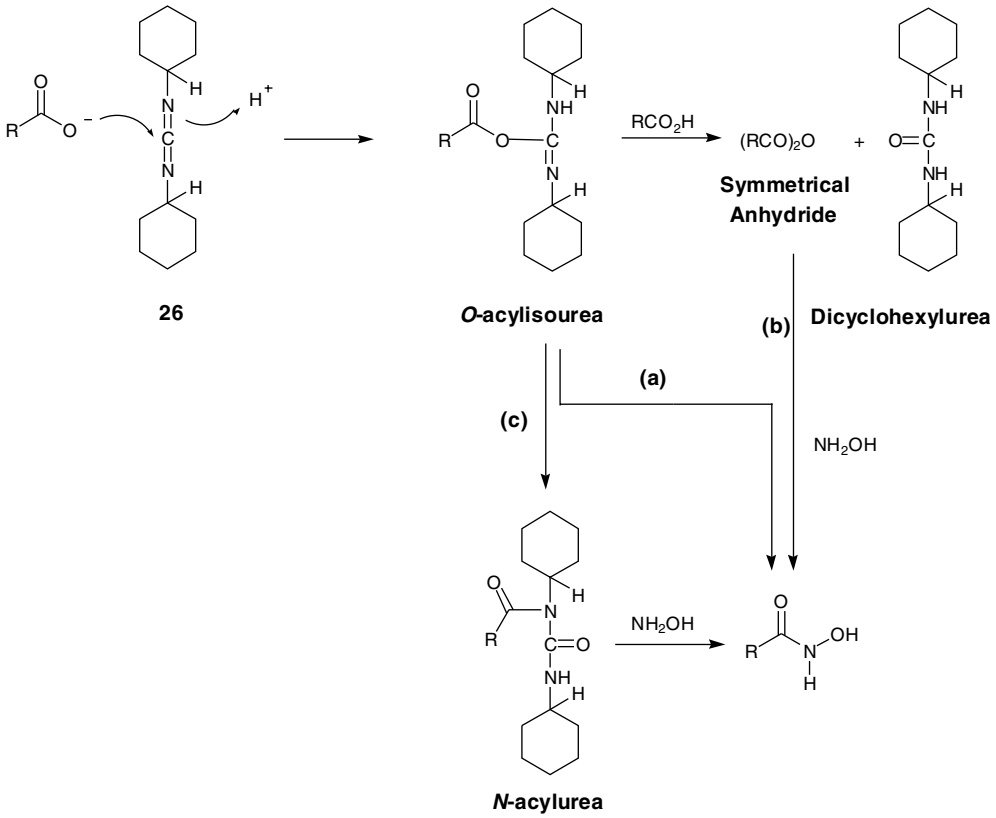
Coupling agents that were developed primarily for peptide bond formation are routinely used to activate carboxylic acids and in turn generate the corresponding hydroxamic acids. One such agent is *N,N′*-dicyclohexylcarbodiimide (DCC, **26**) [108–110]. The first step involves generation of an *O*-acylisourea – a potent acylating agent. This is achieved by reacting equivalent amounts of the carboxylic acid with the carbodiimide in an unreactive solvent [80]. Hydroxamic acid formation can occur via (*a*) hydroxylaminolysis of the *O*-acylisourea or (*b*) attack on a symmetrical anhydride (Scheme 3.10).

As conversion of *O*-acylisoureas to the less reactive *N*-acylurea can reduce subsequent yields (path *c*, Scheme 3.10), coupling in the presence of an additive such as 1-hydroxybenzotriazole (HOBt, **27**, Scheme 3.11) to give an active ester can improve yields and even prevent racemization (if relevant) [80].

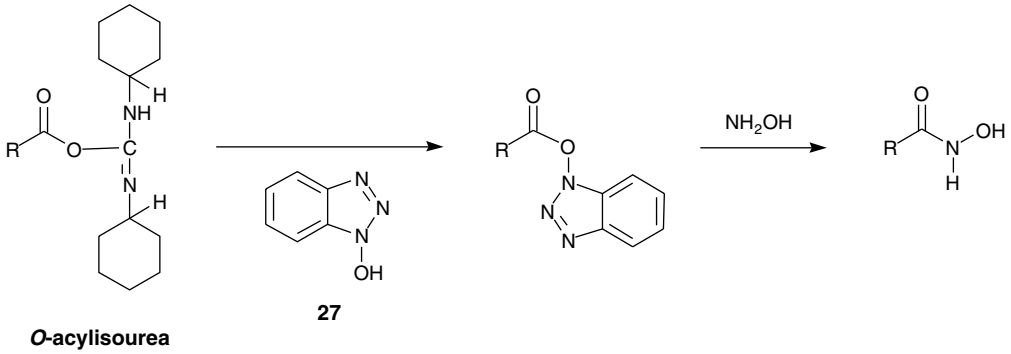
N,N′-Diisopropylcarbodiimide (DIC) [111] and the water-soluble *N*-(3-dimethylaminopropyl)-*N′*-ethylcarbodiimide (EDC) [110, 112–114] are two alternative carbodiimide coupling agents, which can be used to activate carboxylic acids to generate the corresponding hydroxamic acids. A limitation though is that *N,O*-diacylation of hydroxylamine has been observed upon reaction of hydroxylamine with carbodiimide-activated carboxylic acids and even using substoichiometric quantities of acids [115]. An exemplary reaction for the synthesis of hydroxamic acids from a EDC-coupled carboxylic acid is given in Section 3.6.6.

3.3.1.6 From Acyloxyphosphonium Ions

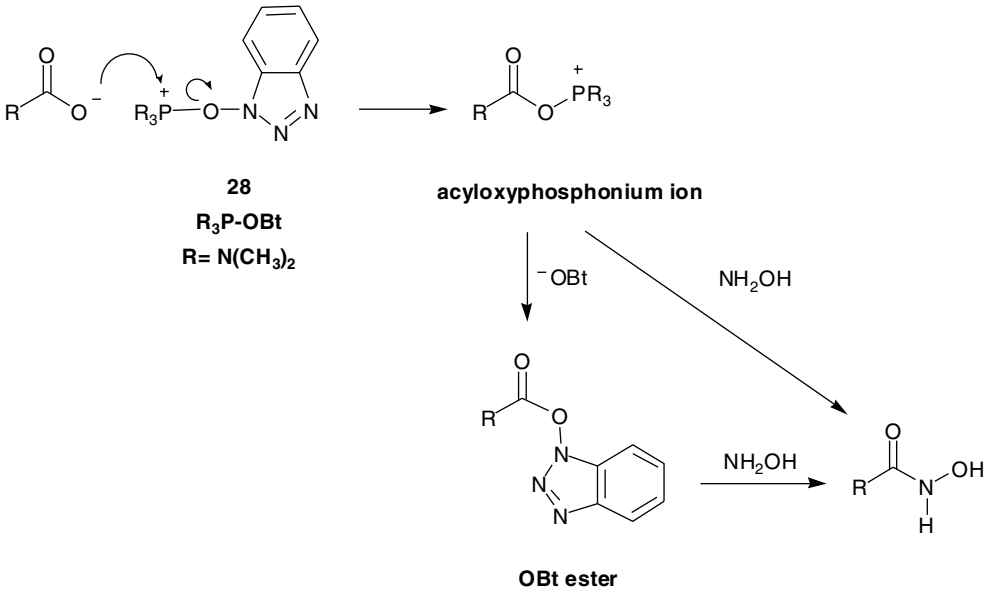
Reaction of phosphonium salts, RP_3-X, and carboxylate anions (generated using a tertiary base) gives the corresponding acyloxyphosphonium ions, which are excellent acylating agents. Benzotriazol-1-yloxytris(dimethylamino)-phosphonium



Scheme 3.10 Synthesis of hydroxamic acids from carboxylic acids, $RCO_2H : NH_2OH : DCC$ (**26**) (1 : 1 : 1).



Scheme 3.11 Synthesis of hydroxamic acids from carboxylic acids, $RCO_2H : NH_2OH : DCC$ (**21**):HOBt (**27**) (1 : 1 : 1 : 1).



Scheme 3.12 Synthesis of hydroxamic acids from carboxylic acids, $RCO_2H : NH_2OH : BOP$ (**28**) : TEA)/DIEA (1 : 1 : 1 : 1).

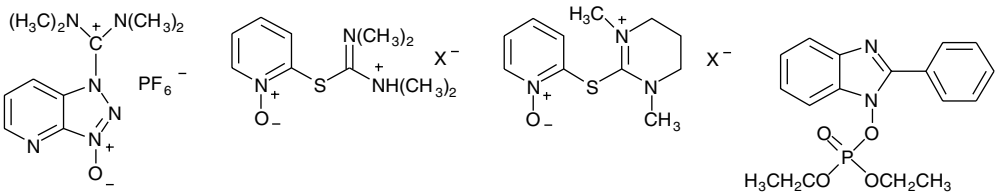
hexafluorophosphate (BOP, Castro's reagent, **28**) is a well-known phosphonium cation coupling agent, which can be employed to synthesize hydroxamic acids from carboxylic acids [116, 117]. Although numerous pathways are possible, the primary pathway is thought to be via the OBt ester (Scheme 3.12) [80].

There are safety reservations in relation to the use of BOP (**28**). Upon its reaction with carboxylic acids, hexamethylphosphoramide, a highly toxic derivative, is generated [80]. Therefore alternative reagents, such as 7-azabenzotriazol-1-yloxytris (dimethyamino)-phosphonium hexafluorophosphate and (7-azabenzotriazol-1-yloxy)tris(pyrrolidino)-phosphonium hexafluorophosphate have been developed for amide bond formation [112, 118]. An exemplary reaction for the synthesis of hydroxamic acids from an acyloxyphosphonium ion is given in Section 3.6.7.

3.3.1.7 From Carboxylic Acids Coupled with other Agents

O-(7-Azabenzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate (HATU, **29**) is a so-called "uronium" coupling reagent. It reacts with carboxylic acids in the presence of base to give the corresponding activated OAt esters [1-hydroxy-7-azabenzotriazole (HOAt)], which are more reactive than the corresponding OBt esters towards aminolysis. HATU was originally assigned the structure of an uronium salt (*O*-HATU), but NMR and crystallographic studies have identified the true structure as a guanidinium derivative (**29**, Figure 3.13) [119].

Bailen *et al.* demonstrated that tetrafluoroborate and hexafluorophosphate thiouronium salts, derived from 2-mercaptopyridone-1-oxide (**30** and **31** respectively Figure 3.13) and tetramethylurea or *N*,*N*′-dimethylpropyleneurea (**32** and **33** respectively



29 X= BF_4 **30** or PF_6 **31** X= BF_4 **32** or PF_6 **33** **34**

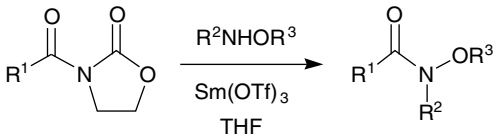
Figure 3.13 Chemical structures of HATU (**29**), tetrafluoroborate and hexafluorophosphate thiouronium salts derived from 2-mercaptopyridone-1-oxide and tetramethylurea (**30** and **31**, respectively) and *N,N′*-dimethylpropyleneurea (**32** and **33**, respectively), and phosphoric acid diethyl ester 2-phenyl-benzimidazol-1-yl-ester (**34**).

Figure 3.13), are efficient reagents for the direct coupling of carboxylic acids and *N*-protected amino acids to *O*-methyl and *O*-benzylhydroxylamine to give *O*-methyl or *O*-benzylhydroxamic acids in high yields, respectively [120].

More recently, Kokare *et al*. designed and synthesized a novel coupling agent, phosphoric acid diethyl ester 2-phenyl-benzimidazol-1-yl-ester (**34**, Figure 3.13), for the synthesis of *O*-alkylhydroxamic acids from carboxylic acids and *N*-protected amino acids [121]. An exemplary reaction for the synthesis of hydroxamic acids derived from a carboxylic acid coupled 2-phenyl-benzimidazol-1-yl-ester (**34**) is given in Section 3.6.8.

3.3.2
Synthesis of Hydroxamic Acids from *N*-acyloxazolidinones

The ongoing development of novel methods for the synthesis of hydroxamic acids is important given the growing interest in their use as therapeutic agents (see Section 3.2.2). One such example is that of the generation of hydroxamic acids from *N*-acyloxazolidinones using the Lewis acid, samarium triflate (Scheme 3.13) [122]. This conversion is straightforward, and proceeds with a high degree of chemoselectivity and without racemization. In addition to *O*-protected and *N*- and *O*-protected hydroxylamines, the parent hydroxylamine can also be utilized. Sibi *et al*. used this procedure to synthesize a known succinate-based MMP inhibitor, BB-1101 (see Section 3.2.2.2.1) [123].

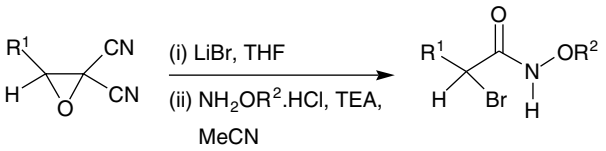


Scheme 3.13 Synthesis of hydroxamic acids from *N*-acyloxazolidinones.

An exemplary reaction for the synthesis of hydroxamic acids from an *N*-acyloxazolidinone is given in Section 3.6.9.

3.3.3
Synthesis of Hydroxamic Acids from *gem*-Dicyanoepoxides

Boukhris *et al.* have demonstrated that reaction of *gem*-dicyanoepoxides with hydroxylamine hydrochloride and triethylamine (TEA) in acetonitrile (MeCN) gives the corresponding α-chlorohydroxamic acids [124]. The corresponding α-bromohydroxamic acids, although less readily synthesized, can be obtained by reacting *gem*-dicyanoepoxides with LiBr followed by a variety of hydroxylamine derivatives (Scheme 3.14). In turn α-alkoxy, α-hydroxyl, and α-aminohydroxamic acids are easily synthesized from the α-halohydroxamic acids [125]. An exemplary reaction for the synthesis of hydroxamic acids from a *gem*-dicyanoepoxide is given in Section 3.6.10.

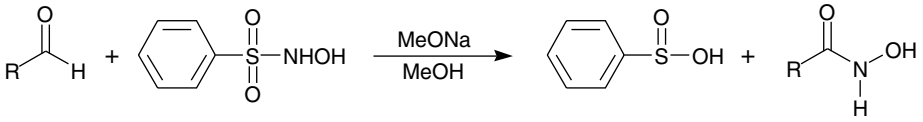


Scheme 3.14 Synthesis of hydroxamic acids from *gem*-dicyanoepoxides.

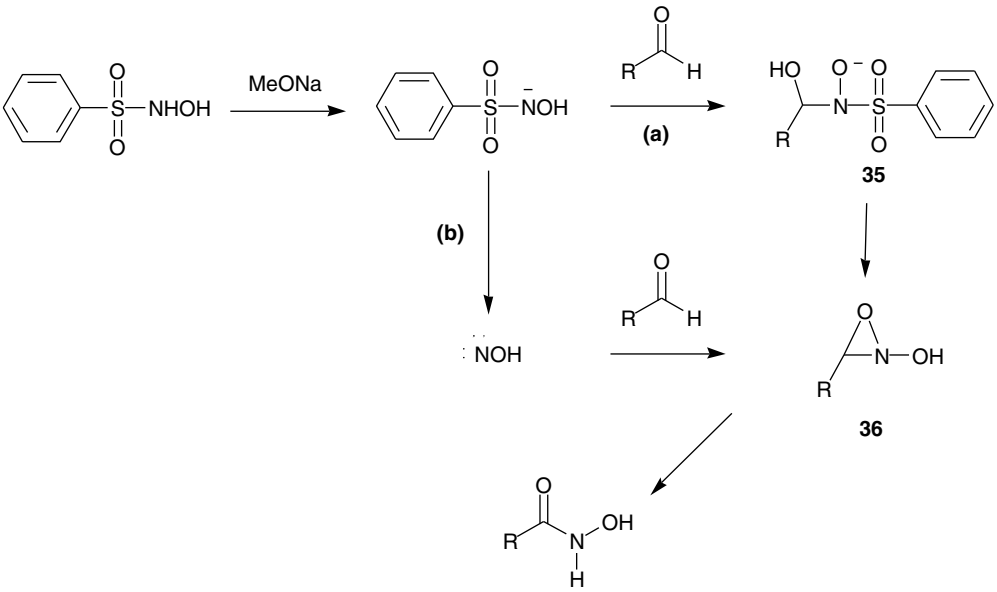
3.3.4
Synthesis of Hydroxamic Acids from Aldehydes

Over 100 years ago Angeli and Rimini discovered a method for synthesizing alkyl or aryl hydroxamic acids from the reaction of aldehydes with *N*-hydroxybenzenesulfonamide in the presence of a strong base in methanol (Scheme 3.15) [126–129]. Synthetic use of this method is limited though due to the difficulty in separating a side product, benzenesulfinic acid, from the desired hydroxamic acid [129].

Hassner *et al.* [130] proposed that the mechanism for hydroxamic acid formation may involve nucleophilic attack of the anion of *N*-hydroxybenzenesulfonamide to give **35**, followed by ring closure to give an *N*-hydroxyaziridine, **36**, which would ring open to give the hydroxamic acid (Scheme 3.16, *a*). Nucleophilic attack of an *N*-hydroxynitrene, on the aldehydic carbonyl carbon was also suggested as a possible route to **36** (Scheme 3.16, *b*).



Scheme 3.15 Angeli–Rimini reaction.



Scheme 3.16 Proposed mechanisms for the synthesis of hydroxamic acids from aldehydes.

An exemplary reaction for the synthesis of hydroxamic acids from an aldehyde is given in Section 3.6.11.

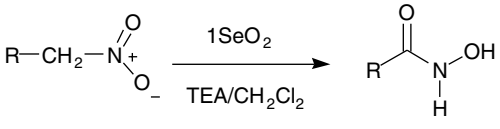
3.3.5
Synthesis of Hydroxamic Acids from Nitro Compounds

Aliphatic hydroxamic acids can be prepared by reacting one equivalent of a primary nitroalkane with one equivalent of selenium dioxide in the presence of TEA (Scheme 3.17, *a*). In the case of nitrooctane if two equivalents of selenium dioxide are used, 50% octanenitrile is formed (Scheme 3.17, *b*) [131]. The photorearrangement of alkane nitronate anions, which can be derived from primary and secondary nitroalkanes, provides an alternative synthetic method for the generation of hydroxamic acid functions (Scheme 3.17, *c*) [132].

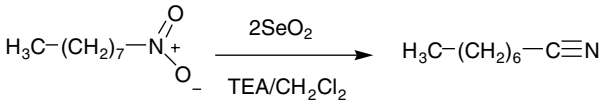
Two exemplary reactions for the synthesis of hydroxamic acids from nitro compounds are given in Sections 3.6.12 and 3.6.13.

3.3.6
Synthesis of Hydroxamic Acids via a Palladium-Catalyzed Cascade Reaction

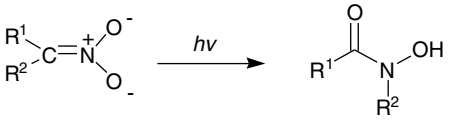
An alternate route to hydroxamic acids, developed by Grigg *et al.*, employs a catalytic cascade reaction. The hydroxamic acid moiety is generated via palladium-catalyzed carbonylation of aryl iodides and trapping of the acyl-Pd(II) intermediate with *O*-alkylhydroxylamines, which can be subsequently deprotected using standard conditions [133].



(a)



(b)

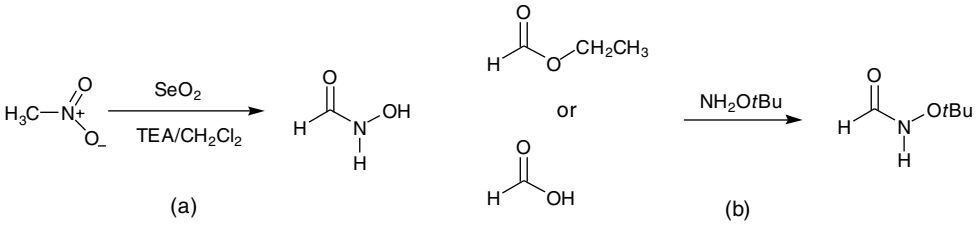


(c)

Scheme 3.17 Synthesis of (*a*) hydroxamic acids from nitroalkanes, (*b*) octanenitrile from nitrooctane, and (*c*) hydroxamic acids from alkane nitronate anions.

3.3.7
Synthesis of *N*-Formylhydroxylamine (Formohydroxamic Acid)

N-Formylhydroxylamine can be synthesized by reaction of nitromethane with selenium dioxide in the presence of TEA (Scheme 3.18) [131]. Alternatively, Stowell and Christenson synthesized *O-tert*-butyl-*N*-formylhydroxylamine by the reaction of formic acid or ethylformate with *O-tert*-butylhydroxylamine (Scheme 3.18), and the details of which are given in Section 3.6.14 [134].

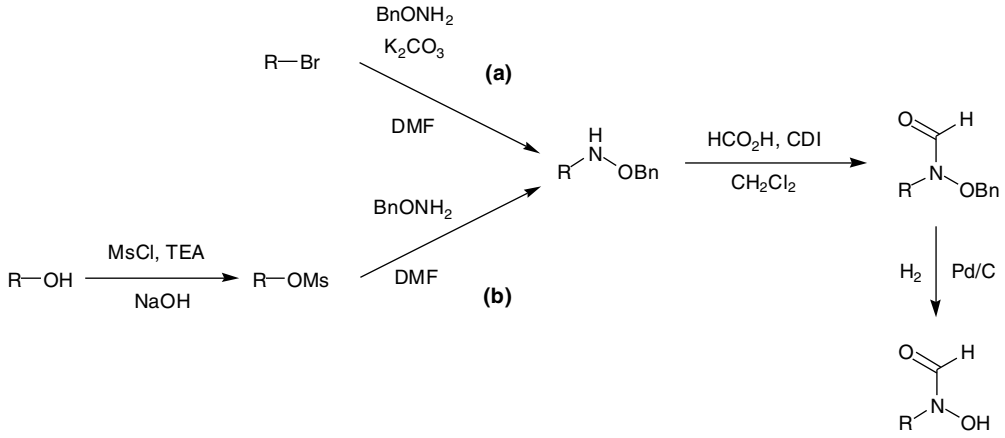


(a) (b)

Scheme 3.18 Synthesis of (*a*) *N*-formylhydroxylamine from nitromethane and (*b*) *O-tert*-butyl-*N*-formylhydroxylamine from formic acid or ethylformate.

3.3.8
Synthesis of Reverse or Retro-Hydroxamates

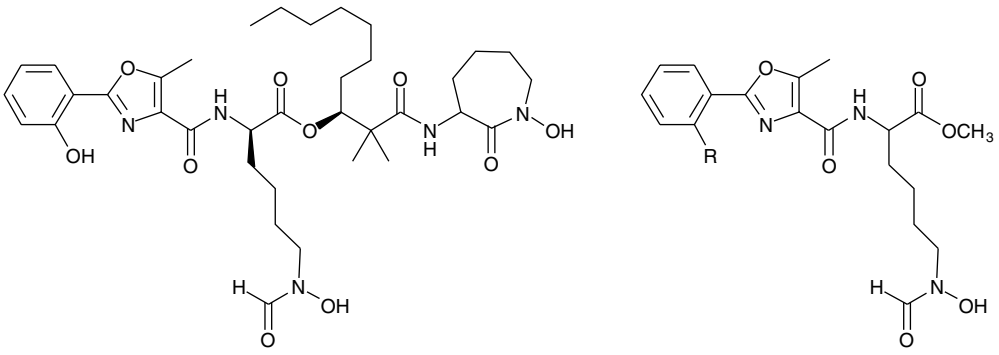
Reverse or retro-hydroxamates (see Section 3.2.1, Figure 3.3) are considered to be *N*-alkyl or *N*-aryl derivatives of *N*-formylhydroxylamine. Incorporation of the



Scheme 3.19 Incorporation of an *N*-formylhydroxylamine group into a target compound via (*a*) an alkyl halide and (*b*) a mesylate.

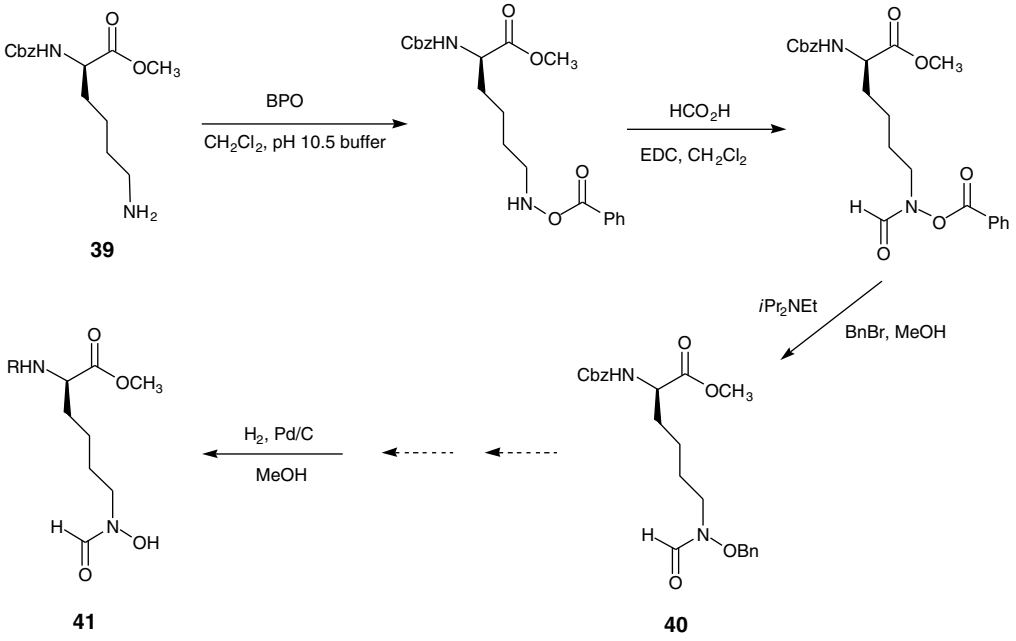
N-formylhydroxylamine group into a target compound, such as a potential metalloenzyme inhibitor, can be achieved in a number of synthetic steps. In one method a primary bromide can be displaced by *O*-benzylhydroxylamine followed by nucleophilic attack on the 1,1′-carbonyldiimidazole (CDI)-activated formic acid and subsequent benzyl deprotection by hydrogenolysis (Scheme 3.19). Alternatively the *N*-substituted *O*-benzylhydroxylamine can be generated via a mesylate as shown (Scheme 3.19) [135].

The natural products amamistatin A and B (**37**, Figure 3.14), which contain an *N*-formylhydroxylamine lysine moiety, have been demonstrated to have antiproliferative properties. Recently, Miller *et al.* successfully synthesized analogs of amamistatin B (**38**) with a view to testing their HDAC inhibitory properties and antitumor activity [136].



37 **38**

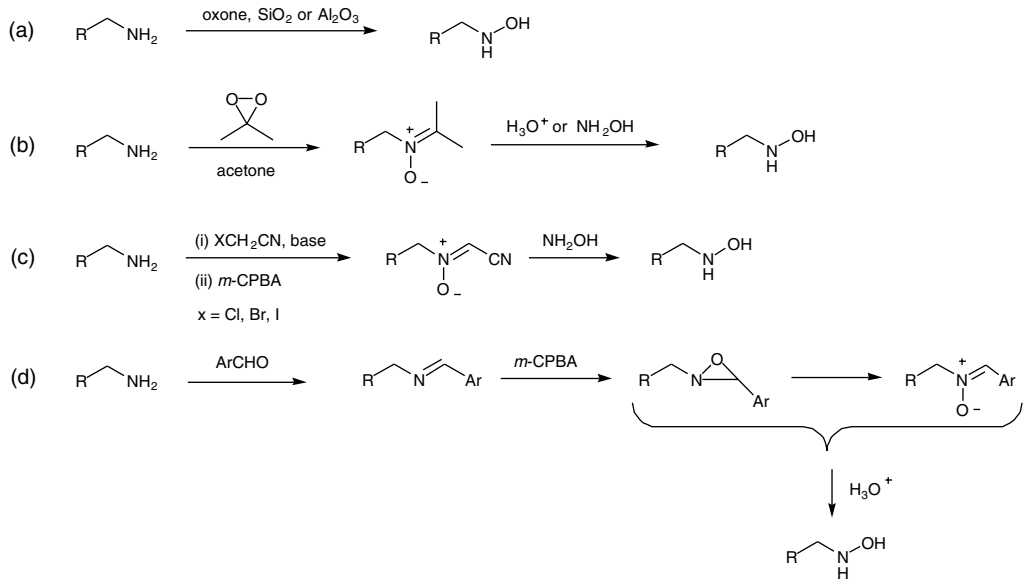
Figure 3.14 Chemical structure of amamistatin B (**37**) and general structure for amamistatin B analogs (**38**).



Scheme 3.20 Synthesis of an N^{ε}-formylhydroxylamine lysine residue.

The *N*-formylhydroxylamine lysine residue (**41**) of the amamistatin B analogs can be synthesized as shown in Scheme 3.20. The ε-amino group of the *N*-(-Cbz-D-lysine derivative (**39**) can be oxidized using benzoyl peroxide (BPO), followed by formylation, removal of the protecting group, and reprotection as the *O*-benzylhydroxamate (**40**) and subsequent deprotection by hydrogenolysis in the final step of the syntheses [136].

Miller *et al*. highlighted that there is significant interest in developing methods for synthesizing hydroxylamines and hydroxamic acids from amines due to the importance of ε-*N*-hydroxylysine and *N*-hydroxyornithine derivatives as precursors for the synthesis of various siderophores [137]. In addition to the oxidation of amines with BPO [136, 138–141], there are four alternative routes for synthesizing hydroxylamines from amines as outlined in Scheme 3.21: (*a*) oxidation of primary and secondary amines using oxone over silica gel or, in certain instances, alumina [142], (*b*) oxidation of amines using dimethyldioxirane in acetone to give the acetone-derived nitrone, followed by acid hydrolysis or hydroxylaminolysis [143], (*c*) selective mono-cyanomethylation of primary amines, followed by *meta*-chloroperoxybenzoic acid (*m*-CPBA) oxidation to give the nitrone and subsequent hydroxylaminolysis with hydroxylamine hydrochloride [144], and (*d*) reaction of an amine and an aromatic aldehyde to give an imine, followed by oxidation with *m*-CPBA to give the corresponding oxaziridine and hydrolysis to the hydroxylamine under acid-catalyzed conditions [145–147].



Scheme 3.21 Synthesis of hydroxylamines from amines.

3.3.9
Synthesis of Acylhydroxamic Acids

Reaction of hydroxamic acids with acid halides or anhydrides can give *O*-acyl- and even *N*,*O*-diacylhydroxamic acids (Figure 3.15) [8]. Reaction of benzohydroxamic acid with acetic anhydride, for example, in the presence of sulfuric acid gives *N*,*O*-diacylbenzohydroxamic acid **42** [74]. An exemplary reaction for the synthesis of an acylhydroxamic acid is given in Section 3.6.15.

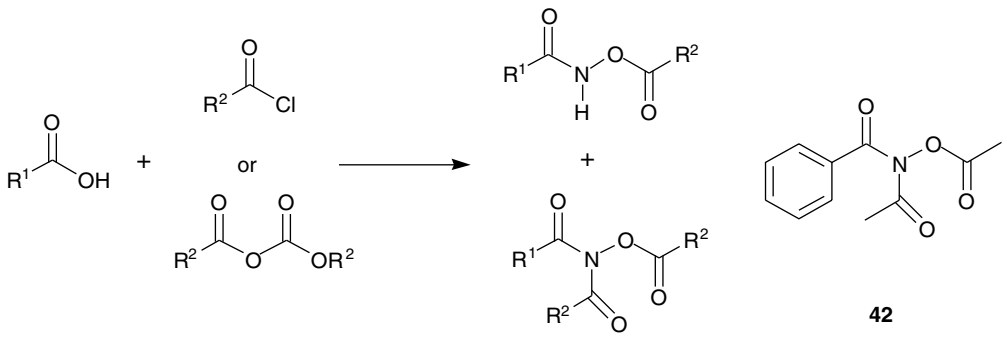


Figure 3.15 Synthesis of *O*-acyl- and *N*,*O*-diacylhydroxamic acids, and structure of *N*,*O*-diacylbenzohydroxamic acid (**42**).

3.4
Solid-Phase Synthesis of Hydroxamic Acids

Since the introduction of solid-phase peptide synthesis by Merrifield in 1963 [148], the application of polymeric supports in organic synthesis has been extended to solid-phase synthesis of other macromolecules, such as oligonucleotides and oligosaccharides, molecules of low molecular weight, and solution-phase synthesis assisted by solid-supported reagents (polymer-assisted solution-phase synthesis) where the use of functional polymers is in a "reverse Merrifield sense" [149]. The use of insoluble resins in synthesis has a number of advantages, such as simplification of product purification, ability to use large excess of reagents to force the reaction to completion, liberation of high-purity products, and possible automation of some synthetic steps. Consequently, solid-phase synthesis is particularly suited for the synthesis of polymeric macromolecules, assembled by iterative reaction cycles, or for parallel or split-mix syntheses. In regard to hydroxamic acids, it is not surprising that most solid-phase methods have been implemented for the synthesis of peptide-based hydroxamic acids and the generation of combinatorial arrays of hydroxamic acids. In contrast, some limitations of solid-phase synthesis are that reaction monitoring and scale are more limited than their solution counterparts. Typical loading of a 1–2% cross-linked polystyrene resin is 0.5–4.0 mmol/g of resin, which translates in a usual synthesis scale to between 0.2 and 0.5 mmol.

Two different strategies can be implemented for the solid-phase synthesis of hydroxamic acids. These involve (i) acidic cleavage and release from insoluble polymers of acylated hydroxylamine, *O*-tethered, or *N*-tethered to the resin support, or (ii) nucleophilic cleavage of resin bound esters, including thioesters, with hydroxylamine derivatives. The choice of method within one of these two strategies should be guided by considerations such as the susceptibility of the molecule to acidic or nucleophilic reagents and compatibility with the protection strategy (e.g., Fmoc versus Boc).

Most of the polymeric supports used for the solid-phase synthesis of hydroxamic acids are polystyrene resins, cross-linked with 1–2% of divinylbenzene (in Schemes 3.22 and 3.23 and Figures 3.16 and 3.17, solid supports represented by shaded circles are such copoly (styrene-1-2% divinylbenzene) resins, unless specifically stated). This has practical implications, in particular limitations on the nature of the solvent which can be used during synthesis. An important requirement of solid-phase synthesis is the swelling of the cross-linked polymer to allow diffusion of the reagents to the polymer-bound reactive sites. Compatible solvents for polystyrene-based resins include DMF, dichloromethane (DCM), and THF. In contrast, polar and protic solvents such as water and methanol shrink polystyrene resins and diethyl ether does not swell gel-type polystyrenes. Mixtures of water or methanol with a compatible solvent are suitable, as long as (i) they are miscible and (ii) the percentage of organic solvent is kept sufficiently high to allow swelling of the resin (e.g., 15% methanol in chloroform). Also, for strategies based on a nucleophilic cleavage of an activated ester, the presence of a polar and protic

solvent can decrease the yield of isolated hydroxamic acids by competing with the hydroxylamine reagent during the release of the polymer-bound ester.

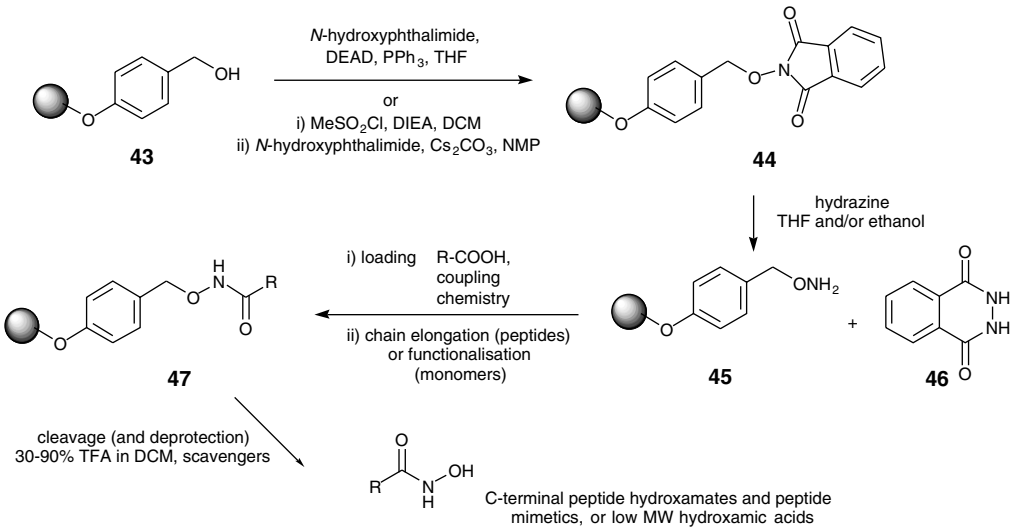
Another practical consideration in solid-phase synthesis is the agitation method, where only gentle magnetic stirring should be used and, when available, alternative equipment such as vortex mixers or orbital shakers should be preferred.

Methods for the solid-phase synthesis of hydroxamic acids developed to date are described in the following sections. The application of these methods to the synthesis of biologically active hydroxamic acids has recently been reviewed [82].

3.4.1 Acidic Cleavage

3.4.1.1 *O*-Tethered Hydroxylamine

3.4.1.1.1 Cleavage with 30–90% TFA One of the first resins to be reported for the solid-phase synthesis of hydroxamic acids was the Wang resin (**43**, Scheme 3.22) [150, 151]. This *p*-alkoxybenzyl alcohol resin is modified to a resin bound *O*-hydroxylamine by formation of a resin-linked hydroximide (**44**), followed by hydrazinolysis of the phthalimido group. Formation of the hydroximide intermediate can be performed with *N*-hydroxyphthalimide under Mitsunobu conditions [150] or displacement of a mesylate intermediate in the presence of cesium carbonate [152] as confirmed by Fourier transform infrared (FTIR) spectroscopy [150]. The hydrazinolysis step is more commonly performed with hydrazine in THF/ethanol (13% v/v) or hydrazine hydrate in THF [153]. For synthesis on a kilogram scale, methylaminolysis in THF is recommended in place of hydrazinolysis on the basis of safety considerations [154]. The resulting hydroxylamine resin (**45**) gives a negative Kaiser test [150] and a positive ninhydrin test [153]. The resin substitution can be quantified by measuring



Scheme 3.22 Solid-phase synthesis of hydroxamic acids from a Wang resin.

phthaldrazide (**46**) absorbance at 346 nm [155] or evaluating the elemental analysis of the resin [150].

Loading of the carboxylic acid substrate for monomeric molecules and peptide mimetics, or of the first amino acid and elongation of the sequence in the case of peptide hydroxamic acids, is carried out by carbodiimide [154, 156] or uranium [150, 157] coupling chemistries, and by standard Fmoc chemistry [150], respectively. In the latter case, deprotection of the Fmoc group can be used to quantify the loading (see Section 3.4.1.1.2). Cleavage from the resin (**47**), with concomitant deprotection of the amino acid side-chains for peptides, yields low-molecular-weight hydroxamic acids or C-terminally modified peptide hydroxamic acids. This final step is performed with cleavage cocktails containing 30–90% TFA in DCM and various scavengers [150–152]. Hydroxamic acids are generally obtained in high purity and good yields. However, the use of solutions containing a high percentage of TFA has been associated with the formation of impurities, attributed to the breakdown of the resin with 70% TFA and water (2%) as scavenger [150], or to the formation of carboxylic acids resulting from hydrolysis of the hydroxamic acids with 90% TFA and anisole as scavenger [152]. Formation of impurities and hydrolysis products can be prevented by replacing the Wang linker by a super acid-sensitive linker and performing the cleavage step under anhydrous conditions.

Conversion of a hydroxy-resin into a hydroxylamine-resin for solid-phase synthesis of hydroxamic acids has also been performed with polyethylene glycol-grafted polystyrene resin, such as the ArgoGel-OH polymeric support [158].

The resin-bound hydroxylamine synthesized from a Wang resin (**43**) and Boc-protected has also been used as a resin-bound capture reagent in a solid-phase synthesis of hydroxamic acids via a palladium-catalyzed cascade reaction (see Section 3.6) [133].

Finally, a polymer-bound *N*-linked hydroxylamine can be prepared from the Merrifield resin (chloromethyl-polystyrene) by introduction of an acid-labile aldehyde linker (4-hydroxy-2-methoxy-benzaldehyde) via nucleophilic displacement of chloride, followed by reductive amination with *O*-2,4-dimethoxylbenzyl-protected hydroxylamine in the presence of sodium cyanoborohydride. This resin can be coupled with carboxylic acids under standard carbodiimide coupling conditions, followed by acidolytic cleavage (95% TFA in DCM), to yield the corresponding hydroxamic acids [159].

The Wang hydroxylamine resin can also be used for the synthesis of *N*-alkyl hydroxamic acids through the formation of an oxime by reaction with an aldehyde or ketone, followed by oxime reduction with borane–pyridine complex in the presence of dichloroacetic acid and acylation of the resulting *N*-substituted hydroxylamine derivatives, the *N*-substituent originating from the carbonyl reagent [160].

Alternatively, activation of the resin-bound hydroxylamine by *N*-sulfonation with 2-nitrobenzenesulfonyl chloride [161], followed by *N*-alkylation under Mitsunobu conditions, deprotection of the 2-nitrophenylsulfonyl group (nosyl) with thiolate anion [2-mercaptoethanol in DMF in the presence of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU)], and acylation of this *N*-alkyl hydroxylamine-resin, can be used to initiate

the solid-phase synthesis of the siderophores DFOB (**1**, see Section 3.2.2.1, Figure 3.4) and methyl carboxymycobactin T7 and their analogs [162, 163]. Assembly of the following hydroxamate units (two to three) utilizes *N*-nosyl-*O*-protected hydroxylamines, reacted with the polymer-bound growing molecule through a hydroxyl group by a Fukuyama reaction, or by alkylation with polymer-bound alkylbromides [163]. This strategy can also be implemented for the solid-phase synthesis of sulfonamide-based hydroxamic acids [164, 165], while alkylation of *N*-nosyl-protected *O*-tethered hydroxylamine has also been performed by Michael addition reaction with α,β-unsaturated carbonyl compounds [166].

3.4.1.1.2 **Super Acid-Sensitive Linkers** The strategy developed with the Wang resin for its conversion into the corresponding *O*-tethered hydroxylamine insoluble support has also been applied to other hydroxy-functionalized resins. The linkers in these resins can release the desired hydroxamic acids under milder acidic conditions than the Wang linker. Hydroxamic acids synthesized from a Sasrin resin (**48**, Figure 3.16) can be released with 5% TFA in DCM [167], while hydroxamic acids can be cleaved from the hydroxylamine derivative of the 4-(4-hydroxymethyl-3-methoxyphenoxy)-butyryl–*p*-methyl-benzhydrylamine (HMPB-MBHA) resin (**49**) with 1% TFA in DCM and triisopropylsilane as a scavenger [150]. This is also the case for the Rink acid resin (**50**), in which cleavage can be effected with as little as 10% acetic acid in DCM [152].

Other resins containing super acid-sensitive linkers have been prepared from chloro-functionalized resins by immobilization of hydroxylamine derivatives, or its precursor *N*-hydroxyphthalimide, by nucleophilic displacement. Polymeric resins based on trityl- and alkoxy-substituted benzhydryl linkers have been reported in this group of synthetic supports [168].

Trityl-based resins can be prepared from 2-chlorotrityl chloride (**51**) or chlorotrityl (**52**, Scheme 3.23) polystyrene resins, by attachment of *N*-protected hydroxylamine reagents or *N*-hydroxyphthalimide, respectively, in the presence of a tertiary amine [169, 170]. In the former case, *N*-Fmoc-hydroxylamine can be prepared and directly immobilized onto the solid support, followed by capping of unreacted chloride sites with methanol. Loading of the resin can be determined by spectrophotometric quantification of the dibenzofulvene adduct formed upon Fmoc deprotection with 20% piperidine in DMF [169]. In the latter case, hydrazinolysis of the resin-bound hydroximide intermediate can subsequently be performed to generate

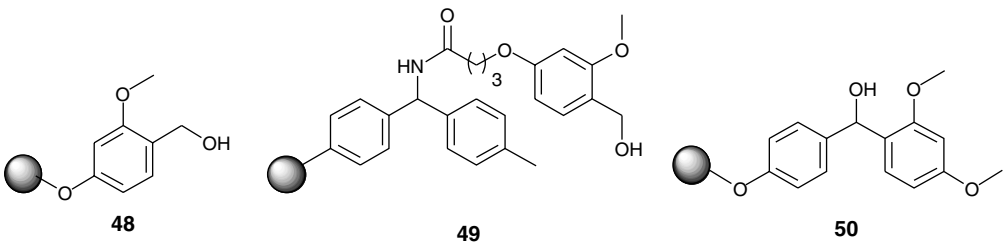
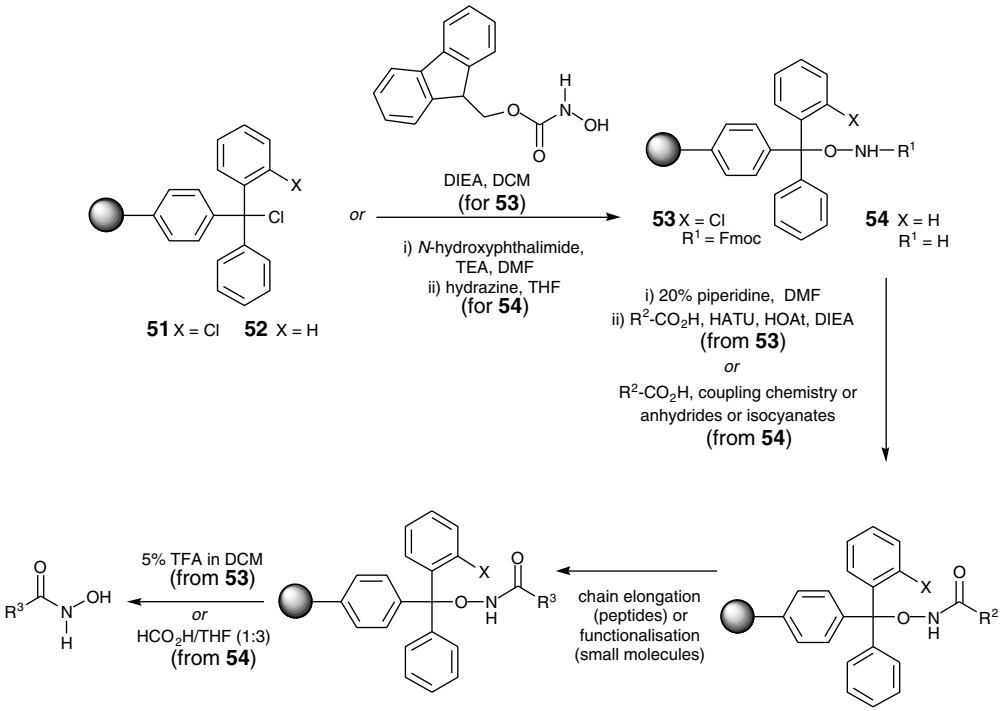


Figure 3.16 Solid supports based on a super acid-sensitive linker.



Scheme 3.23 Solid supports based on a trityl linker.

the hydroxylamine resin. Characterization of the resin-bound intermediates can be performed by FTIR spectroscopy.

C-Terminally modified peptide hydroxamic acids can be assembled from the *N*-Fmoc-aminooxy-2-chlorotrityl resin, using HATU/HOAt/DIEA coupling chemistry for the attachment of the C-terminal residue, followed by elongation of the peptide chain using *O*-(benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium tetrafluoroborate (TBTU)/HOBt activation. Release of the hydroxamic acid can be achieved with 5% TFA in DCM [169]. Consequently, peptides assembled from this resin can be released as fully protected sequences. Subsequent deprotection of the amino acid side-chains requires additional treatment with 30–90% TFA. Direct cleavage and deprotection with high acid concentrations can be associated with decreased purity of the product released [169].

In contrast, low-molecular-weight or peptide hydroxamic acids synthesized from the resin based on a trityl linker, can be released with formic acid in THF (1 : 3) [170].

The polystyrene resin functionalized with a 2-chlorotrityl linker can also be used for the parallel synthesis of hydroxamic acids of low molecular weight [171, 172] and the solid-phase synthesis according to the Miller hydroxamate approach of peptides modified with a C-terminal β-lactam ring [173]. Alternatively, a polystyrene resin carrying a trityl linker can be used for the solid-phase synthesis of nucleoside hydroxamic acids [153].

The polymer-supported *N*-benzyloxy-2-nitrobenzenesulfonamide linker described above with the Wang resin, has also been prepared with the Sasrin and HMPB linkers, for the synthesis of *N*-alkyl hydroxamic acids [166].

Although the attachment of *N*-alkyl-*N*-Fmoc-hydroxylamine reagents can be performed directly on the 2-chlorotrityl chloride polystyrene resin, subsequent loading of the carboxylic acid substrate is hindered by steric effects and a benzhydryl-based resin, 4-[2,4-dimethoxyphenyl(*N*-fluoren-9-ylmethoxycarbonyl-*N*-alkylaminooxy)methyl]-phenoxymethyl polystyrene, has been developed for the synthesis of *N*-alkylhydroxamic acids [168]. A variety of *N*-Fmoc-*N*-alkyl-hydroxylamine reagents have been prepared and immobilized. Acylation of the polymer-bound *N*-alkyl hydroxylamines by HATU coupling chemistry and cleavage from the resin with 5% TFA can be performed as with the 2-chlorotrityl resin.

Some of the resins presented in Section 3.4 are now commercially available (e.g., Wang hydroxylamine-resin, *O*-alkylhydroxylamino-Sasrin, and *N*-Fmoc-aminooxy-2-chlorotrityl [160, 174, 175]).

3.4.1.2 *N*-Tethered Hydroxylamine

As *O*-attachment of hydroxylamine onto a resin involves the initial formation of the hydroxamic acid function, subsequently leaving an acidic NH group which can participate in side-reactions during further transformations carried out on the polymeric support, the anchoring of hydroxylamine reagents by their nitrogen has been evaluated as a means to not only immobilize, but also protect this group during the complete solid-phase synthesis of hydroxamic acids. A polyethylene glycol grafted polystyrene resin (Tentagel S NH_2) and a hypersensitive acid-labile linker [tris(alkoxy)benzyl ester] can be selected for this approach. The linker can be prepared by reductive amination of 4-(4-formyl-3,5-dimethoxy-phenoxy)-butyric acid with *O*-protected (THP or allyl) hydroxylamine, followed by Fmoc protection of the secondary hydroxylamine intermediate, conjugation of this linker to the resin by formation of a stable amide bond (using the butyric acid group), and final Fmoc deprotection. The resulting *O*-THP alkoxyamine and *O*-allyloxyamine resins can be used for solid-phase synthesis of low-molecular-weight hydroxamic acids. Release from the resin followed by deprotection of the hydroxamate can be carried out with 50% TFA and tetrakis(triphenylphosphine)palladium, respectively, for *O*-allyl hydroxamic acids, while for their THP counterparts the order of these reactions can be reversed and THP deprotection (2.5% aqueous TFA in DCM) precedes the cleavage step (50% aqueous TFA in DCM) [176].

3.4.1.3 Other Methods of Solid-Phase Synthesis of Hydroxamic Acids based on an Acidic Cleavage

In solid-phase peptide synthesis, sequences are elongated from the C-terminus to the N-terminus, to prevent racemization of the growing chain. Consequently, polymeric supports carrying hydroxylamine linkers are generally required to yield C-terminally modified peptide hydroxamic acids.

Another strategy for the solid-phase synthesis of peptide hydroxamic acids exploits the β- or γ-carboxylic acid groups of N-terminal aspartate or glutamate residues, by

carrying out their selective deprotection in a resin-bound peptide, followed by coupling with an *O*-protected hydroxylamine. The latter protecting group can be removed during cleavage from the solid support. Conventional resins for solid-phase peptide synthesis, combined with appropriate orthogonal protection strategy for the side-chains of aspartic or glutamic acid, can be used in this context. The original method was developed with a MBHA resin and a Fmoc/*t*Bu protocol for the assembly of the peptide sequence. Phosphonium and carbodiimide coupling reagents were evaluated for the formation of the hydroxamate, the best results being obtained with Castro's reagent. This resin necessitates treatment with HF for the final release of the peptide hydroxamate [177].

A variation of this approach can be performed by reversal of the peptide chain direction, through the use of N-terminal dicarboxylic acid residues, monofunctionalized with a hydroxamate preformed and protected in solution. For example, a succinylhydroxamate can be coupled to a peptide chain elongated from a Rink amide resin and the resulting conjugate subsequently cleaved from this support [178]. Modification of a resin-bound molecule with a diacid has also been exploited for the split-mix synthesis of small molecule inhibitors of HDACs (see Section 3.2.2.2.2). In this case, 600 different resin-bound amino-1,3-dioxanes were modified with diacids by conjugation of monophthalimidomethylester diacids or reaction with pyridine-activated acid anhydrides. Hydrazinolysis of the phthalimidomethylester, followed by conjugation with *O*-2-methoxypropanehydroxylamine yielded protected and supported hydroxamic acids, which were deprotected and cleaved from a silane linker, attached on an amino-functionalized polystyrene resin (polystyrene AM NH_2), by treatment with pyridinium *p*-toluenesulfonate in THF/methanol and HF/pyridine, respectively [179].

Fragment condensation of three protected hydroxamate-containing oligopeptides on a super acid-sensitive linker-based support (2-chlorotrityl chloride resin) can be implemented for the synthesis of trihydroxamate peptides. A *N*-benzyloxy-phenylalanine building block prepared and incorporated in various tripeptides in multiple synthetic steps performed by solution-phase chemistry can be used to assemble the different fragments [180].

Finally, inverse assembly of a peptide chain by solid-phase synthesis (in the N $\rightarrow$ C direction) can also be carried out to functionalize peptides at their C-termini with a hydroxamic acid group. In this case, the first amino acid is anchored on a hydroxymethyl polystyrene resin by a urethane attachment (using phosgene) as a "traceless linker". The chain assembly is consequently performed by inverse solid-phase peptide synthesis, using amino acid *tert*-butyl esters, HATU as coupling reagent, and 2,4,6-trimethylpyridine (TMP, collidine) as a base, to prevent epimerization. Following C-terminal deprotection, *O*-(*tert*-butyl)hydroxylamine hydrochloride can be conjugated by a HATU/TMP coupling protocol, followed by cleavage with 10% trifluoromethanesulfonic acid/TFA to give the corresponding peptide hydroxamic acids. Racemization of individual amino acids in the peptide chain is lower than 2% [181].

Other solid supports for solid-phase synthesis of hydroxamic acids prepared from chloro-functionalized polymeric precursors by nucleophilic displacement with *N*-hydroxyphthalimide, followed by hydrazinolysis, include SynPhase crowns.

Chlorotrityl- and α-chloroethyl-polystyrene precursors have been used to generate *O*-tethered hydroxylamine linkers in these Multipin supports. They release hydroxamic acids upon treatment with 1% or 95% TFA in DCM, respectively [155, 182]. The use of 95% TFA and 5% water as a cleavage cocktail can be associated with substantial hydrolysis to the corresponding carboxylic acid [155].

3.4.2 Nucleophilic Cleavage

Nucleophilic cleavage of resin bound esters with hydroxylamine was first reported with the polyethylene glycol grafted polystyrene resin ArgoGel-OH (**55**, Figure 3.17), esterified with Cbz-protected amino acids by carbodiimide coupling chemistry with 4-(dimethylamino)pyridine (DMAP) catalysis. These polymer-bound esters can be displaced from the resin by treatment with 25 equiv. of 50% aqueous hydroxylamine in THF for 2 days to generate the corresponding hydroxamic acids [183] (the numbers of equivalents described in this section are relative to the loading of the resin). This method can also be applied to the solid-phase synthesis of a library of di- and tripeptidic hydroxamic acids, owing to its compatibility with both Fmoc and Boc chemistries [184–186]. The preparation of hydroxamic acids by hydroxyamination of resin-bound esters has often been described as being associated with poor reproducibility. It has been shown recently that this reaction, performed on a 4-hydroxymethylbenzoic acid AM resin, can be improved by the addition of KCN to the hydroxylamine solution [94].

The oxime (Kaiser) polystyrene resin (**56**) was concurrently implemented for the synthesis of hydroxamic acids by nucleophilic displacement. This strategy allows the use of acid labile protecting groups for the modification of the polymer-bound esters prior to their release as hydroxamic acids from the resin. Various *O*-protected and unprotected hydroxylamines were evaluated for the cleavage step, resulting in the

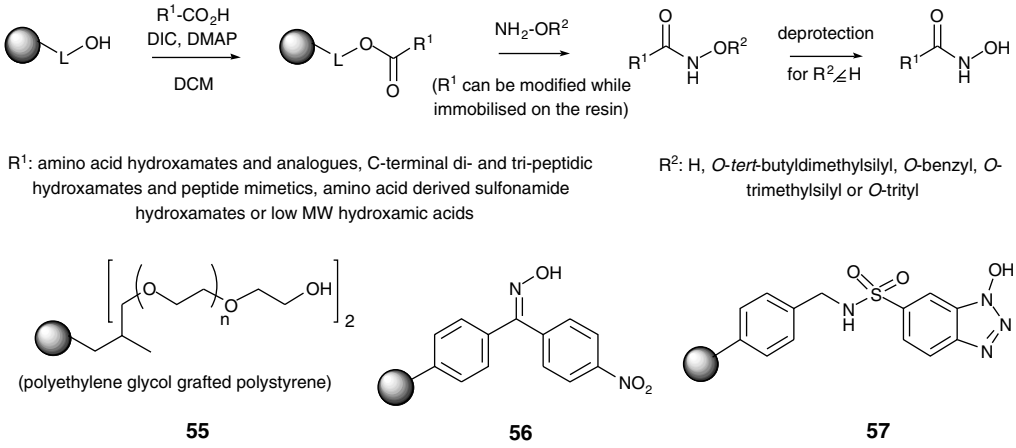


Figure 3.17 Solid-phase synthesis of hydroxamic acids by nucleophilic displacement of resin-bound esters.

formation of *O*-protected hydroxamic acids in the former case. The best results were obtained with an excess (5 equiv.) of *O-tert*-butyldimethylsilyl-hydroxylamine, in 1,2-dichloroethane at reflux for 20 h [187]. The use of anhydrous, free, hydroxylamine (2.7 equiv.), prepared from hydroxylamine hydrochloride and sodium methoxide in methanol, followed by filtration of NaCl and dilution with chloroform, was subsequently shown to circumvent the need for a final deprotection step, resulting in increased isolated yields of hydroxamic acids [188]. Recovery of the hydroxamic acids was also facilitated by the low boiling point of the solvents. Moreover this method was shown to preserve the stereochemical integrity of the carboxylic acid substrate.

While a Boc protecting group can be removed from a polymeric oxime ester, the latter linkage is labile to repeated TFA treatments and therefore unsuitable for the preparation of C-terminal peptide hydroxamic acids assembled by solid-phase synthesis according to the Boc strategy. A thioester linker, introduced as 3-thiopropionic acid on a polyethylene glycol-poly-(*N,N*-dimethylacrylamide) copolymer, through a stable amide bond, has consequently been developed to address this limitation. The resulting 3-mercapto-propionamide-PEGA support used for the elongation of a peptide sequence by Boc solid-phase peptide synthesis, allows deprotection of the amino acid side-chain by treatment with HF and releases various C-terminally modified peptides, depending on the nucleophile selected for the cleavage step. The use of hydroxylamine solutions in DMF, prepared from equimolar quantities of hydroxylamine hydrochloride and DBU, in the cleavage step yields peptide hydroxamic acids [189]. The use of this linker was concurrently developed with PEGA and polystyrene resins [190]. The use of *O*-protected hydroxylamine reagents, soluble in low-boiling-point solvents such as THF, are preferred as they facilitate the recovery of the hydroxamic acid, although subsequent deprotection is required.

A "catch-and-release" strategy has also been developed for the synthesis of hydroxamic acids. It exploits the high reactivity of polymer-bound OBt esters with nucleophiles, which afford short reaction times (the release step can be completed in 1 h using 0.9 equiv. of hydroxylamine; E. Burke and M. Devocelle, unpublished results) and the use of substoichiometric quantities of hydroxylamine during the release step. This advantage, combined with the preparation protocol of free anhydrous hydroxylamine developed by Thouin and Lubell [188], prevents the presence of unreacted hydroxylamine. Consequently, hydroxamic acids are conveniently isolated as the only product in solution and are recovered by simple evaporation of low boiling point solvents. Recycling of the HOBt polystyrene resin (**57**) can also be undertaken [79, 191]. This "catch-and-release" strategy based on the formation of resin-bound activated HOBt esters has also been performed with a polystyrene-HBTU resin for concomitant purification and functionalization of the carboxylic acid substrate, as part of a fully automated multistep (four to five steps) polymer-assisted solution-phase synthesis of HDAC inhibitors (see Section 3.2.2.2.2) [192].

3.4.2.1 **Other Methods**

As illustrated in the latter method (see Section 3.4.2), solution-phase synthesis using polymer-supported reagents and scavengers can been employed to generate arrays of hydroxamic acids [193]. Solid-phase synthesis of C-terminal peptide *O*-methyl

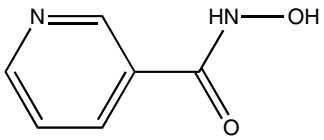
hydroxamic acids can be carried out by the backbone amide linker approach [194], through reductive amination of methoxylamine onto a 5-[(2 or 4)-formyl-3,5-dimethoxyphenoxy]butyramido-polyethyleneglycol-polystyrene resin in the presence of sodium cyanoborohydride. The resulting resin-bound methoxylamine can be acylated by a Fmoc-protected amino acid with bromo-tris(pyrrolidino)-phosphonium hexafluorophosphate coupling chemistry, before assembly of the peptide chain by automated solid-phase peptide synthesis. The peptide hydroxamic acids can be deprotected and released from the resin by treatment with 95% TFA, 2.5% triisopropylsilane, and 2.5% water as the cleavage cocktail [195]. Finally, solid-phase synthesis of hydroxamic acids by the Angeli–Rimini reaction has been developed by adapting its solution-phase counterpart to immobilized *N*-hydroxybenzenesulfonamide (see Section 3.3.4) [129].

3.5 Conclusions

The rich bioactivity and clinical utility of hydroxamic acids has undoubtedly been the impetus behind new developments into efficient methods for their synthesis and that of their derivatives. Both solution- and solid-phase synthetic strategies can be employed to generate a diverse range of hydroxamic acids. The choice of strategy should be guided by the type of hydroxamic acid to be synthesized, its structural complexity, and the synthetic scale-up required. A hybrid synthetic route based on solution- and solid-phase techniques can also be considered in which the advantages of both approaches may be combined. The hydroxamic acid moiety has already proven itself as a key structural element in a diverse range of compounds with therapeutic applications. There is no doubt that these organic acids will continue to fascinate, excite, and challenge us for some time yet.

3.6 Experimental Procedures

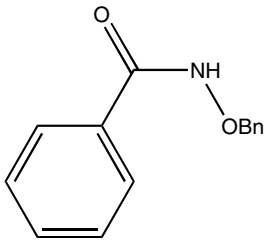
3.6.1 Synthesis of 3-Pyridinehydroxamic Acid [90]



Hydroxylamine hydrochloride (5.10 g, 72 mmol) was added to sodium hydroxide (5.80 g, 146 mmol) in deionized water (37 ml). The resulting solution was then added dropwise to methyl nicotinate (5.00 g, 36 mmol) in methanol (55 ml). The solution was stirred at room temperature for 72 h, after which the solution was acidified to pH 5.5

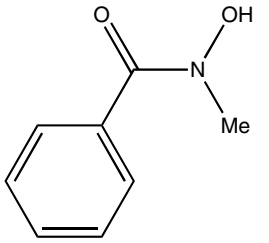
using 5% HCl. The solvent was removed under vacuum yielding a yellow solid. Methanol (60 ml) was added and sodium chloride filtered. The solvent from the filtrate was removed under vacuum yielding a light pink solid, which was recrystallized from water to give, a white solid, 3-pyridinehydroxamic acid in 66% yield.

3.6.2
Synthesis of *O*-benzylbenzohydroxamic Acid [89]



To a stirred suspension of *O*-benzylhydroxylamine hydrochloride (0.16 g, 1 mmol) and ethyl benzoate (0.15 g, 1 mmol) in dry THF (10 ml) at −78 °C and under nitrogen atmosphere was added a 1 M solution of lithium hexamethyldisilazane (3.1 ml, 3.1 mmol). The reaction was stirred for less than 10 min and subsequently quenched with a saturated aqueous solution of NH_4Cl, warmed to room temperature, and extracted with ethyl acetate. The collected organic layers were dried ($NaSO_4$), filtered, and concentrated under reduced pressure. The crude product was purified by flash chromatography ($CHCl_3$/MeOH as eluent) to give *O*-benzylbenzohydroxamic acid in 95% yield.

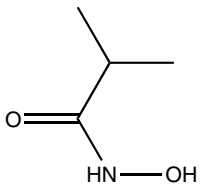
3.6.3
Synthesis of *N*-methylbenzohydroxamic Acid [101]



An aqueous solution of *N*-methylhydroxylamine was prepared under nitrogen by adding *N*-methylhydroxylamine hydrochloride (9.44 g, 0.113 mol) to a solution of sodium carbonate (14.50 g, 0.137 mol) in H_2O (164 ml). The solution was covered with diethyl ether (50 ml) and cooled in an ice-salt bath. Benzoyl chloride (19.70 g, 0.14 mol) in anhydrous ether (50 ml) was added with stirring and dropwise over 14 min. The mixture was stirred and cooled for an additional 30 min, after which 20%

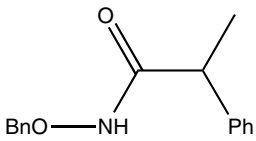
sodium hydroxide (45 ml) was added. The aqueous layer (pH 12) was neutralized to pH 7 with 6 M HCl, saturated with sodium chloride, and extracted 5 times with chloroform. The chloroform was dried ($MgSO_4$), filtered, and concentrated under reduced pressure to give *N*-methylbenzohydroxamic acid in 74% yield.

3.6.4
Synthesis of Isobutyrohydroxamic Acid [83]



Ethyl chloroformate (1.30 g, 12 mmol) and *N*-methylmorpholine (1.30 g, 13 mmol) were added to a solution of isobutyric acid (0.88 g, 10 mmol) in diethyl ether (30 ml) at 0 °C, and the mixture was stirred for 10 min. The solid was filtered off and the filtrate was added to freshly prepared hydroxylamine (0.50 g, 15 mmol) in methanol and stirred at room temperature for 15 min. [Hydroxylamine was prepared by adding hydroxylamine hydrochloride (1.00 g, 15 mmol) dissolved in methanol (10 ml) to a stirred solution of KOH (0.84 g, 15 mmol) in methanol (4 ml) at 0 °C. The mixture was stirred for 15 min, the precipitated potassium chloride removed and the filtrate used as described.] The solvent was evaporated and the crude product was purified by silica gel column chromatography to give isobutyrohydroxamic acid in 81% yield.

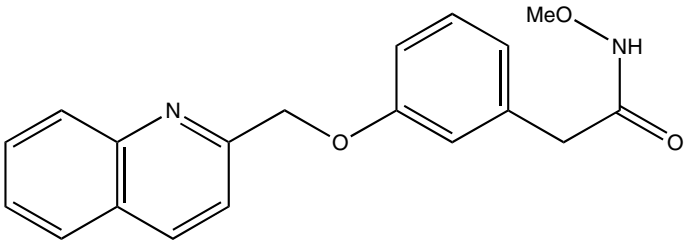
3.6.5
Synthesis of *O*-benzyl-2-phenylpropionohydroxamic Acid [107]



CDMT (0.74 g, 4.4 mmol) and NMM (1.2 ml, 11.1 mmol) were added to a solution of 2-phenylpropionic (0.55 g, 3.7 mmol) in THF (11 ml) at room temperature. The reaction was stirred for 1 hour and a white precipitate formed. *O*-benzylhydroxylamine hydrochloride (0.64 g, 3.7 mmol) was added and the mixture was stirred for 8 h and then quenched with H_2O (15 ml) and extracted with diethyl ether (2 × 7 ml). The combined organic layers were washed twice with a saturated solution of Na_2CO_3 (15 ml), followed by a 1 M solution of HCl (15 ml) and brine (15 ml). The organic layer was dried (Na_2SO_4) and evaporated to give *O*-benzyl-2-phenylpropionohydroxamic acid in 85% yield.

3.6.6

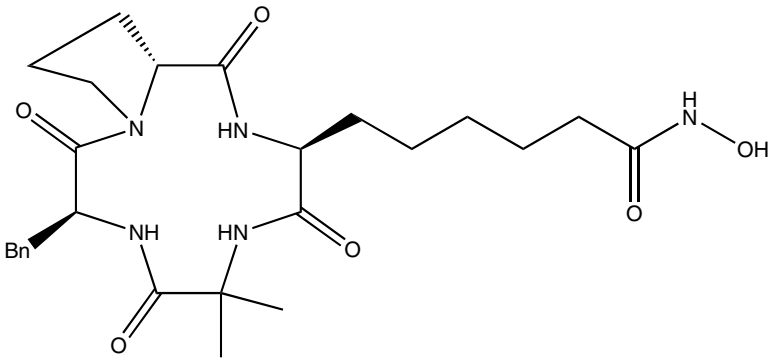
Synthesis of Methyl 3-(2-quinolinylmethoxy)benzeneacetohydroxamic Acid [110]



To a mixture of 3-(2-quinolinylmethoxy)-benzeneacetic acid (5.00 g, 17.4 mmol), *O*-methylhydroxylamine hydrochloride (1.42 g, 17.4 mmol), and *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide (3.26 g, 17.4 mmol) in THF (75 ml) was added TEA (4.7 ml, 33.8 mmol). The reaction was stirred at room temperature overnight. The THF was removed under vacuum and CH_2C1_2 was added to the residue. The mixture was twice washed with water, dried ($MgSO_4$), and concentrated under vacuum to give an oil. The oil was dissolved in acetone, filtered through Celite and silica gel, concentrated at reduced pressure and subsequently purified by preparative high-performance liquid chromatography using ethyl acetate/ethanol 95 : 5 as an eluent to give methyl 3-(2-quinolinylmethoxy)-benzeneacetohydroxamic acid in 29% yield.

3.6.7

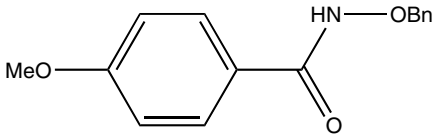
Synthesis of the Chlamydocin Hydroxamic Acid Analog, *cyclo*(L-Asu(NHOH)–Aib-L-Phe–D-Pro) [117]



cyclo(L-Asu–Aib–L-Phe–D-Pro) (0.12 g, 0.25 mmol) was dissolved in DMF (3 ml) at 0 °C and hydroxylamine hydrochloride (0.09 g, 1.24 mmol), $HOBt.H_2O$ (0.06 g, 0.37 mmol), BOP (0.17 g, 0.37 mmol), and TEA (0.24 ml, 1.74 mmol) were added. The mixture was stirred for 2 h. Gel filtration of the reaction mixture using Sephadex LH-20 was followed by lyophilization to give *cyclo*(L-Asu(NHOH)–Aib–L-Phe–D-Pro) in 74% yield.

3.6.8

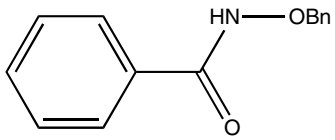
Synthesis of *O*-benzyl-4-methoxybenzohydroxamic Acid [121]



A solution of *N*-hydroxy-2-phenylbenzimidazole (0.02 mol) and *N,N′*-diisopropylethylamine (DIEA) (0.07 mol) was stirred in DMF and diethyl chlorophosphate (0.024 mol) was added with cooling. The reaction mixture was stirred for 10 min, 4-methoxybenzoic acid (0.016 mol) was added and the mixture was stirred for a further 10 min at 0 °C. *O*-benzylhydroxylamine (0.03 mol) was added and the reaction stirred at room temperature for 45 min. Saturated aqueous NaCl solution (20 ml) was then added, the mixture was extracted with ethyl acetate (2 × 15 ml), and the organic layer was washed with 2 M HCl (15 ml), saturated $NaHCO_3$ (15 ml), and water (2 × 20 ml). The organic layer was subsequently dried over sodium sulfate, filtered and evaporated to give *O*-benzyl-4-methoxybenzohydroxamic acid in 95% yield.

3.6.9

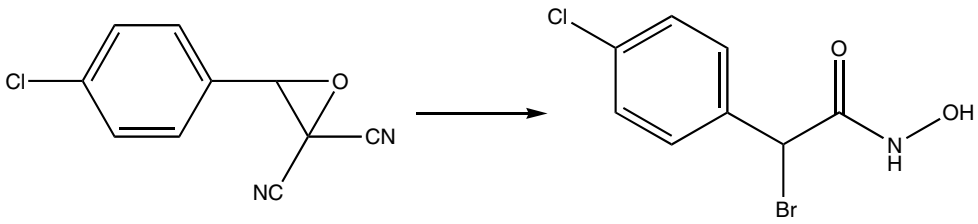
Synthesis of *O*-benzylbenzohydroxamic acid [122]



O-Benzylhydroxylamine (5 equiv.), samarium triflate (1 equiv.) and *N*-benzoyl-4-benzyl-2-oxazolidinone (1 equiv.) were stirred in THF at room temperature for 30 min. The solvent was evaporated and the crude product was purified by silica gel column chromatography to give *O*-benzylbenzohydroxamic in 91% yield.

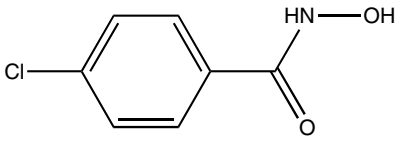
3.6.10

Synthesis of a 4-chlorophenyl Substituted-α-bromohydroxamic acid [125]



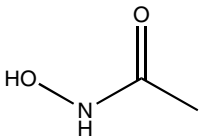
A 4-chlorophenyl substituted *gem*-dicyanoepoxide (1.02 g, 5 mmol) was dissolved in THF (15 ml) and lithium bromide (0.65 g, 7.5 mmol) was added. A solution of hydroxylamine hydrochloride (0.07 g, 10 mmol) and TEA (0.139 ml, 10 mmol) in THF was added dropwise over 1 h to the mixture which was refluxing. The reaction was refluxed for a further hour. The solvent was partially removed under reduced pressure and the residue was extracted with DCM. The combined extracts were dried (Na_2SO_4), filtered, and evaporated to afford the residual α-bromohydroxamic acids, which were purified by silica gel column chromatography (hexane/ethyl acetate 3 : 2 as eluent) to give a 4-chlorophenyl substituted-(-bromohydroxamic acid in 86% yield.

3.6.11
Synthesis of 4-Chlorobenzohydroxamic Acid [130]



A 1.93 M sodium methoxide/methanol solution (2.18 ml, 4.2 mmol) was added dropwise with stirring to an ice-cooled solution of *N*-hydroxybenzenesulfonamide (0.37 g, 2.1 mmol) in methanol. 4-Chlorobenzaldehyde(0.28 g, 2 mmol) dissolved in methanol (2 ml) was added, and the reaction mixture was warmed to room temperature and stirred for an additional 2 h. The solution was concentrated under vacuum, diluted with 100 ml of ether, and extracted twice with 2 M NaOH. The aqueous layer was acidified with concentrated HCl to pH 7–8 and extracted twice with ethyl acetate. The organic layer was dried ($MgSO_4$) and evaporated to give 4-chlorobenzohydroxamic acid in 68% yield.

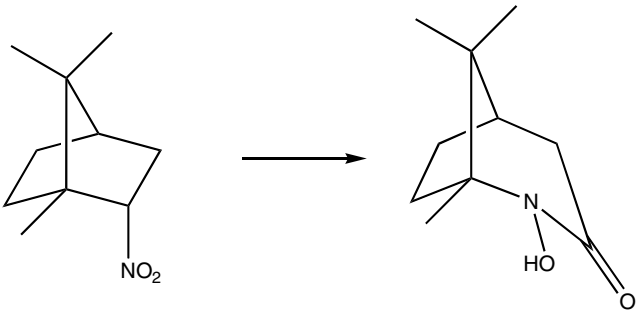
3.6.12
Synthesis of Acetohydroxamic Acid [131]



A solution of nitroethane (1.42 ml, 0.02 mol) in DCM (40 ml) was added dropwise with vigorous stirring to a suspension of selenium dioxide (2.22 g, 0.02 mol) at 0–5 °C under anhydrous conditions in DCM (40 ml). TEA (4.08 g, 0.04 mol) was added dropwise and the temperature maintained at 0–10 °C. The mixture was stirred for 30 min at 20–22 °C, and then refluxed for 1 h, cooled, and filtered

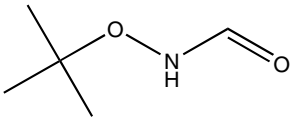
through diatomaceous earth. The filtrate was concentrated using a rotary evaporator. The crude material was dissolved in water (25 ml), to which a solution of Cu(II) acetate (3% w/w in water) was added (~75 ml) until no additional green solid precipitated. The green solid was filtered, washed with water (2 × 10 ml), methanol (2 × 10 ml), and dried in a desiccator. The dried copper complex was suspended in methanol (25 ml) and cooled in an ice bath to which anhydrous hydrogen sulfide was slowly introduced. The green suspension turned black within 5 min and the introduction of the hydrogen sulfide was maintained for a further 5 min. The black precipitate was filtered off and the filtrate concentrated using a rotary evaporator to give acetohydroxamic acid in 66% yield.

3.6.13
Synthesis of *N*-hydroxy Lactam [132]



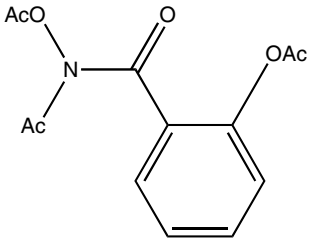
A solution of 2-nitro-1,7,7-trimethylbicyclo[2.2.l]heptane ($[\alpha]_D^{18} + 9.4$ (EtOH)) (1.14 g) in EtOH/EtONa was irradiated for 4.5 h. The solution was subsequently neutralized with 0.1 M acetic acid in EtOH. Solvent removal followed by extraction from the residue with $CHCl_3$ gave a white solid which was purified by silica gel column chromatography to give *N*-hydroxy lactam in 76% yield.

3.6.14
Synthesis of *O-tert*-butyl-*N*-formylhydroxylamine [134]



Formic acid (100 ml, 122 g, 2.65 mol) was added dropwise to *O-tert*-butylhydroxylamine (17.2 g) and the reaction was stirred for 16 h. Methanol was added slowly and stirring was continued for another 12 h. Methyl formate and methanol were distilled at atmospheric pressure and the residue distilled under vacuum to give *O-tert*-butyl-*N*-formylhydroxylamine in 81% yield.

3.6.15
Synthesis of Triacetylsalicylhydroxamic Acid [74]



Acetic anhydride (50 ml, 0.5 mol) and concentrated sulfuric acid (3 drops) were added to salicylhydroxamic acid (10.14 g, 66 mmol). The mixture was stirred at 80 °C under nitrogen for 48 h. Ethyl acetate (100 ml per 10 ml of final reaction volume) was added. The organic phase was washed with 0.5 M NaOH (50 ml per 10 ml of final reaction volume), water (100 ml per 10 ml of final reaction volume), 0.5 M citric acid (50 ml per 10 ml of final reaction volume), and brine (100 ml per 10 ml of final reaction volume). The organic layer was dried ($MgSO_4$) and evaporated. The crude product was purified by silica gel column chromatography (DCM/petroleum ether 3 : 2 as eluent) to give triacetylsalicylhydroxamic acid in 12% yield.

Acknowledgments

The authors are extremely grateful to Science Foundation Ireland and to the Programme for Research in Third Level Institutions, administered by the Higher Education Authority, for financial support. We thank Dr. Anthony Chubb, Molecular Modeling Group, for supplying Scheme 3.3 and Figure 3.10. A special thanks to Professor K.B. Nolan (Royal College of Surgeons in Ireland) who continues to be a source of inspiration and motivation to us in relation to our work on hydroxamic acids.

References

1 Lossen, H. (1869) *Annalen der Chemie und Pharmacie*, **150**, 314–322.
2 Harvey, A.E. and Manning, D.L. (1950) *Journal of the American Chemical Society*, **72**, 4488–4493.
3 Brown, D.A., Glass, W.K., Mageswaran, R. and Girmay, B. (1988) *Magnetic Resonance in Chemistry*, **26**, 970–973.
4 Fitzpatrick, N.J. and Mageswaran, R. (1989) *Polyhedron*, **8**, 2255–2263.
5 Brown, D.A., Coogan, R.A., Fitzpatrick, N.J., Glass, W.K., Abukshima, D.E., Shiels, L., Ahlgran, M., Smolander, K., Pakkanen, T.T., Pakkanen, T.A. and Perakyla, M. (1996) *Journal of the Chemical Society, Perkin Transactions 2*, 2673–2680.
6 Chatterjee, B. (1978) *Coordination Chemistry Reviews*, **26**, 281–303.
7 Yale, H.L. (1943) *Chemical Reviews*, **33**, 209–256.

8 Bauer, L. and Exner, O. (1974) *Angewandte Chemie (International Edition in English)*, **13**, 376–384.

9 Lindberg, B., Berndtsson, A., Nilsson, R., Nyholm, R. and Exner, O. (1978) *Acta Chemica Scandinavica Series A – Physical and Inorganic Chemistry*, **32**, 353–359.

10 Lipczynska-Kochany, E. and Iwamura, H. (1982) *The Journal of Organic Chemistry*, **47**, 5277–5282.

11 Ventura, O.N., Rama, J.B., Turi, L. and Dannenberg, J.J. (1993) *Journal of the American Chemical Society*, **115**, 5754–5761.

12 Edwards, D.C. and Myneni, S.C.B. (2006) *Journal of Physical Chemistry A*, **110**, 11809–11818.

13 Marmion, C.J., Murphy, T., Nolan, K.B. and Docherty, J.R. (2000) *Chemical Communications*, 1153–1154.

14 Low, S.Y. (2005) *Molecular Aspects of Medicine*, **26**, 97–138.

15 Marmion, C.J., Cameron, B., Mulcahy, C. and Fricker, S.P. (2004) *Current Topics in Medicinal Chemistry*, **4**, 1585–1603.

16 Marmion, C.J., Griffith, D. and Nolan, K.B. (2004) *European Journal of Inorganic Chemistry*, 3003–3016.

17 Codd, R. (2008) *Coordination Chemistry Reviews*, **252**, 1387–1408.

18 Kurzak, B., Kozlowski, H. and Farkas, E. (1992) *Coordination Chemistry Reviews*, **114**, 169–200.

19 O'Brien, E.C., Farkas, E., Gil, M.J., Fitzgerald, D., Castineras, A. and Nolan, K.B. (2000) *Journal of Inorganic Biochemistry*, **79**, 47–51.

20 Yang, K. and Lou, B. (2003) *Mini-Reviews in Medicinal Chemistry*, **3**, 349–360.

21 Roosenberg, J.M. J2nd, Lin, Y.M., Lu, Y. and Miller, M.J. (2000) *Current Medicinal Chemistry*, **7**, 159–197.

22 Faraldo-Gómez, J.D. and Sansom, M.S.P. (2003) *Nature Reviews. Molecular Cell Biology*, **4**, 105–116.

23 Dhungana, S., Miller, M.J., Dong, L., Ratledge, C. and Crumbliss, A.L. (2003) *Journal of the American Chemical Society*, **125**, 7654–7663.

24 Bernhardt, P.V. (2007) *Dalton Transactions*, 3214–3220.

25 Cario, H., Janka-Schaub, G., Janssen, G., Jarisch, A., Strauss, G. and Kohne, E. (2007) *Klinische Padiatrie*, **219**, 158–165.

26 Delea, T.E., Edelsberg, J., Sofrygin, O., Thomas, S.K., Baladi, J.-F., Phatak, P.D. and Coates, T.D. (2007) *Transfusion*, **47**, 1919–1929.

27 Kontoghiorghes, G.J., Eracleous, E., Economides, C. and Kolnagou, A. (2005) *Current Medicinal Chemistry*, **12**, 2663–2681.

28 Cohen, A.R. (2006) *Hematology*, 42–47.

29 Roberts, D.J., Brunskill, S.J., Doree, C., Williams, S., Howard, J. and Hyde, C.J. (2007) *Cochrane Database of Systematic Reviews*, 3 (Art. No.: CD004839), doi:10.1002/14651858.CD004839.pub2.

30 Choudhry, V.P. and Naithani, R. (2007) *Indian Journal of Pediatrics*, **74**, 759–764.

31 Kattamis, A. (1054) *Annals of the New York Academy of Sciences*, **2005**, 175–182.

32 Dong, L., Roosenberg, J.M. 2nd and Miller, M.J. (2002) *Journal of the American Chemical Society*, **124**, 15001–15005.

33 Dong, L. and Miller, M.J. (2002) *The Journal of Organic Chemistry*, **67**, 4759–4770.

34 Lou, B. and Yang, K. (2003) *Mini-Reviews in Medicinal Chemistry*, **3**, 609–620.

35 Muri, E.M.F., Nieto, M.J., Sindelar, R.D. and Williamson, J.S. (2002) *Current Medicinal Chemistry*, **9**, 1631–1653.

36 Amtul, Z., Rahman, A.U., Siddiqui, R.A. and Choudhary, M.I. (2002) *Current Medicinal Chemistry*, **9**, 1323–1348.

37 Levin, J.I. (2004) *Current Topics in Medicinal Chemistry*, **4**, 1289–1310.

38 Clark, I.M., Rowan, A.D. and Cawston, T.E. (2000) *Current Opinion in Anti-Inflammatory Immunomodulatory Investigational Drugs*, **2**, 16–25.

39 Handsley, M.M. and Edwards, D.R. (2005) *International Journal of Cancer*, **115**, 849–860.

40 Wallach, J. (2005) *Biochimie*, **87**, 241.

41 Vihinen, P., Ala-aho, R. and Kahari, V.-M. (2005) *Current Cancer Drug Targets*, **5**, 203–220.

42 Raffetto, J.D. and Khalil, R.A. (2008) *Biochemical Pharmacology*, **75**, 346–359.

43 Whittaker, M., Floyd, C.D., Brown, P. and Gearing, A.J.H. (1999) *Chemical Reviews*, **99**, 2735–2776.

44 Mannello, F., Tonti, G. and Papa, S. (2005) *Current Cancer Drug Targets*, **5**, 285–298.

45 Rao, B.G. (2005) *Current Pharmaceutical Design*, **11**, 295–322.

46 Hu, J., Van den Steen, P.E., Sang, Q.X. and Opdenakker, G. (2007) *Nature Reviews. Drug Discovery*, **6**, 480–498.

47 Puerta, D.T. and Cohen, S.M. (2004) *Current Topics in Medicinal Chemistry*, **4**, 1551–1573.

48 Fisher, J.F. and Mobashery, S. (2006) *Cancer and Metastasis Reviews*, **25**, 115–136.

49 Bissett, D., O'Byrne, K.J., von Pawel, J., Gatzemeier, U., Price, A., Nicolson, M., Mercier, R., Mazabel, E., Penning, C., Zhang, M.H., Collier, M.A. and Shepherd, F.A. (2005) *Journal of Clinical Oncology*, **23**, 842–849.

50 Savinov, A.Y., Rozanov, D.V., Golubkov, V.S., Wong, F.S. and Strongin, A.Y. (2005) *The Journal of Biological Chemistry*, **280**, 27755–27758.

51 Ganea, E., Trifan, M., Laslo, A.C., Putina, G. and Cristescu, C. (2007) *Biochemical Society Transactions*, **35**, 689–691.

52 Inche, A.G. and La Thangue, N.B. (2006) *Drug Discovery Today*, **11**, 97–109.

53 Mehnert, J.M. and Kelly, W.K. (2007) *Cancer Journal*, **13**, 23–29.

54 Dokmanovic, M., Clarke, C. and Marks, P.A. (2007) *Molecular Cancer Research*, **5**, 981–989.

55 Ropero, S. and Esteller, M. (2007) *Molecular Oncology*, **1**, 19–25.

56 Xu, W.S., Parmigiani, R.B. and Marks, P.A. (2007) *Oncogene*, **26**, 5541–5552.

57 Bolden, J.E., Peart, M.J. and Johnstone, R.W. (2006) *Nature Reviews. Drug Discovery*, **5**, 769–784.

58 Liu, T., Kuljaca, S., Tee, A. and Marshall, G.M. (2006) *Cancer Treatment Reviews*, **32**, 157–165.

59 Gu, W., Nusinzon, I., Smith, R.D. JJr, Horvath, C.M. and Silverman, R.B. (2006) *Bioorganic and Medicinal Chemistry*, **14**, 3320–3329.

60 Weinmann, H. and Ottow, E. (2004) *Annual Reports in Medicinal Chemistry*, **39**, 185–196.

61 Vanhaecke, T., Papeleu, P., Elaut, G. and Rogiers, V. (2004) *Current Medicinal Chemistry*, **11**, 1629–1643.

62 Mann, B.S., Johnson, J.R., Cohen, M.H., Justice, R. and Pazdur, R. (2007) *Oncologist*, **12**, 1247–1252.

63 Marks, P.A. (2007) *Oncogene*, **26**, 1351–1356.

64 Marks, P.A. and Breslow, R. (2007) *Nature Biotechnology*, **1**, 84–90.

65 Vannini, A., Volpari, C., Filocamo, G., Casavola, E.C., Brunetti, M., Renzoni, D., Chakravarty, P., Paolini, C., De Francesco, R., Gallinari, P., Steinkuhler, C. and Di Marco, S. (2004) *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15064–15069.

66 Finnin, M.S., Donigan, J.R., Cohen, A., Richon, V.M., Rifkind, R.A., Marks, P.A., Breslow, R. and Pavletich, N.P. (1999) *Nature*, **401**, 188–193.

67 Remiszewski, S.W. (2003) *Current Medicinal Chemistry*, **10**, 2393–2402.

68 Glaser, K.B. (2007) *Biochemical Pharmacology*, **74**, 659–671.

69 Beckers, T., Burkhardt, C., Wieland, H., Gimmnich, P., Ciossek, T., Maier, T. and Sanders, K. (2007) *International Journal of Cancer*, **121**, 1138–1148.

70 Smith, W.L., DeWitt, D.L. and Garavito, R.M. (2000) *Annual Review of Biochemistry*, **69**, 145–182.

71 Bingham, S., Beswick, P.J., Blum, D.E., Gray, N.M. and Chessell, I.P. (2006) *Seminars in Cell & Developmental Biology*, **17**, 544–554.

72 Gupta, K., Selinsky, B.S., Kaub, C.J., Katz, A.K. and Loll, P.J. (2004) *Journal of Molecular Biology*, **335**, 503–518.

73 Loll, P.J., Sharkey, C.T., O'Connor, S.J., Dooley, C.M., O'Brien, E., Devocelle, M.,

Nolan, K.B., Selinsky, B.S. and Fitzgerald, D.F. (2001) *Molecular Pharmacology*, **60**, 1407–1413.

74 Dooley, C.M., Devocelle, M., McLoughlin, B., Nolan, K.B., Fitzgerald, D.J. and Sharkey, C.T. (2003) *Molecular Pharmacology*, **63**, 450–455.

75 Ouellet, M., Aitken, S.M., English, A.M. and Percival, M.D. (2004) *Archives of Biochemistry and Biophysics*, **431**, 107–118.

76 Itakura, H., Oda, Y. and Fukuyama, K. (1997) *FEBS Letters*, **412**, 107–110.

77 Henriksen, A., Schuller, D.J., Meno, K., Welinder, K.G., Smith, A.T. and Gajhede, M. (1998) *Biochemistry*, **37**, 8054–8060.

78 Indiani, C., Santoni, E., Becucci, M., Boffi, A., Fukuyama, K. and Smulevich, G. (2003) *Biochemistry*, **42**, 14066–14074.

79 Lee, J., Chubb, A.J., Moman, E., McLoughlin, B.M., Sharkey, C.T., Kelly, J.G., Nolan, K.B., Devocelle, M. and Fitzgerald, D.J. (2005) *Organic and Biomolecular Chemistry*, **3**, 3678–3685.

80 Jones, J. (2002) *Amino Acid and Peptide Synthesis*, 2nd edn, Oxford Science Publications, Oxford.

81 Munson, J.W. (1992) *The Chemistry of Acid Derivatives (Chemistry of Functional Groups)*, John Wiley & Sons, Inc., New York.

82 Krchnak, V. (2006) *Mini-Reviews in Medicinal Chemistry*, **6**, 27–36.

83 Reddy, A.S., Kumar, M.S. and Reddy, G.R. (2000) *Tetrahedron Letters*, **41**, 6285–6288.

84 Barlaam, B., Hamon, A. and Maudet, M. (1998) *Tetrahedron Letters*, **39**, 7865–7868.

85 Renfrow, W.B. and Hauser, C.R. (1937) *Journal of the American Chemical Society*, **59**, 2308–2314.

86 Bade, M.L. (1971) *Journal of the American Chemical Society*, **93**, 949–953.

87 Van't Riet, B., Wampler, G.L. and Elford, H.L. (1979) *Journal of Medicinal Chemistry*, **22**, 589–592.

88 Duan, J.J.W., Chen, L., Wasserman, Z.R., Lu, Z., Liu, R.-Q., Covington, M.B., Qian, M., Hardman, K.D., Magolda, R.L., Newton, R.C., Christ, D.D., Wexler, R.R. and Decicco, C.P. (2002) *Journal of Medicinal Chemistry*, **45**, 4954–4957.

89 Gissot, A., Volonterio, A. and Zanda, M. (2005) *The Journal of Organic Chemistry*, **70**, 6925–6928.

90 Griffith, D., Bergamo, A., Pin, S., Vadori, M., Muller-Bunz, H., Sava, G. and Marmion, C.J. (2007) *Polyhedron*, **26**, 4697–4706.

91 McMurry, J. (2004) *Organic Chemistry*, 6th edn, Thomson, Brookes/Cole, Belmont, CA.

92 Gediya, L.K., Chopra, P., Purushottamachar, P., Maheshwari, N. and Njar, V.C.O. (2005) *Journal of Medicinal Chemistry*, **48**, 5047–5051.

93 Massaro, A., Mordini, A., Reginato, G., Russo, F. and Taddei, M. (2007) *Synthesis*, **20**, 3201–3204.

94 Ho, C.Y., Strobel, E., Ralbovsky, J. and Galemmo, R.A. (2005) *The Journal of Organic Chemistry*, **70**, 4873–4875.

95 Hoegberg, T., Stroem, P., Ebner, M. and Raemsby, S. (1987) *The Journal of Organic Chemistry*, **52**, 2033–2036.

96 Jang, D.O., Park, D.J. and Kim, J. (1999) *Tetrahedron Letters*, **40**, 5323–5326.

97 Levin, J.I., Du, M.T., DiJoseph, J.F., Killar, L.M., Sung, A., Walter, T., Sharr, M.A., Roth, C.E., Moy, F.J., Powers, R., Jin, G., Cowling, R. and Skotnicki, J.S. (2001) *Bioorganic & Medicinal Chemistry Letters*, **11**, 235–238.

98 Levin, J.I., Gu, Y., Nelson, F.C., Zask, A., DiJoseph, J.F., Sharr, M.A., Sung, A., Jin, G., Cowling, R., Chanda, P., Cosmi, S., Hsiao, C.-L., Edris, W., Wilhelm, J., Killar, L.M. and Skotnicki, J.S. (2001) *Bioorganic & Medicinal Chemistry Letters*, **11**, 239–242.

99 Levin, J.I., Chen, J., Du, M., Hogan, M., Kincaid, S., Nelson, F.C., Venkatesan, A.M., Wehr, T., Zask, A., DiJoseph, J., Killar, L.M., Skala, S., Sung, A., Sharr, M., Roth, C., Jin, G., Cowling, R., Mohler, K.M., Black, R.A., March, C.J. and Skotnicki, J.S. (2001) *Bioorganic & Medicinal Chemistry Letters*, **11**, 2189–2192.

100 Levin, J.I., Chen, J.M., Du, M.T., Nelson, F.C., Wehr, T., DiJoseph, J.F., Killar, L.M., Skala, S., Sung, A., Sharr, M.A., Roth, C.E., Jin, G., Cowling, R., Di, L., Sherman, M., Xu, Z.B., March, C.J., Mohler, K.M., Black, R.A. and Skotnicki, J.S. (2001) *Bioorganic & Medicinal Chemistry Letters*, **11**, 2975–2978.

101 Coates, R.M. and Firsan, S.J. (1986) *The Journal of Organic Chemistry*, **51**, 5198–5209.

102 Murray, W., Wachter, M., Barton, D. and Forero-Kelly, Y. (1991) *Synthesis*, 18–20.

103 Flipo, M., Beghyn, T., Charton, J., Leroux, V.A., Deprez, B.P. and Deprez-Poulain, R.F. (2007) *Bioorganic and Medicinal Chemistry*, **15**, 63–76.

104 Giacomelli, G., Porcheddu, A. and Salaris, M. (2003) *Organic Letters*, **5**, 2715–2717.

105 Mori, K. and Koseki, K. (1988) *Tetrahedron*, **44**, 6013–6020.

106 Mai, A., Esposito, M., Sbardella, G. and Massa, S. (2001) *Organic Preparations and Procedures International*, **33**, 391–394.

107 De Luca, L., Giacomelli, G. and Taddei, M. (2001) *The Journal of Organic Chemistry*, **66**, 2534–2537.

108 Geffken, D. (1976) *Archiv der Pharmazie*, **309**, 413–421.

109 Corbett, M.D. and Chipko, B.R. (1978) *Experientia*, **34**, 1125–1126.

110 Musser, J.H., Kreft, A.F., Bender, R.H.W., Kubrak, D.M., Grimes, D., Carlson, R.P., Hand, J.M. and Chang, J. (1990) *Journal of Medicinal Chemistry*, **33**, 240–245.

111 Volonterio, A., Bellosta, S., Bravo, P., Canavesi, M., Corradi, E., Meille, S.V., Monetti, M., Moussier, N. and Zanda, M. (2002) *European Journal of Organic Chemistry*, 428–438.

112 Fray, M.J. and Dickinson, R.P. (2001) *Bioorganic & Medicinal Chemistry Letters*, **11**, 571–574.

113 Walz, A.J. and Miller, M.J. (2007) *Tetrahedron Letters*, **48**, 5103–5105.

114 Bouchain, G. and Delorme, D. (2003) *Current Medicinal Chemistry*, **10**, 2359–2372.

115 Ech-Chahad, A., Minassi, A., Berton, L. and Appendino, G. (2005) *Tetrahedron Letters*, **46**, 5113.

116 Josef, K.A., Kauer, F.W. and Bihovsky, R. (2001) *Bioorganic & Medicinal Chemistry Letters*, **11**, 2615–2617.

117 Nishino, N., Jose, B., Shinta, R., Kato, T., Komatsu, Y. and Yoshida, M. (2004) *Bioorganic and Medicinal Chemistry*, **12**, 5777–5784.

118 Coste, J. (2004) In *Houben–Weyl Methods of Organic Chemistry: Synthesis of Peptides and Peptidomimetics*, vol. **E22a**, 538–540. (eds M. Goodman, A. Felix, L. Moroder and C. Toniolo), Thieme, Stuttgart.

119 Carpino, L.A., Imazumi, H., El-Faham, A., Ferrer, F.J., Zhang, C., Lee, Y., Foxman, B.M., Henklein, P., Hanay, C., Mugge, C., Wenschuh, H., Klose, J., Beyermann, M. and Bienert, M. (2002) *Angewandte Chemie (International Edition in English)*, **41**, 441–445.

120 Bailen, M.A., Chinchilla, R., Dodsworth, D.J. and Najera, C. (2001) *Tetrahedron Letters*, **42**, 5013–5016.

121 Kokare, N.D., Nagawade, R.R., Rane, V.P. and Shinde, D.B. (2007) *Tetrahedron Letters*, **48**, 4437–4440.

122 Sibi, M.P., Hasegawa, H. and Ghorpade, S.R. (2002) *Organic Letters*, **4**, 3343–3346.

123 Sibi, M.P. and Hasegawa, H. (2002) *Organic Letters*, **4**, 3347–3349.

124 Boukhris, S., Souizi, A. and Robert, A. (1996) *Tetrahedron Letters*, **37**, 179–182.

125 Boukhris, S. and Souizi, A. (2000) *Tetrahedron Letters*, **41**, 2559–2562.

126 Angeli, A. (1896) *Gazzetta Chimica Italiana*, **26**, 17–25.

127 Rimini, E. (1901) *Gazzetta Chimica Italiana*, **31**, 84.

128 Angeli, A. and Angelico, E. (1904) *Gazzetta Chimica Italiana*, **34**, 50–57.

129 Porcheddu, A. and Giacomelli, G. (2006) *The Journal of Organic Chemistry*, **71**, 7057–7059.

130 Hassner, A., Wiederkehr, R. and Kascheres, A.J. (1970) *Organic Chemistry* **35** 1962–1964.

131 Sosnovsky, G. and Krogh, J.A. (1980) *Synthesis*, 654–656.

132 Yamada, K., Kanekiyo, T., Tanaka, S., Naruchi, K. and Yamamoto, M. (1981) *Journal of the American Chemical Society*, **103**, 7003–7005.

133 Grigg, R., Major, J.P., Martin, F.M. and Whittaker, M. (1999) *Tetrahedron Letters*, **40**, 7709–7711.

134 Stowell, J.C. and Christenson, J.R. (1980) *Liebigs Annalen der Chemie*, 791–797.

135 Wu, T.Y.H., Hassig, C., Wu, Y., Ding, S. and Schultz, P.G. (2004) *Bioorganic & Medicinal Chemistry Letters*, **14**, 449–453.

136 Fennell, K.A. and Miller, M.J. (2007) *Organic Letters*, **9**, 1683–1685.

137 Walz, A.J., Möllmann, U. and Miller, M.J. (2007) *Organic and Biomolecular Chemistry*, **5**, 1621–1628.

138 Milewska, M.J. and Chimiak, A. (1990) *Synthesis*, 233–234.

139 Bergeron, R.J. and Phanstiel, O. (1992) *The Journal of Organic Chemistry*, **57**, 7140–7143.

140 Wang, Q.X., King, J. and Phanstiel, O. (1997) *The Journal of Organic Chemistry*, **62**, 8104–8108.

141 Phanstiel, O., Wang, Q.X., Powell, D.H., Ospina, M.P. and Leeson, B.A. (1999) *The Journal of Organic Chemistry*, **64**, 803–806.

142 Fields, J.D. and Kropp, P.J. (2000) *The Journal of Organic Chemistry*, **65**, 5937–5941.

143 Hu, J. and Miller, M.J. (1994) *The Journal of Organic Chemistry*, **59**, 4858–4861.

144 Tokuyama, H., Kuboyama, T., Amano, A., Yamashita, T. and Fukuyama, T. (2000) *Synthesis*, **9**, 1299–1304.

145 Polonski, T. and Chimiak, A. (1974) *Tetrahedron Letters*, 2453–2456.

146 Naegeli, H.-U. and Keller-Schierlein, W. (1978) *Helvetica Chimica Acta*, **61**, 2088–2095.

147 Lin, Y.-M. and Miller, M.J. (1999) *The Journal of Organic Chemistry*, **64**, 7451–7458.

148 Merrifield, R.B. and Stewart, J.M. (1965) *Nature*, **207**, 522–523.

149 Ley, S.V. and Baxendale, I.R. (2002) *Nature Reviews. Drug Discovery*, **1**, 573–586.

150 Floyd, C.D., Lewis, C.N., Patel, S.R. and Whittaker, M. (1996) *Tetrahedron Letters*, **37**, 8045–8048.

151 Salvino, J.M., Mathew, R., Kiesow, T., Narensingh, R., Mason, H.J., Dodd, A., Groneberg, R., Burns, C.J., McGeehan, G., Kline, J., Orton, E., Tang, S.-Y., Morrisette, M. and Labaudininiere, R. (2000) *Bioorganic & Medicinal Chemistry Letters*, **10**, 1637–1640.

152 Richter, L.S. and Desai, M.C. (1997) *Tetrahedron Letters*, **38**, 321–322.

153 Khan, S.I. and Grinstaff, M.W. (1998) *Tetrahedron Letters*, **39**, 8031–8034.

154 Salvino, J.M., Mervic, M., Mason, H.J., Kiesow, T., Teager, D., Airey, J. and Labaudiniere, R. (1999) *The Journal of Organic Chemistry*, **64**, 1823–1830.

155 Ede, N.J., James, I.W., Krywult, B.M., Griffiths, R.M., Eagle, S.N., Gubbins, B., Leitch, J.A., Sampson, W.R. and Bray, A.M. (1999) *Letters in Peptide Science*, **6**, 157–163.

156 Zhang, Y., Li, D., Houtman, J.C., Witiak, D.T., Seltzer, J., Bertics, P.J. and Lauhon, C.T. (1999) *Bioorganic & Medicinal Chemistry Letters*, **9**, 2823–2826.

157 Somoza, J.R., Skene, R.J., Katz, B.A., Mol, C., Ho, J.D., Jennings, A.J., Luong, C., Arvai, A., Buggy, J.J., Chi, E., Tang, J., Sang, B.-C., Verner, E., Wynands, R., Leahy, E.M., Dougan, D.R., Snell, G., Navre, M., Knuth, M.W., Swanson, R.V., McRee, D.E. and Tari, L.W. (2004) *Structure*, **12**, 1325–1334.

158 Brennan, T., Biddison, G., Frauendorf, A., Schwarcz, L., Keen, B., Ecker, D.J., Davis, P.W., Tinder, R. and Swayze, E.E. (1998) *Biotechnology and Bioengineering*, **61**, 33–45.

159 Yin, Z., Low, K. and Lye, P. (2005) *Synthetic Communications*, **35**, 2945–2950.

160 Robinson, D.E. and Holladay, M.W. (2000) *Organic Letters*, **2**, 2777–2779.

161 Reddy, P.A., Schall, O.F., Wheatley, J.R., Rosik, L.O., McClurg, J.P., Marshall, G.R.

and Slomczynsk, U. (2001) *Synthesis*, 1086–1092.

162 Poreddy, A.R., Schall, O.F., Marshall, G.R., Ratledge, C. and Slomczynska, U. (2003) *Bioorganic & Medicinal Chemistry Letters*, **13**, 2553–2556.

163 Poreddy, A.R., Schall, O.F., Osiek, T.A., Wheatley, J.R., Beusen, D.D., Marshall, G.R. and Slomczynska, U. (2004) *Journal of Combinatorial Chemistry*, **6**, 239–254.

164 Stanger, K.J. and Krchnak, V. (2006) *Journal of Combinatorial Chemistry*, **8**, 435–439.

165 Stanger, K.J., Sliva, D., Jiang, J. and Krchnak, V. (2006) *Combinatorial Chemistry & High Throughput Screening*, **9**, 651–661.

166 Krchnak, V. and Slough, G.A. (2004) *Tetrahedron Letters*, **45**, 4649–4652.

167 Barlaam, B., Koza, P. and Berriot, J. (1999) *Tetrahedron*, **55**, 7221–7232.

168 Mellor, S.L. and Chan, W.C. (1997) *Chemical Communications*, 2005–2006.

169 Mellor, S.L., McGuire, C. and Chan, W.C. (1997) *Tetrahedron Letters*, **38**, 3311–3314.

170 Bauer, U., Ho, W.-B. and Koskinen, A.M.P. (1997) *Tetrahedron Letters*, **38**, 7233–7236.

171 Chollet, A.-M., Le Diguarher, T., Kucharczyk, N., Loynel, A., Bertrand, M., Tucker, G., Guilbaud, N., Burbridge, M., Pastoureau, P., Fradin, A., Sabatini, M., Fauchere, J.-L. and Casara, P. (2002) *Bioorganic and Medicinal Chemistry*, **10**, 531–544.

172 Robinson, L.A., Wilson, D.M., Delaet, N.G.J., Bradley, E.K., Dankwardt, S.M., Campbell, J.A., Martin, R.L., Van Wart, H.E., Walker, K.A.M. and Sullivan, R.W. (2003) *Bioorganic & Medicinal Chemistry Letters*, **13**, 2381–2384.

173 Meloni, M.M. and Taddei, M. (2001) *Organic Letters*, **3**, 337–340.

174 Delaet, N.G.J., Robinson, L.A., Wilson, D.M., Sullivan, R.W., Bradley, E.K., Dankwardt, S.M., Martin, R.L., Van Wart, H.E. and Walker, K.A.M. (2003) *Bioorganic & Medicinal Chemistry Letters*, **13**, 2101–2104.

175 Lampariello, L.R., Piras, D., Rodriquez, M. and Taddei, M. (2003) *The Journal of Organic Chemistry*, **68**, 7893–7895.

176 Ngu, K. and Patel, D.V. (1997) *The Journal of Organic Chemistry*, **62**, 7088–7089.

177 Chen, J.J. and Spatola, A.F. (1997) *Tetrahedron Letters*, **38**, 1511–1514.

178 Leeuwenburgh, M.A., Geurink, P.P., Klein, T., Kauffman, H.F., van der Marel, G.A., Bischoff, R. and Overkleeft, H.S. (2006) *Organic Letters*, **8**, 1705–1708.

179 Sternson, S.M., Wong, J.C., Grozinger, C.M. and Schreiber, S.L. (2001) *Organic Letters*, **3**, 4239–4242.

180 Ye, Y., Liu, M., Kao, J.L.-F. and Marshall, G.R. (2006) *Biopolymers*, **84**, 472–489.

181 Sasubilli, R. and Gutheil, W.G. (2004) *Journal of Combinatorial Chemistry*, **6**, 911–915.

182 Bui, C.T., Maeji, N.J. and Bray, A.M. (2000/2001) *Biotechnology and Bioengineering*, **71**, 91–93.

183 Dankwardt, S.M. (1998) *Synlett*, 761.

184 Dankwardt, S.M., Billedeau, R.J., Lawley, L.K., Abbot, S.C., Martin, R.L., Chan, C.S., Van Wart, H.E. and Walker, K.A.M. (2000) *Bioorganic & Medicinal Chemistry Letters*, **10**, 2513–2516.

185 Dankwardt, S.M., Martin, R.L., Chan, C.S., Van Wart, H.E., Walker, K.A.M., Delaet, N.G. and Robinson, L.A. (2001) *Bioorganic & Medicinal Chemistry Letters*, **11**, 2085–2088.

186 Dankwardt, S.M., Abbot, S.C., Broka, C.A., Martin, R.L., Chan, C.S., Springman, E.B., Van Wart, H.E. and Walker, K.A.M. (2002) *Bioorganic & Medicinal Chemistry Letters*, **12**, 1233–1235.

187 Golebiowski, A. and Klopfenstein, S. (1998) *Tetrahedron Letters*, **39**, 3397–3400.

188 Thouin, E. and Lubell, W.D. (2000) *Tetrahedron Letters*, **41**, 457–460.

189 Camarero, J.A., Adeva, A. and Muir, T.W. (2000) *Letters in Peptide Science*, **7**, 17–21.

190 Zhang, W., Zhang, L., Li, X., Weigel, J.A., Hall, S.E. and Mayer, J.P. (2001) *Journal of Combinatorial Chemistry*, **3**, 151–153.

191 Devocelle, M., McLoughlin, B.M., Sharkey, C.T., Fitzgerald, D.J. and Nolan,

K.B. (2003) *Organic and Biomolecular Chemistry*, **1**, 850–853.

192 Vickerstaffe, E., Warrington, B.H., Ladlow, M. and Ley, S.V. (2003) *Organic and Biomolecular Chemistry*, **1**, 2419–2422.

193 Caldarelli, M., Habermann, J. and Ley, S.V. (1999) *Bioorganic & Medicinal Chemistry Letters*, **9**, 2049–2052.

194 Jensen, K.J., Alsina, J., Songster, M.F., Vágner, J.V., Albericio, F. and Barany, G. (1998) *Journal of the American Chemical Society*, **120**, 5441–5452.

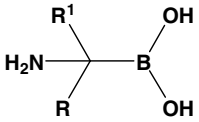
195 Gazal, S., Masterson, L.R. and Barany, G. (2005) *The Journal of Peptide Research*, **66**, 324–332.

4
Chemistry of α-Aminoboronic Acids and their Derivatives

Valery M. Dembitsky and Morris Srebnik

4.1
Introduction

In recent years there has been increasing interest in new practical methods to prepare novel non-natural (*R*) amino acid derivatives to serve as building blocks in combinatorial chemistry and drug discovery [1–5]. Although many routes to amino acids have been developed, there is still a need for concise and convergent approaches that allow structure variability and facile incorporation of functional groups and ring systems [6–10]. As non-natural (*R*) amino acid derivatives, boronic acids (**1**) have assumed great importance since they serve as transition state analogs of natural amino acids. Their ease of preparation and their relative stability have further increased their value.



1

The bioinorganic chemistry of boron-containing compounds is therefore an area of growing interest, and has recently expanded to include purine nucleosides, psuedocryptands (which mimic the naturally occurring antibiotics boromycin and aplasmomycin), steroids, calixarenes, carbohydrates, fatty acids, porphyrins, and amino acids [11–17]. Boronic acids [$RB(OH)_2$] and boronate esters [$RB(OR^1)_2$] have been found to facilitate the transport of various ribonucleosides in and out of liposomes – an important attribute in the area of drug design [18–20]. Simple aminoboron compounds have also found some utility in boron neutron capture therapy [21, 22] and other forms of cancer therapy [1–3]. As a result, much effort has been focused on the synthesis of boron-containing amino acid and peptide derivatives [23, 24]. Unfortunately, the incorporation of a Lewis basic amine and a Lewis

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

acidic boronic acid functionality into the same molecule is a notoriously difficult procedure, and synthetic routes to these compounds are scarce. New effective methods to prepare amino and related boron compounds will have a tremendous impact on bioinorganic and pharmaceutical chemistry [25–30].

Serine proteases – a large and functionally diverse class of proteolytic enzymes – are prominent therapeutic targets because of their involvement in a host of physiological processes [31]. They catalyze peptide bond cleavage by acylation and deacylation of the active-site serine residue in a sequence that involves two tetrahedral intermediates [32]. Most small-molecule inhibitors of these enzymes form covalent adducts with the active-site serine that mimic to some degree these tetrahedral intermediates. Peptide derivatives with electron-deficient ketones, aldehydes, boronic acids, and phosphonylating agents have been devised as analogs of the second tetrahedral intermediate [33] with their selectivity among the various proteases related to the substrate specificity these enzymes manifest at the S1, S2, and higher, binding sites [34].

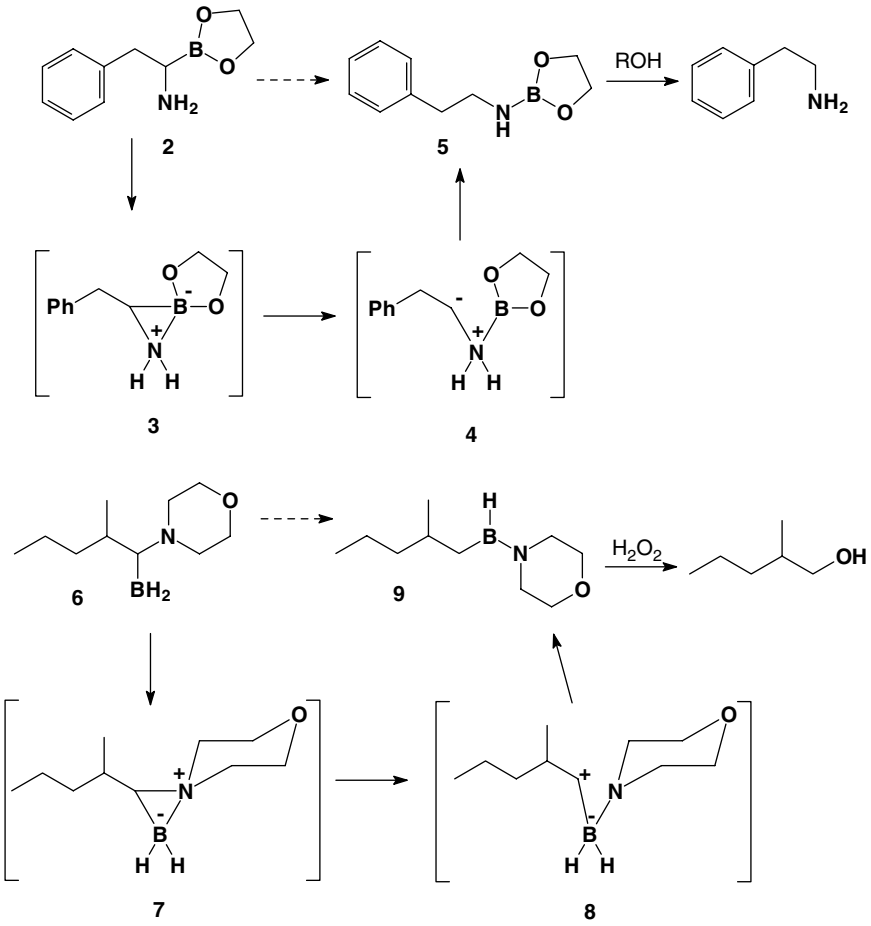
4.2 Synthesis of α-Aminoboronic Acids

α-Aminoboronic acids or esters are unstable toward migration of boron from carbon to nitrogen and yield the corresponding amines in protic solvents [35–37]. A typical example of migration of boron from carbon to nitrogen in α-aminoboranes is rearrangement of the 1-amino-2-phenylethylboronic ester (**2**) to the 2-phenylethylamine derivative (**3**) (Scheme 4.1) [35]. Boronic esters bearing tertiary amino groups in the α-position do not undergo this type of deboronation [36, 37]. It therefore appears likely that the rearrangement occurs via intramolecular nucleophilic attack of the amino group of (**2**) on boron. Ring opening of (**3**) to (**4**) might then occur or proton migration to form (**5**) might be concerted with ring opening [38, 39] In an analogous manner, migration of the α-aminoborane (**6**) would lead to the intermediate (**7**), which could undergo ring opening to the zwitterion (**8**), in this example concerted with, or followed by, hydride migration to form (**9**).

4.3 Synthesis of α-Amidoboronic Acid Derivatives

α-Haloboronic esters are usually useful for the synthesis of α-amidoboronic acid derivatives [40]. Nucleophilic reactions of α-haloboronic esters with carbon nucleophiles are utilized in asymmetric synthesis with displacements of the halide atom. The asymmetric conversion of a CHCl group into a B–C bond can be controlled with very high precision by the use of chiral ligands on the boron atom.

The first synthesis of the unnatural α-amidoboronic ester (**13**) was studied by Matteson *et al.* [41]. (*S*)-Pinanediol (*S*)-(1-chloro-2-phenylethyl)boronate (**10**) was prepared from the reaction of (*S*)-pinanediol benzylboronate with (dichloromethyl)

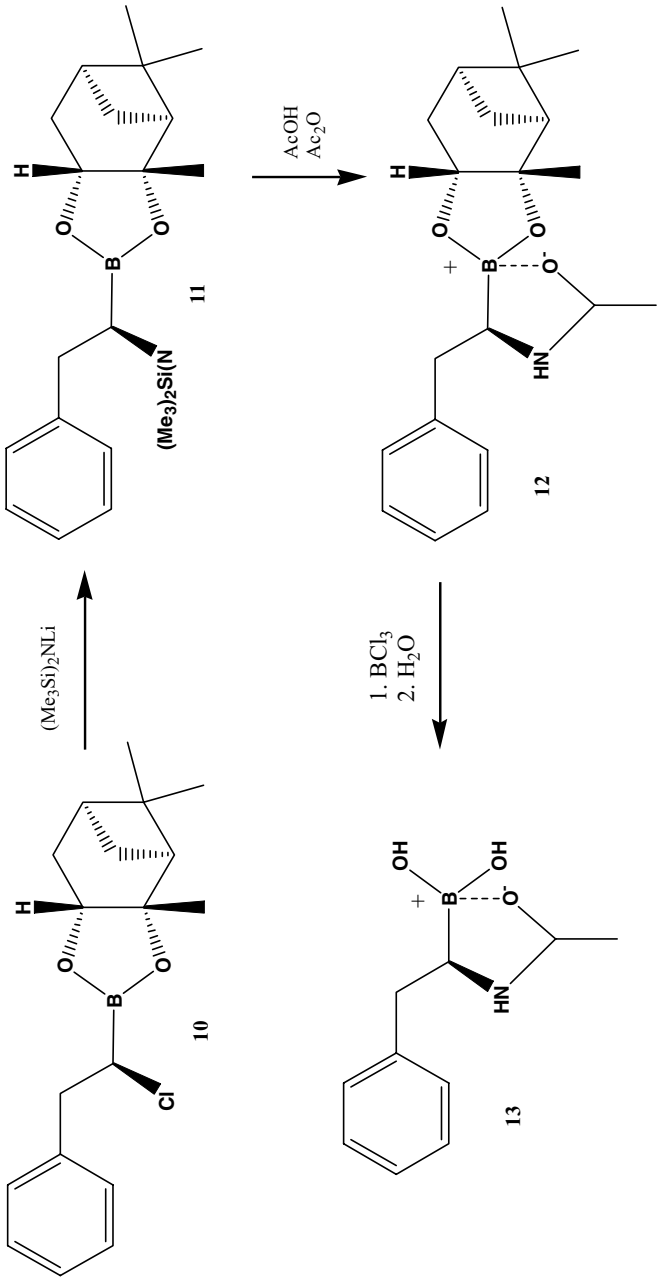


Scheme 4.1

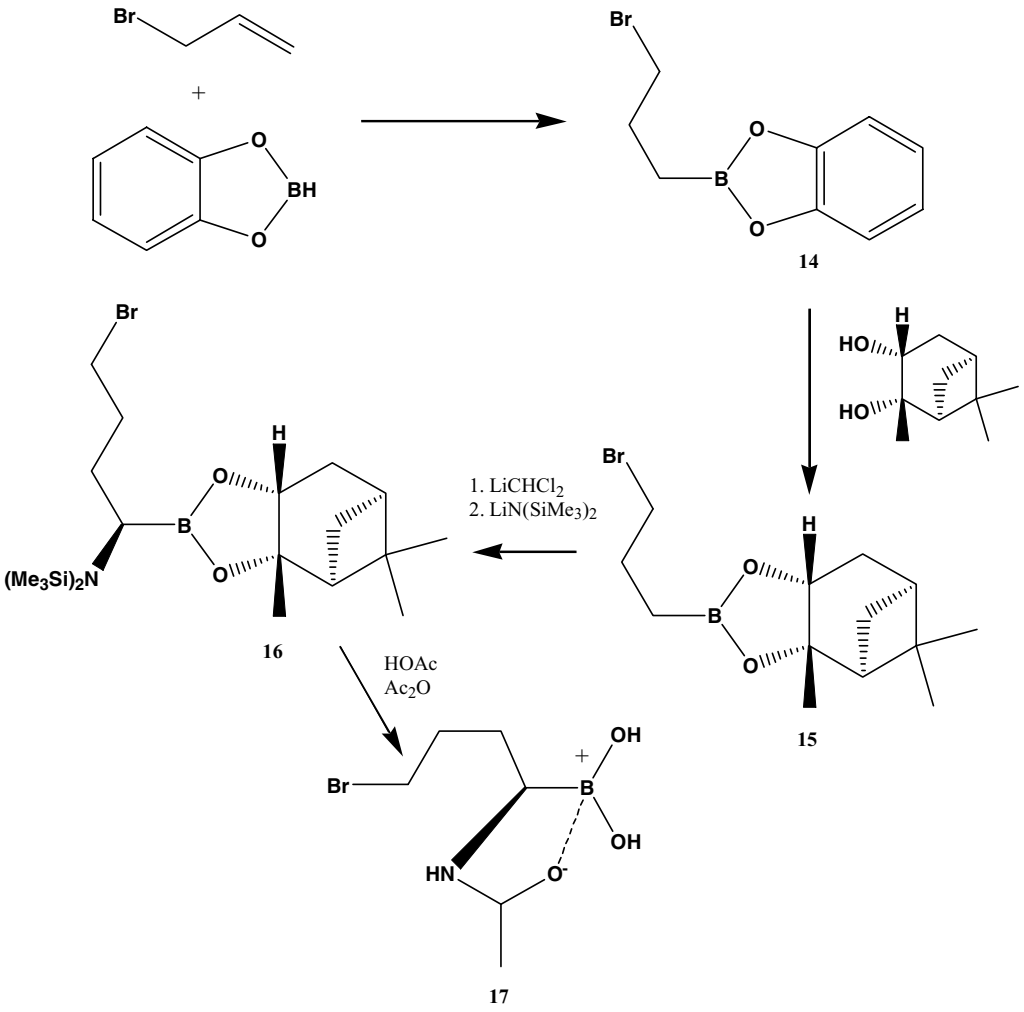
lithium [35], followed by displacement of the chloride ion from (**10**) with lithium hexamethyldisilazane (LiHMDS), to provide (**11**). Treatment of (**11**) with $AcOH/Ac_2O$ yielded the acetoamidoboronic ester (**12**), which was cleaved with BCl_3 to yield (*S*)-*N*-acetylboraphenylalanine (**13**) (Scheme 4.2). Compound (**13**) was a potent inhibitor of certain enzymes [2]. Some other routes for the synthesis of (**13**) have been described [8, 40].

The synthesis of (*S*)-pinanediol (*R*)-(1-acetamido-4-bromobutyl)boronate (**17**) started from allyl bromide and catecholborane via (**14**), which was transesterified with (*S*)-pinanediol to (**15**), and then converted to the silylated aminoboronic ester (**16**) by treatment with acetic anhydride and acetic acid to form (**17**) (Scheme 4.3) [42].

The pinane amidoboronic esters (**18**, **19** and **20a**,**b**) could be synthesized by using similar chemistry (Scheme 4.4) [42]. Enzyme inhibition studies have shown that the D-amino acid analogs (**20a**,**b**) were active inhibitors of *Bacillus cereus* β-lactamase, with $K_i = 44$ and 49 μM at pH 7, respectively [43].



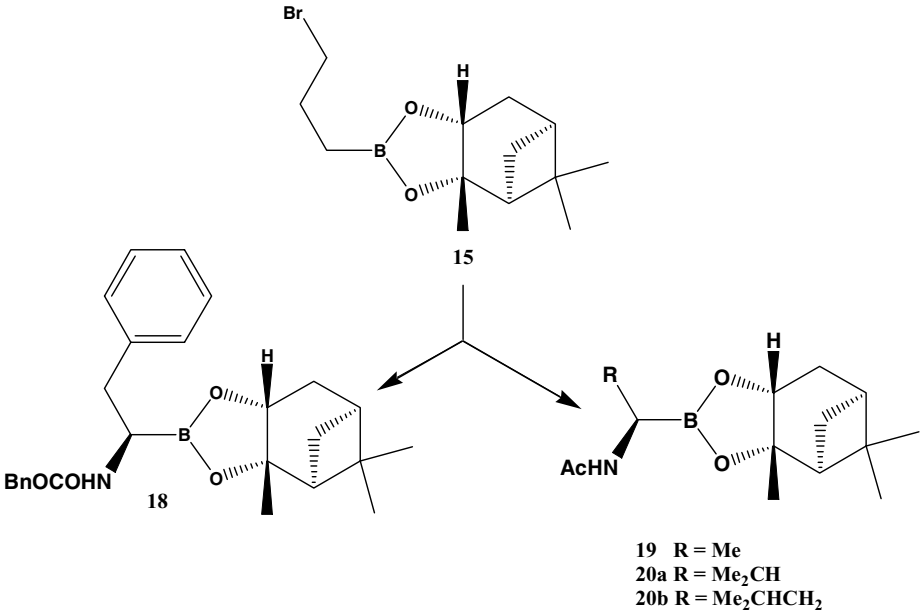
Scheme 4.2



Scheme 4.3

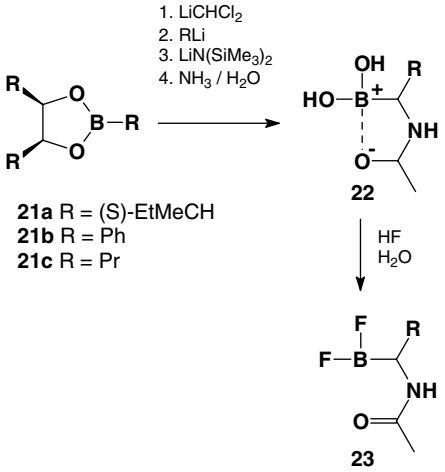
The racemic α-acetamidoboronic acids (**22**) have been obtained using similar chemistry. This reaction was used as the starting point for the corresponding *meso*-butanediol esters (**21**) (Scheme 4.5), and (**22**) were found to inhibit elastase and chymotrypsin [44]. The fluoro derivatives (**23**) could be obtained by the treatment of (**22**) with aqueous hydrofluoric acid [40].

(*S*)-Pinanediol (*S*)-(1-chloroallyl)boronate (**24**) reacted with LiHMDS and gave compound (**25**), which after desilylation/acetylation gave (*S*)-pinanediol (*R*)-(1-acetamidoallyl)boronate (**26**) (Scheme 4.6). Addition of methyl mercaptan to the double bond of (**26**) under ultraviolet light yielded the crystalline boronic ester (**27**) [45]. Treatment of (**27**) with BCl_3 led to (**28**), which was esterified by ethylene glycol to give the crystalline product (**29**).

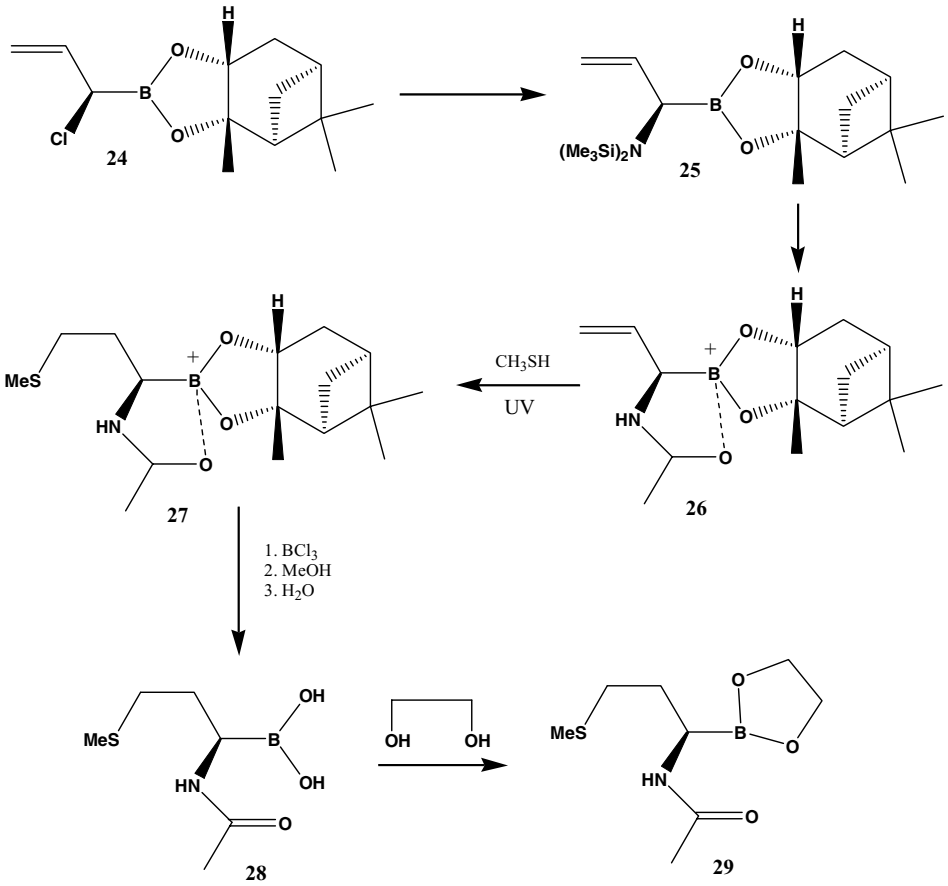


Scheme 4.4

Pinacol (iodomethyl)boronate reacts with tertiary amines to give the quaternary ammonium salts (**30**) (Scheme 4.7) and dibutyl (iodomethyl)boronate with morpholine forms (**31**) [39]. Many attempts to react (1-halo-2-phenylethyl)boronic esters with ammonia or ammonia derivatives failed. The reaction of (1-iodo-2-phenylethyl)-boronate (**32**) with aqueous ammonia gave 2-phenylethylamine (Scheme 4.7) [39].



Scheme 4.5

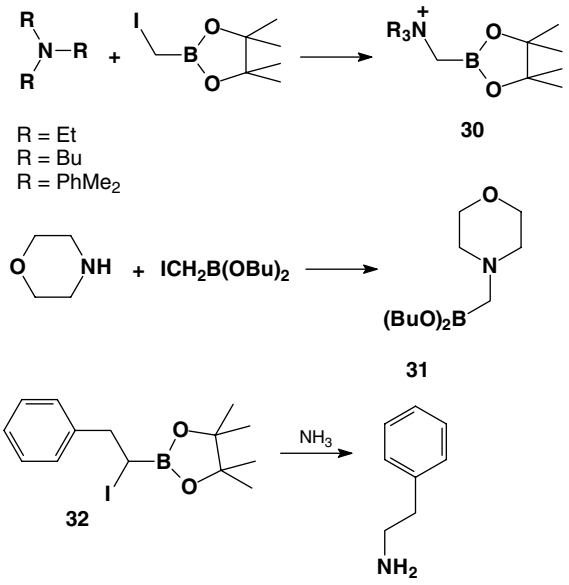


Scheme 4.6

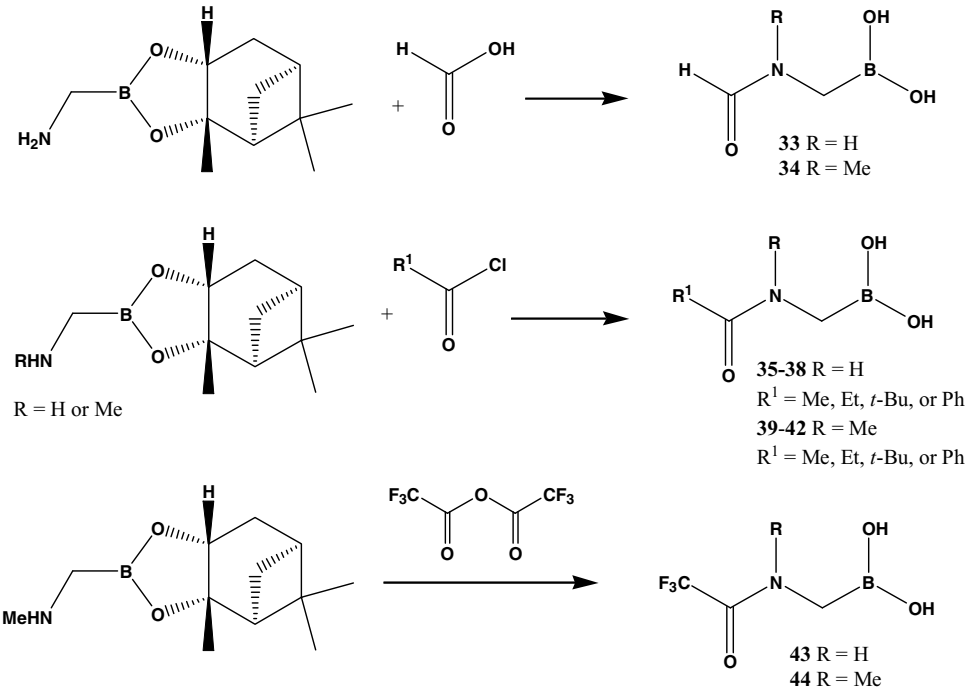
N-Acyl-boroGly and *N*-acyl-boroSar derivatives (**33–44**) were recently synthesized (Scheme 4.8) [46]. Structural characterization of these α-amidoboronic acids was accomplished by extensive use of ^{11}B- and ^{1}H-nuclear magnetic resonance (NMR) spectroscopy. These compounds were prepared to detect the extent of intramolecular B−O dative bond formation within the context of a five-membered (:O=C−N−C−B) ring motif. It was shown that the formation of such dative bonds depends on the nature of the substituents at both the acyl carbon and the nitrogen atoms. A proposed mechanism for alkali hydrolysis of (**39**) is shown in Scheme 4.9.

4.4
Asymmetric Synthesis via α-Haloalkylboronic Esters

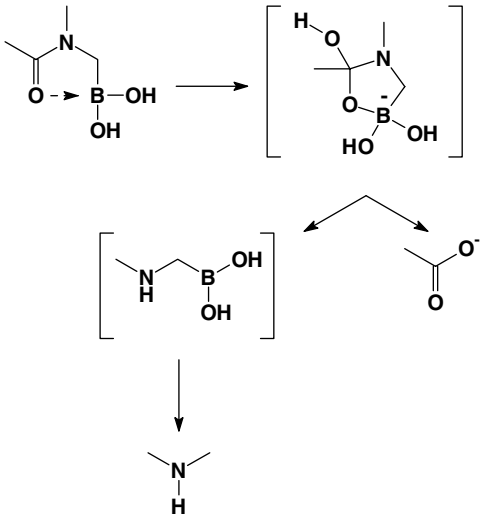
Free α-aminoboronic acids were synthesized and tested as potential enzyme inhibitors. The racemic boraalanine (**46**) was obtained in solution by hydrolysis of the



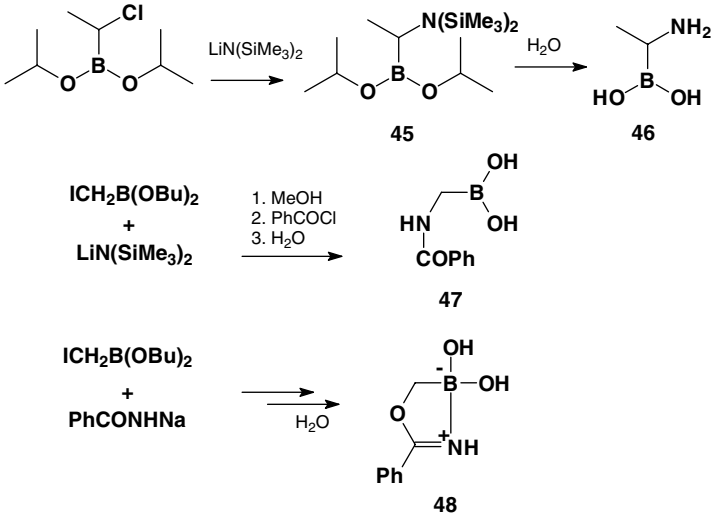
Scheme 4.7



Scheme 4.8

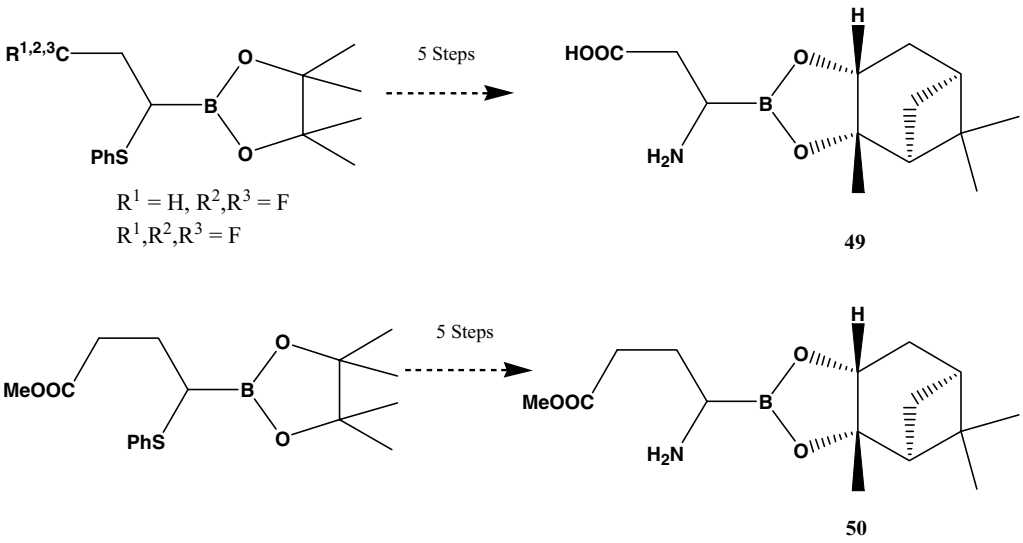


Scheme 4.9



Scheme 4.10

boronic ester (**45**) (Scheme 4.10). It is a good inhibitor of alanine racemase from *Bacillus stearothermophilus* with $K_i = 20$ mM (it was slow binding at $K_i = 0.15$–$0.35\ min^{-1}$). For D-alanine/D-alanine ligase from *Salmonella typhimurium*, two binding constants for different enzyme sites were found: $K_i = 35$ and 18 μM [37], respectively. Benz



Scheme 4.11

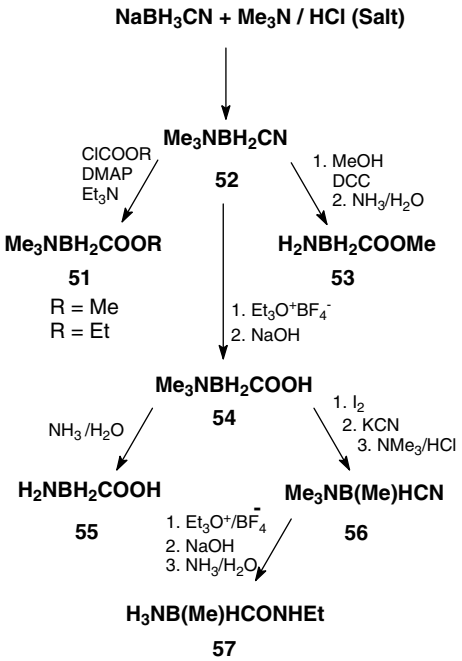
(amidomethyl)-boronic (**47**) acid was synthesized from dibutyl (iodomethyl)boronate and LiHMDS followed by benzoylation [47]. The product from reaction with sodiobenzamide was the imido ester isomer (**48**) [36].

α-Aminoboronic acid esters (**49** and **50**) were prepared and they are useful as inhibitors of the serine proteases, leukocyte and pancreatic elastases, cathepsin G, chymotrypsin, and hepatitis C virus (HCV) protease (Scheme 4.11) [48].

4.5 Synthesis of Glycine α-Aminoboronic Acids

In the simplest amino acid, glycine, replacement of the central methylene by boron, as depicted in (**55**), would give an isoelectronic and isostructural analog. The synthesis of (**55**) was performed by the reaction of sodium cyanoborohydride with trimethylammonium hydrochloride to give (**52**). Since direct hydrolysis of the nitrile group could not be achieved, its conversion to a carboxylic acid was performed in two steps comprising the a reaction with Meerwein's reagent, followed by alkaline hydrolysis of the intermediate nitrilium salt [49–51]. The metal-complexing capabilities and basicity of this betaine (**54**) have been described [52]. Upon displacement of the trimethylamine with a large excess of liquid ammonia, the desired glycine analog (**55**) was isolated [53]. The methyl ester (**53**) was formed from (**52**) in the presence of dicyclohexylcarbodiimide (DCC), which was then submitted to an amine exchange to give the glycine analog (**53**) [54] (Scheme 4.12).

Compound (**51**) can be obtained in varying yields upon reaction of (**62**) with chloroformates in the presence of triethylamine and 4-dimethyl-aminopyridine (DMAP). These conditions were found to be more advantageous than using

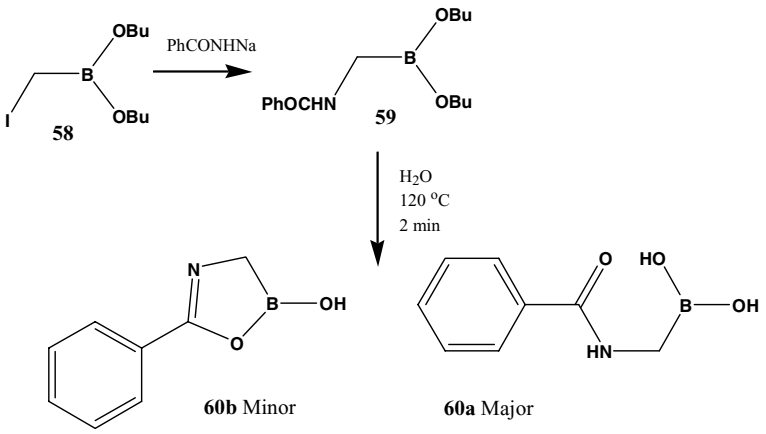


Scheme 4.12

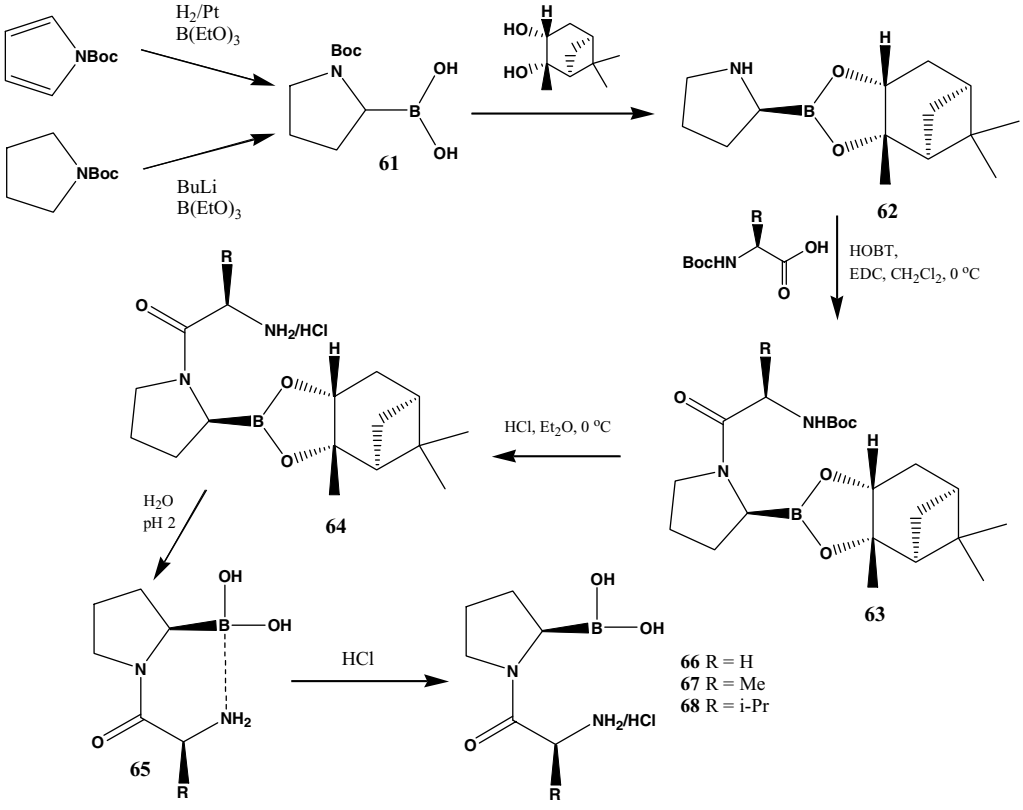
DCC [55]. Lithium aluminum hydride reduction of (**54**) followed by quenching with trimethylamine hydrochloride afforded (**56**). The conversion of (**54**) into (**55**) has also reported [55]. The action of Meerwein's reagent followed by basic hydrolysis and amine exchange gave the desired amide (**57**) [56]. The initial synthetic efforts to obtain such amino acid-based inhibitors used *N*-acylated analogs of glycine. As an example, dibutyl iodomethane-boronate (**58**) was alkylated with the sodium salt of benzamide to give (**59**) (Scheme 4.13) [47], which was shown to be a potent inhibitor of α-chymotrypsin. Hydrolysis of (**59**) gave two isomers (**60a**,**b**).

4.6 Synthesis of Proline α-Aminoboronic Acids

A series of prolineboronic acid-containing dipeptides was synthesized and assayed for their ability to inhibit the serine protease dipeptidyl peptidase IV (DPPIV) [57]. The synthesis of boroproline (**61**) was developed by Matteson's procedure [42, 58] for the preparation of aminoboronic acids. The amino boronic ester (**62**) was coupled with the desired Boc-amino acids in the presence of *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide (EDC) hydrochloride to generate the fully protected dipeptides (**63**), and treatment with HCl gave proline α-aminoboronic acids (**66–68**) via intermediates (**64**) and (**65**) (Scheme 4.14). The abilities of the peptides to inhibit the enzyme [57] are shown in Table 4.1.

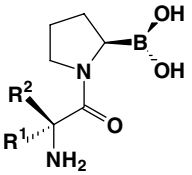


Scheme 4.13



Scheme 4.14

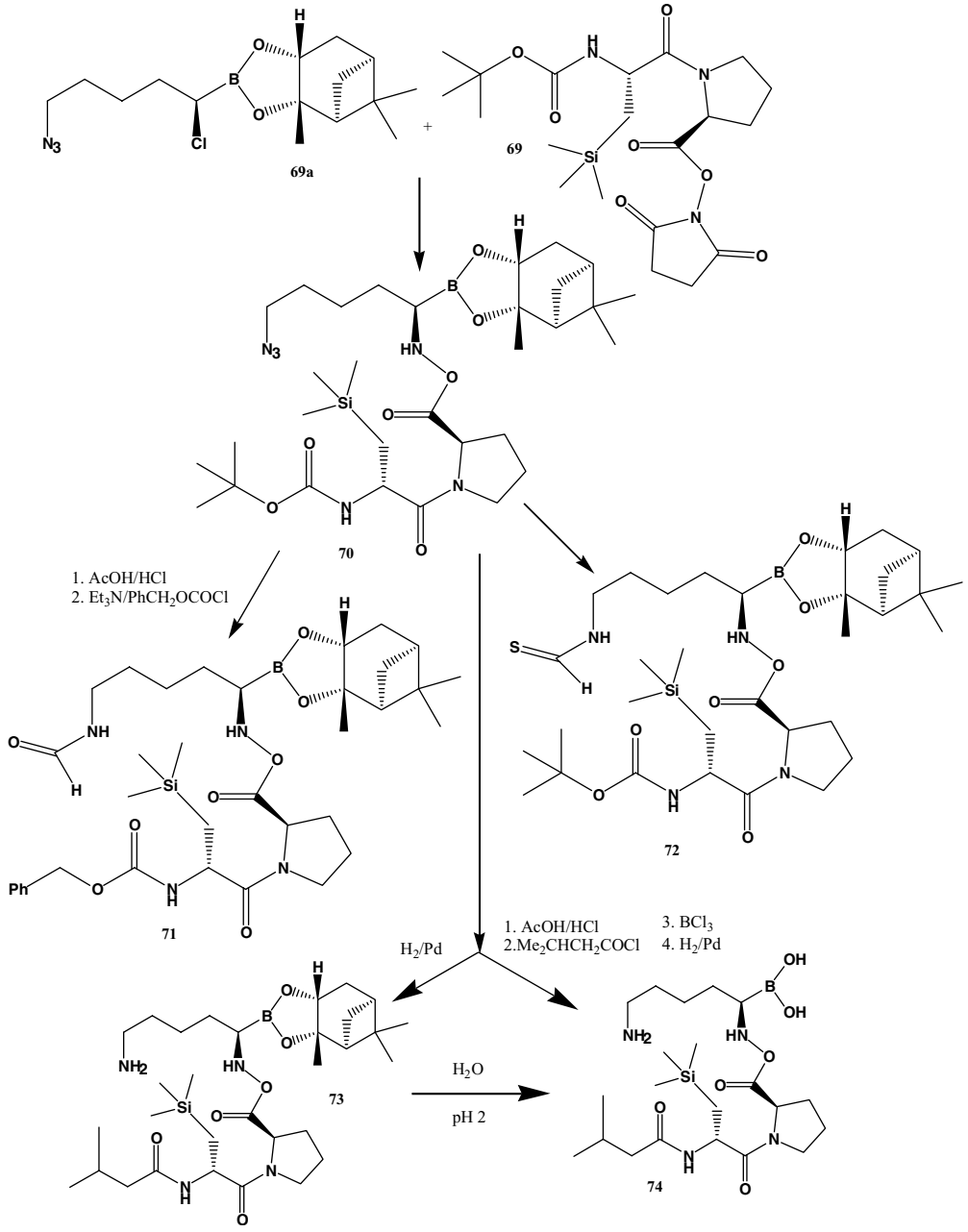
Table 4.1 Structure–activity relationships for H_2N–X_{aa}-boroPro dipeptides versus DPPIV.



67

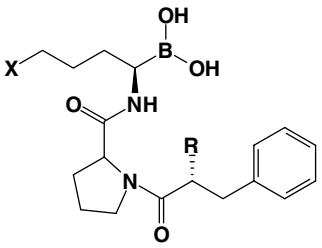
Compound	Amino acid	R^1	R^2	Boron configuration	IC_{50} (nM)	+SE (nM)
67a	L-Val	H	*i*Pr	*R*	26	1
67b	L-Val	H	*i*Pr	*S*	4000	600
67c	D-Val	*i*Pr	H	*R*	116000	15000
67d	L-Ala	H	Me	*R*	15	3
67e	Aib	Me	Me	*R*	30000	8000
67f	L-Gly	H	H	*R*	16000	2400
67g	L-Abu	H	Et	*R*	11	1
67h	L-Leu	H	*i*Bu	*R*	44	2
67i	L-Ile	H	2-Bu	*R*	25	1
67j	L-*t*Leu	H	*t*Bu	*R*	60	7
67k	L-Phe	H	CH_2Ph	*R*	70	7
67l	L-Phg	H	Ph	*R*	63	5
67m	L-Tyr	H	CH_2(Ph-4-OH)	*R*	32	1
67n	L-Lys	H	$(CH_2)_4NH_2$	*R*	95	19
67o	L-Thr	H	CH_3CHOH	*R*	190	13
67p	L-Pro	H	-$(CH_2)_3$-	*R*	20	5
67q	L-Azet	H	-$(CH_2)_2$-	*R*	250	13
67r	L-His	H	CH_2Im	*R*	17000	1800

A series of Boc-D-trimethylsilylalanine–proline–boro-X pinanediol derivatives (**73**) and (**74**) that are active as thrombin inhibitors has been synthesized by von Matt *et al.* [59, 60] (Scheme 4.15). All of the thrombin inhibitors were synthesized starting from the common intermediate (**70**), which was obtained by the reaction of (**69a**) with (**69**). Hydrogenation of the azido group of (**70**) catalyzed by palladium on carbon led to the borolysine derivatives [61]. The free amino group could be derivatized by reaction with an anhydride, acid chlorides, isocyanates, and/or potassium cyanate. The thioformamide (**72**) could be obtained by treatment of the corresponding formamide with Lawesson's reagent. Removal of the Boc protecting group followed by reaction with benzyl chloroformate led to the compound (**71**). Hydrolysis of (**73**) generated (**74**). The X-ray structure analysis of (**71**) and (**74**) have shown that these inhibitors bound to the active side of thrombin.


69a
69
70
1. AcOH/HCl
2. Et3N/PhCH2OCOCl
71
72
H2/Pd
1. AcOH/HCl
2.Me2CHCH2COCl
3. BCl3
4. H2/Pd
73
H2O
pH 2
74

Scheme 4.15

The synthesis of the thrombin inhibitor DuPont 714 (**75**) was achieved starting from 3-bromopropylboronic ester [40]. This enzyme inhibitor was active in rabbits at dose levels of 0.1 mg/kg/h [62, 63]. A methoxy group in place of guanidine on (**75**) also provides a potent thrombin inhibitor [40]. A series of conformationally restricted boropeptide thrombin inhibitors (**76**) and (**77**) have also been synthesized [64].

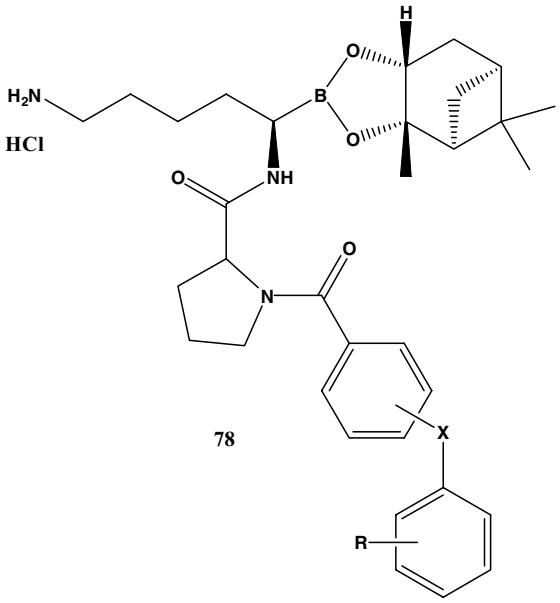


75 Du Pont 714: R = NHAc
X = $NHC(NH)NH_2 \bullet$ HCl
76 R = NHAc, X = $CH_2NH_2 \bullet$ HCl
77 R = H, X = $CH_2NH_2 \bullet$ HCl

The potent binding affinity of the resulting inhibitors (**78**), such as (**78f**), may be due in part to a unique mode of binding at the thrombin active site. The thrombin-binding activity data for a series of inhibitors are shown in Table 4.2. The synthesis of these inhibitors involved the initial preparation of the P_3 benzoic acids followed by elaboration to their corresponding boropeptides [64].

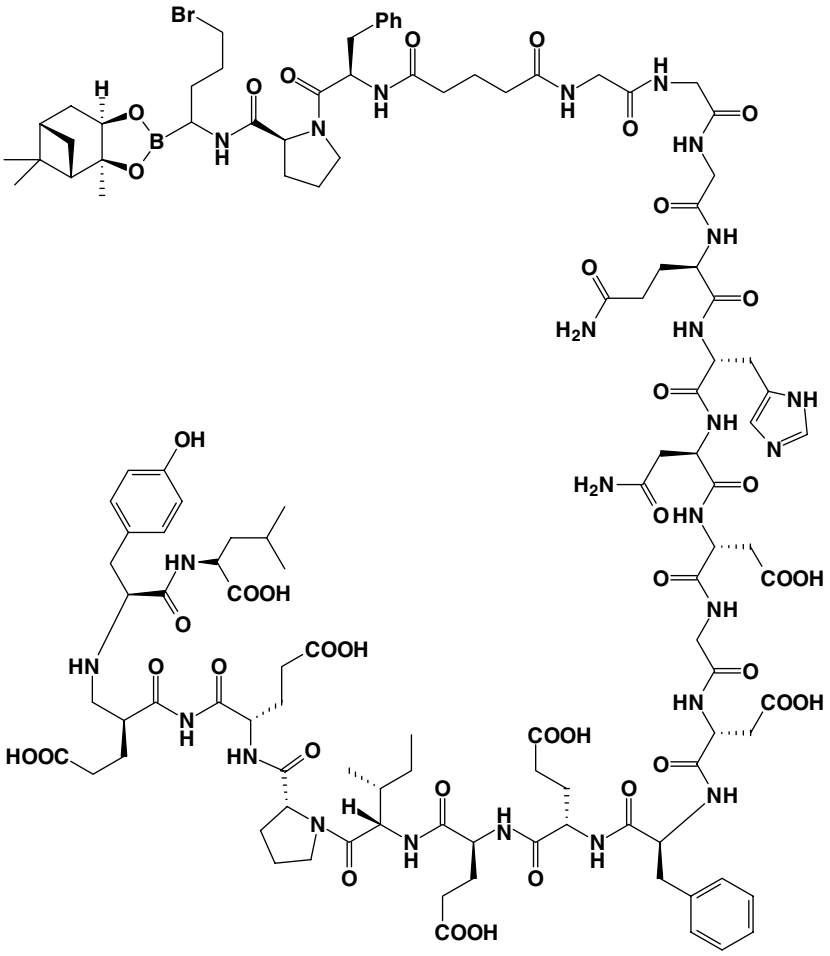
Bifunctional serine protease inhibitors of boron-containing peptides were prepared [65]. These serine protease inhibitors comprise a catalytic site-directed moiety, which binds to and inhibits the active site of a serine protease, and an exosite-associating moiety, which are joined by a connector. The catalytic site-directed moiety and the exosite-associating moiety are capable of binding simultaneously to a molecule of the serine protease. Thus, boron-containing peptide conjugate (**79**) was prepared by standard solid-phase methods using Fmoc chemistry to assemble the hirudin segment, followed by amidation with glutaric anhydride and peptide coupling with the boron-containing tripeptide segment. Compound (**79**) inhibited human α-thrombin with $K_i = 0.649$ nM in an *in vitro* assay.

Table 4.2 Boropeptide thrombin inhibitors with benzoic acid-derived residues.



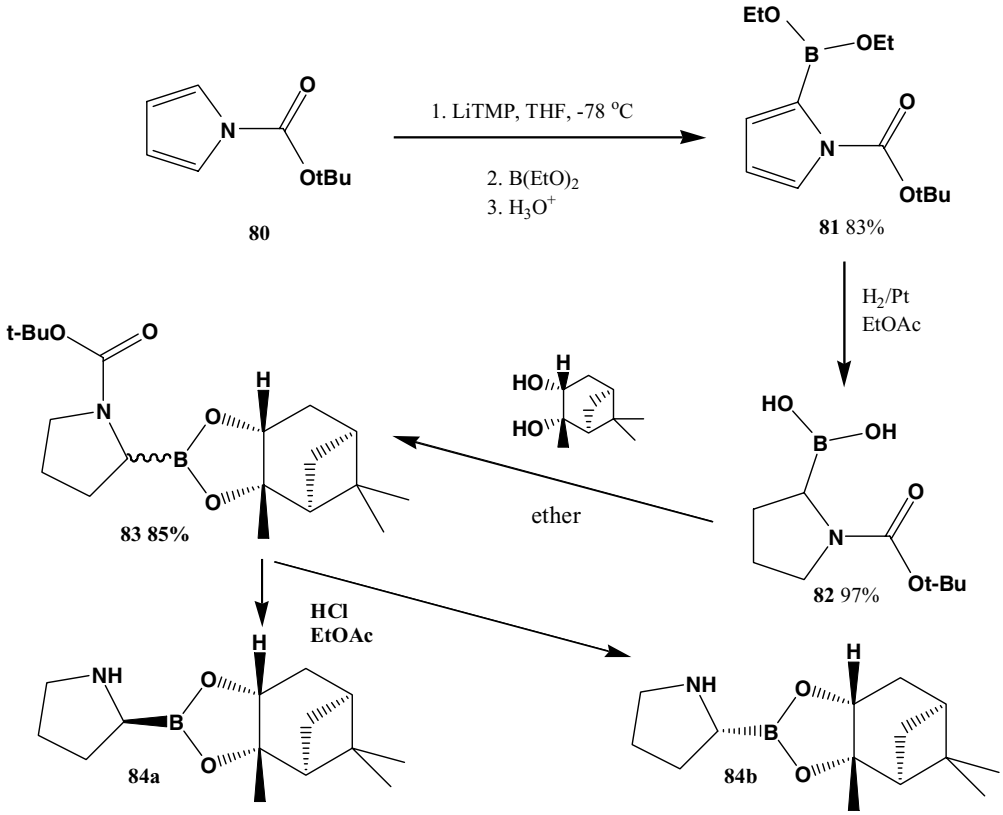
78

Compound	X	R, position	K_i (nM)
78a	CH_2	H, *ortho*	0.29
78b	CH_2	H, *meta*	0.19
78c	CH_2	H, *para*	1.80
78d	O	H, *ortho*	0.27
78e	O	H, *meta*	0.36
78f	CH_2	2-CF_3, *meta*	0.07
78g	CH_2	2-CH_3, *meta*	0.25
78h	CH_2	2-SCH_3, *meta*	0.50
78i	CH_2	2-Br, *meta*	0.23
78j	CH_2	3-F, *meta*	0.43
78k	CH_2	3-CF_3, *meta*	0.16
78l	CH_2	4-CF_4, *meta*	0.22
78m	CH_2	3,4-(-OCH_2-O-), *meta*	0.09
78n	S	H, *meta*	<0.1
78o	S	2-CF_3, *meta*	0.45
78p	S	2-OCH_3, *meta*	0.19
78q	S	4-OCH_3, *meta*	0.42
78r	SO_2	H, *meta*	0.85
78s	SO_2	2-OCH_3, *meta*	0.58



79

The synthesis of the pinanediol ester of prolineboronic acids was described by Kelly *et al*. (Scheme 4.16) [66]. Boc-pyrrole (**80**) after treatment with lithiotetramethylpiperidide gave Boc-pyrrole-2-boronic acid (**81**). Hydrogenation generated Boc-prolineboronic acid (**82**), which was easily esterified with (1*S*,2*S*,3*R*,5*S*)-(+)-pinanediol to give (**83**). The compound (**83**) was deprotected by HCl in EtOAc to form two diastereomers (**84a**) and (**84b**). Diastereomers were separable by high-perofrmance liquid chromatography, thus providing a means to resolve chiral center α- to boron. The diastereomeric purity of **84a** and **84b** could be assayed by intergrating the ^{1}H-NMR spectra of these compounds obtained in C_6D_6. In this solvent the resonances of the methine protons α to the oxygen in the pinane system are clearly discernible AB-quartets appearing at δ 4.33 ppm (**84b**) and δ 4.25 ppm (**84a**). Due to conformational restriction in these molecules, each stereoisomer contains 30% of a minor rotamer. The resonances for the minor rotamers are observed at δ 4.01 ppm and δ 4.11 ppm for **84b** and **84a** respectively. This explanation is supported by a variable-temperature NMR experiment on **84b**.



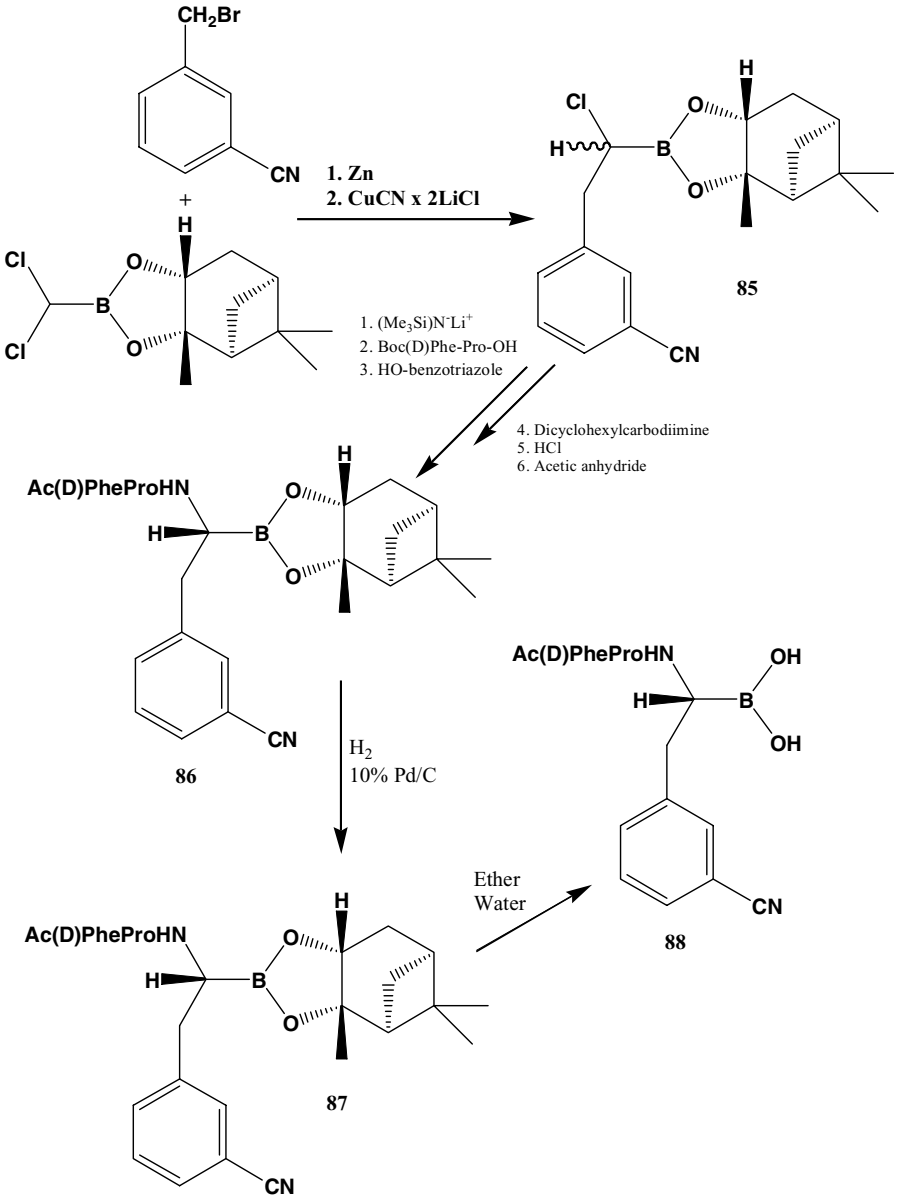
Scheme 4.16

4.7 Synthesis of Alanine α-Aminoboronic Acids

Novel highly effective thrombin inhibitors have been obtained by preparing boronic acid analogs (**85**–**87**) of the *m*-cyanoborophenylalanine analog (**86**) [67a] (Scheme 4.17). The free boronic acid (**88**) was isolated from the aqueous phase as a single component. The interaction of trypsin with a series of inhibitors (**89**) is shown in Table 4.3.

Potent and selective dipeptidyl boronic acid inhibitors have been demonstrated by Adams *et al.* [67]. The synthesis of the boronated dipeptide (**90**) was described in a US patent [68] (see Scheme 4.18). It was found that the boronic acid derivative (**91**) has inhibition properties as indicated in Table 4.4.

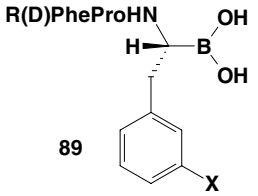
The solid-phase synthesis of the aminoboronic acids (**93**) and (**94**) – potent inhibitors of the HCV NS3 proteinase – was demonstrated by Dunsdon *et al.* (Scheme 4.19) [69] using well known peptide boronic acid derivatives (**92**). [70, 71] HCV is the cause of the majority of cases of transfusion-associated hepatitis. The target which requires new compounds for antiviral therapy against HCV is the NS3 serine proteinase [72]. The amidoboronic acids (**93**) and (**94**) were found [73] to be highly potent inhibitors of the HCV NS3 proteinase.



Scheme 4.17

Enantiomeric 1-acetamidoboronic acids, which are *N*-acetyl *trans*-state inhibitor analogs of the L- and D-forms of the amino acids alanine, phenylalanine, *p*-fluorophenylalanine, *p*-chlorophenylalanine, and 1-naphthylalanine were synthesized (Scheme 4.20), and tested as inhibitors of the serine proteases subtilisin Carlsberg and α-chymotrypsin [74]. All L-(*R*)- and D-(*S*)-1-acetamidoboronic acids were prepared according to the basic strategy developed by Matteson *et al.* [35, 45, 75]. The pinanediol esters (**95**) gave the α-chloroboronic acids (**96**) in 75–95% yields with

Table 4.3 Binding of peptide boronic acids to serine proteases.



89

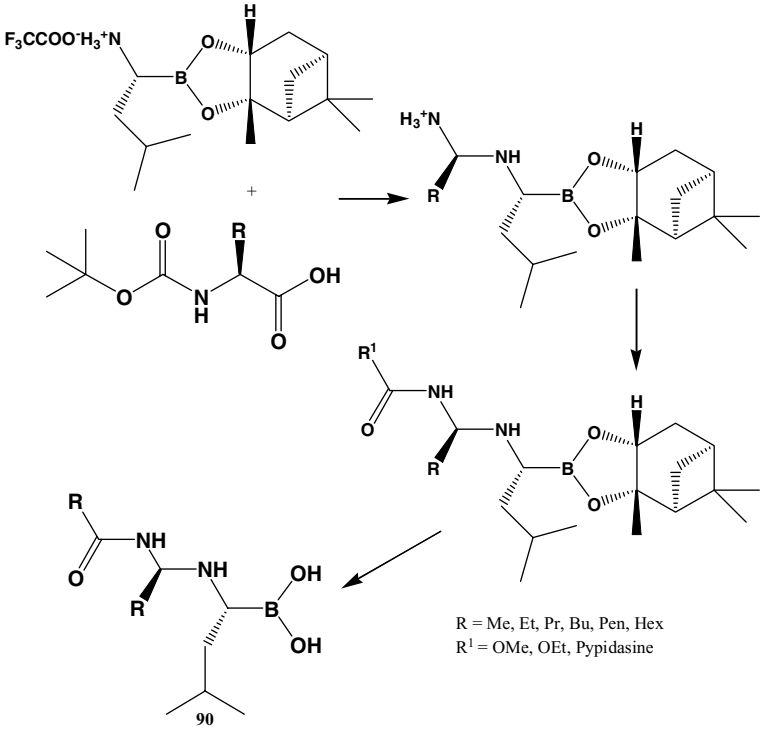
Protease	R	X	K_i for X (nM)
Thrombin	Ac	CN	0.79
Thrombin	Boc	CN	0.51
Thrombin	H	CN	0.48
Thrombin	Ac	CH_2NH_2	4.80
Trypsin	Ac	CN	130.0
Trypsin	Boc	CN	73.0
Trypsin	H	CN	5.6
Trypsin	Ac	CH_2NH_2	5.6
Pancreatic	Ac	CN	8600
Chymotrypsin	Ac	CN	49.0
Thrombin	Ac	H	320
Thrombin	Boc	H	59
Thrombin	H	H	6.3
Thrombin	Ac	H	ND
Trypsin	Ac	H	11000
Trypsin	Boc	H	5400
Trypsin	H	H	1800
Trypsin	Ac	H	ND
Pancreatic	Ac	H	39000
Chymotrypsin	Ac	H	100

ND, not detected.

diastereoselectivites greater than 98%. Treatment the α-chloroboronic acids (**96**) with LiHMDS afforded the corresponding silylated aminoboronic esters, which when heated with Ac_2O and AcOH formed the 1-acetamidoboronic esters (**97**). Hydrolysis of (**97**) with boron trichloride gave the 1-acetamidoboronic acids (**98**). Both the anhydride forms of (**98**) and the diethanolamine derivatives (**99**) were hydrolyzed to the corresponding free boronic acids (**100a–e**). All of the boronic acids (**100a–e**) are powerful competitive inhibitors of both enzymes.

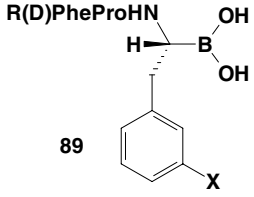
4.8 Synthesis of Ornithine α-Aminoboronic Acids

The asymmetric syntheses of (*R*)-1,4-diaminobutane-1-boronic acid dihydrochloride (**106**) and the aminoboronic acid analog of L-ornithine have been described [76]

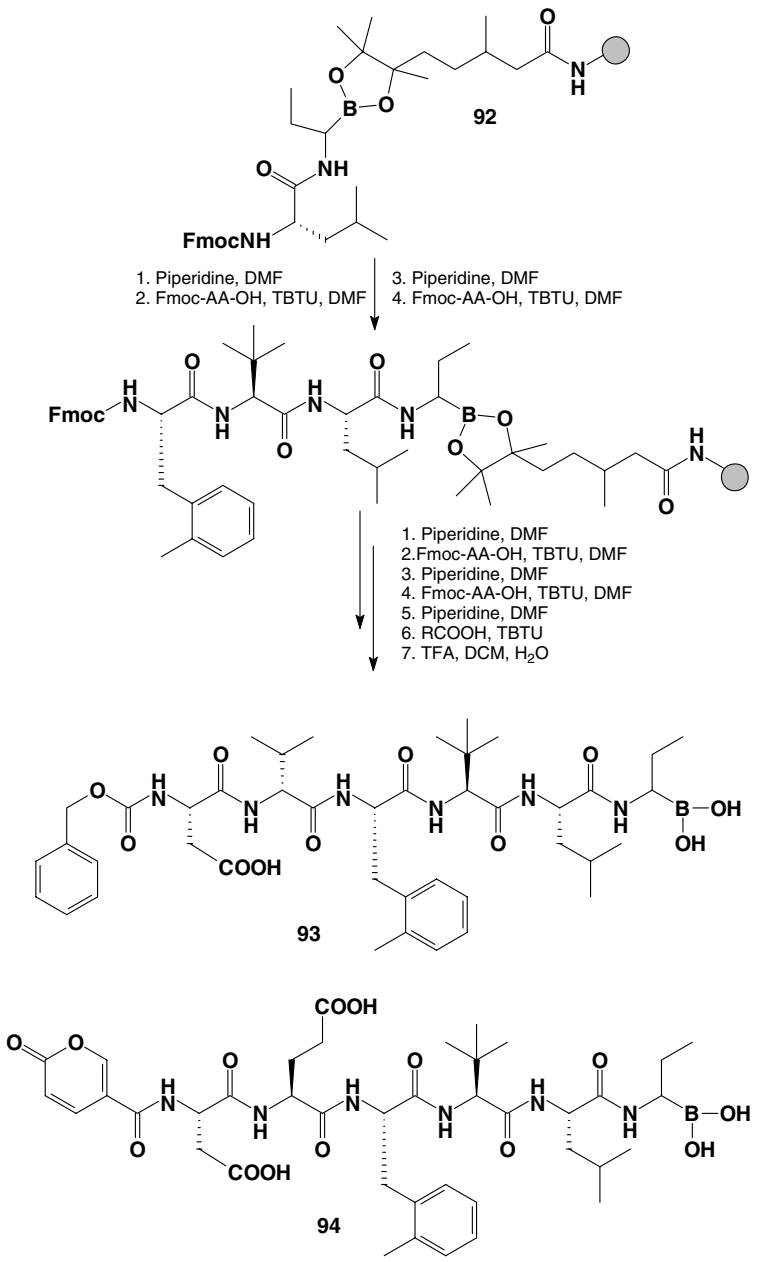


Scheme 4.18

Table 4.4 Enzyme inhibitory profile of the peptide boronic acid derivative (**91**).

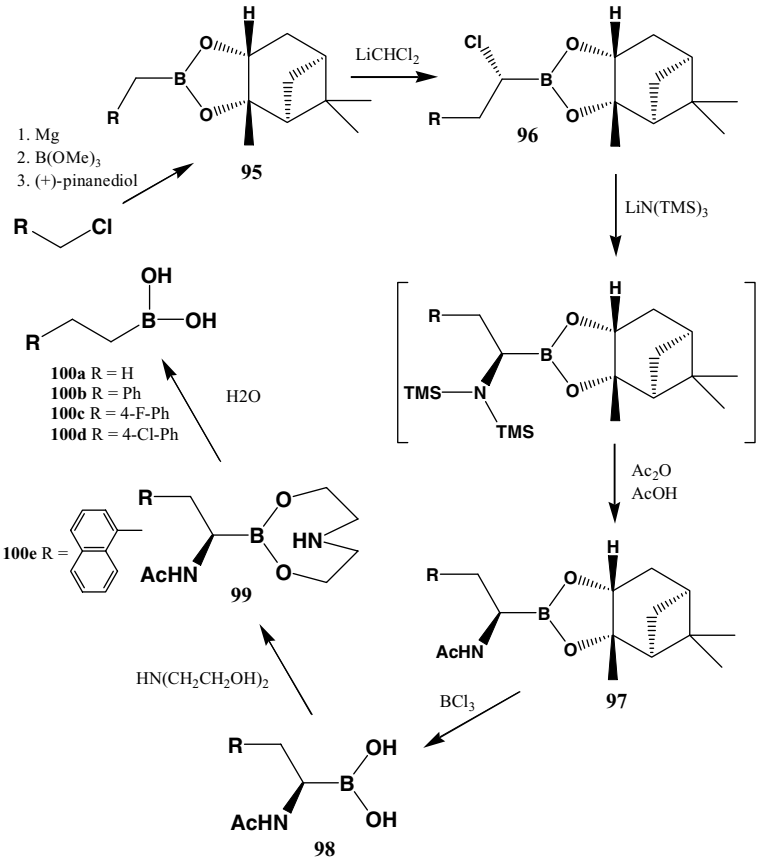


Enzyme	K_i (nM)
20S protease	0.62
Human leukocyte elastase	2300
Human cathepsin G	630
Human chymotrypsin	320
Thrombin	13000



Scheme 4.19

(Scheme 4.21). The 3-azidopropaneboronic ester (**101**) was obtained from allyl bromide and converted to the optically active (+)-pinanediol derivative (**102**), which could be transformed to two compounds (**103**) and (**104**). Attempts to obtain the (*R*)-1,4-diaminobutane-1-boronic acid (**108**) from (**105**) and (**106**) were unsuccessful.

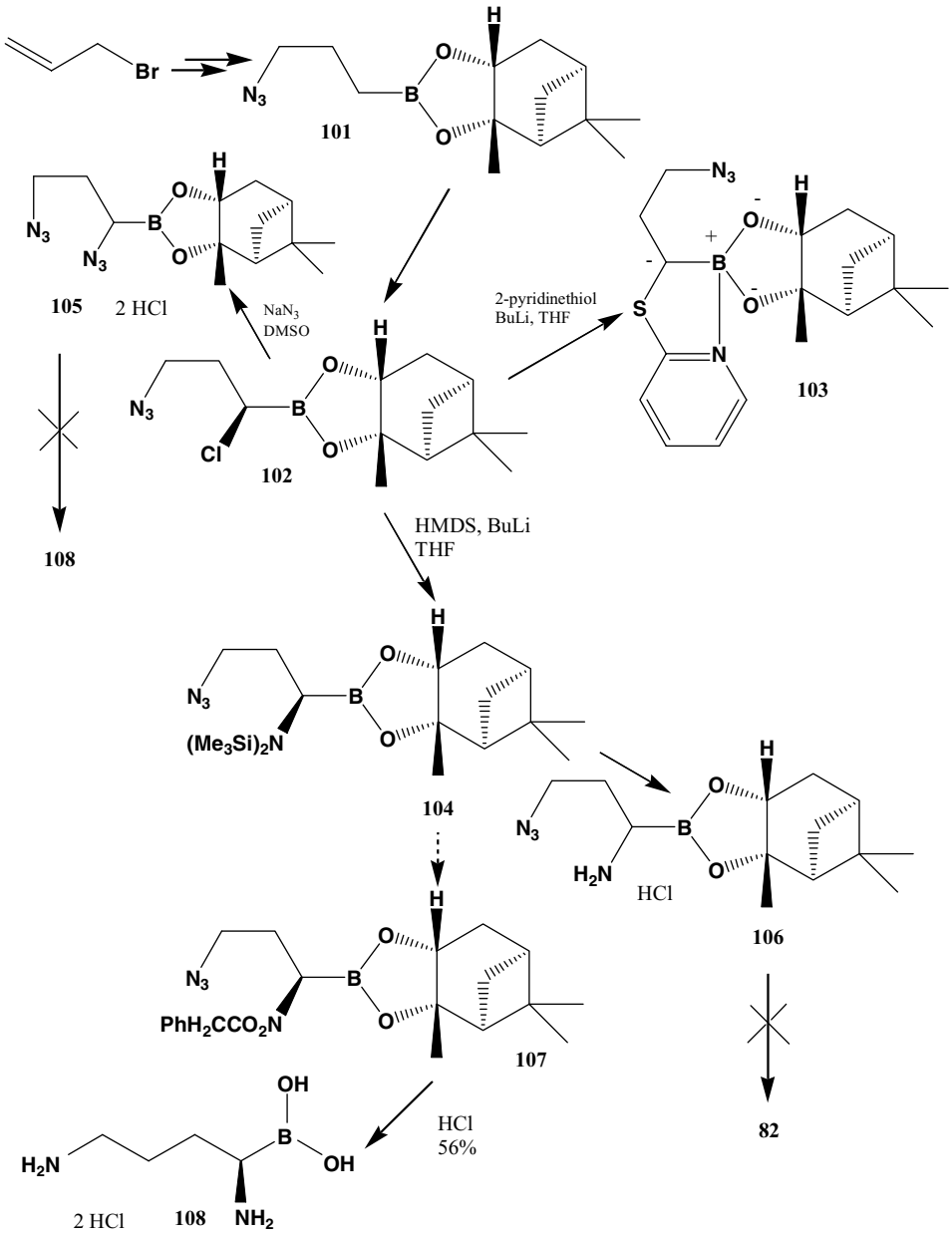


Scheme 4.20

The *N*-protective group in (**104**) could be desilylated when treated with benzyl chloroformate to give the monoprotected boroornithine derivative (**107**), which should be a valuable precursor of the arginine boronic acid analogs. Peptides containing C-terminal boronic acid derivatives of ornithine, lysine, arginine, or homoarginine and corresponding isothiuronium analogs are reversible inhibitors of trypsin-like serine proteases such as thrombin, plasma kallikrein, and plasmin, and are useful in treatment of blood coagulation disorders and inflammation [77].

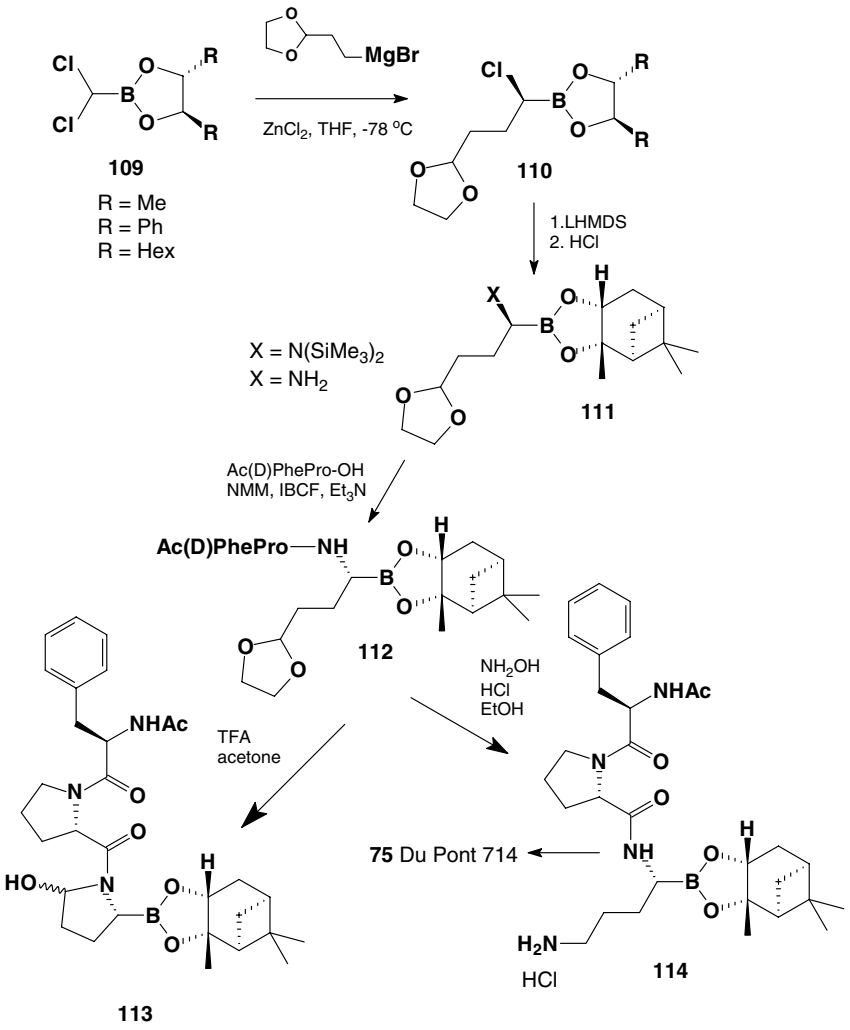
4.9
Synthesis of Arginine α-Aminoboronic Acids

Peptides such as DuPont 714 (**75**), containing boroarginine at the primary residue are potent thrombin inhibitors [78]. The synthesis of boropeptide (**114**) containing a basic α-aminoboronic ester started from the dichloromethylboronic ester (**109**) [79],



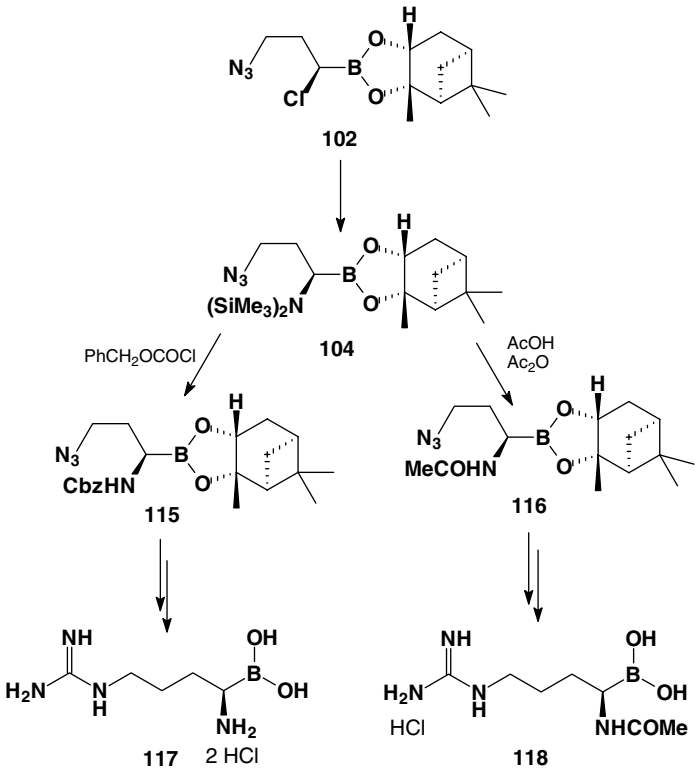
Scheme 4.21

and addition of the Grignard to (**109**) to give (**110**). Reaction with (+)-pinanediol then provided the boronic ester (**111**). The boropeptide (**112**) was obtained in 80% yield and it was transformed to (**113**) and (**114**). DuPont 714 (**75**) obtained from (**114**) was stable in water (Scheme 4.22).



Scheme 4.22

The asymmetric synthesis of an unprotected α-aminoboronic acid analog of L-arginine (**104**) and its *N*-acetyl derivatives (**115** and **116**) provided alternative substrates or inhibitors of nitric oxide synthase (Scheme 4.23) [80]. Nitric oxide displays activities in the cardiovascular system as well as in the central and peripheral nervous systems and has considerable attracted in the past few years [81, 82]. The general synthetic sequence is based on the asymmetric methodology developed by Matteson [40]. In order to introduce the amino group in the α-position of boron atom, hexamethyldisilazane (HMDS) was used to displace the chloride in compound (**102**) of (**117**), and (**118**) were obtained using benzyl-chloroformate and a mixture of AcOH and Ac_2O, respectively.



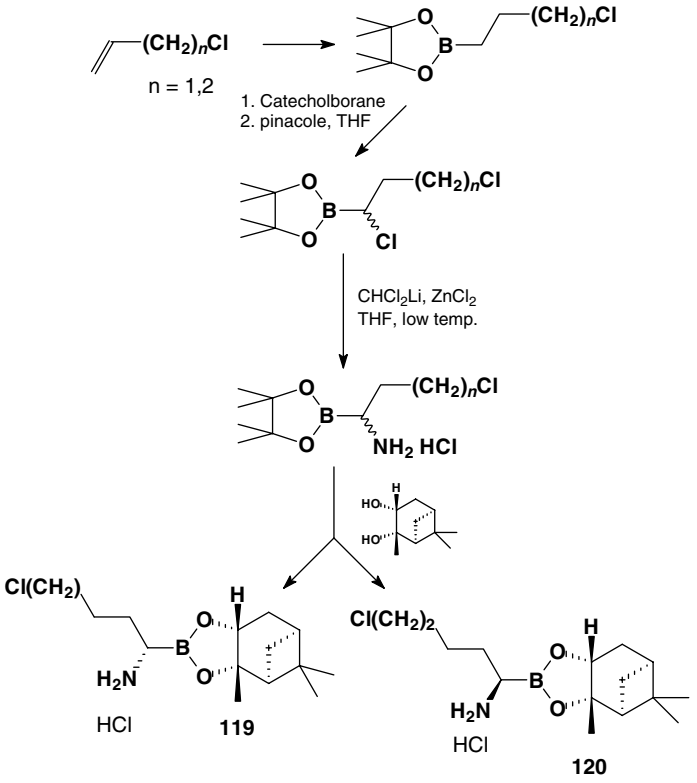
Scheme 4.23

Diastereoselective crystallization with (+)- and (−)-pinanediols was demonstrated by Brosz *et al.* [83]. Using this method, the two stereoisomers (**119**) and (**120**) (Scheme 4.24) could be obtained.

Some novel α-aminoboronic acids (**121–124**) which act as serine protease inhibitors have been synthesized [84] (Scheme 4.25). The methodology affords α-amino boronic acids (**121–124**) with the general formula (*R*)-NHCH(R)BO_2-pinanediol where R = CH_2CHF_2, CH_2CO_2tBu, and $(CH_2)_2$COMe. In the peptide Pz-CO-Val–Val–Hyp(Bzl)-OH, Pz is pyrazine and Hyp(Bzl) is 4-hydroxyproline with the hydroxyl group protected as a benzyl ether.

4.10
Synthesis of Phenethyl Peptide Boronic Acids

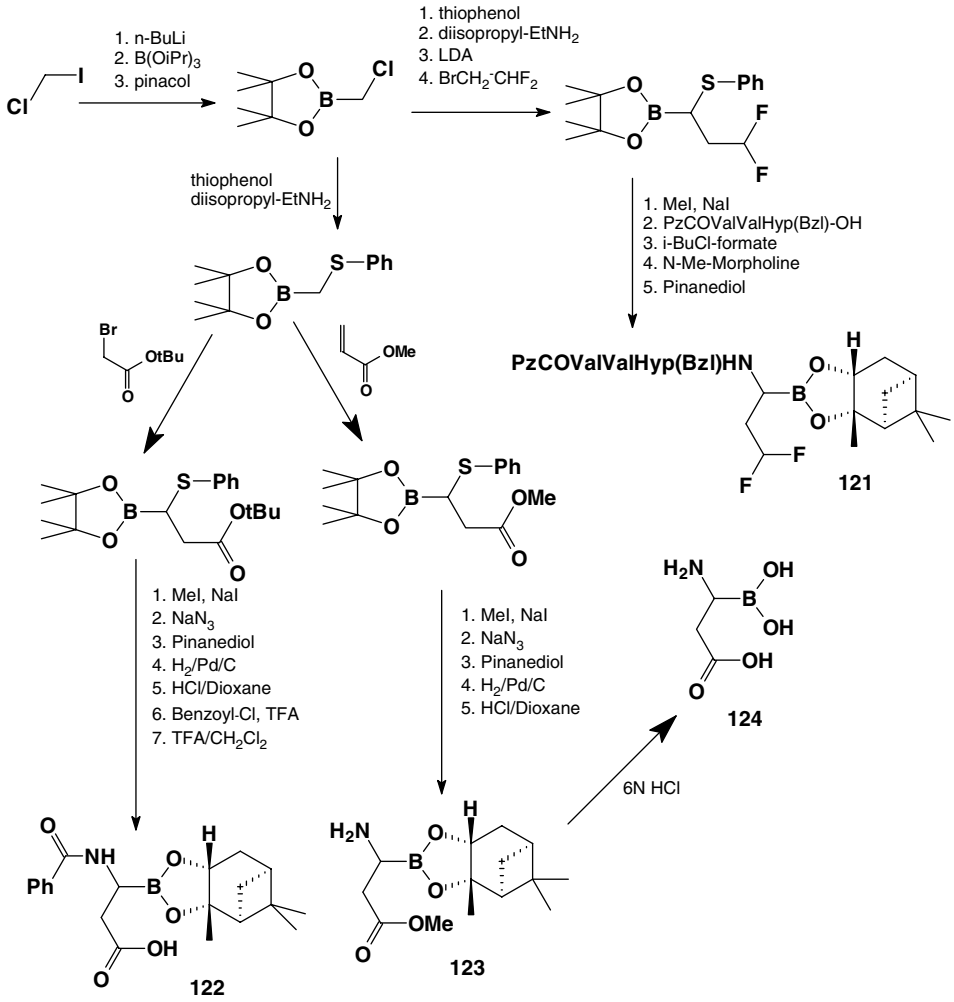
A series of peptide boronic acids containing extended, hydrophobic P_1 residues have been synthesized to probe the shallow, hydrophobic S_1 region of HCV NS3 protease [85]. Peptide boronic acid inhibitors were synthesized using the methodology for asymmetric homologation of boronic acid pinanediol esters developed by



Scheme 4.24

Matteson *et al.* [86]. Reaction of a Grignard reagent with triisopropyl borate, followed by esterification with (+)-pinanediol affords a boronic ester as shown in Scheme 4.26. The reaction was done at −100 °C. Homologation with dichloromethyllithium [87] diastereoselectively provides an (*S*)-α-chloroboronic ester. Displacement of chloride by lithium bis(trimethylsilyl)amide [41] followed by acidolysis gives the (*R*)-α-aminoboronic ester as a stable hydrochloride salt. Coupling to the protected pentapeptide Boc-Asp(O*t*Bu)–Glu(O*t*Bu)–Val–Val–Pro-OH and deprotection with trifluoroacetic acid afforded the desired peptide boronic acids (**125a–l**). The inhibition activity of peptide boronic acids (**125a–l**) has been demonstrated against NS3 protease, human leukocyte elastase, and human pancreatic chymotrypsin, as shown in Table 4.5.

A series of substituted phenethyl containing peptides was prepared as shown in Scheme 4.27. In this case, the required boronates were prepared by hydroboration of a substituted styrene with catecholborane, followed by transesterification with (+)-pinanediol [85, 88]. Subsequent homologation, nitrogen substitution, and peptide coupling afforded hexapeptides (**126a–u**). Inhibition of NS3 protease, human leukocyte elastase, and human pancreatic chymotrypsin by P_1 phenethyl peptide boronic acids (**126a–u**) was observed [85] (Table 4.6).

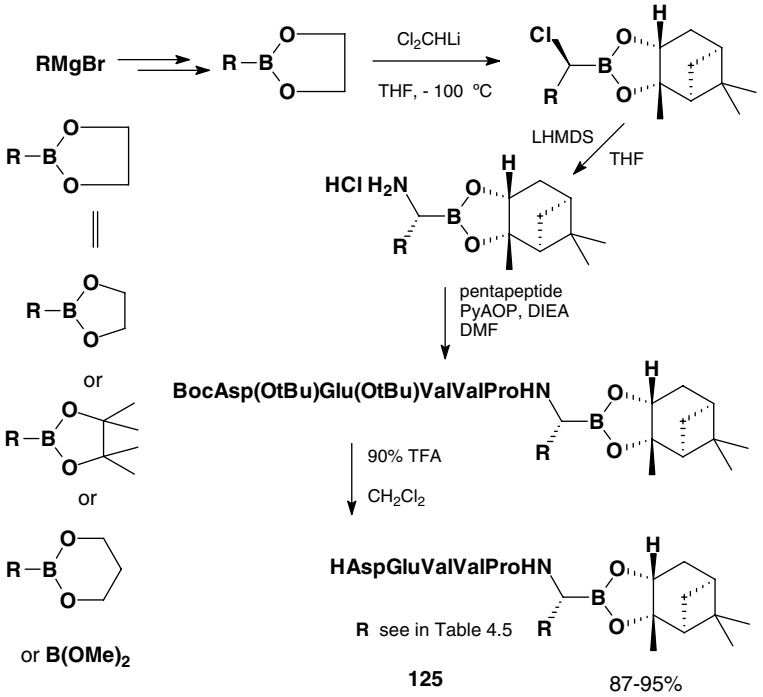


Scheme 4.25

Within the P1 phenethyl series, substantial effects on inhibitor potency and selectivity were observed with changes in the position and identity of the aromatic ring substituents. The 4-trifluoromethylphenethyl (**126l**) P_1 was identified as optimal with respect to inhibitor potency for NS3 and selectivity against elastase and chymotrypsin [85].

4.11
Synthesis via Zirconocene Species

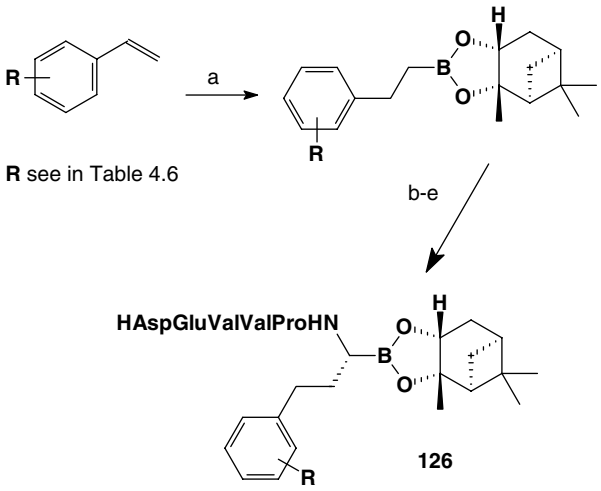
α-Aminoboronic esters could be synthesized via zirconocene species [89, 90]. Conversion of organozirconium compounds to amines from organozirconocene



Scheme 4.26

Table 4.5 Inhibition of NS3 protease, human leukocyte elastase, and human pancreatic chymotrypsin by peptide boronic acids (**125a–l**).

Compound	R	NS3 K_i (μM)	Elastase IC_{50} (μM)	Chymotrypsin IC_{50} (μM)
125a	ethyl	0.008	0.020	>60
125b	*n*-butyl	0.011	NT	2.1
125c	*n*-pentyl	0.012	NT	0.38
125d	*n*-hexyl	0.013	NT	0.42
125e	*i*-butyl	0.008	0.060	NT
125f	*i*-amyl	0.039	NT	0.30
125g	4-methylpentyl	0.007	7.3	0.28
125h	phenyl	0.900	NT	NT
125i	benzyl	0.500	NT	0.070
125j	phenethyl	0.080	3.5	0.075
125k	phenpropyl	0.200	NT	NT
125l	phenbutyl	0.010	0.4	1.9



Scheme 4.27 Reagents and conditions: (a) catecholborane, 70 °C; (+)-pinanediol, THF; (32–82%); (b) Cl_2CHLi, THF, −100 °C; $ZnCl_2$, −100 to 25 °C; (14–92%); (c) LiHMDS, THF, −78 to 25 °C; 4 N HCl/dioxane, −78 °C; (10–95%); (d) Boc-Asp(OtBu)–Glu (OtBu)–Val–Val–Pro-OH, 7-azabenzotriazol-1-yloxy)tris(pyrrolidino)-phosphonium hexafluorophosphate (PyAOP), *N*,*N*′-diisopropylethylamine (DIEA), *N*,*N*-dimethylformamide (DMF), or EDC, HOAt, $NaHCO_3$, CH_2Cl_2/DMF (5 : 1), 0 °C; (8–56%); (e) 90% trifluoroacetic acid (TFA), 5% triisopropylsilane, 5% CH_2Cl_2; (87–95%).

chlorides has been observed [89] and, of the various electrophilic aminating reagents available for reaction with organometallic compounds [89], use of the *O*-sulfonylhydroxylamines (MSHs) was found to be the most effective. They are readily available from easily accessible starting materials in a number of high-yielding steps [91, 92]. MSH has been shown to be superior to the other agents in terms of solubility in organic solvents and reactivity as an electrophilic aminating reagent [93, 94]. The amination of *gem*-borazirconocene alkanes such as (**127**) with MSH gave compound (**128**) and subsequent treatment with Ac_2O gave the α-aminoboronic ester derivatives (**129**) (Scheme 4.28).

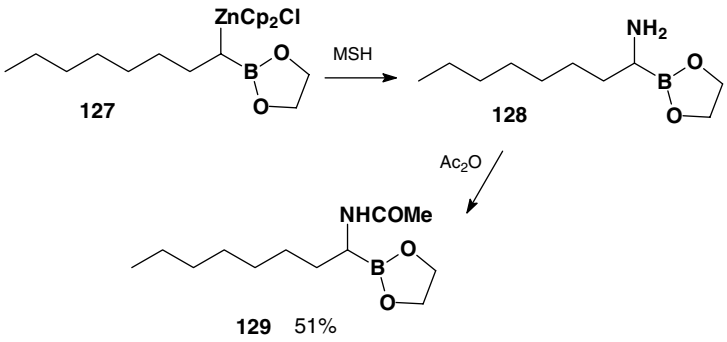
4.12
Synthesis and Activity of Amine-Carboxyboranes and their Derivatives

Amine-carboxyboranes with the common structure (**130**) can be regarded as isoelectronic analogs of protonated α-amino acids [58] or, more correctly, aliphatic carboxylic acids [95]. This resemblance has inspired extensive biological screening of these molecules, and the promising early results led to the syntheses of a large number of ester [96–99], amide [100, 101], peptide [102, 103], hydroxamic acid [104], and transition metal [105–107] derivatives of amine-carboxyboranes (A-BH_2COX, X = OR, NR^1R^2, or NHOH) containing a broad range substitutes, among other amine-boranes, which have been reviewed recently [2, 108, 109].

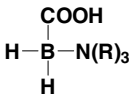
Table 4.6 Inhibition of NS3 protease, human leukocyte elastase, and human pancreatic chymotrypsin by P_1 phenethyl peptide boronic acids (**126a–u**).

Compound	R	NS3 K_i (μM)	Elastase IC_{50} (μM)	Chymotrypsin IC_{50} (μM)
126a	H	0.008	3.5	0.075
126b	2-methyl	0.82	NT[a]	NT
126c	3-methyl	0.034	5.7	NT
126d	4-methyl	0.017	5.0	3.7
126e	2,4-dimethyl	0.53	NT	NT
126f	2,5-dimethyl	1.0	NT	NT
126g	2-fluoro	0.018	NT	NT
126h	3-fluoro	0.009	NT	NT
126i	4-fluoro	0.006	0.8	0.050
126j	2,6-difluoro	0.930	NT	NT
126k	3-trifluoro-Me	0.025	NT	NT
126l	4-trifluoro-Me	0.002	1.8	16.0
126m	4-cloro	0.002	1.4	0.065
126n	4-bromo	0.004	1.6	NT
126o	4-phenyl	0.007	0.9	48.0
126p	4-isopropyl	0.005	0.45	>60
126q	4-cyclohexyl	0.003	0.40	>60
126r	4-*tert*-butyl	0.003	0.34	>60
126s	4-hydroxy	0.008	0.90	NT
126t	4-methoxy	0.003	0.56	20.0
126u	4-phenoxy	0.003	0.22	>60

[a]NT, not tested.

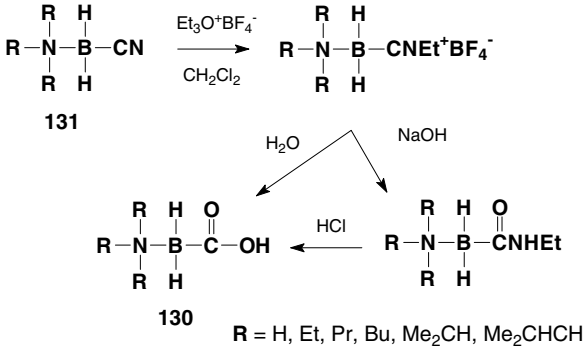


Scheme 4.28



130

R = H
R = Alkyl



Scheme 4.29

Today, many of these molecules are known to possess remarkable antitumor [110, 111], antiosteoporotic ([112] and references therein), anti-inflammatory ([113] and references therein), and hypolipidemic activities [114], and their mode of action is under investigation [115].

Amine-carboxyboranes with the general structure (**130**) (R = H or alkyl) may be regarded as boron analogs of substituted α-amino acids in their dipolar forms. As boron has one less positive charge on its nucleus than carbon, BH_2^- will be isoelectronic with CH_2 and consequently the boron-analog counterparts of the α-amino acids would exist in their protonated forms in the free state.

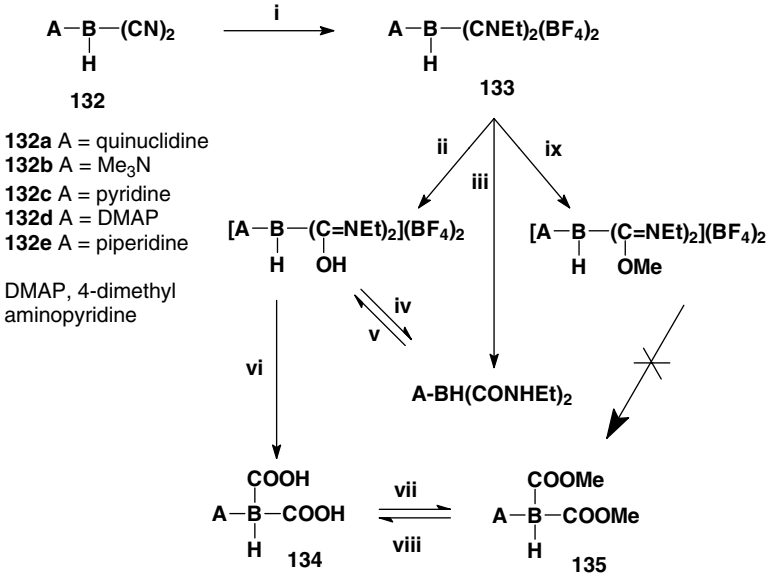
Several routes for the synthesis of amine-carboxyboranes have been described. In one study, Das and Mukherjee [115] have demonstrated that the acid- or base-catalyzed hydrolysis of the amine-cyanoboranes (**131**) always yields the acid (**130**) (Scheme 4.29).

The amine-dicarboxyboranes (**134**) and their dimethyl esters (**135**) have been synthesized from amine-dicyanoboranes via the [amine-bis(ethylnitrilium)-hydro-boron(2^+)]-tetrafluoroborates (**133**), [amine-bis(*C*-hydroxy-*N*-ethylimi-nium)-hydro-boron(2^+)] cations and amine-bis(*N*-ethylcarbamoyl)boranes. The *C*-methoxy-*N*-ethyliminium groups adjacent to boron undergo an unusual hydro-lysis [116]. The synthetic sequence is outlined in Scheme 4.30. First, the introduction of new functional groups (a second carboxylic group) offers potential biological activity. Second, the substitution of a hydrogen for an electron-withdrawing substituent on the boron is expected to increase the stability of the B−H bond and brings the electron distribution of the carboxylic group closer to that in aliphatic carboxylic acids.

The amine-dicyanoboranes (**132a–d**) were readily synthesized by base exchange.

These routes were used for the synthesis of acyclic amine-carboxyboranes which showed anti-inflammatory activity (see Table 4.7). The heterocyclic amine derivatives as well as amine-carbamoylboranes, carboalkoxyboranes, and cyanoboranes were generally less active. Those derivatives that demonstrated good anti-inflammatory activity (**136–173**) were effective inhibitors of hydrolytic lysosomal and proteolytic enzyme activities with IC_{50} values equal to 10^{-6} M in mouse macrophages, human leukocytes, and Be Sal osteofibrolytic cells [117].

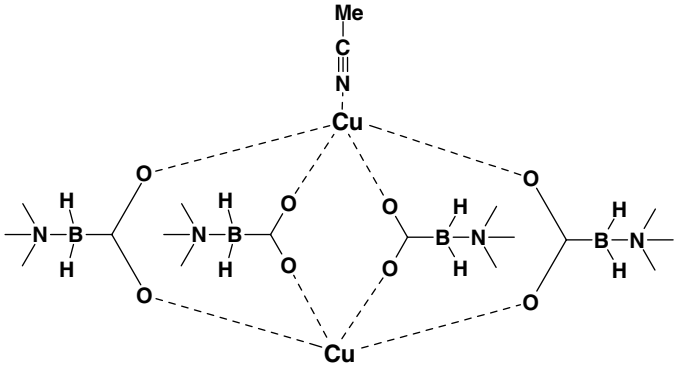
A number of metal complexes of amine-carboxyborane adducts having antitumor activity have been synthesized. Bis-μ-(morpholine-boranecarboxylato)zinc dihydrate



Scheme 4.30 Reagents and conditions: (i) 2.5 mol equiv. Et_3OBF_4, CH_2Cl_2, reflux, 6–10 h; (ii) H_2O, RT, 50–70 min; (iii and iv), 1 M NaOH, RT, immediate; (v) HBF_4, H_2O RT, immediate; (vi) 1 M HCl, 120 °C, 1.5 atm, 10 min; (vii) 7.5 mM HBr–MeOH, RT, 5–10 min; (viii), 0.05 M HCl, 60 °C, 15 min; (ix) MeOH, RT, 5–10 min.

(**174**) demonstrated cytotoxic activity against human $Tmolt_3$, HeLa-S^3, and MB-9812 cell growth [118]. The synthesis of the compound (**174**) is shown in Scheme 4.31.

Another amine-carboxyborane metal complex, tetrakis-μ-(trimethylamine-boranecarboxylato)-acetonitrile dicopper (**175**) [118] and also (**174**) inhibited L_{1210} DNA, RNA, and protein syntheses, with greatest inhibitory effects on DNA. The reduction in DNA synthesis correlates well with the inhibition of *de novo* purine synthesis and the key enzymes involved in this pathway (i.e., inosine monophosphate dehydrogenase and phosphoribosyl-α-1-diphosphate amidotransferase).

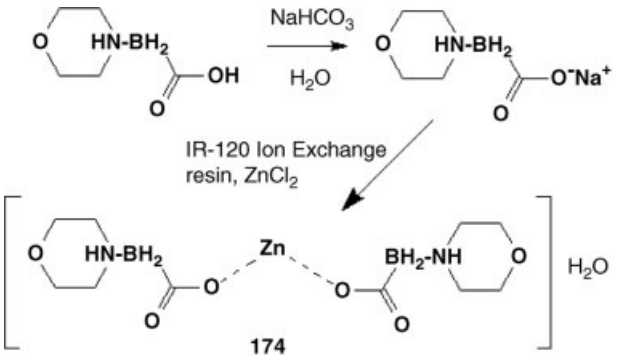


175

Table 4.7 Reaction data for R-BH_2X products and the anti-inflammatory activity of boron derivatives in CF-1 mice at 8 mg/kg.

No.	Compounds	Days	Yield (%)	Activity (ED_{50}, microgram/mL)
Amine-BH_2-COOH derivatives				
136	H_3NBH_2COOH	7	75	55
137	$MeNH_2BH_2COOH$	3	88	54
138	Me_2NHBH_2COOH	NI	NI	59
139	Me_3NBH_2COOH	2	90	59
140	$EtNH_2BH_2COOH$	2	92	59
141	$Me_3CNH_2BH_2COOH$	2	88	55
142	$H_2{=}CHCH_2NH_2BH_2COOH$	2	82	52
143	$PhCHNH_2BH_2COOH$	2	72	51
144	$EtO_2CCH_2NH_2BH_2COOH$	2	60	48
145	$(n\text{-}C_3H_7)_2NHBH_2COOH$	NI	NI	81
146	$C_{16}H_{33}NMe_2BH_2COOH$	NI	NI	40
147	$C_{18}H_{37}NMe_2BH_2COOH$	NI	NI	58
148	$H_2NNH_2BH_2COOH$	NI	NI	82
149	$[CH_2NMe_2BH_2COOH]_2$	NI	NI	58
150	$C_5H_5NBH_2COOH$	5	90	40
Amine-BH_2COOMe derivatives				
151	H_3NBH_2COOMe	12	86	66
152	$MeNH_2BH_2COOMe$	2	92	79
153	Me_2NHBH_2COOMe	2	80	58
154	Me3NBH2COOMe	2	72	58
155	$H_2C{=}CHCH_2NH_2BH_2COOMe$	2	89	56
156	$PhCH_2NH_2BH_2COOMe$	2	78	53
157	$C_5H_5NBH_2COOMe$	5	93	58
158	$EtO_2CCH_2NH_2BH_2COOMe$	2	52	48
159	$Me(CH_2)_3NH_2BH_2COOMe$	NI	NI	71
160	$C_{16}H_{33}NMe_2BH_2COOMe$	NI	NI	71
Amine-BH_2COOEt				
161	$H_3NBH_2CONHEt$	8	90	58
162	$C_{16}H_{33}NMe_2BH_2COOEt$	NI	NI	53
163	Me_2NHBH_2COOEt	2	82	76
164	$(BH_2CONHEt)_2$	NI	NI	94
165	$H_3NBH_2CONHEt$	8	90	76
166	$MeNH_2BH_2CONHEt$	2	90	51
167	$Me_2NHBH_2CONHEt$	2	91	84
168	$Me_3NBH_2CONHEt$	2	80	61
Amine-BH_2-CONHR derivatives				
169	$Me_3NBH_2CONH(n\text{-}Pr)$	NI	NI	61
170	$Me_3NBH_2CONH(n\text{-}Bu)$	NI	NI	86
171	$Me_3NBH_2CONH(n\text{-}Oct)$	NI	NI	86
172	$Me_3NBH_2CONHC_6H_5$	NI	NI	72
173	$[Me_2NCH_2BH_2CONHEt]_2$	NI	NI	65

ED_{50}, the dose of a drug that is pharmacologically effective for 50% of the population exposed to the drug or a 50% response in a biological system that is exposed to the drug.
NI, not indicated in origin paper.



Scheme 4.31

A series of boron-containing nicotine (NIC) analogs have been synthesized and evaluated for binding to α4β2 and α7 neuronal nicotinic receptors [119]. The compounds (**177**) and (**178**) were prepared according to earlier published methods [120–122]. The boron-containing analogs (**179–181**) were synthesized by refluxing a suspension of $NaBH_3CN$ and the hydrochloride salts of the corresponding precursors (**176–178**) in tetrahydrofuran (THF) under N_2 overnight, according to Scheme 4.32. All three boron-containing analogs were found to be very stable in water. The compound (**180**) inhibited [^{3}H]methyllycaconitine (MLA) binding to rat brain membranes with a similar potency compared to NIC (**179**). Methyllycaconitine (MLA) is a plant alkaloid found in the larkspur, and has been identified as an antagonist of nicotinic acetylcholine receptors (nAChrs) in the muscle and brain.

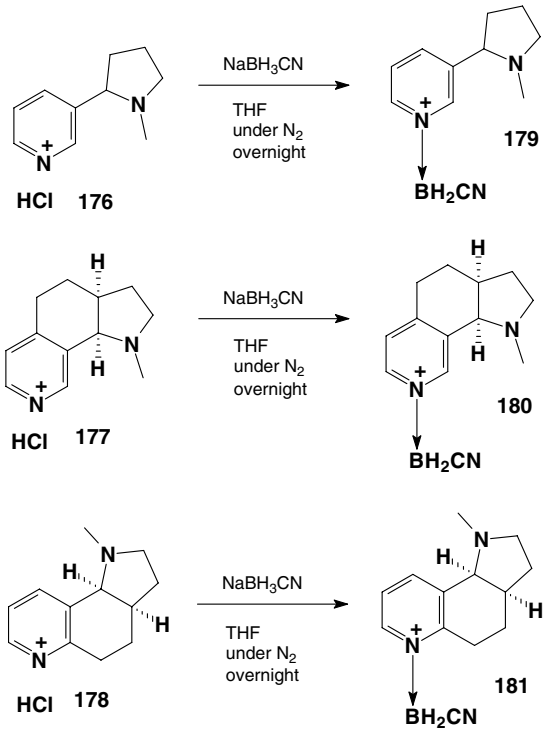
All three boron-containing analogs were found to be very stable in water. Hydrolytic stability at pH **179** is critical for biological testing. No decomposition was observed after a 0.1 M solution of **179** in D_2O was kept at room temperature (RT) for 8 days, as indicated by ^{1}H-NMR. Hydrolysis of **179** also occurred very slowly in alkali. However, significant decomposition (8%) was observed when **179** was exposed to 1 N HCl for 1 week.

The borane-containing analogs **179–181** were evaluated as their hydrobromide or hydrochloride salts for their binding affinities for $\alpha_4\beta_2$ and α_7 nAChRs. The affinity of the analogs for the $\alpha_4\beta_2$ receptor subtype was assayed by analog-induced inhibition of [^{3}H]NIC binding to rat striatal membranes.

The affinity of the analogs for the α_7 receptor subtype was assayed by inhibition of [^{3}H]MLA binding to rat brain membranes. Results were reported as K_i values (Table 4.8). Both MLA and α-bungarotoxin (α-BTX) bind with very high affinity to the α_7 receptor ($K_i = 0.0023$ and 0.0072 mM, respectively), and MLA also exhibited low affinity for the $\alpha_4\beta_2$ receptor ($K_i = 1.56$ mM). In contrast, NIC showed about 100-fold higher affinity for the $\alpha_4\beta_2$ receptor than for the α_7 receptor.

4.13
Synthesis of Boron Analogs of Phosphonoacetates

A general procedure has been described for exchange reactions between *N*-methylmorpholine-borane derivatives and various organic bases involving a simple work-up

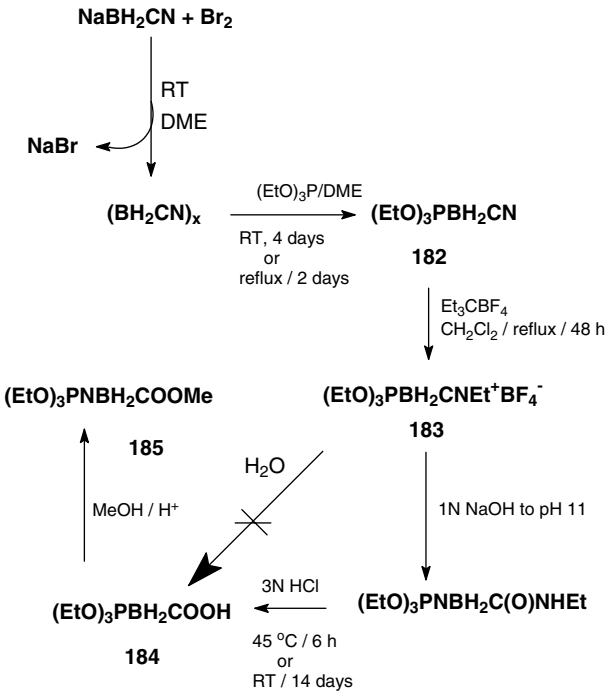


Scheme 4.32

to produce the exchanged products. Aqueous as well as liquid amines in the exchange reactions with *N*-methylmorpholine-BH_2X (X = COOH, COOMe, or CONHEt) gave the corresponding amine-BH_2X derivatives in good yields [123]. The method also utilizes other organic bases, such as phosphine, phosphate, and amino acid esters in the exchange reaction, which have been obtained as their borane derivatives.

Table 4.8 K_i values for MLA, α-BTX, and compounds **176–181** in the [^{3}H]MLA and [^{3}H]NIC binding assays.

Compound	K_i [^{3}H]MLA binding assay (μM)	K_i [^{3}H]NIC binding assay (μM)
176	0.77	0.001
177	0.59	0.40
178	>100	10.5
179	15.2	0.041
180	2.4	0.60
181	>100	>100
MLA	0.0023	1.56
α-BTX	0.0072	>10



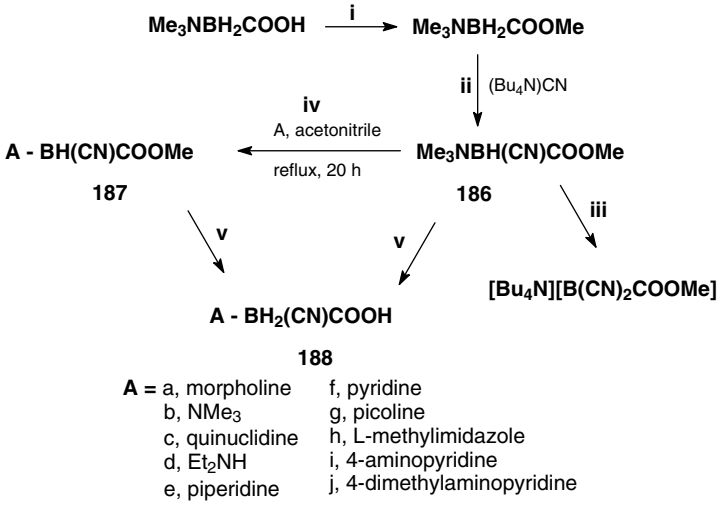
Scheme 4.33

Boron analogs of phosphonoacetates have been synthesized (Scheme 4.33), and their antitumor and anti-inflammatory activities were studied [123, 124]. Cytotoxicity data for the compounds (**182–185**) are shown in Table 4.9.

Table 4.9 Cytotoxicity of boron analogs of phosphonoacetates.[a]

	Compound			
Cancer cell line	**182**	**183**	**184**	**185**
Ehrlich carcinoma[b]	66.0	80.0	39.0	97.0
Murine L_{1210}	3.61	4.45	3.47	4.15
Murine P388	NT	6.46	7.26	4.98
Human HeLa	2.12	NT	4.41	3.96
Human KB	2.86	NT	3.87	1.75
Human glioma	4.26	NT	5.29	1.60
Human osteosarcoma	2.88	NT	4.18	7.25
Human lung	6.88	NT	4.48	4.80
Human colon	6.56	NT	2.65	2.42
Human $Tmolt_3$	2.42	NT	6.16	7.64

[a]Activity of boron compounds (ED_{50}, μg/ml).
[b]Percent *in vivo* inhibition; NT, not tested;



Scheme 4.34 Reagents and conditions: (i) N-bromosuccinimide, methanol, RT, 30 min; (ii) 1.1 mol equiv. $[Bu_4N]CN$, acetonitrile, RT, 25 h; (iii) 2 mol equiv. $[Bu_4N]CN$, acetonitrile, reflux, 20 h; (iv) 3 mol equiv. A, acetonitrile, reflux, 2–12 h, yields 48–79%; (v) 0.45 mol equiv. HCl in 0.05–0.2 M HCl in water or water–acetone mixture, 55–95 °C, 0.5–2.5 h, yields 27–86%.

A number of amine-carboxyborane esters have been studied for their antihyperlipidemic activity [125] and hypolipidemic activity in rodents [96].

The synthesis of amine-cyanocarboxyboranes-isoelectronic analogs of α-cyanocarboxylic acids has been reported [126]. In the first step, trimethyl-carboxyborane was brominated and simultaneously esterified using *N*-bromosuccinimide in methanol [95]. The synthetic sequence outlined in Scheme 4.34, employing activation and then nucleophilic substitution of the boron, subsequently resulted in the preparation of several novel compounds (**186**, **187** and **188a–j**). Boron analogs of phosphonoacetates proved to be potent hypolipidemic agents in rodents, lowering both serum cholesterol and triglyceride levels. Et_3PBH_2COOMe proved to be the most effective agent in mice, lowering serum cholesterol to 46% and serum triglycerides to 54% after 16 days. Et_3PBH_2COOH and $Na^+H^+(EtO)_2PBH_2COO^-$ caused a greater than 40% reduction in lipids. The cyanoborane adducts of aminomethylphosphonates were generally less effective; $(C_6H_5O)_2P(O)CH_2NH_2BH_2CN$ was the most effective, lowering serum cholesterol to 32% and serum triglycerides to 43% after 16 days. The phosphonoacetates appeared to lower lipid concentrations by several mechanisms. (i) They lowered the *de novo* synthesis of cholesterol and triglycerides in the liver. (ii) They accelerated the excretion of lipids into the bile and feces. (iii) They modulated low- and high-density lipoprotein-cholesterol contents in a manner that suggests reduction of the deposition of lipids in peripheral tissues, and accelerated the movement of cholesterol from tissues (e.g., plaques) to the liver for excretion into the bile [127].

4.14 Conclusions

A remarkable diversity of reactions has been discussed in this chapter, which has also described and summarized the development of a new boron-based methodology for applications in organic synthesis of α-aminoboronic acids, amine-carboxyboranes and their derivatives. These compounds are unique among boron-containing compounds and they have high biological activity in different fields. The investigation of the chemistry of these compounds is continuing, especially in the areas of selective reactions, synthesis, catalysis, and coordination chemistry [128–130].

References

1 Dembitsky, V.M. and Srebnik, M. (2003) *Tetrahedron*, **59**, 579.

2 Dembitsky, V.M., Quntar, A.A.A. and Srebnik, M. (2004) *Mini-Reviews in Medicinal Chemistry*, **4**, 1001.

3 Abu Ali, H., Dembitsky, V.M. and Srebnik, M. (2005) In *Contemporary Aspects of Boron: Chemistry and Biological Applications, Studies in Inorganic Chemistry*, vol. 22, Elsevier, Amsterdam, p. 618.

4 (a) Ma, J.S. (2003) *Chimica Oggi – Chemistry Today*, **21**, 65; (b) Saito, H. and Shiba, K. (2004) *Biobencha*, **4**, 27.

5 Cuerrier, D.T., Moldoveanu, T., Campbell, R.L., Kelly, J., Yoruk, B., Verhelst, S.H.L., Greenbaum, D., Bogyo, M. and Davies, P.L. (2007) *The Journal of Biological Chemistry*, **282**, 9600.

6 Duthaler, R.O. (1994) *Tetrahedron*, **50**, 1539.

7 Williams, R.M. (1992) *Aldrichimica Acta*, **25**, 11.

8 Williams, R.M. (1994) *Synthesis of Optically Active R-Amino Acids*, Pergamon, Oxford.

9 Oishi, T., Lee, Y.K. and Onimura, K. (2004) *Kobunshi Ronbunshu*, **61**, 216.

10 Lustenberger, P., Martinborough, E., Denti, T.M. and Diederich, F. (1998) *Journal of the Chemical Society, Perkin Transactions 2*, 747.

11 Graf, E. and Hosseini, M.W. (1998) *Coordination Chemistry Reviews*, **178–180**, 1193.

12 Dembitsky, V.M., Smoum, R., Al-Quntar, A.A., Abu Ali, H., Pergament, I. and Srebnik, M. (2002) *Plant Science*, **163**, 931.

13 Lyer, R.P., Roland, A., Zhou, W. and Ghosh, K. (1999) *Current Opinion in Molecular Therapeutics*, **1**, 344.

14 Imamura, M., Sawada, S., Kasahara-Imamura, M., Harima, K. and Harada, K. (2002) *International Journal of Molecular Medicine*, **9**, 11.

15 Otkidach, D.S. and Pletnev, I.V. (2001) *Theochem – Journal of Molecular Structure*, **536**, 65.

16 Morin, C. (1994) *Tetrahedron*, **50**, 12521.

17 Braunschweig, H. and Colling, M. (2001) *Coordination Chemistry Reviews*, **223**, 1.

18 Dordas, C. and Brown, P.H. (2000) *The Journal of Membrane Biology*, **175**, 95.

19 Feakes, D.A., Spinler, J.K. and Harris, F.R. (1999) *Tetrahedron*, **55**, 11177.

20 Malinin, V.S., Haque, M.E. and Lentz, B.R. (2001) *Biochemistry*, **40**, 8292.

21 Shull, B.K., Spielvogel, D.E., Gopalaswamy, R., Sankar, S., Boyle, P.D., Head, G. and Devito, K. (2000) *Journal of the Chemical Society, Perkin Transactions 2*, 557.

22 Srivastava, R.R., Singhaus, R.R. and Kabalka, G.W. (1999) *The Journal of Organic Chemistry*, **64**, 8495.

23 Faber, K. (1995) *Biotransformations in Organic Chemistry*, 2nd edn, Springer, New York.

24 Hiratake, J. and Oda, J. (1997) *Bioscience, Biotechnology, and Biochemistry*, **61**, 211.

25 (a) Wong, C.H. and Whitesides, G.M. (eds) (1994) *Enzymes in Synthetic Organic Chemistry*, Pergamon, New York; (b) Dembitsky, V.M. (2003) *Tetrahedron*, **59**, 4701.

26 Verlinde, C.L.M.J. and Hol, W.G.J. (1994) *Structure*, **2**, 577.

27 Kurinov, I.V. and Harrison, R.W. (1994) *Nature Structural Biology*, **1**, 735.

28 Grutter, M.G., Priestle, J.P., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J. and Stone, S.R. (1990) *The EMBO Journal*, **9**, 2361.

29 Hynes, T.R., Randal, M., Kennedy, L.A., Eigenbrot, E. and Kossiakoff, A.A. (1990) *Biochemistry*, **29**, 10018.

30 Yamamoto, M., Sawaya, R., Mohanam, S., Rao, V.H., Bruner, J.M., Nicolson, G.L., Ohshima, K. and Rao, J.S. (1994) *Journal of Neuro-Oncology*, **22**, 139.

31 (a) Craik, C.S. and Debouck, C. (1995) *Perspectives in Drug Discovery and Design*, **2**, 257; (b) Bernkop-Schnurch, A. (1998) *Journal of Controlled Release*, **52**, 1.

32 (a) Kraut, J. (1977) *Annual Review of Biochemistry*, **46**, 331; (b) Polgar, L. and Halasz, P. (1982) *The Biochemical Journal*, **207**, 1.

33 Powers, J.C. and Harper, J.W. (1986) In *Proteinase Inhibitors*, 2nd edn (eds A.J. Barret and G. Salvesen), Research Monographs in Cell and Tissue Physiology, vol. 12, Elsevier, Amsterdam, p. 55.

34 Schechter, I. and Berger, A. (1967) *Biochemical and Biophysical Research Communications*, **27**, 157.

35 Matteson, D.S. and Sadhu, K.M. (1984) *Organometallics*, **3**, 614.

36 Amiri, P., Lindquist, R.N., Matteson, D.S. and Sadhu, K.M. (1984) *Archives of Biochemistry and Biophysics*, **234**, 531.

37 Duncan, K., Faraci, W.S., Matteson, D.S. and Walsh, C.T. (1989) *Biochemistry*, **28**, 3541.

38 Matteson, D.S. and Cheng, T.-C. (1968) *The Journal of Organic Chemistry*, **33**, 3055.

39 Matteson, D.S. and Majumdar, D. (1979) *Journal of Organometallic Chemistry*, **170**, 259.

40 Matteson, D.S. (1995) *Stereodirected Synthesis with Organoboranes*, Springer, Berlin.

41 Matteson, D.S., Sadhu, K.M. and Lienhard, G.E. (1981) *Journal of the American Chemical Society*, **103**, 5241.

42 Matteson, D.S., Jesthi, P.K. and Sadhu, K.M. (1984) *Organometallics*, **3**, 1284.

43 Sadhu, K.M. (1983) *Chemical Abstracts*, **99**, 195047.

44 Kinder, D.H. and Katzenellenbogen, J.A. (1985) *Journal of Medicinal Chemistry*, **28**, 1917.

45 Matteson, D.S., Michnick, T.J., Willett, R.D. and Patterson, C.D. (1989) *Organometallics*, **8**, 726.

46 Lai, J.H., Liu, Y., Wu, W., Zhou, Y., Maw, H.H., Bachovchin, W.W., Bhat, K.L. and Bock, C.W. (2006) *The Journal of Organic Chemistry*, **71**, 512.

47 Lindquist, R.N. and Nguyen, A. (1977) *Journal of the American Chemical Society*, **99**, 6435.

48 Kettner, C.A., Jagannathan, S. and Forsyth, T.P. (2003) *Chemical Abstracts*, **139**, 69373.

49 Spielvogel, B.F., McPhail, A.T. and Hall, I.H. (1982) *Chemical Abstracts*, **97**, 85322.

50 Hall, I.H., Starnes, C.O., McPhail, A.T., Wisian-Neilson, P., Das, M.K., Harchelroad, R. and Spielvogel, B.F. (1980) *Journal of Pharmaceutical Sciences*, **69**, 1025.

51 Spielvogel, B.F., Wojnowich, L., Das, M.K., McPhail, A.T. and Hargrave, K.D. (1976) *Journal of the American Chemical Society*, **98**, 5702.

52 Scheller, K.H., Martin, R.B., Spielvogel, B.F. and McPhail, A.T. (1982) *Inorganica Chimica Acta*, **57**, 227.

53 Spielvogel, B.F., Das, M.K., McPhail, A.T., Onan, K.D. and Hall, I.H. (1980) *Journal of the American Chemical Society*, **102**, 6343.

54 Spielvogel, B.F., Ahmed, F.U., Silvey, G.L., Wisian-Neilson, P. and McPhail, A.T. (1984) *Inorganic Chemistry*, **23**, 4322.

55 Spielvogel, B.F., Ahmed, F.U. and McPhail, A.T. (1986) *Synthesis*, 833.

56 Sood, A. and Spielvogel, B.F. (1989) *Main Group Metal Chemistry*, **12**, 143.

57 Coutts, S.J., Kelly, T.A., Snow, R.J., Kennedy, C.A., Barton, R.W., Adams, J., Krolikowski, D.A., Freeman, D.M., Campbell, S.J., Ksiazek, J.F. and Bachovcin, W.W. (1996) *Journal of Medicinal Chemistry*, **39**, 2087.

58 Matteson, D.S. (1989) *Tetrahedron*, **45**, 1859.

59 von Matt, A., Ehrhardt, C., Burkhard, P., Metternich, R., Walkinshaw, M. and Tapparelli, C. (2000) *Bioorganic and Medicinal Chemistry*, **8**, 2291.

60 Wienand, A., Ehrhardt, C., Metternich, R. and Tapparelli, C. (1999) *Bioorganic and Medicinal Chemistry*, **7**, 1295.

61 Metternich, R. (1992) *Chemical Abstracts*, **117**, 49266.

62 Knabb, R.M., Kettner, C.A., Timmermans, P.B.M.W.M. and Reilly, T.M. (1992) *Thrombosis and Haemostasis*, **67**, 56.

63 Hussain, M.A., Knabb, R., Aungst, B.J. and Kettner, C. (1991) *Peptides*, **12**, 1153.

64 Cacciola, J., Fevig, J.M., Alexander, R.S., Brittelli, D.R., Kettner, C.A., Knabb, R.M. and Weber, P.C. (1996) *Bioorganic & Medicinal Chemistry Letters*, **6**, 301.

65 Deadman, J.J., Elgendy, S., Green, D., Skordalakes, E., Scully, M.F., Goodwin, C.A. and Kakkar, V.V. (1998) *Chemical Abstracts*, **128**, 128287.

66 Kelly, T.A., Fuchs, V.U., Perry, C.W. and Snow, R.J. (1993) *Tetrahedron*, **49**, 1009.

67 (a) Lee, S.L., Alexander, R., Smallwood, A., Trievel, R., Mersinger, L., Weber, P.C. and Kettner, C. (1997) *Biochemistry*, **36**, 13180; (b) Adams, J., Behnke, M., Chen, S., Cruickshank, A.A., Dick, L.R., Grenier, L., Klunder, J.M., Ma, Y.-T., Plamondon, L. and Stein, R.L. (1998) *Bioorganic & Medicinal Chemistry Letters*, **8**, 333; Adams, J. (2002) *Current Opinion in Chemical Biology*, **6**, 493.

68 Shenvi, A.B. and Kettner, C.A. (1985) *Chemical Abstracts*, **103**, 71709.

69 Dunsdon, R.M., Greening, J.R., Jones, P.S., Jordan, S. and Wilson, F.X. (2000) *Bioorganic & Medicinal Chemistry Letters*, **10**, 1577.

70 Carboni, B., Pourbaix, C., Carreaux, F., Delueze, H. and Maillard, B. (1999) *Tetrahedron Letters*, **40**, 7979.

71 Kwong, A.D. (1997) *Current Opinion in Infectious Diseases*, **10**, 485.

72 Martichonok, V. and Jones, J.B. (1996) *Journal of the American Chemical Society*, **118**, 950.

73 Elgendy, S., Patel, G., Green, D., Goodwin, C.A., Scully, M.F., Husman, W., Skordalakes, E., Kakkar, V.V. and Deadman, J.J. (1997) *Tetrahedron Letters*, **38**, 3305.

74 Martichonok, V. and Jones, J.B. (1997) *Bioorganic and Medicinal Chemistry*, **5**, 679.

75 Matteson, D.S. and Michnick, T.J. (1992) *Journal of Labelled Compounds & Radiopharmaceuticals*, **31**, 567.

76 Lebarbier, C., Carreaux, F. and Carboni, B. (1996) *Synthesis*, 1371.

77 Kettner, C.A. and Shenvi, A.B. (1988) *Chemical Abstracts*, **112**, 91790.

78 Kettner, C., Mersinger, L. and Knabb, R. (1990) *The Journal of Biological Chemistry*, **265**, 18289.

79 Mantri, P., Duffy, D.E. and Kettner, C.A. (1996) *The Journal of Organic Chemistry*, **61**, 5690.

80 Lebarbier, C., Carreaux, F., Carboni, B. and Boucher, J.L. (1998) *Bioorganic & Medicinal Chemistry Letters*, **8**, 2573.

81 Kerwin, J.F. Jr, Lancaster, J.R. Jr and Feldman, P.L. (1995) *Journal of Medicinal Chemistry*, **38**, 4343.

82 Feldman, P.L., Griffith, O.W. and Stuehr, D. (1993) *Chemical & Engineering News*, **71**, 26.

83 Brosz, C.S., Calabrese, J.C., Kettner, C.A. and Teleha, C.A. (1997) *Tetrahedron-Asymmetry*, **8**, 1435.

84 Jagannathan, S., Forsyth, T.P. and Kettner, C.A. (2001) *The Journal of Organic Chemistry*, **66**, 6375.

85 Priestley, E.S., De Lucca, I., Ghavimi, B., Erickson-Viitanen, S. and Decicco, C.P. (2002) *Bioorganic & Medicinal Chemistry Letters*, **12**, 3199.

86 (a) Matteson, D.S. and Majumdar, D.J. (1980) *Journal of the American Chemical Society*, **102**, 7588; (b) Matteson, D.S. and Ray, R. (1980) *Journal of the American Chemical Society*, **102**, 7590.

87 Matteson, D.S. and Majumdar, D. (1983) *Organometallics*, **2**, 1529.

88 Brown, H.C. and Gupta, S.K. (1975) *Journal of the American Chemical Society*, **97**, 5249.

89 Dembitsky, V.M. and Srebnik, M. (2002) In *Titanium and Zirconium in Organic Synthesis* (ed. I. Marek), Wiley-VCH Verlag GmbH, Weinheim, pp. 230–281.

90 Zheng, B., Deloux, L., Pereira, S., Skrzypckak-Jankun, E., Cheesman, B.V., Sabat, M. and Srebnik, M. (1996) *Applied Organometallic Chemistry*, **10**, 267.

91 Tamura, Y., Minamikawa, J., Sumoto, K., Fujii, S. and Ikeda, M. (1973) *The Journal of Organic Chemistry*, **38**, 1239.

92 Houben, J. and Schmidt, E. (1913) *Chemische Berichte*, **46**, 3616.

93 Midland, M.M., Greer, S., Tramontano, A. and Zderic, S.A. (1979) *Journal of the American Chemical Society*, **101**, 2352.

94 Zheng, B. and Srebnik, M. (1995) *The Journal of Organic Chemistry*, **60**, 1912.

95 Györi, B., Berente, Z. and Kovács, Z. (1998) *Inorganic Chemistry*, **37**, 5131.

96 Hall, I.H., Spielvogel, B.F., Sood, A., Ahmed, F. and Jafri, S. (1987) *Journal of Pharmaceutical Sciences*, **76**, 359.

97 Gyoeri, B. and Berente, Z. (2001) *Inorganic Chemistry*, **40**, 1770.

98 Mittakanti, M., Feakes, D.A. and Morse, K.W. (1992) *Synthesis*, 380.

99 Gyori, B., Berente, Z., Emri, J. and Lazar, I. (1995) *Synthesis*, 191.

100 Mittakanti, M. and Morse, K.W. (1990) *Inorganic Chemistry*, **29**, 554.

101 Das, M.K., Mukherjee, P. and Roy, S. (1990) *Bulletin of the Chemical Society of Japan*, **63**, 3658.

102 Sood, A., Sood, C.K., Spielvogel, B.F. and Hall, I.H. (1990) *European Journal of Medicinal Chemistry*, **25**, 301.

103 Karthikeyan, S., Sood, A., Tomasz, J., Spielvogel, B.F. and Hall, I.H. (1995) *Amino Acids*, **8**, 323.

104 Norwood, V.M. and Morse, K.W. (1988) *Inorganic Chemistry*, **27**, 302.

105 Norwood, V.M. III and Morse, K.W. (1987) *Inorganic Chemistry*, **26**, 284.

106 Hall, I.H., Spielvogel, B.F. and McPhail, A.T. (1984) *Journal of Pharmaceutical Sciences*, **73**, 222.

107 Carboni, B. and Monnier, L. (1999) *Tetrahedron*, **55**, 1197.

108 Norwood, V.M. III and Morse, K.W. (1986) *Inorganic Chemistry*, **25**, 3690.

109 Jabbour, A., Smoum, R., Takrouri, K., Shalom, E., Zaks, B., Steinberg, D., Rubinstein, A., Goldberg, I., Katzhendler, J. and Srebnik, M. (2006) *Pure and Applied Chemistry*, **78**, 1425.

110 Spielvogel, B.F., Sood, A. and Hall, I.H. (1993) *Chemical Abstracts*, **119**, 203618.

111 Hall, I.H., Hall, E.S., Miller, M.C. III, Sood, A. and Spielvogel, B.F. (1993) *Amino Acids*, **4**, 287.

112 Murphy, M.E., Elkins, A.L., Shrewsbury, R.P., Sood, A., Spielvogel, B.F. and Hall, I.H. (1996) *Metal-Based Drugs*, **3**, 31, n.

113 Hall, I.H., Rajendran, K.G., Chen, S.Y., Wong, O.T., Sood, A. and Spielvogel, B.F. (1995) *Archiv der Pharmazie*, **328**, 39.

114 Miller, M.C. III, Woods, C.M., Murphy, M.E., Elkins, A., Spielvogel, B.F. and Hall, I.H. (1998) *Biomedicine & Pharmacotherapy*, **52**, 169.

115 Das, M.K. and Mukherjee, P. (1985) *Journal of Chemical Research (S)*, 66.

116 Berente, Z. and Gyori, B. (1997) *Chemical Communications*, 1799.

117 Hall, I.H., Burnham, B.S., Chen, S.Y., Sood, A., Spielvogel, B.F. and Morse, K.W. (1995) *Metal-Based Drugs*, **2**, 1–15.

118 Miller, M.C., Sood, A., Spielvogel, B.F. and Hall, I.H. (1998) *Applied Organometallic Chemistry*, **12**, 87.

119 Xu, R., Dwoskin, L.P., Grinevich, V.P., Deaciuc, G. and Crooks, P.A. (2001) *Bioorganic & Medicinal Chemistry Letters*, **11**, 1245.

120 Chavdarian, C.G., Seeman, J.I. and Wooten, J.B. (1983) *The Journal of Organic Chemistry*, **48**, 492.

121 Glassco, W., Suchocki, J., George, C., Martin, B.R. and May, E.L. (1993) *Journal of Medicinal Chemistry*, **36**, 3381.

122 Vernier, J.-M., Holsenback, H., Cosford, N.D.P., Whitten, J.P., Menzaghi, F., Reid, R., Rao, T.S., Sacaan, A.I., Lloyd, G.K., Suto, C.M., Chavez-Noriega, L.E., Washburn, M.S., Urrutia, A. and McDonald, I.A. (1998) *Bioorganic & Medicinal Chemistry Letters*, **8**, 2173.

123 Mittakanti, M. and Morse, K.W. (1996) *Main Group Metal Chemistry*, **19**, 727.

124 Sood, A., Sood, C.K., Hall, I.H. and Spielvogel, B.F. (1991) *Tetrahedron*, **47**, 6915.

125 Hall, I.H., Das, M.K., Spielvogel, B.F. and McPhail, A.T. (1981) *Journal of Pharmaceutical Sciences*, **70**, 339.

126 Gyori, B., Berente, Z., Kiraly, R. and Lazar, I. (2002) *Journal of the Chemical Society, Perkin Transactions 1*, 300.

127 Hall, I.H., Wong, O.T., Sood, A., Sood, C.K., Spielvogel, B.F., Shrewsbury, R.P. and Morse, K.W. (1992) *Pharmaceutical Research*, **25**, 259.

128 Abu Ali, H., Dembitsky, V.M. and Srebnik, M. (2004) In *The Science of Synthesis. Organometallics: Boron Compounds* (eds M. Regitz, D. Kaufmann and D.S. Matteson), Houben–Weyl Methods of Molecular Transformation, vol. 6, Thieme, Stuttgart, p. 855.

129 Abu Ali, H., Dembitsky, V.M. and Srebnik, M. (2004) In *The Science of Synthesis. Organometallics: Boron Compounds* (eds M. Regitz, D. Kaufmann and D.S. Matteson), Houben–Weyl Methods of Molecular Transformation, vol. 6, Thieme, Stuttgart, p. 867.

130 Abu Ali, H., Dembitsky, V.M. and Srebnik, M. (2004) In *The Science of Synthesis. Organometallics: Boron Compounds* (eds M. Regitz, D. Kaufmann and D.S. Matteson), Houben–Weyl Methods of Molecular Transformation, vol. 6, Thieme, Stuttgart, p. 881.

5
Chemistry of Aminophosphonic Acids and Phosphonopeptides

Valery P. Kukhar and Vadim D. Romanenko

5.1
Introduction

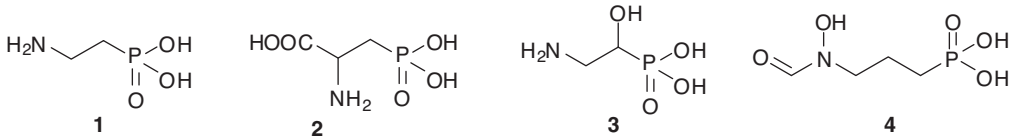
Aminophosphonic (and aminophosphinic) acids occupy a prominent position in the understanding of bioprocesses in living organisms and in the construction of new bioregulators – pharmaceuticals and agrochemicals. The so-called "phosphorus analogs" of the amino acids, aminophosphonic and aminophosphinic acids, are amino acid mimetics in which the carboxylic group is substituted by a phosphonic acid residue $P(O)(OH)_2$ or phosphinic acid group P(O)(OH)R. The replacement of the carboxylic group by a phosphonic or related moiety results in a number of important consequences: the central atom of a phosphonic acid function consists of an additional substituent that has a tetrahedral configuration in contrast to the planar carbonyl atom of carboxylic group. There are also significant differences in the acidity and steric bulk of these residues. The tetrahedral configuration of the pentavalent phosphorus atom mimics the transition state of the peptide bond cleavage reaction in some enzyme–substrate interactions that is used in the synthesis of new inhibitors. The hydrolytic stability of C−P bonds is often used in order to prepare stable analogs of bioactive phosphates $O-P(O)(OH)_2$. In addition, aminophosphonic acids and their derivatives are important as metal-complexing agents, which have diagnostic and therapeutic applications, and as industrial chemicals in water treatment, metal extraction, or pollution control.

Synthesis, and the chemical, physical, and biological properties of aminophosphonic acids and some directions for their practical application are presented in a number of reviews and book chapters. The most comprehensive account of the chemistry and biology of aminophosphonates and aminophosphinate which covers literature up to 2000 is presented in Kukhar and Hudson ([1] and references therein). Readers are referred to this reference for more detailed information concerning specific aspects of the chemistry and biological activity of aminophosphonic acid derivatives. The purpose of this chapter is to give an updated report on aminophosphonic acids as analogs of amino acids. Numerous heterocyclic compounds that contain both phosphonic and amino groups will not be covered in this report.

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

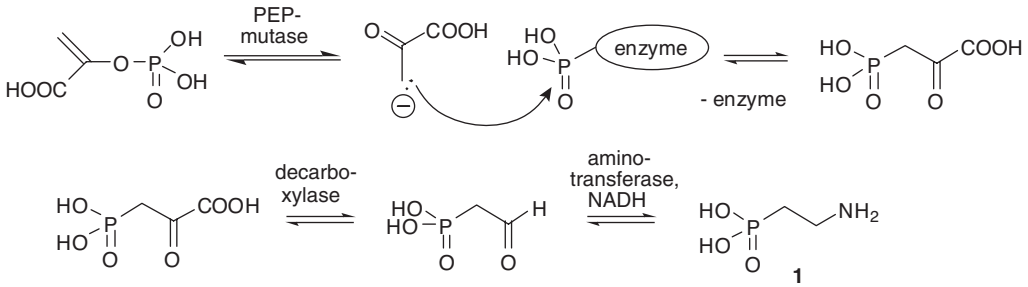
ISBN: 978-3-527-32098-1

Aminophosphonic acids were unknown until 1959 when Horiguchi and Kandatsu discovered 2-aminoethanephosphonic acid (AEP; "ciliatine") in ciliated sheep rumen protozoa [2]. To date, AEP **1** as well as aminophosphonates **2** and **3** have been demonstrated to be present in bacteria, protozoa, insects, vertebrates, and human tissues, occasionally as a significant part of total phosphorus. Some aminophosphonates (e.g., fosmidomycin **4** and related compounds) are a part of natural phosphonopeptide antibiotics and were found in *Streptomyces*.

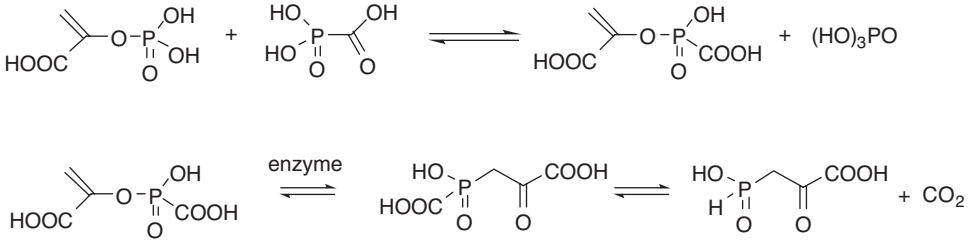


Phosphonolipids containing an aminophosphonate residue instead the phosphate ester moiety, most frequently AEP, have been detected in many organisms, sometimes at a remarkably high level. For example, in some species of Coelenterata the content of phosphonate phosphorus approaches 50% of total phosphorus. Significant concentrations of AEP have been found in glycoproteins (lipid-free protein fraction) from sea anemones. The most exciting discovery was the detection of AEP in the human body, including the brain, heart, liver, kidney, spleen, and aorta. More detailed information on naturally occurring aminophosphonates and biosynthesis can be found in recent publications [3, 4].

It is now well established that the P−C bond is formed biosynthetically by the phosphoenolpyruvate (PEP)–phosphonopyruvate rearrangement catalyzed by two phosphomutase enzymes [5]. One of these enzymes, PEP mutase, catalyzes the direct transfer of the phosphono group from oxygen to the carbon of PEP. The latter is a key substrate in further biosynthetic transformations of phosphonates. In the case of AEP, the next step involves decarboxylation of phosphonopyruvate into phosphonoacetaldehyde catalyzed by a specific decarboxylase. Then phosphonoacetaldehyde is converted to AEP by a transamination route (Scheme 5.1). The second isolated enzyme, carboxyphosphonoenolpyruvate phosphomutase, induces the rearrangement of carboxyphosphonoenolpyruvate in the presence of phosphonoformic acid into 3-phosphinopyruvate (Scheme 5.2). This process was supposed to be a key step in the biosynthesis of herbicide bialaphos by *Streptomyces hygroscopicus*.



Scheme 5.1

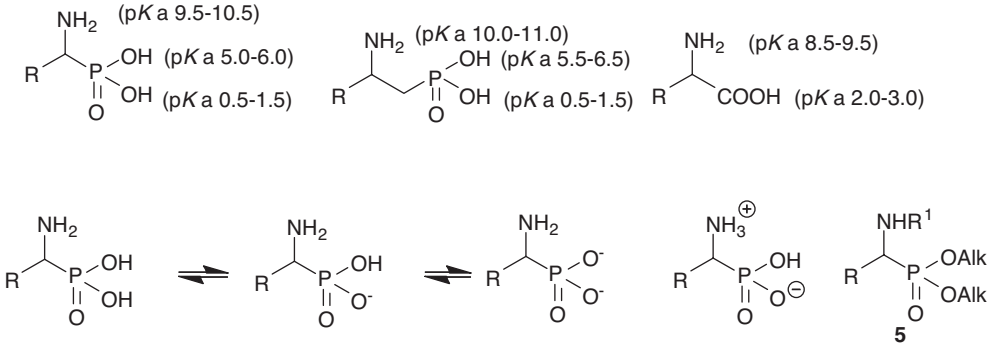


Scheme 5.2

There is also evidence of the existence of other enzymes involved both in the biosynthesis of C−P bond-containing compounds and in the degradation of phosphonates. Although the C−P bond is relatively stable to chemical splitting, there are numerous microorganisms capable of assimilating organophosphonates and transforming them into inorganic phosphates after cleavage of the C−P bond. These facts, together with data on the wide distribution of AEP and similar phosphonates, enables the assumption by Mastalerz and Kafarski [3] that phosphonates were present in primordial reducing environments and were used to build the first forms of life.

5.2
Physical/Chemical Properties and Analysis

Phosphonic and carboxylic acids differ significantly in many respects. The carboxylate function is flat while the phosphonate group is tetrahedral and considerably larger in size. Aminophosphonic acids have two acidic OH groups in contrast to amino carboxylic acids. The first OH group in phosphonic acids is very acidic, $pK_a \sim 0.5$–1.5. The second OH group has significantly lower acidity, $pK_a \sim 5.0$–6.5 (Scheme 5.3). Similar to α-amino acids, aminophosphonic acids have a zwitterionic form due to internal hydrogen bonding between the phosphonate and ammonium groups



Scheme 5.3

a situation that is strongly supported by nuclear magnetic resonance (NMR) and X-ray studies [6]. The acid–base properties of aminophosphonates depend on many factors such as electronic effects of the substituents, the number of phosphonic groups in a molecule, and the presence of other functional groups (e.g., COOH, C=O, NH, OH). The basicity of an amino group in dialkyl *N*-alkylaminomethylphosphonates **5** was determined to be around pK_a 5–6, and was significantly decreased for *N*-alkyl- and *N*-aryl-benzylphosphonates due to the different conformation for sterically hindered and nonhindered aminophosphonates [7]. Partition constants between water and some organic solvents ($CHCl_3$, toluene, *n*-octanol, etc.) were determined for a series of α-aminophosphonates. In the case of simple dialkyl *N*,*N*-dialkylaminomethanephosphonic acids, $Alk_2NCH_2P(O)(OH)_2$, a linear correlation between the partition constants and the number of carbon atoms in the molecule was observed [8].

The crystal structures of numerous aminophosphonic acids and their derivatives have been reported. The P−O bond lengths for *N*-unsubstituted aminophosphonic acids correspond to the zwitterionic character of the compounds and two categories of bonds are clearly distinguished. The single P−OH bond lengths have a value of 1.522–1.603 Å; the second type of bond, a partial double P−O bond, is shorter and has a mean value of 1.501 Å. In *N*-acylated aminophosphonic acids, PO_3H_2 groups are fully protonated and exhibit one short P=O bond (1.477 Å) and two long P−OH bonds (mean length 1.542 Å) [9]. NMR spectral characteristics of aminophosphonic acids and their derivatives are an important and common tool both for determination of structure peculiarities of aminophosphonic acids and identification of new and known compounds in reaction mixtures and biological samples [10]. For example, ^{31}P-NMR analysis has been successfully used to detect and discriminate aminophosphonic acids AEP **1** and $MeNHCH_2CH_2P(O)(OH)_2$ from other phosphoric acid derivatives in sea animals [11]. The solid-state ^{31}P-NMR cross-polarization magic angle spinning technique enables determination of the amount of C−P forms in the total phosphorus in sea anemones [12]. NMR-controlled titration methods can be applied to study protonation and metal complex formation of aminophosphonates [13]. Some limitations in the use of ^{31}P-NMR spectroscopy for the analysis of aminophosphonic acids and their derivatives originate from the relatively small differences in chemical shift for homologous aminoalkanephosphonic acids and the complex dependence of these parameters on the pH at which the spectra are recorded.

NMR spectroscopy represents a valuable tool for the determination of enantiomeric composition of aminophosphonates. In the majority of cases, additional *N*-derivatization of free aminophosphonates with chiral reagents to form covalent diastereoisomers has been required prior to analysis (e.g., [14]). The NMR methodology for a rapid assessment of the enantiomeric composition of various *N*-benzyloxycarbonyl-α-aminophosphonates has recently been elaborated with the use of commercially available cyclodextrins as a chiral solvating agent [15, 16].

High-performance liquid chromatography on chiral columns [17, 18] and capillary electrophoresis using quinine [19] or cyclodextrins [20, 21] as chiral discriminators have also been used for the analysis of enantiomeric composition

of chiral aminophosphonic acids. Derivatization of aminophosphonates is a necessary procedure to analyze compounds by gas chromatography methods (either alone or coupled with mass spectrometry). The development of new mass spectrometry techniques has also made possible the direct identification of aminophosphonic acids and their derivatives [22].

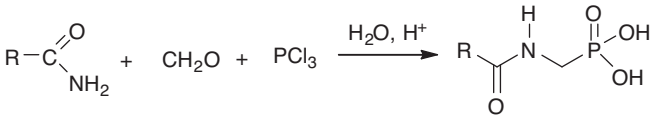
Aminophosphonates and, specifically, poly-aminophosphonic acids are potent chelating agents for a variety of metal ions including the alkaline earth ions, and the divalent and trivalent metal ions. The coordination chemistry of aminophosphonic acids has developed significantly in the last three decades due to their versatility in adopting monodentate, bridging, and chelating modes of coordination [6]. The mentioned properties of aminophosphonic acids and their derivatives are especially important from the point of view of their high biological activity, and wide range of uses for industrial, agricultural, and pharmacological purposes.

5.3
Synthesis of α-Aminophosphonic Acids

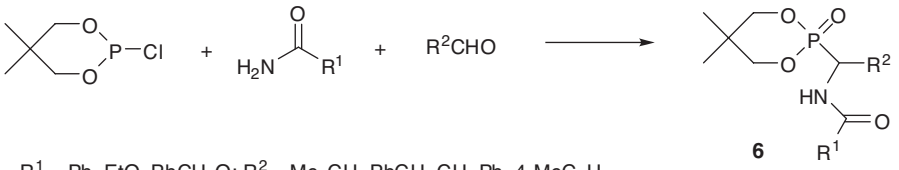
5.3.1
Amidoalkylation in the "Carbonyl Compound–Amine–Phosphite" Three-Component System

The three-component condensation involving an amide, formaldehyde, and phosphorus trichloride was first mentioned in the literature in the early 1940s in a patent describing the synthesis of *N*-acylaminophosphonic acids (Scheme 5.4) [23]. Since that time, the range of functionalized α-aminophosphonic acids has been considerably extended by the use of three-component systems in which the amide is replaced by a urea or a carbamate, and phosphorus trichloride by various P(III)−Cl and P(III)−OR derivatives [24–27].

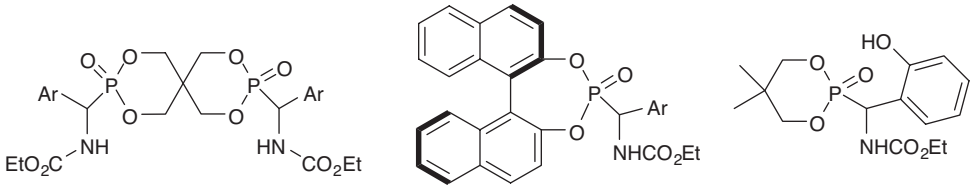
In recent years several improved procedures for the preparation of *N*-protected α-aminoalkylphosphonates have been reported [28–33]. For example, diphenyl α-aminomethylphosphonates $R^3OC(O)NHCR^1R^2P(O)(OPh)_2$ have been synthesized in good yields by a Lewis acid-catalyzed reaction of a carbamate, a carbonyl compound, and triphenyl phosphite [34]. By using acetyl chloride both as solvent and reagent, benzyl carbamate was found to react with $RPCl_2$ (R = MeO, EtO, *n*PrO, *n*BuO, PhO, or Ph) and aromatic aldehydes to give *N*-protected aminoalkanephosphonic acids, $PhCH_2OC(O)NHCHArP(O)(OR)(OH)$, in good yields [35]. Treatment of diethyl thiophosphoramidate, $(EtO)_2P(S)NH_2$, with triphenyl phosphite and



Scheme 5.4



R^1 = Ph, EtO, $PhCH_2O$; R^2 = Me_2CH, PhCH=CH, Ph, 4-MeC_6H_4, 4-$MeOC_6H_4$, 4-ClC_6H_4, 3-MeC_6H_4,1-naphthyl, 9-antryl

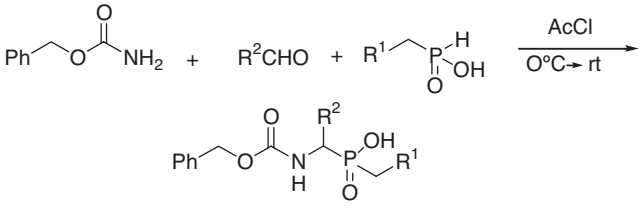


Scheme 5.5

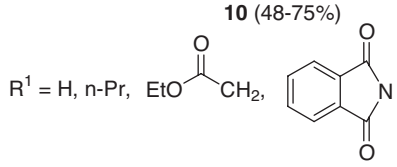
a substituted benzaldehyde or ketone in acetyl chloride leads to α-(diethoxythiophosphorylamino)methylphosphonate, $(EtO)_2P(S)NHCR^1R^2P(O)(OEt)_2$ [36]. Arenesulfonamides can also be used in the preparation of α-aminoalkanephosphonic acids. For example, a vast range of cyclic esters of *N*-toluenesulfonyl α-aminophosphonic acid have been prepared by the three-component reaction of *p*-toluenesulfonamide, an aromatic aldehyde, and 2-chloro-1,3,2-benzodioxaphosphole [32]. The related reactions of *p*-toluenesulfonamide and an aldehyde with PCl_3 and $RPCl_2$ (R = MeO, EtO, *n*Pr, *n*Bu, or Ph) were also studied [35]. The use of cyclic chlorophosphites as scaffolds for the one-pot synthesis of α-aminophosphonates have been described by Kumara Swamy *et al*. [37]. α-Aminophosphonates **6** have been synthesized in high yields by a three-component reaction using a cyclic chlorophosphite, benzamide (or urethane or a carbamate), and an aldehyde under solvent-free conditions. This route has been adopted for bisaminophosphonates **7** and optically active binaphthoxy aminophosphonates **8**. The method tolerates the phenolic OH group as shown by the synthesis of hydroxy-functionalized aminophosphonates **9** (Scheme 5.5) [37].

The condensations of an aldehyde, benzyl carbamate, and an alkylphosphonous acid in acetyl chloride provide a general method for the direct synthesis of α-aminophosphinic acids **10** (Scheme 5.6) [38]. This approach has recently been used for the preparation of Fmoc-protected phosphinic pseudodipeptidic compounds [39].

The diastereoselectivity of one-pot three component reactions employing chiral cyclic carbamates or ureas depends on the nature of the chiral auxiliary. The highest diastereomeric excess in the synthesis of α-aminophosphonates from diethyl phosphite, benzaldehyde, and amides **A**–**C** was reached for derivatives of camphorsulfonic acid **C** (Scheme 5.7) [40].



10 (48-75%)

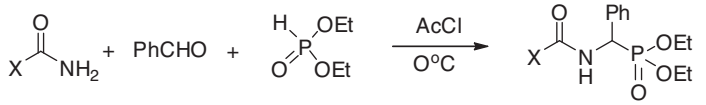


R^2 = H, Me, Et, i-Bu, Ph, 4-$MeOC_6H_4$

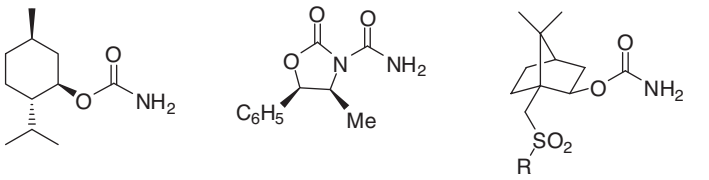
Scheme 5.6

5.3.2
Kabachnik–Fields Reaction

A three-component reaction of an aldehyde/ketone, an amine and a dialkyl phosphite is commonly known as the Kabachnik–Fields reaction (Scheme 5.8). Due to the immense potential scaffold variability, this reaction has become very popular in the synthesis of arrays of α-aminophosphonates [27, 41]. Its reaction mechanism is



A, **B** or **C** **AP**

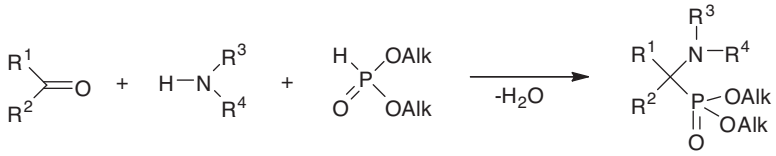


A **B** **C**

a, R = (cyclo-$C_6H_{11})_2N$
b, R = i-Pr_2N
c, R = $MeOC_6H_4$

Urea/carbamate	**A**	**B**	**C**(a)	**C**(b)	**C**(c)
Yield of AP	71	84	75	73	77
de (%)	14.4	34.1	99	96.4	96.2

Scheme 5.7



R^1, R^2, R^3, R^4 = Alk, Ar or Het
Alk = Me, Et, i-Pr, n-Bu

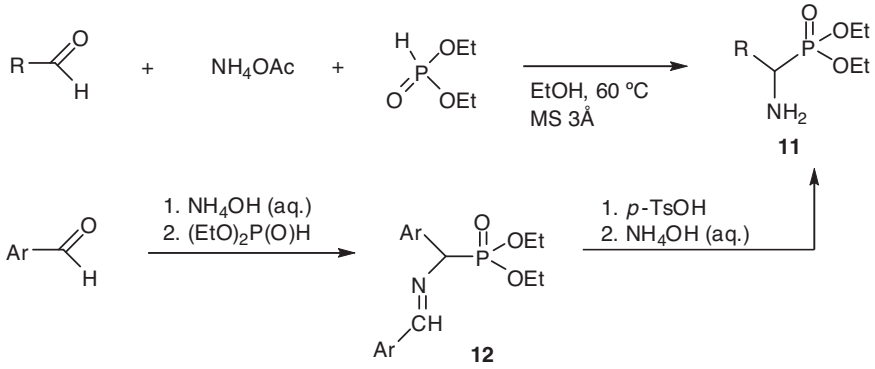
Scheme 5.8

believed to involve either nucleophilic addition of phosphite to an imine, formed from a carbonyl reagent and an amine, or nucleophilic substitution of a hydroxyl group by an amine moiety in an intermediary formed α-hydroxyalkanephosphonate. A review by Cherkasov and Galkin demonstrates the diversity of factors that determine the pathway of the Kabachnik–Fields reaction [42].

Typically, α-aminoalkanephosphonates are formed in good yield by treating a carbonyl compound and an amine with a dialkyl phosphite in the presence of an organic base, an alkali metal alkoxide, or a Brønsted/Lewis acid. Recent preparative modifications of the reaction which significantly simplify the original method and increase the yield of the product concern the use of microwave irradiation [44–49], silica gel-supported reagents [43, 46, 49, 50], surface-mediated reactions on Al_2O_3 [51, 52], ionic liquids as solvents [53], solvent-free conditions [43, 44, 47, 54, 55], and water-tolerant catalysts (magnesium perchlorate [56, 57], lithium perchlorate [58, 59], metal triflates [60–62], indium trichloride [63], samarium diiodide [64], lanthanide trichloride [65], $TaCl_5/SiO_2$ [66], bismuth nitrate pentahydrate [67], scandium tris (dodecyl sulfate) [68], tetrabromomethane [54], and montmorillonite KSF [69]).

Perhaps the most direct and simple method for the preparation of α-aminophosphonates **11** bearing a primary amino group is the treatment of an aldehyde with an ethanolic solution of ammonium acetate and dialkyl phosphite in the presence of molecular sieves [70]. When aromatic aldehydes were allowed to react with aqueous ammonia and diethyl phosphite, benzylidene derivatives of α-aminophosphonates **12** were obtained. The latter can be easily hydrolyzed to diethyl α-aminophosphonates **11** (Scheme 5.9) [71]. The reaction of aromatic aldehydes with diethyl phosphite and hexamethyldisilazane (HMDS) on the surface of acidic Al_2O_3 goes in a similar manner [72]. It was also found that silica-supported NH_4HCO_3 is an effective reagent for the synthesis of α-aminophosphonates under microwave irradiation in solvent-free conditions [50]. The treatment of aromatic aldehydes with ammonia followed by reaction with diethyl phosphite in the presence of acetyl chloride as catalyst gives bis[aryl(diethoxyphosphoryl)methyl] amines, $[(EtO)_2P(O)CHAr]_2NH$ [73].

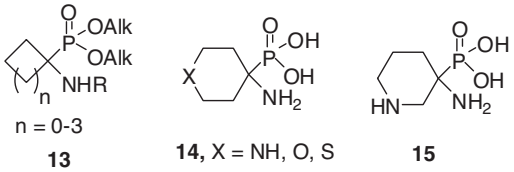
The applicability of the Kabachnik–Fields reaction to primary and secondary amines is well documented [74–78]. For further details of this subject, readers are referred to the leading references cited in a monograph [27] and reviews [41, 42]. In recent years, many new cyclic α-aminophosphonic acids have been prepared, such as α-aminocyclopropylphosphonic acids, as well as their cyclobutyl, cyclopentyl,



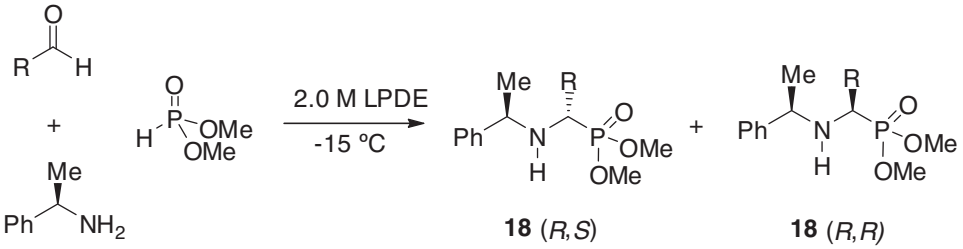
R = Ph, 4-MeOC$_4$H$_4$, i-Pr, i-Bu (yield 32-59%)
Ar = Ph, 4-ClC$_6$H$_4$, 4-Me$_2$NC$_6$H$_4$, 4-MeOC$_6$H$_4$, 4-BrC$_6$H$_4$, 3-ClC$_6$H$_4$, 3-MeC$_6$H$_4$, 1-naphthyl, 2-naphthyl (yield 51-81%)

Scheme 5.9

and cyclohexyl analogs **13** [60, 79–82]. 4-Heterocyclohexyl phosphonates **14** and 3-heterocyclohexylphosphonates **15** were efficiently prepared from heterocyclic ketones via phosphite addition to iminium ions formed by *in situ* condensation of these ketones with benzylic amines [83]. Phosphorylated oxazolidine **16** was synthesized in a one-pot procedure from (*R*)-(−)phenylglycinol, formaldehyde, and dimethyl phosphite in boiling methanol [84]. Potentially bioactive phosphonomethylaminocyclopentane-1-carboxylic acid **17** was formed in good yield from aminocyclopentane-1-carboxylic acid (cycloleucine), paraformaldehyde and diethyl phosphite in the presence of triethylamine in methanol [85].



The broad diversity of possibilities for the synthesis of α-aminophosphonic acids by the Kabachnik–Fields reaction is due to the variability of the structures – not only carbonyl and amine components, but also phosphite derivatives. The simplest variant of *P*-modified reactions is the use of a three-component system in which a dialkyl phosphite is replaced by a neutral trivalent phosphite. As in a classical version of the Kabachnik–Fields reaction, in this case it is possible to carry out the transformation by a one-pot technique. The reactions occurred in a solution or in solvent-free



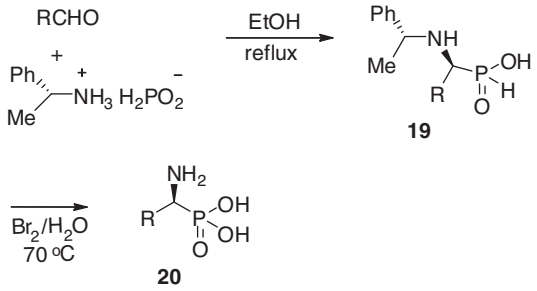
R	yield (%)	(*R*,*S*):(*R*,*R*)
i-Pr	95	79:21
t-Bu	96	82:18
c-Hex	92	80:20
Bn	90	83:17

Scheme 5.10

conditions in the presence of catalysts [86–90] or without them at all [91, 92], sometimes at room temperature.

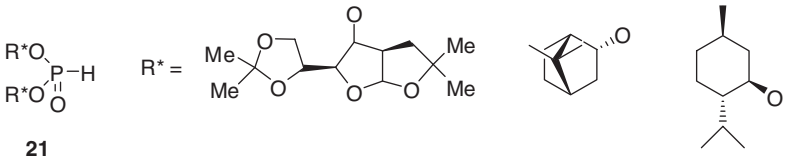
A few examples of the asymmetric synthesis of α-aminophosphonic acids based on the Kabachnik–Fields reaction have been published [27]. Dimethyl phosphite undergoes condensation with various aldehydes and (*R*)-(+)-phenylethylamine in the presence of lithium perchlorate-diethyl ether (LPDE) solution to give α-aminophosphonates **18** in high yields and with good diastereoselectivity (Scheme 5.10) [59]. A highly convenient synthesis of α-aminophosphonic acids in high optical purity was described by Hamilton *et al.* [93]. Chiral salts of hypophosphorous acid with (*R*)-(+)- or (*S*)-(−)-1-phenylethylamine reacted with aldehydes in boiling ethanol to give homochiral α-aminophosphonates **19**. These can be simultaneously deprotected and oxidized by bromine-water in one step to provide α-aminophosphonic acids **20** as single enantiomers (Scheme 5.11).

A route to diastereomeric α-aminophosphonates starting from chiral dialkyl phosphites has also been reported. The reactions of dialkyl phosphites **21** with benzaldehyde and benzylamine proceed with modest to good stereoselectivity (50–90% d.e.) [94].



R = n-Pr, i-Pr, cycloxehyl, Ph, Bn

Scheme 5.11



21

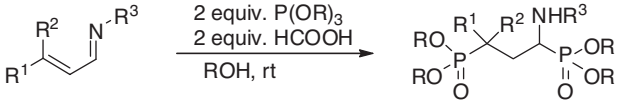
5.3.3
Direct Hydrophosphonylation of C=N Bonds

Among the numerous methods for the preparation of α-aminophosphonic acids, the most general route is the addition of dialkyl phosphites to Schiff's bases. The synthetic potential of the method has been covered in a series of reviews [27, 42, 95]. During the last 10 years, direct hydrophosphonylation of imines has been widely used for the preparation of multifunctional α-aminophosphonic acids without the excessive use of protective groups [96–103].

Trialkyl phosphites were evaluated for addition reactions to α,β-unsaturated imines. An acidic medium is required to allow tandem addition to occur. In this manner, 3-phosphonyl-1-aminophosphonates **22**, phosphonic acid analogs of glutamates, are obtained in good yields (Scheme 5.12) [97]. Analogous results were also obtained with α,β-unsaturated hydrazones, which gave 3-phosphonyl-1-hydrazinoalkylphosphonates [96].

Diethyl aryl[(arylmethylidene)amino]methylphosphonates **12**, prepared from aromatic aldehydes, ammonia, and diethyl phosphite (see Scheme 5.9), easily react with diethyl phosphite in the presence of chlorotrimethylsilane to give bis(1-diethoxyphosphorylalkyl)amines **23** [98]. Another route to the formation of compounds **23** which has recently been investigated is the reaction of aromatic diimines **24** with diethyl phosphite (Scheme 5.13) [104].

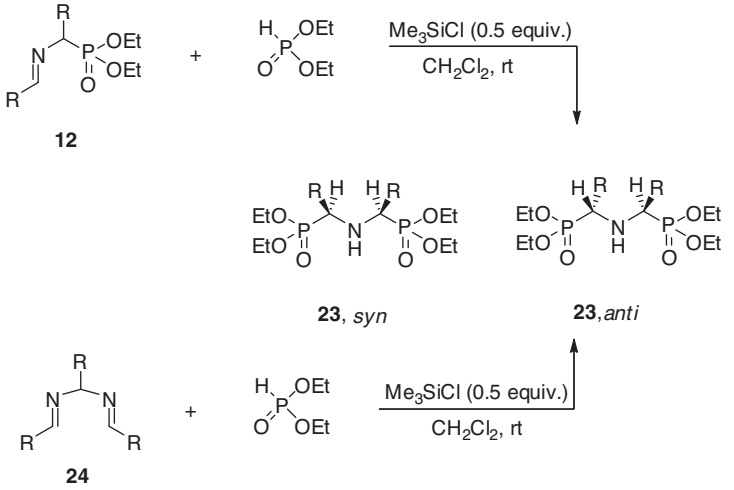
An interesting modification of the hydrophosphonylation reactions was described Vasella *et al.* who used readily available nitrones as the C=N-containing substrates [105]. The scope of the method has been now extended to diastereoselective synthesis of α-(hydroxyamino)phosphonates and α-aminophosphonates. It was



22 (32-90 %)

R = Me, Et
R^1 = Ph, furan-2-yl, Me
R^2 = H, Me
R^3 = Ph, i-Pr, allyl, Bn, 4-$MeOC_6H_4CH_2$

Scheme 5.12

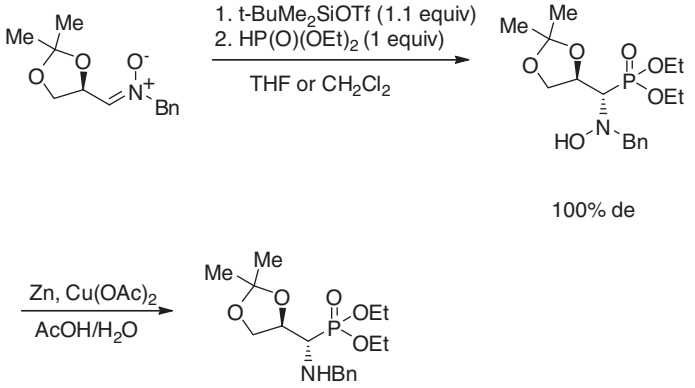


R = Ph, 4-MeOC$_6$H$_4$, 4-ClC$_6$H$_4$, 4-FC$_6$H$_4$, 4-MeC$_6$H$_4$, 3-FC$_6$H$_4$, 3-NO$_2$C$_6$H$_4$, 2-naphthyl

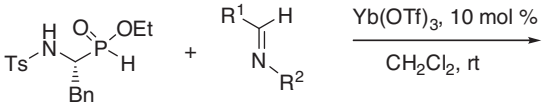
Scheme 5.13

shown that the stereoselective C−P bond-forming reaction between diethyl phosphite and *N*-benzyl nitrones derived from chiral α-alkoxy and α-(Boc-amino) aldehydes proceeded smoothly to give α-(hydroxyamino)phosphonate intermediates, which were subsequently converted into the corresponding polyhydroxylated α-amino- and α,β-diaminophosphonates. A representative example of such methodology is shown in Scheme 5.14 [106].

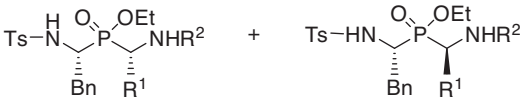
Relatively few papers have reported the stereoselective addition of chiral hydrophosphoryl compounds to imines. In general, such reactions give desired products with low-to-moderate diastereoselectivity [27]. Kaboudin *et al.* used α-substituted α-amino-*H*-phosphinates **25** in the Lewis acid-catalyzed hydrophosphinylation of



Scheme 5.14



25



26A **26B**

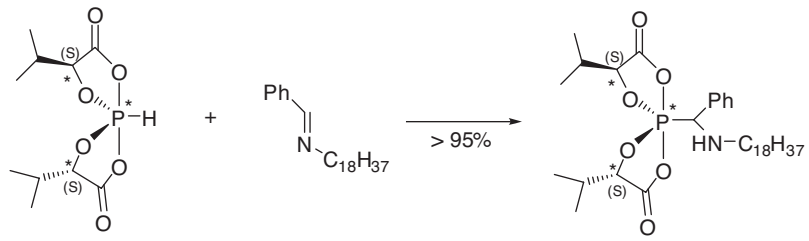
R^1 = i-Pr, t-Bu, Ph, 1-naphthyl, 2-naphthyl, 2-furyl

R^2 = Ph, Bn

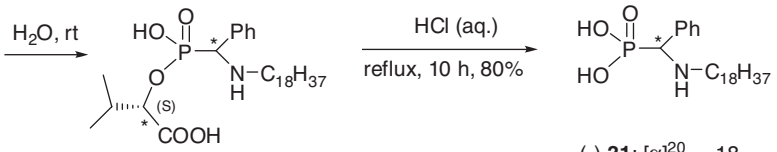
Scheme 5.15

imines. In the presence of $Yb(OTf)_3$, α,α′-diaminophosphinic derivatives **26A** and **26B** were obtained with diastereomeric excesses ranging from 10 to 95% (Scheme 5.15). Ytterbium triflate afforded the highest diastereomieric excess and yield in comparison to other catalysts such as scandium triflate or $BF_3{\cdot}Et_2O$ [107].

The asymmetric synthesis of α-aminophosphonic acid amphiphiles using chiral P−H spirophosphoranes has been developed by French chemists [108]. An illustrative example of this strategy is described in Scheme 5.16. The addition reaction between spirophosphorane **27** derived from (*S*)-α-hydroxyisovaleric acid and long-



27 (two diastereoisomers) **28** **29** (four diastereoisomers)



30A, **30B** (two diastereoisomers)

(-)-**31**: $[\alpha]_D^{20}$ = -18

(+)-**31**: $[\alpha]_D^{20}$ = +18

Scheme 5.16

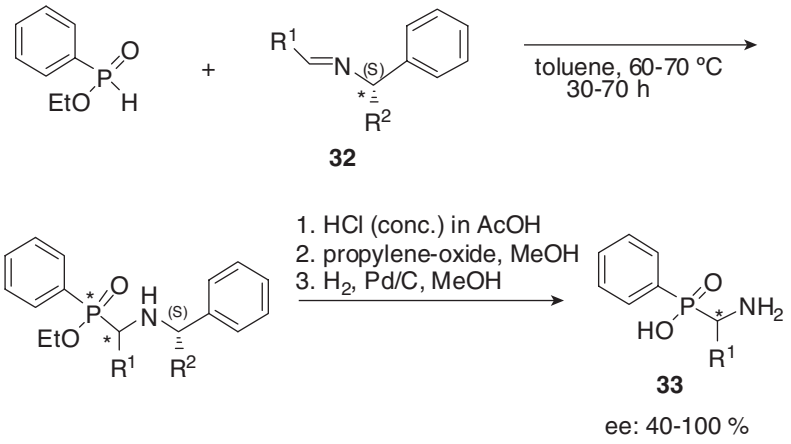
chain aldimine **28** occurred instantaneously to give the corresponding spirophosphorane **29** as a mixture of four diastereoisomers. Smooth hydrolysis of this mixture furnished, nearly quantitatively, two diastereoisomers **30A** and **30B**, which were separated and purified by selective crystallization, yielding enantiopure forms. The enantiopure monoesters **30** were then converted into the corresponding enantiopure free α-aminoalkanephosphonic acids **31**.

Dialkyl phosphites add to aldimines derived from chiral amines to afford α-aminophosphonates bearing an enantioenriched α-carbon center. The first application of this strategy demonstrated low or moderate diastereoselectivity, but subsequent studies allowed the stereochemical results of the method to be substantially improved [27]. Diastereoselective phosphonylation of the chiral nonracemic imine PhCH=NCH(Me)Ph with diethyl phosphite in the presence of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) and with triethylphosphite in the presence of $BF_3{\cdot}Et_2O$ was applied to prepare (*S*)- and (*R*)-α-aminoalkanephosphonic acids $(HO)_2P(O)CH(NH_2)Ph$ [109]. The same strategy was used for the asymmetric synthesis of (*R*)- and (*S*)-α-amino-3-piperidinyl-phosphonic acids via phosphite addition to iminium ions [110]. The asymmetric synthesis of α-aminophosphonothionates via diastereoselective addition of dimethyl thiophosphite (DMTP), $(MeO)_2P(S)H$, to benzaldimines bearing a chiral auxiliary group has also been examined. The addition of DMTP to the benzaldimines derived from (*S*)-phenylglycinol, (*S*)-$H_2NCH(Ph)CH_2OH$, afforded the highest diastereoselectivity (83 : 17), whereas additions of DMTP to the benzaldimines derived from (2*S*,3*R*)-L-threonine methyl ester and (*S*)-L-alanine methyl ester were less diastereoselective, affording 38 : 62 and 61 : 39 ratios, respectively [111].

The first enantioselective synthesis of α-aminophosphinic acids **33** was carried out by the addition of ethyl phenylphosphinate to chiral imines **32** in the absence of base or other catalyst (Scheme 5.17) [112].

Chiral enantiopure sulfinimines have been found to be especially useful substrates for stereoselective synthesis of α-aminophosphonates [113–117]. Thus, the use of dialkyl or diamido phosphite anions $R_2P(O)M$ (R = MeO, EtO, *i*PrO, Me_2N,or Et_2N; M = Li or Na) in reaction with enantiopure sulfinimines *p*-Tol-S(O)−N=C(H)R results in the formation of the diastereomeric adducts **35** that may be separated and converted into enantiomerically pure α-aminophosphonic acids. The highest diastereomeric ratio (about 94 : 6) was observed in the addition of lithium dimethyl phosphite to the sulfinimines (+)-(*S*)-**34a**, (+)-(*S*)-**34b**, and (+)-(*S*)-**34c** (Scheme 5.18) [118, 119]. Further improvement of this methodology is based on the asymmetric synthesis of α-aminophosphonic acids via reaction of (*S*)- and (*R*)-sulfinimines with a lithiated diaminophosphine borane complex (Scheme 5.19). The diastereomeric addition products **36** were converted into the corresponding free α-aminophosphonic acids **37** by heating in a refluxing mixture of acetic acid and hydrochloric acid. The α-aminophosphonic acids (−)- and (+)-**37a**, (−)-**37c**, and (−)-**37d** were obtained as practically pure enantiomers (≥98% e.e.) [120].

Catalytic asymmetric hydrophosphonylation of imines was first reported by Shibasaki *et al.* in 1995 [121]. The optically active α-aminophosphonates **39** have been prepared by addition of dimethyl phosphite to imines using lanthanoid–

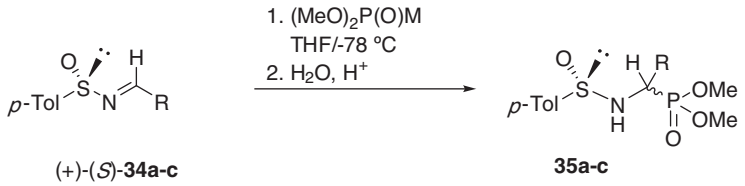


R^1 = substituted Ph, i-Bu
R^2 = Me, $MeOCH_2$

Scheme 5.17

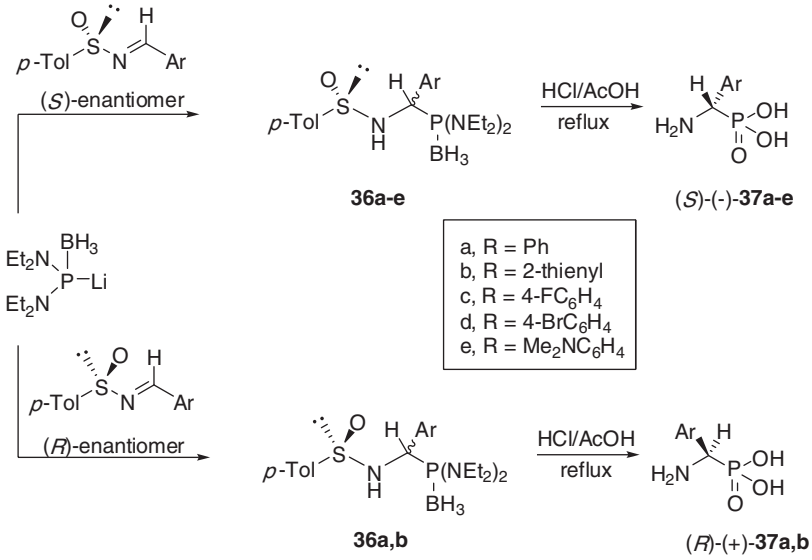
potassium–(*R*)-1,1′-bi-2-naphtol (BINOL) complexes (lanthanoid–potassium–(*R*)-1,1′-bi-2-naphtolLnPB) (Scheme 5.20). Optical yields in this reaction depend on the nature of the lanthanoid, solvent, and substituents in the imine. LnPB complexes were also successfully applied in the asymmetric synthesis of thiazolidinyl-phosphonates [122, 123]. Furthermore, a cyclic phosphoric acid derivative **38**, derived from (*R*)-BINOL, was used as a chiral Brønsted acid (10 mol%) in the hydrophosphonylation of aldimines with diisopropyl phosphite at room temperature. α-Aminophosphonates $R^1CH(NHR^2)P(O)(OiPr)_2$ (R^1 = Ph, 2-MeC_6H_4, 2-$NO_2C_6H_4$, ArCH=CH, or 1-NaphCH=CH; R^2 = 4-$MeOC_6H_4$) were obtained with good to high enantioselectivities (52–90% e.e.) [124].

A very high enantioselectivity in catalytic hydrophosphonylation of imines was observed when certain chiral thioureas were used as organocatalysts. In particular, highly enantiomerically enriched α-aminophosphonates **41** could be efficiently prepared by the addition of phosphite **40** to *N*-benzyl imines catalyzed by chiral thiourea **43**. Deprotection of the phosphonates **41** under mild conditions provided the corresponding α-aminophosphonic acids **42** (Scheme 5.21) [125].

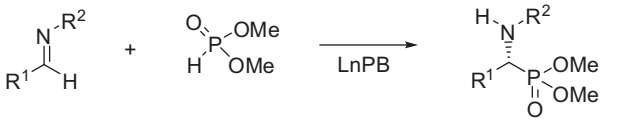


R = Ph (a), α-furyl (b), α-thienyl (c)

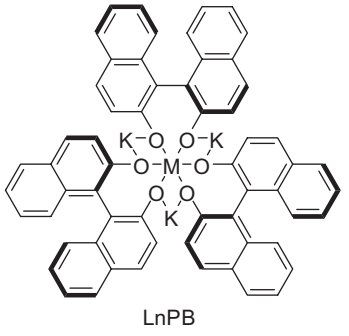
Scheme 5.18



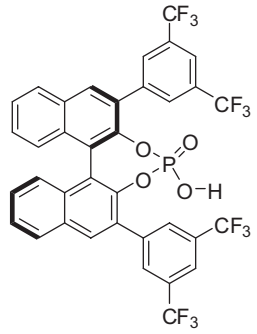
Scheme 5.19



R^1 = Me, Et, i-Pr, C_5H_{11}, *cyclo*-Hex, (*E*)-PhCH=CH
R^2 = Tr, Ph_2CH, *p*-$MeOC_6H_4$

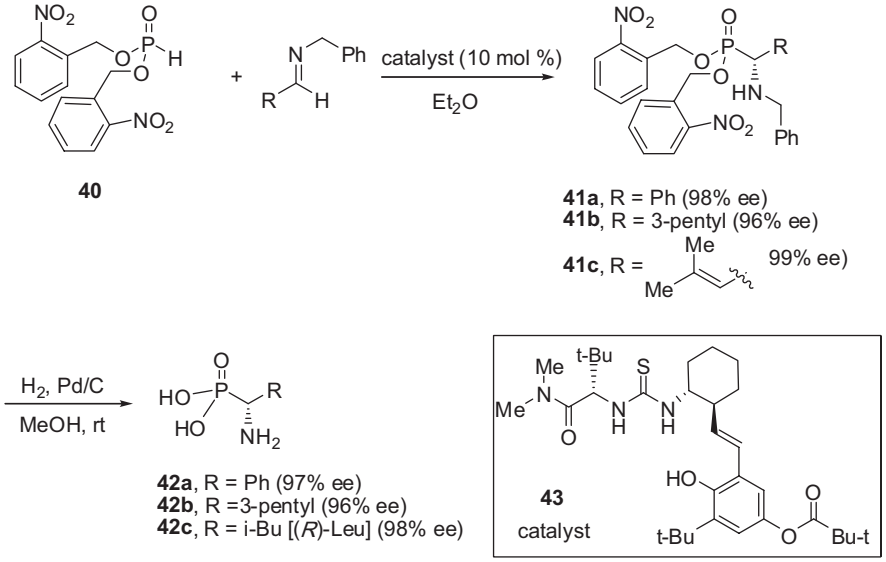


M = La, Pr, Gd, Sm, Yb



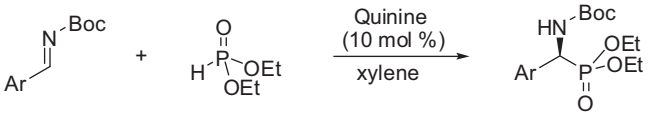
38

Scheme 5.20



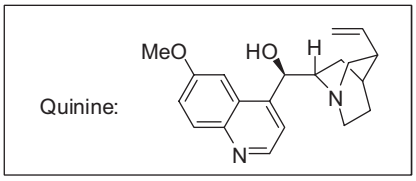
Scheme 5.21

Particularly noteworthy is a new organocatalytic approach for hydrophosphonylation of imines using commercially available and nonexpensive cinchona alkaloids as catalysts. For example, by using 10 mol% of quinine as the catalyst in the enantioselective addition of dimethyl phosphite to *N*-Boc-protected imines, α-aminophosphonates **44** have been obtained in moderate to good yields and with up to 94% e.e. (Scheme 5.22) [126].



44 (72-94% ee)

Ar = Ph, 3-MeC_6H_4, 4-MeC_6H_4, 2,5-Me$_2C_6H_3$, 4-MeOC_6H_4, 1-naphthyl, 2-naphthyl, 3-pyridyl

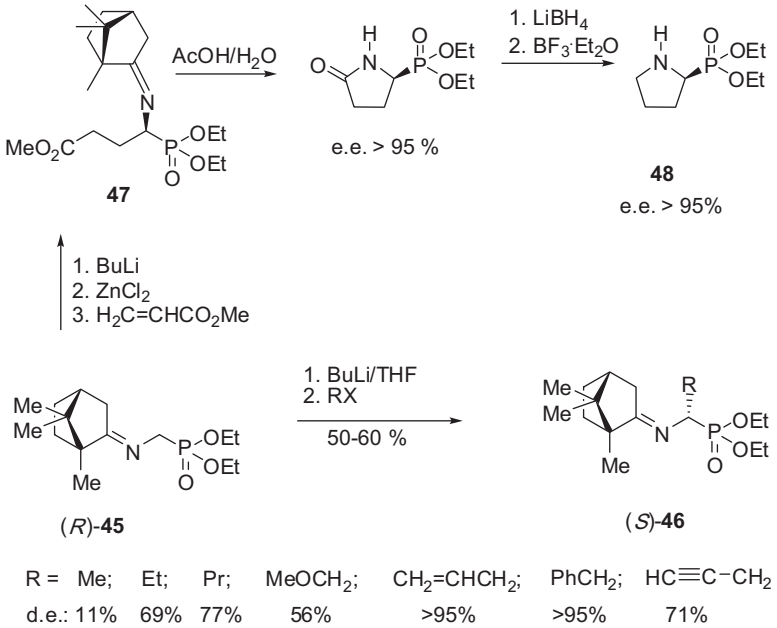


Scheme 5.22

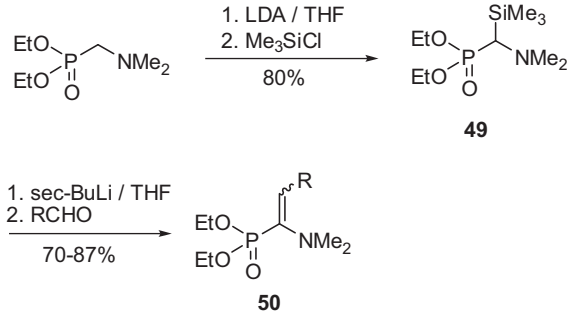
5.3.4
Syntheses using C–N and C–C Bond-Forming Reactions

The alkylation of chiral Schiff's bases derived from esters of aminomethylphosphonic acid and appropriate carbonyl compounds represents one of the most widely used approaches for the asymmetric synthesis of α-aminophosphonic acids. In most cases the Schiff's base is deprotonated with *n*-butyl-lithium or lithium diisopropylamide and the carbanion formed is made to react with alkyl halides, carbonyl compounds, or Michael acceptors. This topic has been summarized [27, 95]. Illustrative examples are shown in Scheme 5.23. Thus, the lithium derivative of phosphonate **45** reacts with alkyl, allyl, or benzyl halides to give the corresponding diethyl (*S*)-α-aminophosphonates **46** with 11 to greater than 95% d.e. [127]. The additional reaction of the lithium derivative with methyl acrylate proceeds with high diastereoselectivity to give the ester of the α-mono-phosphonic analog of glutamic acid **47**. The latter was transformed into (*S*)-pyrrolidine-2-phosphonic acid ester **48** [128].

Phosphonate carbanions bearing a carbanion-stabilizing group on the α-carbon are effective partners in the reactions with aromatic and aliphatic aldehydes. The synthetic utility of functionalized α-silyl α-amino carbanions, in particular, has been demonstrated to be a stereoselective method for the synthesis of functionally rich α-phosphonoenamines based on a modified Peterson olefination. The carbanion derived from α-silyl-substituted methylphosphonate **49** reacts with aldehydes to



Scheme 5.23



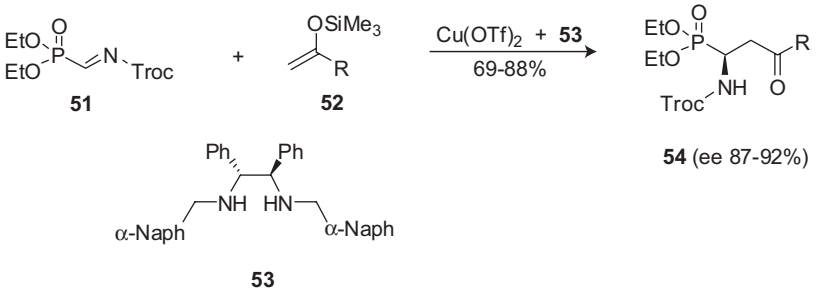
R = Ph, substituted Ph, $PhCH_2CH_2$, C_5H_{11}, i-Bu, α-furyl

Scheme 5.24

give enamines **50** in 70–87% isolated yield and with high (*E*) stereoselectivity (Scheme 5.24) [129].

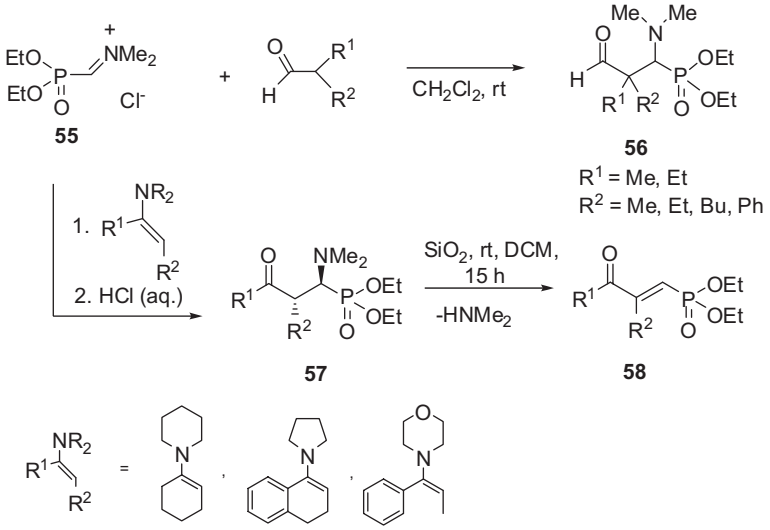
A novel, catalytic asymmetric synthesis of α-aminophosphonates using enantioselective C−C bond-forming reactions was recently reported by Kobayashi *et al.* [130]. Treatment of *N*-acyl-α-iminophosphonate **51** with silyl enol ethers **52** in the presence of a Cu(II) complex derived from $Cu(OTf)_2$ and diamine **53** lead to α-aminophosphonates **54** in high yields and selectivities (Scheme 5.25). Interestingly, the direct aminoalkylation of carbonyl compounds can be performed with the *C*-phosphoryl aldiminium salt **55**. Thus, by reacting **55** with α-branched aldehydes at room temperature products **56** have been isolated in good yields. However, the diastereoselectivity of this reaction is very low. In contrast, aminoalkylation of enamines proceeded with very high diastereoselectivity as only one diastereoisomer **57** was observed by NMR spectroscopy. Mannich bases **57** easily eliminate the amino group by simple stirring a solution of **57** in dichloromethane in the presence of silica gel to give substituted vinylphosphonates **58** (Scheme 5.26) [131].

Another example of C−C bond formation for the synthesis of α-aminophosphonates is the addition of terminal alkynes to α-iminophosphonates **59**. An



R = Me, Ph, substituted Ph, 1-naphthyl, 2-naphthyl

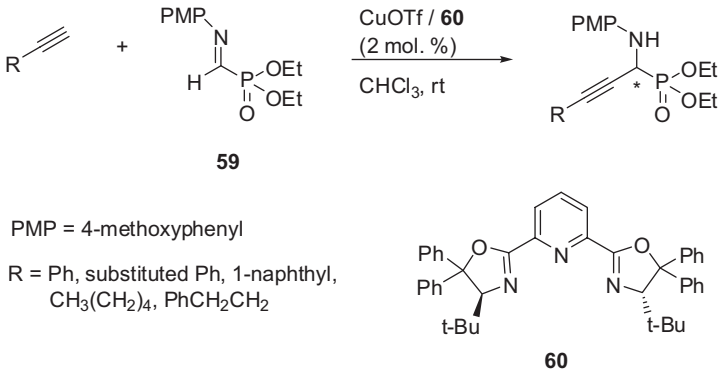
Scheme 5.25 Troc = 2,2,2-trichloro-ethoxycarbonyl.



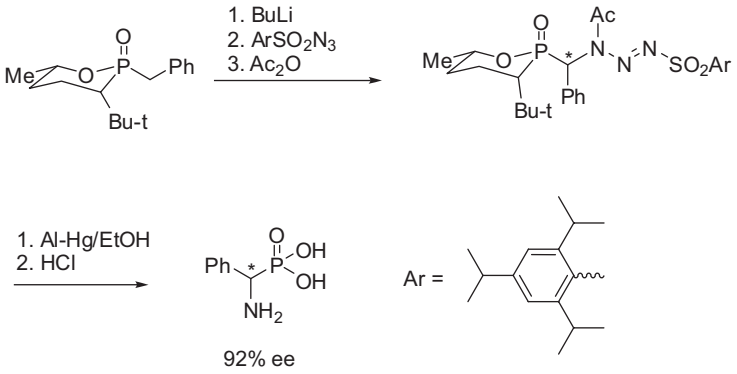
Scheme 5.26

enantioselective version of this reaction has been achieved by using Cu(I) triflate in the presence of the ligand **60** (Scheme 5.27) [132]. A Ag(I) triflate-catalyzed one-pot, three-component reaction between *p*-anisidine, diethyl formylphosphonate hydrate, and terminal alkynes with aromatic substituents for the direct synthesis of *N*-*p*-methoxyphenyl-protected α-aminopropargylphosphonates has also been described [133].

The syntheses of α-aminophosphonates based on the introduction of an amino function on the α-carbon have also developed quickly in recent times [27]. Both traditional methods for the formation of the C−N bond such as electrophilic amination of a formally α-phosphonate carbanion and specific methods of functional group transformations have been applied. The route via the electrophilic amination of chiral α-phosphonate carbanions is connected with the search for methods for the



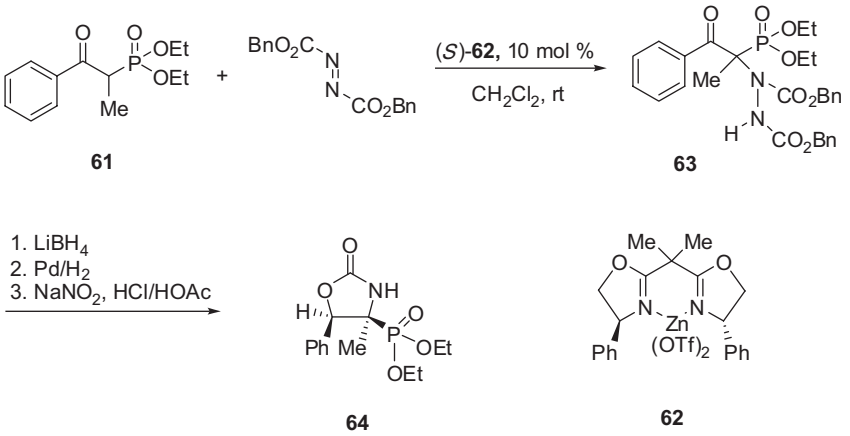
Scheme 5.27



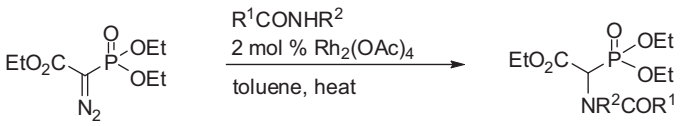
Scheme 5.28

stereoselective synthesis of α-aminophosphonic acid derivatives. A representative example of such strategy is shown in Scheme 5.28 [134].

A direct enantioselective amination of β-keto phosphonates giving optically active α-aminophosphonic acid derivatives in good yields and very high enantioselectivities has been proposed by Jørgensen *et al.* [135]. Acyclic β-keto phosphonates bearing alkyl, benzyl, naphthyl, or phenyl substituents reacted smoothly with dibenzyl azodicarboxylate in the presence of chiral Zn(II) complexes to give the corresponding α-aminated adducts with high enantiomeric excesses. For example, the reaction of the β-keto phosphonate **61** with dibenzyl azodicarboxylate in the presence of (*S*)-**62** afforded the corresponding aminated adducts **63** in 85% yield and with high enantioselectivity (92% e.e.). Reduction of the β-keto functionality in **63**, further deprotection, and N−N bond cleavage afforded oxazolidinyl-phosphonic acid derivative **64** in 60% overall yield and with a diastereomeric ratio of 10 : 1 (Scheme 5.29).



Scheme 5.29



65 **66**

R^1 = t-BuO, $PhCH_2O$, Me, Et, MeNH

R^2 = H, Pr

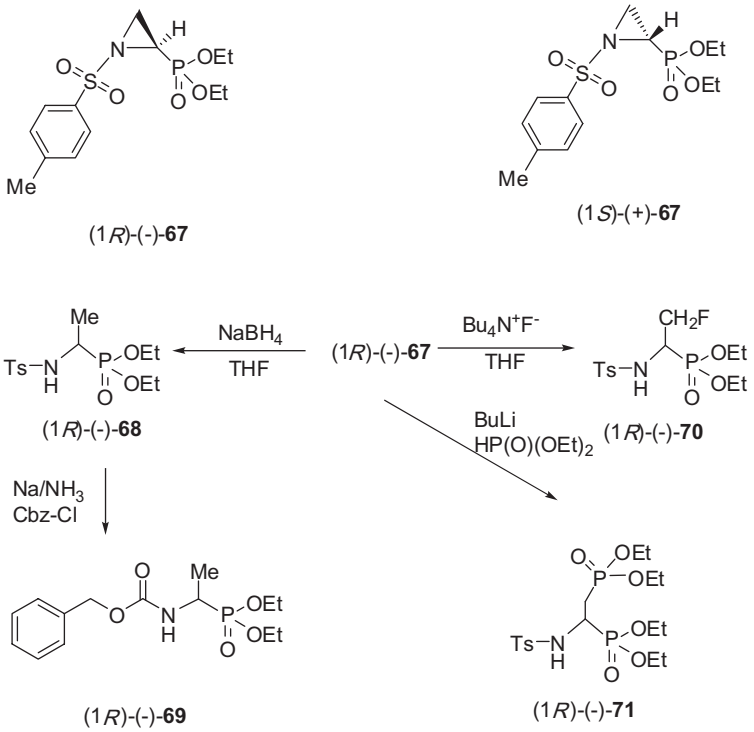
Scheme 5.30

The possibility of preparation of α-aminophosphonates via the insertion of carbenoids into the NH bonds of amines, amides, and carbamates has received special attention. It has been shown that the Rh(II) acetate-catalyzed decomposition of dimethyl α-diazobenzylphosphonate, $PhC(N_2)P(O)(OMe)_2$, in the presence of a range of N−H compounds results in an N−H insertion reaction of the intermediate carbenoid and formation of α-aminophosphonates $PhCH(NHR)P(O)(OMe)_2$ [136]. Similarly, the $Rh_2(OAc)_4$-catalyzed reaction of the diazaphosphonate **65** with *tert*-butyl or benzyl carbamates gave the corresponding N−H insertion products **66**. Several other N−H compounds were investigated: primary amides (acetamide, propionamide), *N*-Boc-propylamine, and *N*-methylurea all gave the corresponding *N*-acylaminophosphonoacetates (Scheme 5.30) [137]. In a related area, 1-trifluoromethyl-1-diethoxyphosphoryl carbine generated via a catalytic decomposition of 1-diaza-2,2,2-trifluoroethylphosphonate, $CF_3C(N_2)P(O)(OMe)_2$, constitutes another good synthon for the preparation of CF_3-containing α-aminophosphonic acid derivatives [138]. A rhodium-catalyzed N−H insertion reaction of an α-diazophosphonoacetate on a polymer support has also been reported [139].

Recent work by Dolence and Roylance utilizes both enantiomers of protected aziridine 2-phosphonate **67** as chiral synthons for the synthesis of β-substituted α-aminophosphonates. The aziridines arise from either (*R*)- or (*S*)-phosphonoserine diethyl esters followed by *N*-tosylation, *O*-mesylation, and cyclization with sodium hydride. These highly enantio-enriched aziridine 2-phosphonates react with various nucleophiles to give β-substituted α-aminophosphonates **68–71** in either the (*R*)- or (*S*)-configurations (Scheme 5.31) [140].

Another efficient synthesis of 1,2-diamino-, 1-amino-2-hydroxy-, and 1-amino-2-chloro-2-phenylethylphosphonates **73–76** utilizes the formation of an aziridinium ion from dimethyl ($1R^*,2S^*$)-2-(*N,N*-dibenzylamino)-1-hydroxy-2-phenylethylphosphonate **72** and its reactions with various nucleophiles (Scheme 5.32) [141].

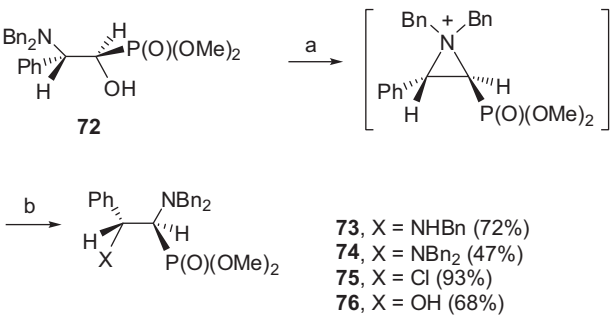
Among the common routes to α-aminoalkanephosphonates are those based on reduction of iminoalkanephosphonates. Thus, the conversion of 1-oxoalkanephosphonates, $R^1C(O)\text{-}P(O)(OR^2)_2$, into 1-oximinoalkanephosphonates, $R^1C(=NOH)P(O)(OR^2)_2$, followed by their reduction was found to be a simple and practical synthesis of α-aminophosphonic acid derivatives. The reduction can be achieved with aluminum amalgam in ethanol, zinc dust in formic acid, a Zn/Cu couple in aqueous ethanol, diborane in tetrahydrofuran (THF), or by catalytic hydrogenation



Scheme 5.31

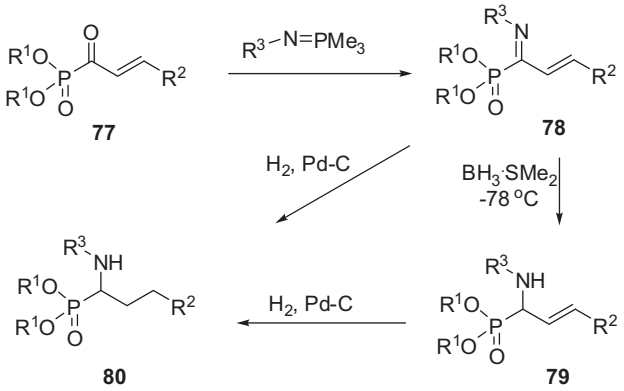
over Raney nickel [27]. A good yield of the respective α-aminophosphonate is obtained while treating compounds $RC(=NOH)P(O)(OEt)_2$ with $NaBH_4$ in the presence of MoO_3 or $NiCl_2$ [142].

Recently, a new example of synthesis of α-phosphorylated imines and their selective reduction to α-aminophosphonates has been reported. β,γ-Unsaturated



Reagents and conditions: (a) MsCl, NEt_3, CH_2Cl_2; (b) $BnNH_2$, NEt_3; or Bn_2NH, NEt_3; or $Et_4N^+Cl^-$; or H_2O-SiO_2; or wash out Cl^-

Scheme 5.32



Scheme 5.33

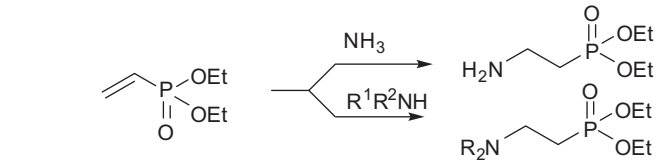
α-ketophosphonates **77** react with phosphinimines affording good yields of α-phosphorylated α,β-unsaturated imines **78**. Selective 1,2-reduction of such 1-azadienes yields β,γ-unsaturated α-aminophosphonates **79**, which are hydrogenated to give saturated α-aminophosphonate derivatives **80** (Scheme 5.33) [143].

5.4 Synthesis of β-Aminophosphonates

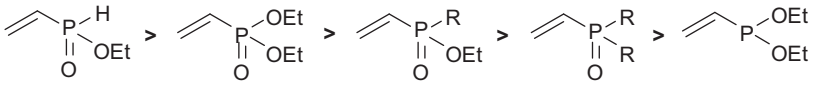
The first synthesis of β-aminophosphonates was realized in 1953 before the isolation of AEP from natural sources. Findings of AEP structural elements in various living organisms have stimulated active developments in this area. At present there are several general strategies to synthesize β-aminophosphonates, which are based on formation of P–C, C–C, and C–N bonds or transformation of precursors [144, 145].

The simplest way to β-aminophosphonates is Michael addition of ammonia or an amine to the activated C=C bond of vinylphosphonates. Pudovik and Denisova first showed that ammonia adds to diethyl vinylphosphonate in the presence of EtONa in the "normal" manner (Scheme 5.34) [146]. The reaction with primary amines is usually activated by basic catalysts under heating [147]; secondary amines add to the vinylphosphonate easily without a catalyst, whereas aniline reacts only in the presence of MeONa and with only a fair yield [148].

Reactivity of vinyl phosphorus derivatives in amine addition reactions decreases from tetracoordinate to tricoordinate phosphorus derivatives [149]. In all cases the addition of the amine is directed to the β-position of the C=C bond. The phosphonate group shows control of the regiodirecting effect even in the reaction of perfluoroalkylvinylphosphonates containing highly electronegative R_f groups (Scheme 5.35) [150].



R^1R^2NH = Me_2NH, $(CH_2)_5NH$, $PhNH_2$

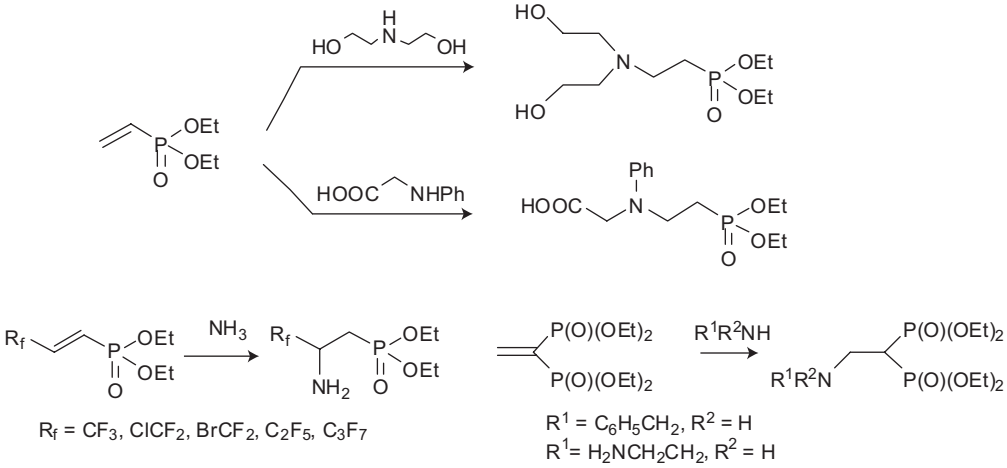


Scheme 5.34

Amines with additional functional groups can be used to synthesize desired functionalized aminophosphonates (Scheme 5.35) [151].

Benzylamine and ethylenediamine give relatively stable products by "normal" 1 : 1 Michael addition to tetraethyl ethenylidene-bis(phosphonate) **81** at 20 °C in $CHCl_3$ solution. Methylamine, *tert*-butylamine and isopropylamine also yield products of 1 : 1 addition, which undergo retro-Michael reaction during evaporation of the solvent and any excess amine. In the case of methylamine the reaction results in the formation of a 1 : 2 addition product **82** [152] resulting from a second Michael reaction (Scheme 5.36).

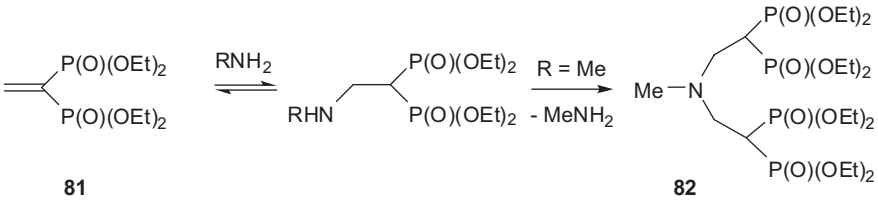
The disodium salt of ethenylidene-bis(phosphonic) acid also reacts with primary amines and diethanolamine in water, alcohols or acetic acid solutions at 100–120 °C and pH 7–8 to give 2-amino-1,1-bis(phosphonic) acids in 50–90% yields [153]. The addition reaction between aza-macrocyclic compounds with NH groups and two equivalents of vinylphosphonate yields new phosphonoethylated aza-macrocycles (Scheme 5.37) [154].



R_f = CF_3, $ClCF_2$, $BrCF_2$, C_2F_5, C_3F_7

R^1 = $C_6H_5CH_2$, R^2 = H
R^1= $H_2NCH_2CH_2$, R^2 = H

Scheme 5.35



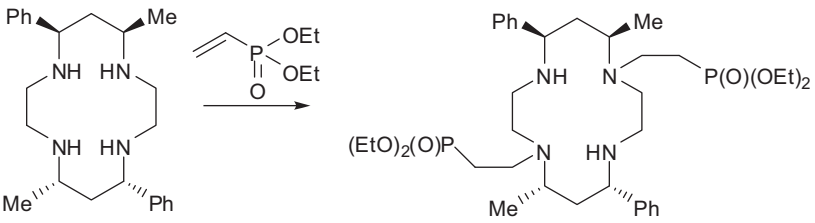
81 **82**

Scheme 5.36

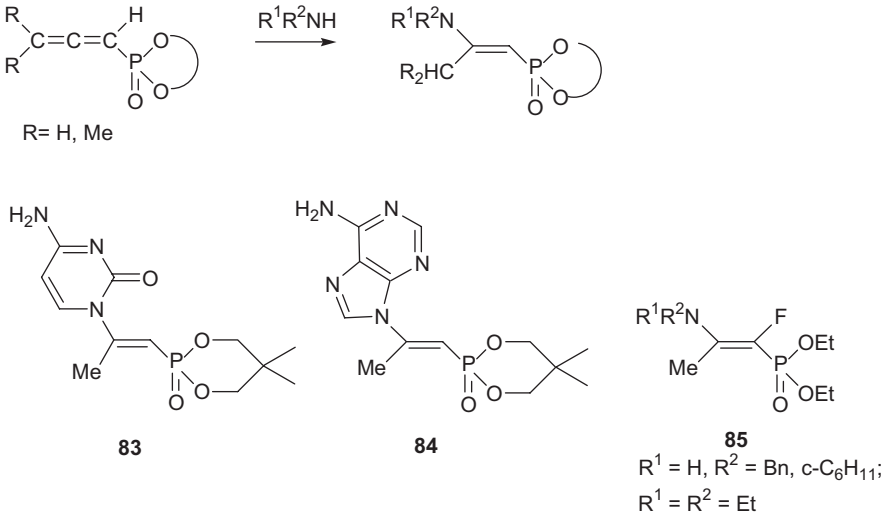
Allenylphosphonates constitute a readily accessible family of allenes that can be used as versatile building blocks in aminophosphonate chemistry. Various NH nucleophiles, including ammonia and some NH heterocycles, are easily added to dialkyl allenylphosphonate to give β-aminovinyl [155] or nucleobase-containing phosphonates such as compounds **83** and **84** [156]. The reaction was also used to synthesize β-amino-α-fluorovinylphosphonates **85** from $CH_2{=}C{=}CF{-}P(O)(OEt)_2$ and amines (Scheme 5.38) [157].

Vinylphosphonates are starting compounds in the asymmetric synthesis of α-amino-β-hydroxyphosphonic acids by Sharpless catalytic aminohydroxylation. Thus, *N*-chloro-*N*-sodio toluenesulfonamide in the presence of the catalytic system 4% $K_2OsO_2(OH)_4$/5% hydroquinidine 1,4-phthalazinediyl diether [$(DHQD)_2$-PHAL] at 20 °C regioselectively converted styrylphosphonates **86** into α-amino-β-hydroxyphosphonic acids **87** with moderate to high enantioselectivity (Scheme 5.39) [158, 159]. *N*-Chloro-*N*-sodio ethyl carbamate in *n*PrOH gave products with higher enantioselectivity. *N*-Bromoacetamide as the nitrogen source in the same catalytic conditions reacted with electron-rich styrylphosphonates to afford only racemic α-amino-β-bromophosphonates **88** in low yields [160].

Arbuzov and Michaelis–Bekker reactions, which are classical transformations in organophosphorus chemistry, have been used in the synthesis of β-aminophosphonic acids. However, these reactions are restricted by the limited availability of β-haloethylamino derivatives. As a rule, *N*-(-haloalkyl)phthalimides are used for these purposes (Schemes 5.40 and 5.41). Deprotection of the amino and ester groups in the final reaction leads to free aminophosphonic or aminophosphinic acids. A drawback of the Arbuzov reaction consists of the relatively hard conditions used, which sometimes decrease the yields of target compounds. On the other hand, *N*-phthaloyl (Phth)-protected β-aminophosphonates can be used directly in the



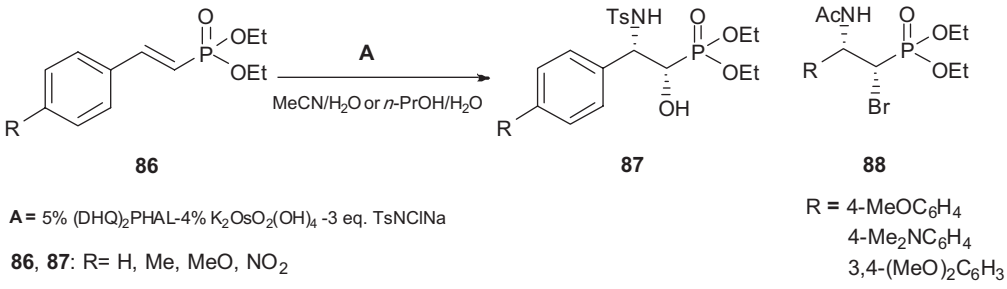
Scheme 5.37



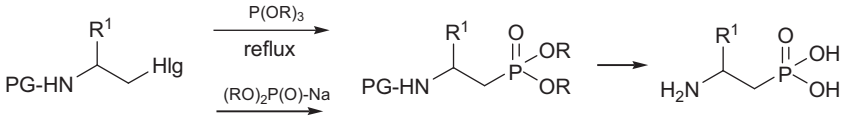
Scheme 5.38

synthesis of phosphonolipids and dipeptides containing the 2-aminoethanephosphonate fragment with subsequent deprotection of the amino function [161, 162]. The reaction has been used in syntheses of some sophisticated aminophosphonates using the specially constructed and protected synthons [163, 164]. Chiral *N*-Phth-protected amino acid chlorides were used in the Arbuzov reaction to prepare β-phthalimido-α-ketophosphonates. The latter were reduced catalytically with BH_3 to diastereomeric β-amino-α-ketophosphonates in good yields and high diastereoselectivity [165]. The Michaelis–Bekker reaction proceeds as a rule at milder conditions due to the high reactivity of the $(RO)_2P(O)^-$ anion (Scheme 5.42). Moreover, in the reaction, tosylated compounds can be used instead of halo-derivatives that was demonstrated by a chiral β-aminophosphonate synthesis [166].

β-Aminophosphonates and their derivatives can be prepared from β-halophosphonates by displacement of the halogen with amines. This approach is widely used in synthesis of biologically interesting compounds derived from AEP **1** [167–172]. In xmany cases the substitution proceeds in an appropriate solvent at room



Scheme 5.39



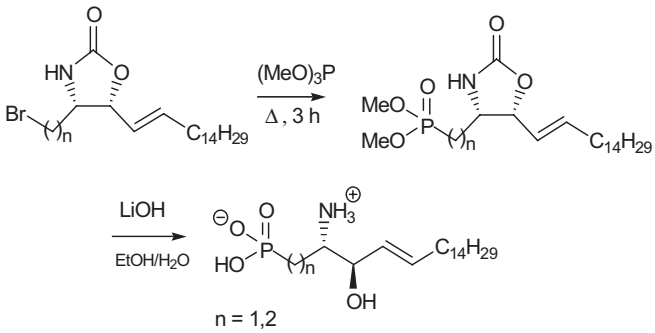
R = Me, Et; R^1 = H, Me, Pr, Bu; PG = Phth, Bn, Cbz, Tos, Gly, Ala

Scheme 5.40

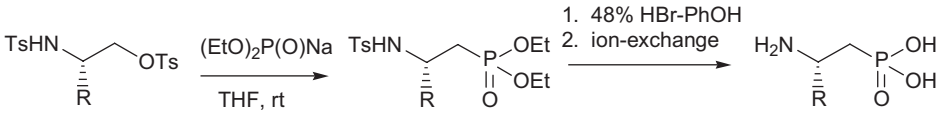
temperature, but sometimes the use of an *N*-silylated amino reagent is more convenient. Scheme 5.43 shows some β-aminophosphonates prepared by this route.

Two general strategies have been used for the preparation of β-aminophosphonates by the regioselective ring opening of three- or four-membered nitrogen heterocyclic compounds. The first consists of the nucleophilic addition of phosphites or sodium phosphonates to *N*-acyl or *N*-tosyl aziridines [173]. The second route consists of regioselective ring openings of readily accessible aziridinyl and epoxy phosphonates. It is useful in the asymmetric synthesis of β-aminophosphonates (Scheme 5.44) [174–178].

The nucleophilic addition of phosphites and alkyl alkylphosphinates to C=O groups of α-amino aldehydes or ketones is an efficient method to synthesize β-amino-α-hydroxyphosphonates. The reaction is catalyzed by base (Et_3N, DBU, KF, etc.) [179, 180] or Lewis acids ($TiCl_4$, $SnCl_4$) [181, 182]. The use of the chiral catalyst [(*R*)-(Al-Li-bis(binaphthoxide)] [183] enables an asymmetric version of the process. Applied to chiral *N*-protected α-aminoaldehydes **96** derived from amino acids, the reaction presents another route to diastereomeric β-amino-α-hydroxyphosphonates. The stereoselectivity of this process is dependent on the reaction conditions and the nature of the solvent and base. Solid-phase synthesis of β-amino-α-hydroxyphosphonates using of resin-bound chiral *N*-acylated aminoaldehydes has also been reported [184]. When chiral keto-phosphonates of the type **97** derived from naturally occurring 1-amino acids were introduced into the reaction, 2-amino-1-hydroxy-alkane-1,1-bisphosphonic acids **98** were prepared in moderate yields [185]. These may result from a phosphonate–phosphate rearrangement induced by the base. In fact, phosphonate–phosphate products **99** are obtained in a one-pot proce-



Scheme 5.41

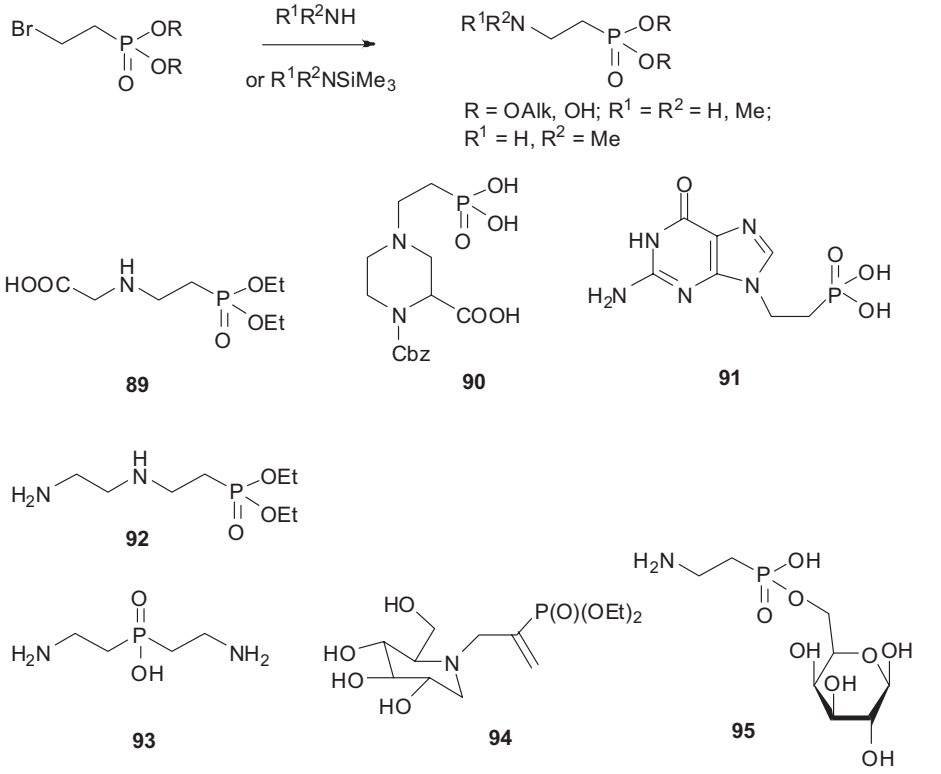


R = H, Me, Bn, $PhCH_2CH_2$, i-Pr, i-$PrCH_2$

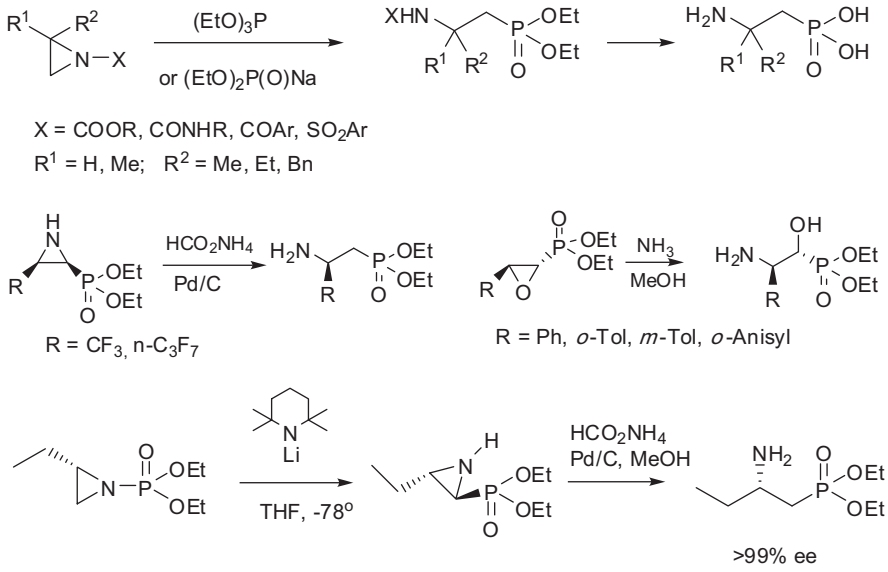
Scheme 5.42

dure from amino acid thioesters and diethyl phosphite in the presence of *t*BuOK in 70–80% yields (Scheme 5.45) [186].

α-Phosphonate carbanions generated from phosphonates and strong bases (i.e., BuLi, lithium diisopropylamide, LiHMDS, or NaH) present straightforward access to β-aminophosphonates through a C–C bond-forming process (Scheme 5.46). In particular, this strategy has been used for β-aminophosphonate synthesis using *N*-bromomethylphthalimide as an amino group source [187]. Numerous reports have also demonstrated the effective application of α-phosphonate carbanion addition to the C=N bond of imines, especially in the asymmetric synthesis of β-amino-α-substituted phosphonates. For example, the method was used in a high-yielding

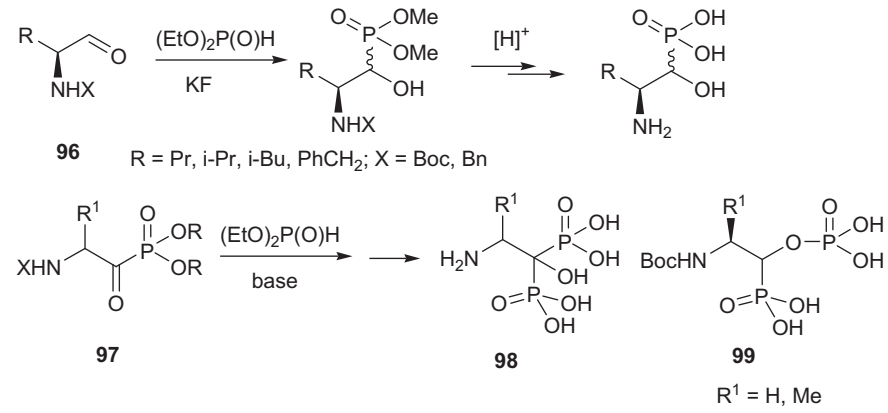


Scheme 5.43

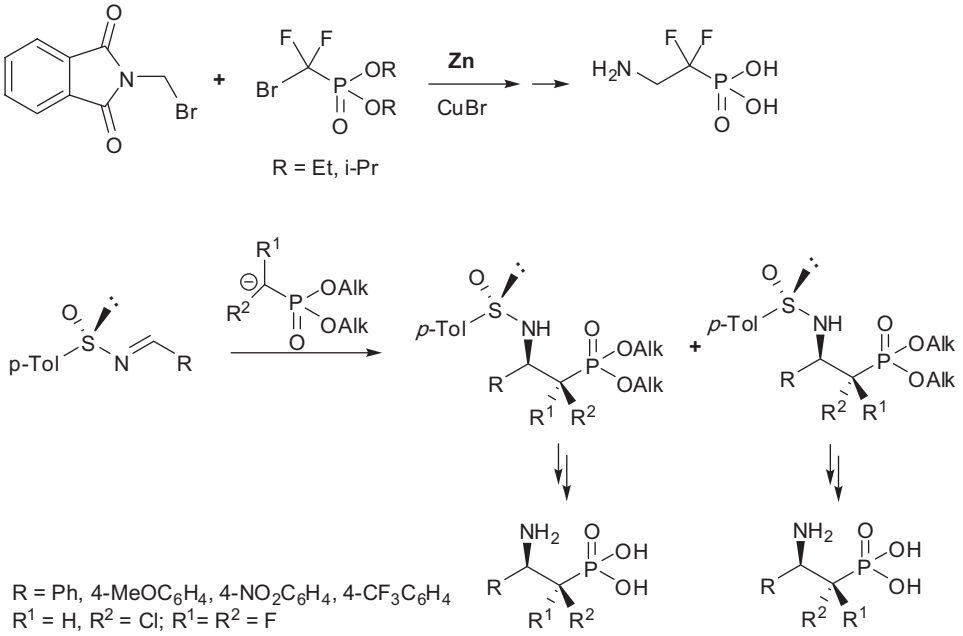


Scheme 5.44

asymmetric synthesis of β-amino-α-chloro-, α-cyano-, α-methyl-, and α,α-difluoro-alkanephosphonic acids [119, 188]. Diastereoselectivity of the reaction depends on the structure of the starting compounds, temperature and nature of the solvent. A chiral *p*-tolylsulfinyl group on an imine nitrogen simultaneously plays the roles of a stereo inductor and protective group [189, 190]. Obviously, this strategy can be especially useful in synthesis of various α,β-functionalized β-aminophosphonates. For example, substituted 2-aminoethylidene-1,1-bisphosphonates have been prepared from sodium-methylene-bisphosphonates and *N*-Boc-imines in high yields [191].



Scheme 5.45

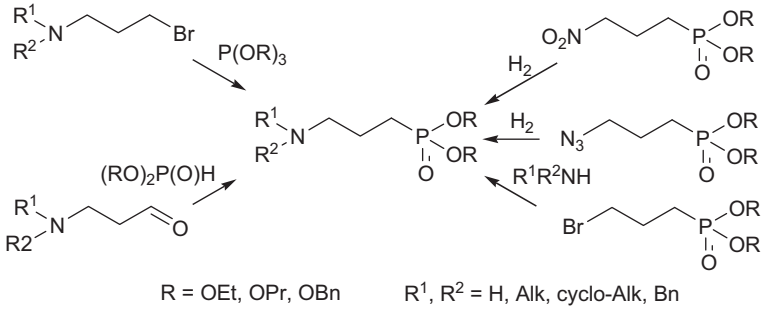


Scheme 5.46

β-Aminophosphonic acids can be also prepared by reduction of phosphonylated enamines [192], β-azido-, β-nitro-, or β-hydroxylaminophosphonates derivatives [193], and reductive amination of β-carbonylalkylphosphonates [194]. For the preparation of enantiomerically pure β-aminophosphonates, besides asymmetric syntheses, enzymatic methods are also used. The enzymatic process is attractive as a route to prepare both enantiomers from one racemic substrate. For example, a number of diethyl α- and β-aminoalkanephosphonates were resolved with high enantioselectivity in the presence of *Candida antarctica* lipase B by catalytic acetylation with ethyl acetate. It was found that diisopropyl ether was the best reaction medium to obtain (*S*)-aminophosphonates with high enantiomeric excess (97–100%). On the other hand, the enantioselectivity of the resolution is very sensitive to the nature of the alkyl residue. After separation, enantiomeric (*R*)-*N*-acylaminophosphonates could be hydrolyzed to the corresponding free amino acids [195].

5.5
Synthesis of γ-Aminophosphonates and Higher Homologs

Synthetic chemistry of γ-aminophosphonic acids and higher homologs often uses the same "classic reactions" for aminophosphonate preparation as described above. They continue to be useful in the synthesis of desired compounds, such as

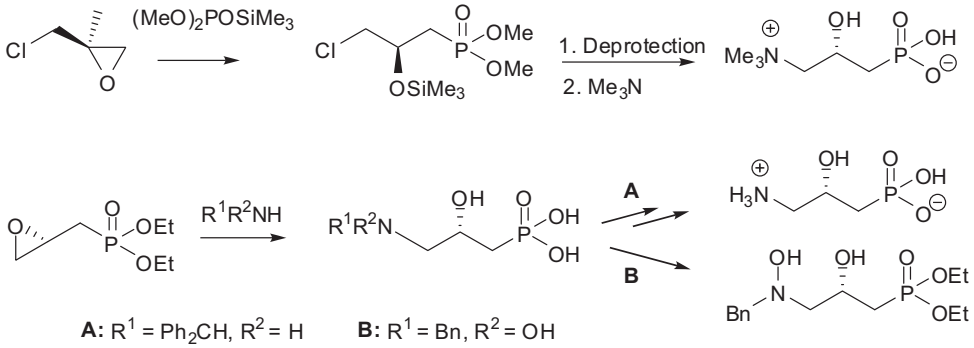


Scheme 5.47

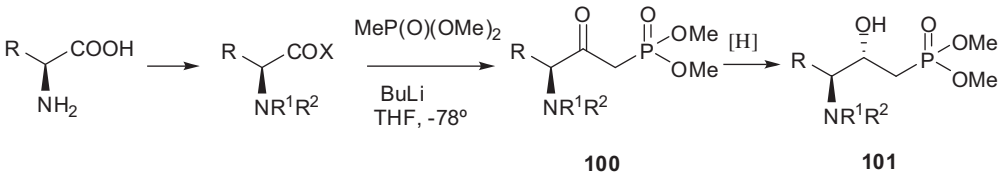
Arbuzov reactions of corresponding halo-amine derivatives [196, 197], amination of haloalkylphosphonates [198], addition of H-phosphonates to β-aminoaldehydes [199], reduction of γ-azido- and nitrophosphonates [200–202], and so on (Scheme 5.47). Allylphosphonates can be converted to chiral γ-aminophosphonates via asymmetric dihydroxylation with AD-mix-α and regioselective amination of the cyclic sulfates [203, 204].

The special interest in γ-aminophosphonic acids is driven by three biomedical research areas: γ-aminobutyric acid (GABA) and carnitine phosphonic analogs [205], amino bisphosphonates, and fosmidomycin derivatives. The diversity of structural requirements for these compounds demands special strategies and methods to build the needed molecules. Thus, an optimal route to phosphocarnitine derivatives consists of appropriate regio- and stereoselective oxirane ring opening with trimethylsilyl (TMS)-phosphite to introduce phosphorus residue [206]. Opening of an oxirane in a chiral 2,3-epoxypropylphosphonate was used to place in the 3-position the amino group [207, 208] and the *N*-benzyloxyamino residue (Scheme 5.48) [209].

(1*S*,2*S*)-3-Amino-1,2-dihydroxypropylphosphonic acid and (1*R*,2*S*)-3-amino-1,2-dihydroxypropylphosphonic acid were synthesized from (*S*)-3-azido-2-benzyloxypropanal in two steps by addition of dibenzyl phosphite, separation of dibenzyl diastereomeric 3-azido-2-benzyloxy-1-hydroxypropylphosphonates, followed by



Scheme 5.48



X = OMe, OEt, OBn, N(OMe)Me

[H]= $NaBH_4$, $Zn(BH_4)_2$, catecholborane

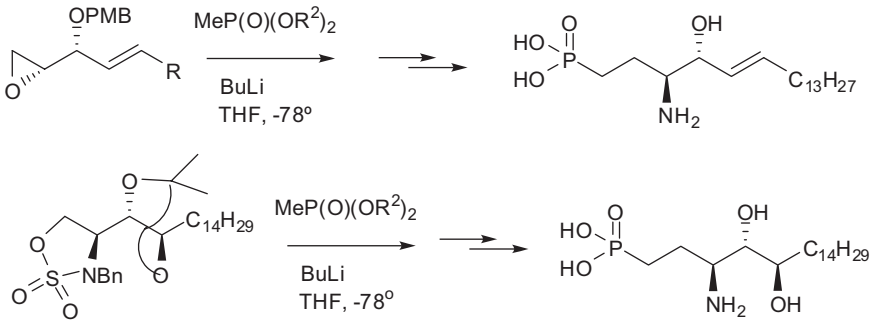
R^1, R^2 = H, Bn; H, Boc; Bn_2; $(PhMe\overset{*}{C}H)_2$

Scheme 5.49

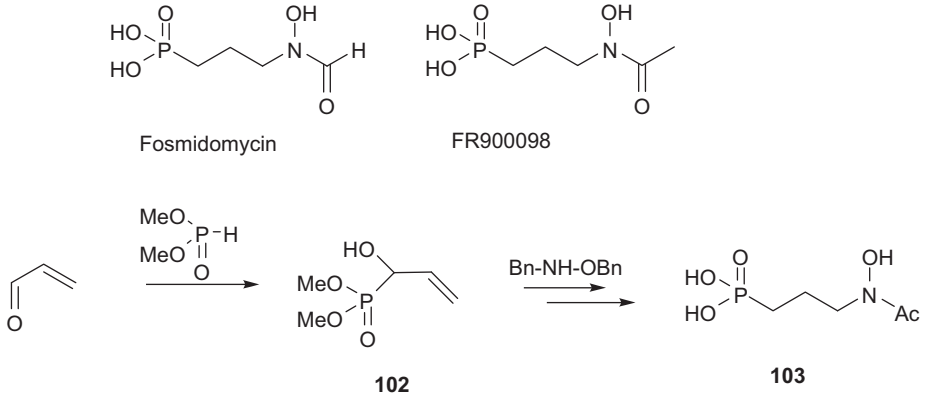
azide reduction with simultaneous hydrogenolysis of the benzyl protecting groups [210].

One of the most effective strategies in the synthesis of complex γ-aminophosphonic acids is based on the use of phosphonate carbanions to form C−P bonds in the corresponding protected substrate. Thus, chiral amino acid esters can be introduced through the reaction with lithium dimethyl methylphosphonate easily generated from BuLi in THF to yield γ-amino-β-keto-phosphonic acids with high stereocontrol [211–213]. The compounds **100** were used as reagents in the Horner–Wadsworth–Emmons reaction to synthesize stereoselectively unsaturated amino keto acids [211] and aminoalkohols [214]. Reduction of carbonyl group in **100** with $NaBH_4$ or $Zn(BH_4)_2$ enables a diastereoselective route to carnitine analogs **101** (Scheme 5.49) [215, 216]. The highest diastereoselectivity (>98% d.e.) was achieved in the case of a catecholborane reductive reagent and (*S,S*)- and (*R,R*)-bis(α-methylbenzyl)amino derivatives **100** at −78 °C [217]. The methodology was applied to synthesis of (3*S*)-amino-(2*R*)-hydroxy-5-methylhexylphosphonic acid (phosphostatine) and (3*S*)-amino-(2*S*)-hydroxy-5-methylhexylphosphonic acid (epiphosphostatine) via diastereoselective reduction of dimethyl (3*S*)-[3-(*N,N*-dibenzyl-amino)- and (3*S*)-[3-(*N*-benzylamino)-5-methyl-2-oxohexylphosphonates derived from L-leucine [218].

Phosphonate carbanion reactions were applied successfully to synthesize phosphonate analogs of sphingolipids via regioselective ring opening of an oxirane



Scheme 5.50



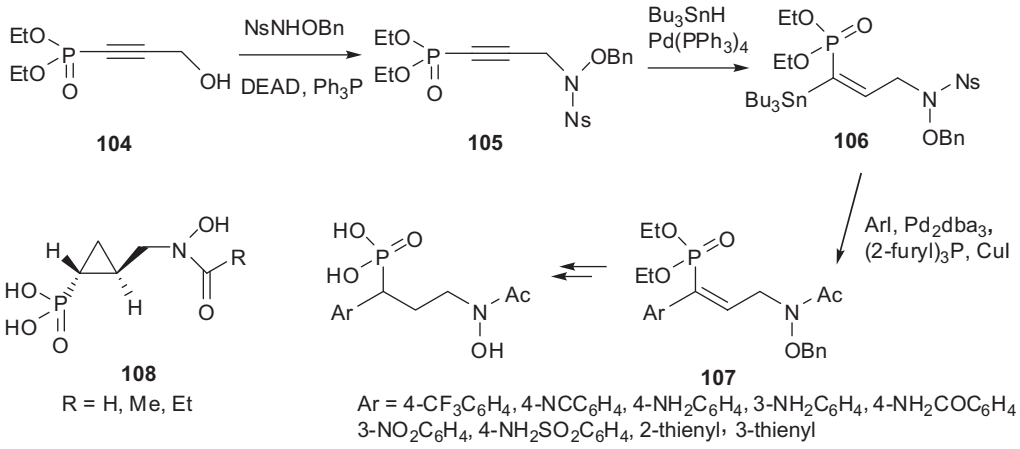
Scheme 5.51

and a cyclic sulfamide (see Scheme 5.50) [219, 220] and via interaction with a corresponding bromo derivative [197]. The anion generated from a bisphosphonate has been used to form new C=C bonds and simultaneously to introduce a phosphonate residue via the reaction with a chiral aminoaldehyde [221].

Numerous publications are devoted to syntheses of fosmidomycin and its analogs in a search for antimalarial drugs. Some analogs of fosmidomycin were prepared by addition of H-phosphonates to α,β-enones. Thus, subsequent acylation and palladium-catalyzed amination of phosphonate **102** with protected hydroxylamine afforded compound **103** (Scheme 5.51) [222]. Other syntheses of fosmidomycin analogs use a phosphonate carbanion approach to construct phosphoryl-substituted halo and aldehyde synthons for further introduction of a hydroxyamino function [223–225].

The Mitsunobu reaction using diethyl (3-hydroxypropynyl)phosphonate **104** and *N*-benzyloxy-2-nitrobenzenesulfonamide is a superior method for obtaining the highly functionalized phosphonate **105** as a precursor for the synthesis of fosmidomycin analogs (Scheme 5.52). Pd(0)-catalyzed addition of Bu_3SnH to the triple bond of the compound **105** afforded 1-tributylstannylpropenylphosphonate **106** as a single isomer in 90% yield. Subsequent Stille coupling led to alkenylphosphonates **107** in good to excellent yields. The method proved to be highly compatible with a wide range of diverse functionalities including heteroaromatic rings [226]. A series of conformationally restricted cyclopropyl analogs **108** of fosmidomycin has been synthesized starting from methylphosphonate carbanion [227]. Recently, a new approach to fosmidomycin was developed involving, as the key step, the nitrosoene reaction of commercially available diethyl allylphosphonate with *in situ* prepared nitrosocarbonyl methane [228].

Recently, a [4 + 2] reaction of a butadienylphosphonate with 2-nitrosotoluene was described as a new strategy to form δ-aminophosphonate derivatives [229]. Reaction of diethyl butadienylphosphonate with *o*-nitrosotoluene results in dihydro-1,2-oxazine **109** which was cleaved with zinc in AcOH to give 4-(*o*-tolylamino)-1-hydroxybut-2-enylphosphonate **110** in quantitative yield. Dihydroxylation of **110**

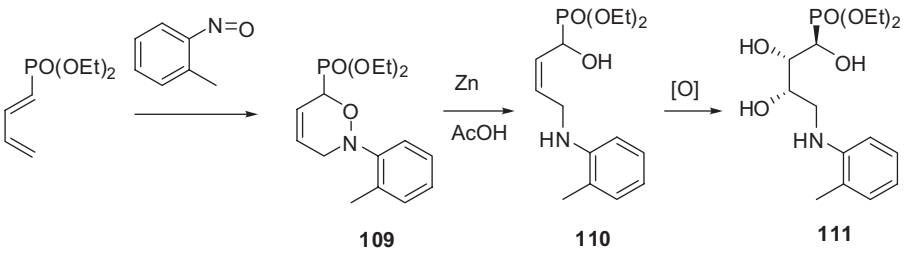


Scheme 5.52

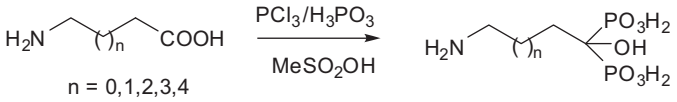
with $K_2OsO_2(OH)_2$ afforded 4-tolylamino-1,2,3-trihydroxybutylphosphonate **111** (Scheme 5.53). Obviously, the strategy may be useful in a synthesis of other polyfunctional δ-aminophosphonates.

ω-Aminoalkylidene-bisphosphonic acids are another group of compounds that are interesting as therapeutic agents for the treatment of a number of diseases which are characterized by abnormal calcium metabolism. Thus, 4-amino-1-hydroxybutylidene-1,1-bisphosphonic acid (alendronate; Fosamax) has shown efficacy in postmenopausal osteoporosis, Paget's disease, and malignant hypercalcemia [230]. 4-Amino-l-hydroxybutylidene-1,1-bisphosphonic acid can be prepared by the reaction of 4-aminobutyric acid with a mixture of H_3PO_3 and PCl_3 in the presence of $MeSO_3H$ solvent in 90% yield [231]. The method has also been used in the synthesis of higher ω-amino acid homologs (Scheme 5.54). A similar approach was applied to prepare 3-amino-l-hydroxybutylidene-1,1-bisphosphonic acids from *N*-substituted β-aminopropionic acids, PCl_3 and H_3PO_3 in moderate yields (40–46%) [232]. Numerous patents in this area differ only in the details concerning halophosphorus reagents, solvent, temperature, and so on.

The most commonly utilized strategy for the preparation of 1-hydroxy-1,1-bisphosphonates involves the Arbuzov reaction between trialkyl phosphites and an acyl



Scheme 5.53

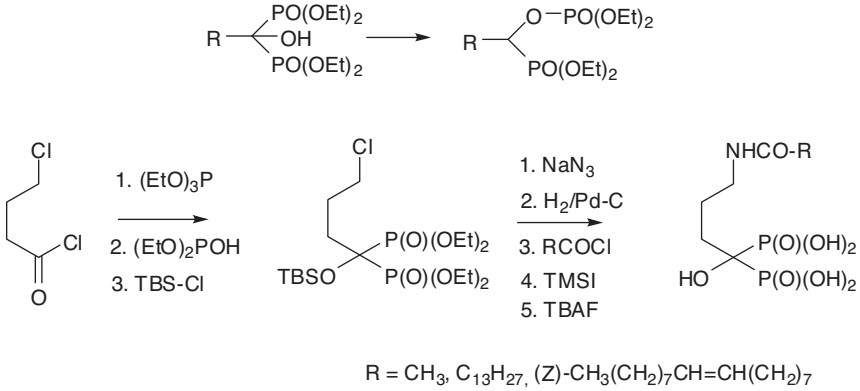


Scheme 5.54

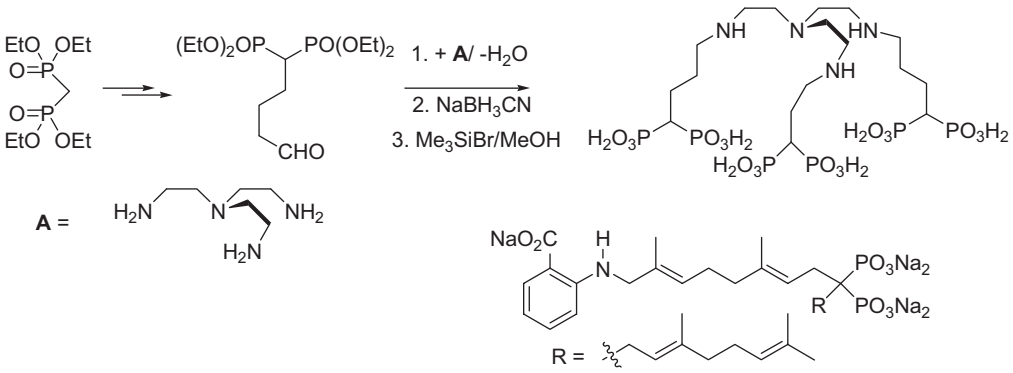
chloride followed by the addition of a dialkyl phosphite to the α-ketophosphonate. In many cases, the product is formed in low overall yield due to the phosphonate-phosphate rearrangement initiated by a base. Combining both steps into a single one-pot reaction sequence without solvent in the presence of weak base simplified the overall synthetic process. This modified approach has been used to prepare *N*-acyl alendronate derivatives via a ω-azide intermediate, reduction, acylation and finally deprotection (Scheme 5.55) [233]. Due to the poor solubility of alendronate in organic solvents, its direct *N*-acylation is carried out in higher yields by reaction of the disodium salt with carboxylic acid anhydrides or acyl chlorides in water [234, 235].

Tetraethyl ω-aminoalkylidene-1,1-bisphosphonates have been prepared from the tetraethyl methylene-bisphosphonate by a multistep procedure involving the synthesis of the corresponding ω-hydroxyalkylidene-1,1-bisphosphonates and subsequent transformations. Modified methodology uses the preparation of ω-formyl-bisphosphonate conjugates from methylene-bisphosphonate and subsequent introduction of an amino residue by reductive amination with $NaBH_3CN$ [236]. Alkylation of methylene-bisphosphonates was also used in the synthesis of fluorescent anthranilate analogs of isoprenoid bisphosphonates using *N*-anthranilate geraniol and farnesol analogs as building blocks (Scheme 5.56) [237].

Cyclic analogs of bioactive compounds with rigid structure or limited flexibility are of considerable interest for biochemical and pharmaceutical studies. As a result, numerous cyclic aminophosphonic and aminophosphinic acids and heterocycles containing amino and phosphonate residues have been prepared and studied [238]. Thus, (piperidinyl-4)-methylphosphinic acid **112** derived from *N*-benzyl-4-iodo-3,4-dehydropiperidine and MePH(O)(OMe) is a potent competitive $GABA_C$ receptor



Scheme 5.55

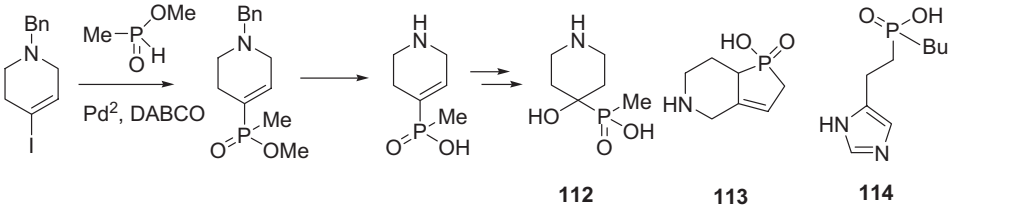


Scheme 5.56

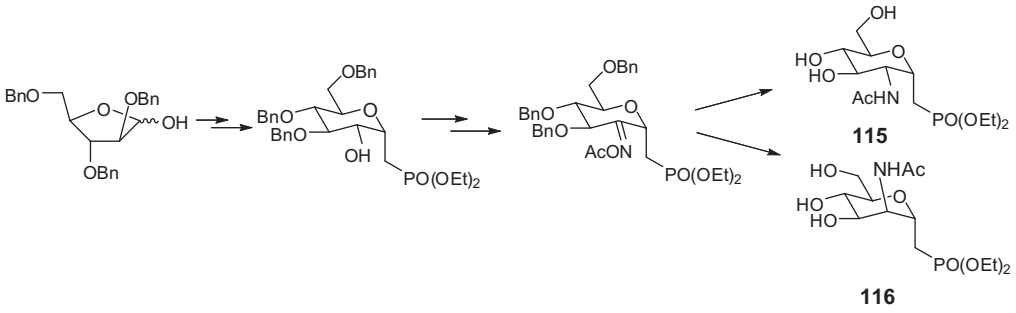
antagonist [239]. Some analogs of **112** such as **113** and **114** have been synthesized to compare their neuroactivity with the activity of "open-chain" γ-aminophosponic acids (Scheme 5.57) [240–242].

The isosteric phosphono analogs of *N*-acetyl-α-D-glucosamine 1-phosphate and *N*-acetyl-α-D-mannosamine 1-phosphate (**115** and **116**) are stereoselectively synthesized starting from 2,3,5-tri-*O*-benzyl-D-arabinofuranose. The free hydroxyl group of 2,3,5-tri-*O*-benzyl-D-arabinose is converted into an acetamido group by oxidation, oximation, acetylation of the oxime, reduction, and subsequent acetylation. The reduction of the oxime with diborane stereoselectively affords the *gluco* isomer, whereas catalytic hydrogenation gives the *manno* isomer. Acetylation and deprotection afford the desired products **115** and **116** (Scheme 5.58) [243]. A short synthesis of these compounds was also developed using *N*-acetyl D-glucosamine as the starting compound [244].

Phosphonic acid analogs of *N*-acetylneuraminic acid (Neu5Ac) and 3-deoxy-D-*glycero*-D-*galacto*-2-nonulosonic acid have been synthesized using an indium-mediated coupling of *N*-acetylmannosamine or mannose with dimethyl (3-bromopropen-2-yl) phosphonate in aqueous media [245]. The potential of these phosphonate acid analogs of sialic acids as sialidase inhibitors was also evaluated [245]. A series of cyclohexenephosphonates **117** was designed as core structures for the development of high-affinity mimics of sialic acid (Scheme 5.59). The use of both D- and L-xylose enantiomers gave access to two series of cyclohexenephosphonates with regioisomeric double bonds [246].



Scheme 5.57



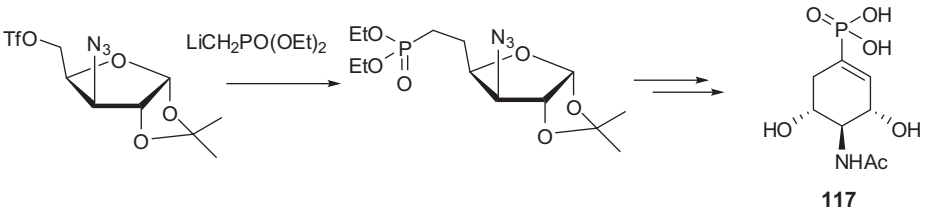
Scheme 5.58

An original strategy for the synthesis of precursors of constrained γ-aminophosphonic compounds as glutamate analogs has been developed via the regio- and stereocontrolled Diels–Alder cycloaddition of *N*-protected 1-amino-dienes and β-substituted vinylphosphonates (Scheme 5.60). By using chiral 1-amino-dienes, a major *endo* diastereoisomer was obtained with a good facial selectivity of 80–95% which was easily purified by chromatography. Further transformations of the cycloadducts, such as selective hydrolysis of phosphonic and carboxylic esters, epoxidation, dihydroxylation, and oxidative cleavage of the cyclohexene double bond could be realized by known reactions [247].

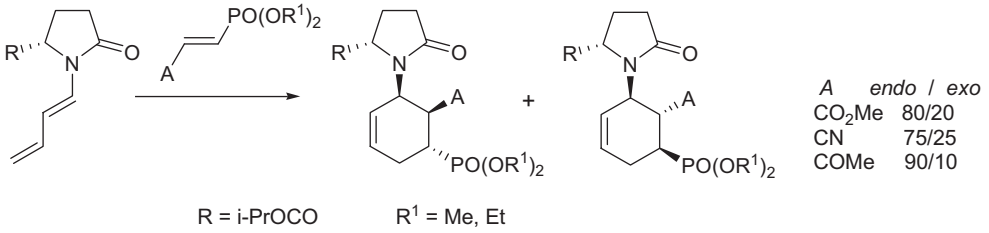
Another representative of this series, (2-amino-4,5-dimethylcyclohexyl) phosphonic acid **120** was prepared starting from nitrocyclohexenylphosphonate **118** derived from nitroethenylphosphonate and 2,3-dimethyl-1,3-butadiene under Diels–Alder reaction conditions (Scheme 5.61). Alkaline hydrolysis of phosphonate **119** with LiOH followed by acidification resulted in 2-aminocyclohexyl-1-phosphonic acid **120** (yield 60%) [248].

A one-pot reaction of diethyl 5-chloro-1-pentynylphosphonate **121** with primary and secondary amines produces novel 2-amino-1-cyclopentenylphosphonates **122** in excellent isolated yields (79–88%). A two-step mechanism was proposed: an initial amine addition to give a zwitterionic intermediate followed by cyclization and proton loss. The reaction proceeds smoothly in the absence of catalysts and inorganic additives. The reaction is general for primary and secondary amines, and tolerates hydroxy groups (Scheme 5.62) [249].

The synthesis of the *rac trans* and *cis* isomers of 1-amino-2-phosphonomethyl-cyclobutanecarboxylic acid **123** and 1-amino-2-phosphonomethyl-cyclopentanecar-



Scheme 5.59



Scheme 5.60

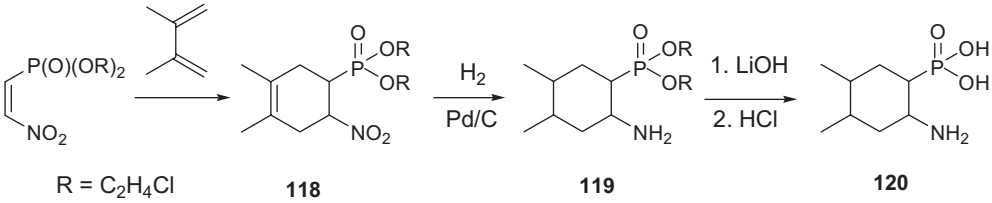
boxylic acid **124** was carried out in good yields as shown in Scheme 5.63. Amino nitriles prepared by a Strecker reaction were benzoylated with benzoyl chloride and the product *N*-benzoylamino nitriles were then separated by column chromatography on silica [250].

The synthesis of (2*S*)- and (2*R*)-2-(3-phosphonobicyclo[1.1.1]pentyl)glycine isomers (**125A** and **125B**) was accomplished by a stereoselective Ugi condensation of a phosphonoaldehyde, 2,3,4-tri-*O*-pivaloyl-α- and β-D-arabinopyranosylamine, *t*BuNC, and HCO_2H (Scheme 5.64). The enantiomeric aminophosphonates showed activity as group III mGluR agonists with significant differences in potency and subtype, and the (*S*) isomer **125A** is found to be a moderately potent and selective mGluR4 agonist [251].

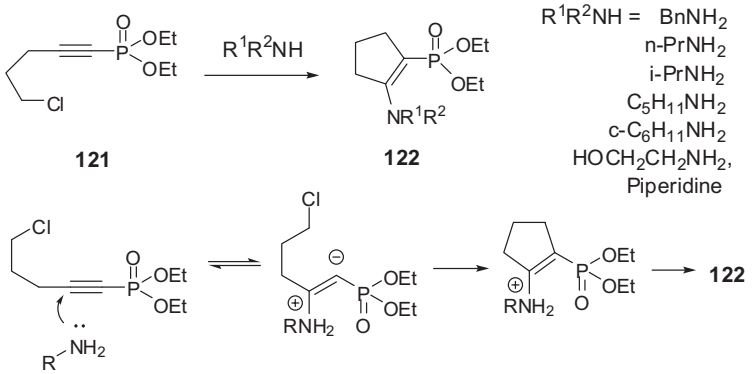
5.6 Phosphono- and Phosphinopeptides

Phosphono- and phosphinopeptides can be broadly defined as mimetics of peptides in which an aminocarboxylic acid residue is replaced by an aminophosphonic or related (aminophosphinic, aminophosphonous) moiety at any position of the peptide chain. The milestones of this chemistry have been presented in several recent review articles [252–258].

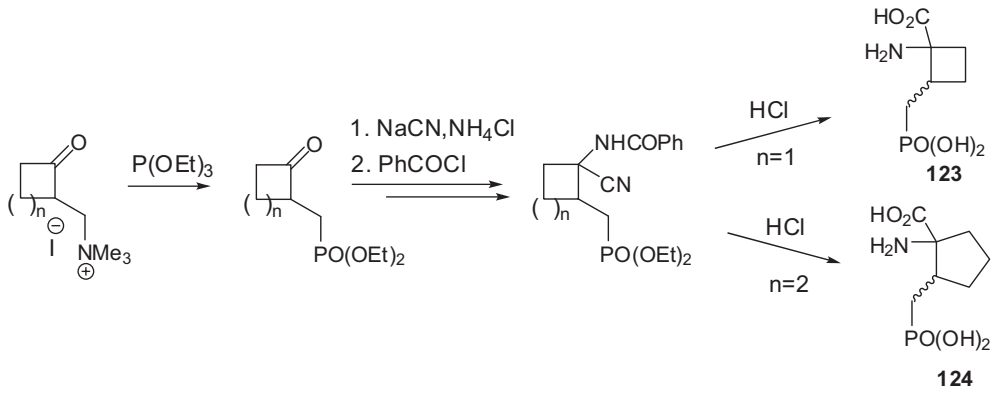
The presence of phosphonopeptide fragments in several naturally occurring products has been established. Among these the most important is bialaphos – an antibacterial metabolite produced by *S. hydroscopicus* and *Streptomyces viridochromogenes* [3]. The tripeptide phosphonic acid from *Actinomycete* sp. K-26 is another representative of a C–P bond-containing natural peptide [259, 260].



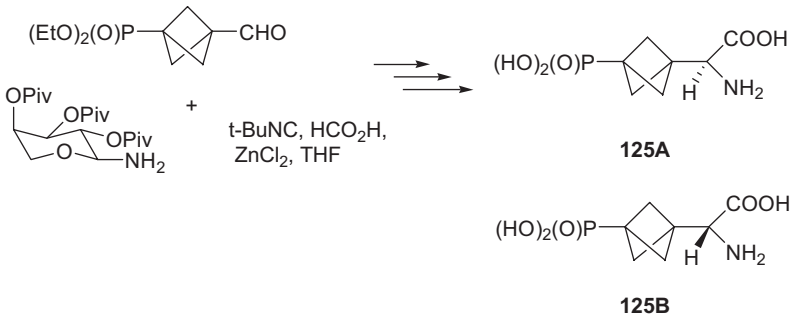
Scheme 5.61



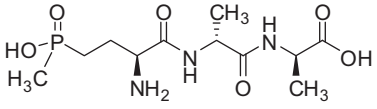
Scheme 5.62



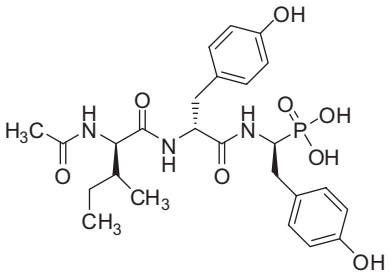
Scheme 5.63



Scheme 5.64

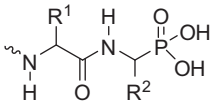


Bialaphos

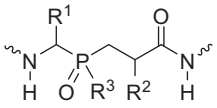


K-26

The main structure types of phosphonopeptides are (i) peptides containing a P-terminal aminophosphonate moiety (**A**); (ii) peptides in which an aminophosphonic or phosphinic acid unit mimics a dipeptide fragment (**B**); (iii) peptides containing a phosphonamidate (**C**) or phosphonoester bond (**D**); and (iv) peptides in which a phosphonic or phosphinic moiety is present in a side-chain (**E**).

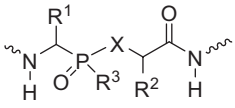


A

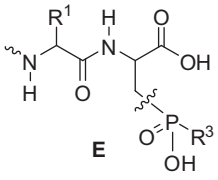


B

R = OH, H, Alk or Ar



C, X = NH (phosphonamidate pseudopeptides)
D, X = O (phosphonodepsipeptides)



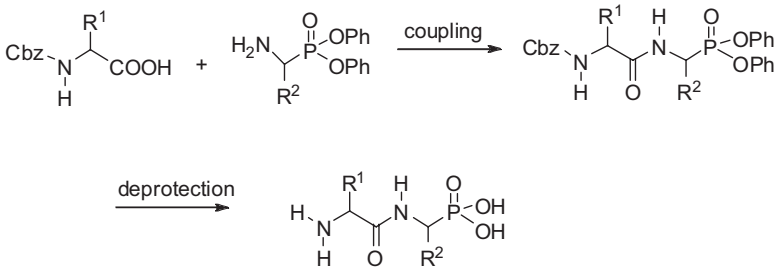
E

R^1, R^2 = H, Alk, Ar
R^3 = OH, H, Alk or Ar

5.6.1

General Strategies for the Phosphonopeptide Synthesis

As in the case of classical peptides, the first steps in the synthesis of phosphono- and phosphinopeptides involve the protection of functional groups. The formation of an amide or phosphonamide bond is the following step of the synthesis. Classical coupling procedures such as mixed carboxylic–carbonic anhydride (MCA), dicyclohexylcarbodiimide (DCC), active ester, and acid chloride methods are widely used for the preparation of the peptides **A**, **B**, and **E**. Peptide mimetics **C** and **D** in which a phosphonamide or phosphonate ester bond replaces the dipeptide fragment are generally synthesized by the reaction of aminoalkanecarboxylates or hydroxyalkanecarboxylates with phosphorochloridates or by other standard phosphorylation



R^1, R^2 = H, Alk, Ar

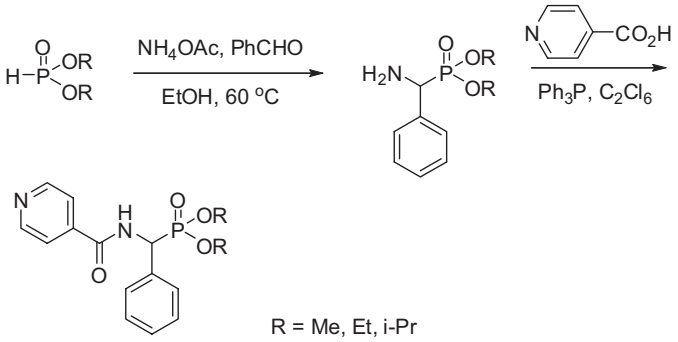
Scheme 5.65

procedures. The final steps of the phosphonopeptide synthesis involve total or selective removal of protecting groups. The scope and limitations of the reactions have been well delineated in an excellent review by Kafarski and Lejczak [252].

5.6.2
Peptides Containing P-terminal Aminophosphonate or Aminophosphinate Moiety

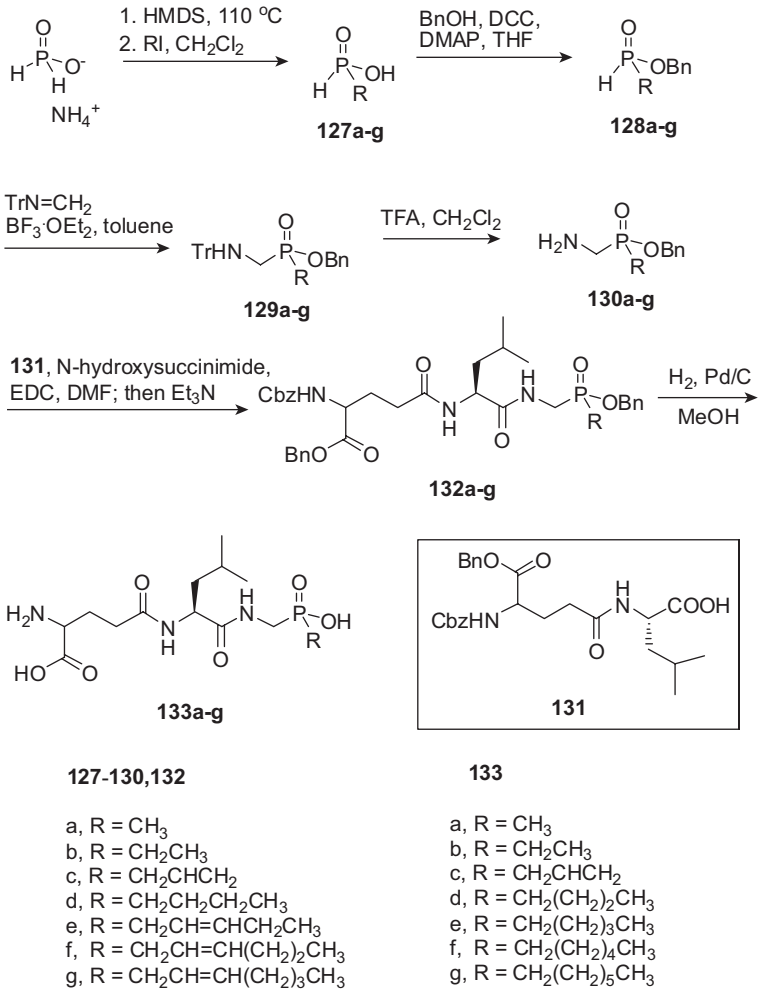
Standard procedures for the synthesis of pseudopeptides containing P-terminal aminophosphonate units include coupling *N*-protected amino acids or small peptides with free aminoalkanephosphonic acids and their ethers (Scheme 5.65). It should be noted that the direct acylation of aminoalkanephosphonic acids and their salts is relatively inefficient. The yields are usually moderate or low and dependent on the nature of the amino acid. Much more satisfying results are obtained when dialkyl (or diphenyl) aminophosphonates are used as substrates [252].

A typical example is the efficient synthesis of phosphonopeptides **126** using peptide coupling in the presence of the Appel reagents (C_2Cl_6, Ph_3P) (Scheme 5.66) [261].



126 (76-85%)

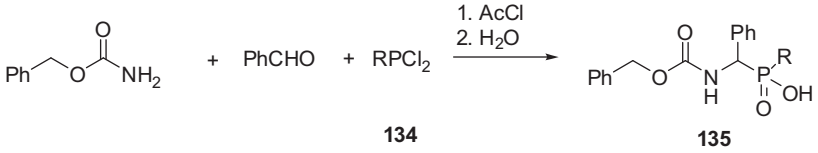
Scheme 5.66



Scheme 5.67

A series of phosphinotripeptides **133**, structurally related to glutathione, was synthesized starting from *O*-benzyl alkyl(aminomethyl)phosphinates **130** and *N*-benzyloxycarbonyl-L-glutamyl(α-benzylester)-L-leucine **131** (Scheme 5.67). The key step of the synthesis is coupling phosphinates **130** with dipeptide **131** in the presence of *N*-hydroxysuccinimide and 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide. Protecting group cleavage by catalytic hydrogenation afforded the desired pseudopeptides of which the compounds **133d** and **133e** were found to be potent inhibitors against *Trypanosoma cruzi* [262].

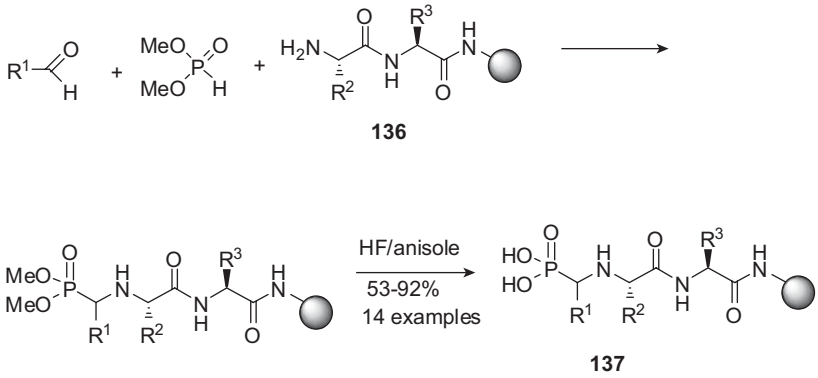
There are several reports on the use of multicomponent condensations for the one-pot preparation of phosphonopeptides. Thus, the reaction of benzyl carbamate with dichlorophosphites **134** and benzaldehyde in acetyl chloride as a solvent



R = MeO, EtO, n-PrO, n-BuO, PhO, Ph

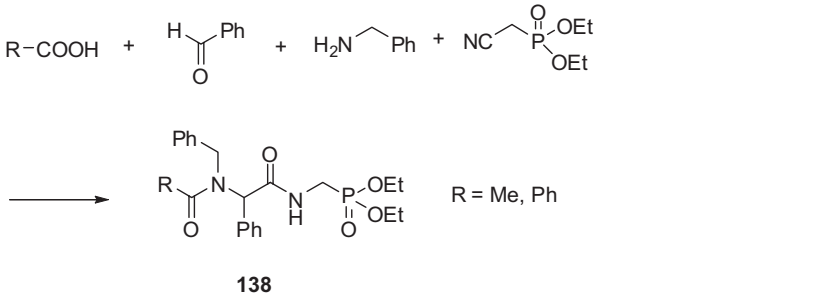
Scheme 5.68

followed by treatment with water afforded the phosphonopeptides **135** in satisfactory yields (Scheme 5.68) [35]. The solid-phase synthesis of α-aminoalkyl phosphonopeptides **137** has been achieved by the three-component reaction of resin-bound dipeptide **136** with an aldehyde and dimethyl phosphite (Scheme 5.69) [263]. A four-component Ugi condensation was successfully used for the preparation of phosphonopeptides **138** (Scheme 5.70) [264].



R^1 = Pr, Bu, Ph
R^2 = Me, Bn, $HOOCCH_2$
R^3 = H, Me, Bn

Scheme 5.69

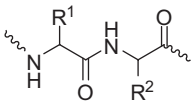


Scheme 5.70

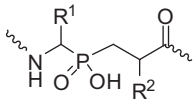
There are also examples of the use of biocatalysts for the preparation of phosphonopeptides. For example, papain was applied in order to synthesize a peptide bond [265]. Removal of protecting groups was achieved by using some proteases, such as penicillin acylase and phosphodiesterase [266, 267].

5.6.3
Peptides Containing an Aminophosphinic Acid Unit

These pseudopeptides are designed in such a way that an aminophosphinic acid unit replaces a dipeptide fragment of the corresponding peptide.



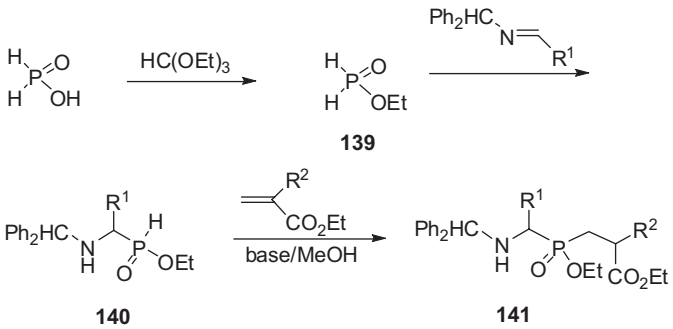
Dipeptide



Phosphinopeptide

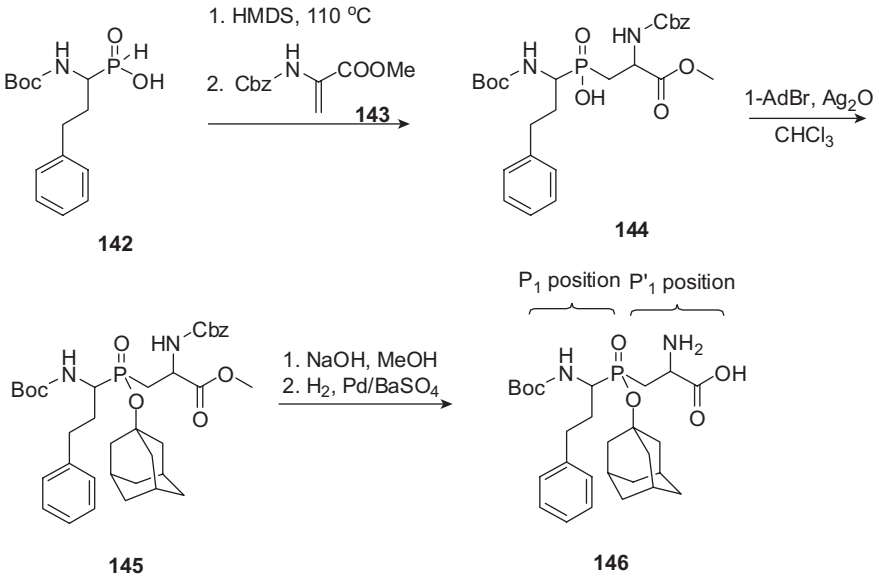
The standard strategy for the synthesis of peptidomimetic phosphinates includes an addition of phosphorus nucleophiles (suitably protected aminophosphonous acids and their analogs) to appropriate carbon electrophiles (usually Michael acceptors or alkyl halides) [252]. The 1-aminophosphonous acids themselves can be synthesized by numerous methods: (i) Kabachnik–Fields reactions involving addition of hypophosphorous acid or its derivatives to a C=N double bond [268–270]; (ii) Michaelis–Arbuzov reaction with the bis(TMS)phosphonite [271, 272]; (iii) Mitsunobu reaction on α-hydroxyalkylphosphinates [252]; (iv) the oxime procedure [252]; (v) alkylation of a suitably protected α-aminomethylphosphinic acid according to the Schöllkopf procedure [273]; and (vi) Michael reaction of ethyl diethoxymethylphosphonite with ethyl acetamidomethylenemalonate [274].

A general and simple synthetic route affording the possibility of variation of R^1 and R^2 groups in phosphinopeptides is shown in Scheme 5.71. The addition of ethyl hypophosphite **139** to various imines affords the corresponding ethyl



R^1 = H, Me, Ph; R^2 = Me, n-Bu, Bn; Base = MeONa or t-BuOK

Scheme 5.71



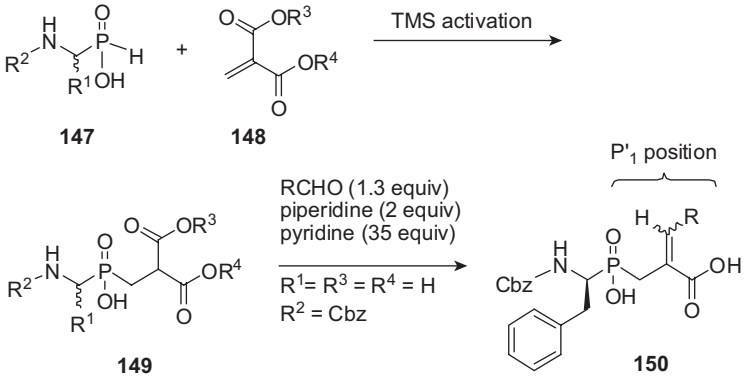
Scheme 5.72

phosphinates **140**. The Michael additions to α-substituted acrylates are performed using basic activation and afford the phosphinopeptides **141** [275].

The phosphinopeptide building block **144** was prepared by Michael addition of Boc-protected phosphonic homophenylalanine analog **142** to aminoacrylate **143**. Alkylation of the hydroxyphosphinyl function in **144** with 1-adamantyl bromide and subsequent deprotection of amino acid residue in **145** afforded the phosphinopeptide **146**. The availability of the C-terminal ester **145** and the corresponding acid **146** allows diversification (postmodification) of the P_1' position as shown in Scheme 5.72 using cross-coupling and alkylation reactions [276].

Yiotakis *et al*. have reported an efficient strategy for diversification of phosphinopeptides, using active methylene phosphinic scaffolds, which give access to a wide variety of P_1'-substituted phosphinopeptides [277]. β-Phosphinoyl malonates **149** were prepared via addition of HMDS or TMS-activated H-phosphinic acids **147** to methylenemalonates **148** serving as the Michael acceptors. Compounds **149** are easily alkylated with alkylbromides to give various diverse phosphinopeptide mimetics. Moreover, they have been submitted to Knoevenagel-type condensation reactions, providing peptides of type **150** (Scheme 5.73). These are of great interest as intermediates for enantioselective hydrogenations [277].

Another practical approach to prepare a series of P_1'-diversified phosphinic peptides is shown in Scheme 5.74. 2-(Bromomethyl)acrylic ester was subjected to Michael addition of Cbz-protected α-aminophosphonate, leading to the dehydroalaninyl phosphonic synthon **151** via a tandem Arbuzov addition/allylic rearrangement. Deprotection and coupling with L-tryptophanylamide gave the dehydroalaninyl precursor **153** in 69% overall yield. The latter was used as

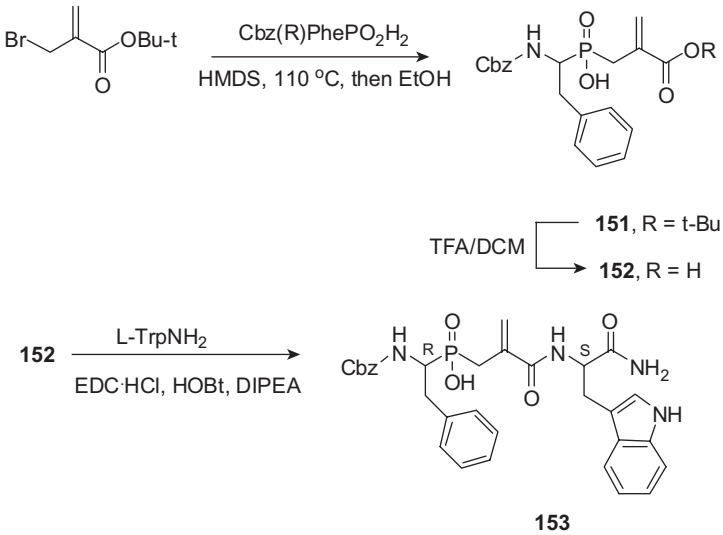


R^1 = $PhCH_2$, i-Pr; R^2 = Cbz, Boc; R^3 = Et; R^4 = Et, t-Bu;
R = H, Ph, $Ph(CH_2)_2$, $PhCH_2OCH_2$, Me_2CHCH_2, 4-ClC_6H_4, 4-$HO_2CC_6H_4$, 4-t-$BuO_2CC_6H_4$

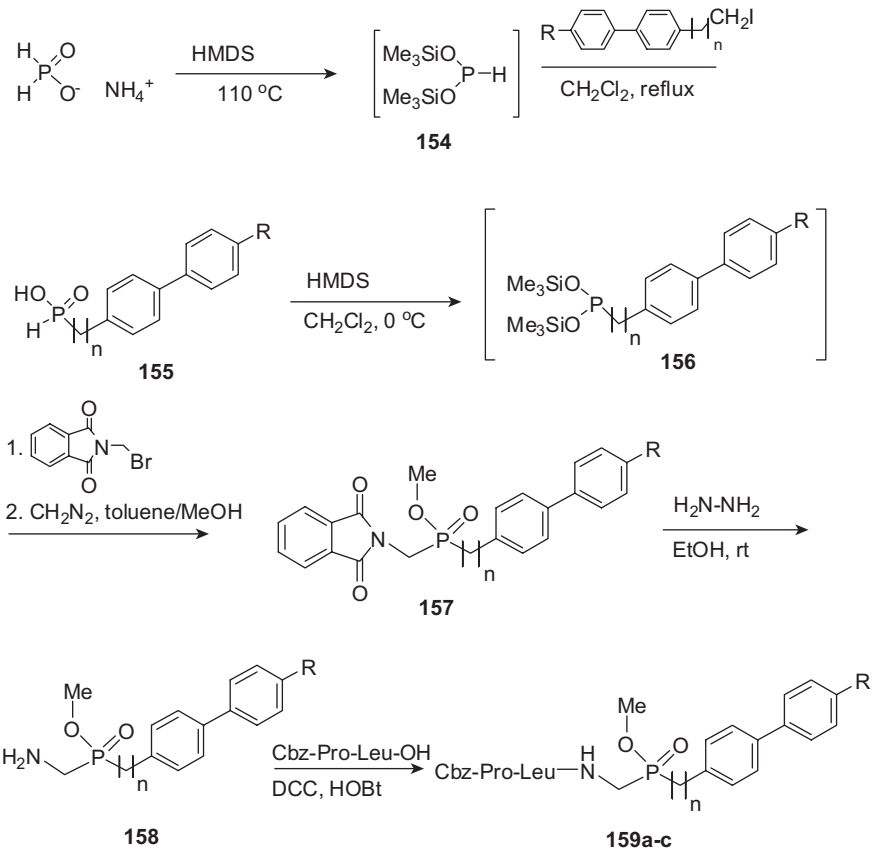
Scheme 5.73

a template for parallel diversification of the P_1' position with a wide variety of side-chains [278].

An important alternative to the addition of hydrophosphoryl compounds to Michael acceptors is the alkylation of $RP(OSiMe_3)_2$ species giving a phosphinate core [279, 280]. Thus, Bianchini *et al.* have developed the synthesis of peptidomimetic phosphinates **159** making use of the high reactivity of bis(TMS)phosphonite **154**.



Scheme 5.74



(a) n = 0, R = H; (b) n = 0, R = OH; (c) n = 1, R = H

Scheme 5.75

Alkylation of **154** with biphenylalkyl iodides afforded the monoalkylphosphinic acids **155**. To perform the second alkylation step at the phosphorus atom, compounds **155** were converted into the bis(TMS)phosphonites **156**, and then they were reacted with phthalimidomethyl bromide and diazomethane. After removal of the phthalimido protecting group with hydrazine, the α-aminophosphinates **158** were coupled with Cbz-Pro–Leu-OH mediated by DCC in the presence of α-1-hydroxybenzotriazole (HOBt) to give the peptidomimetics **159** (Scheme 5.75) [281].

5.6.4
Peptides Containing a Phosphonamide or Phosphinamide Bond

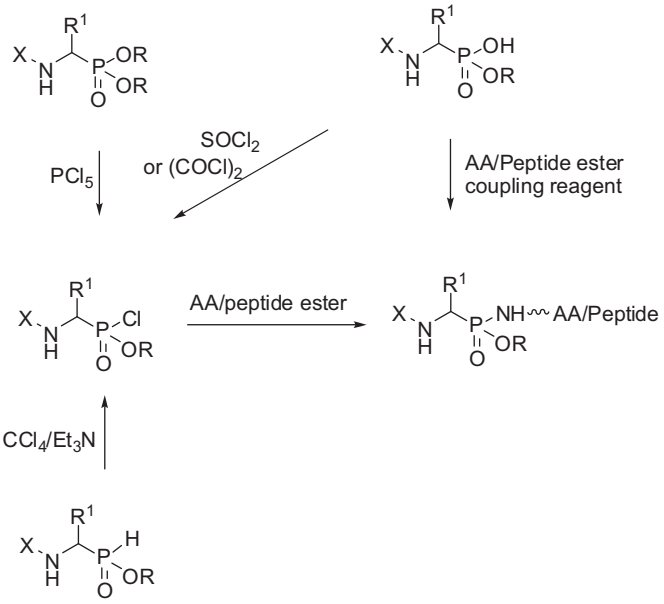
Phosphonamidate pseudopeptides represent a class of peptide mimetics in which a phosphonamide or phosphinamide bond replaces an amide bond in peptides. In contrast to the C–N bond of classical peptides, the P–N bond of phosphonamidate peptides is unstable at acidic and even physiological pH. On the other hand,

phosphonamidate peptides are transition state analogs that show a resemblance to the transition state of amide bond hydrolysis. Consequently, they represent a class of enzyme inhibitors and have played a role in the development of catalytic antibodies [253].

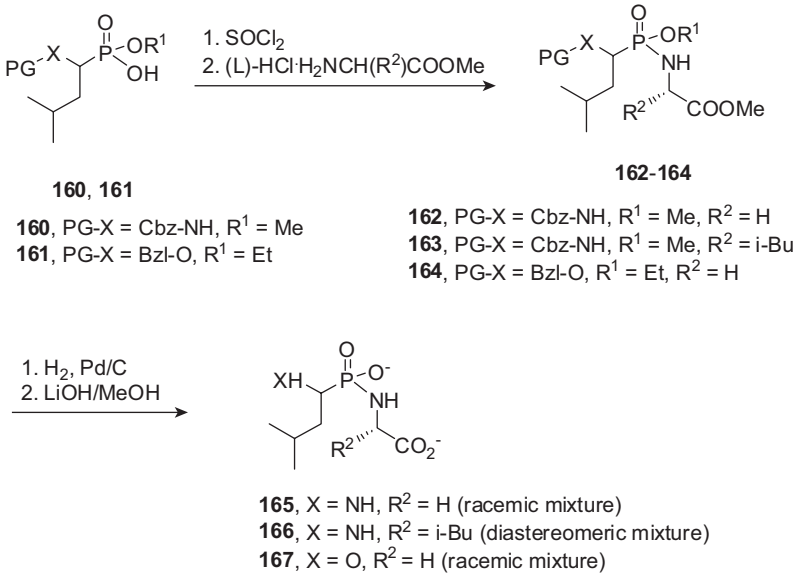
Generally, phosphonamidate pseudopeptides are prepared by the reactions of *N*-protected phosphonochloridates with a suitable amino acid derivative or peptide fragment [282, 283] or by the direct condensation of phosphonate monoesters with amino acid esters or peptide esters in the presence of coupling reagents [284, 285]. In turn, phosphonochloridates are accessible by the reaction of dialkyl phosphonates with one equivalent of phosphorus pentachloride [286–289], by reaction of monoesters with thionyl chloride or oxalyl chloride [290–292], and by oxidative chlorination of phosphonate esters (Scheme 5.76) [293, 294].

Synthesis of phosphonamidopeptides based on aminolysis of the appropriate phosphonochloridates with an amino ester is illustrated in Scheme 5.77. The starting α-amino- and α-hydroxyphosphonate monoesters (**160** and **161**) were converted into the P−N peptides **162–164** by reaction with thionyl chloride followed by treatment with a glycine or leucine methyl ester. After purification and deprotection, the final compounds **165–167** were obtained as a mixture of enantiomers or diastereomers [295].

There are several reports on the DCC-induced condensation of aminophosphonic acids with α-amino acids [284, 295]. Synthesis of ferrocenyl-containing pseudopeptide mimetics derived from α-aminophosphonous acids is shown in Scheme 5.78. Condensation of (ferrocene)-*N*-benzhydrylaminomethanephosphonous acid **168**



Scheme 5.76



160, **161**

160, PG-X = Cbz-NH, R^1 = Me
161, PG-X = Bzl-O, R^1 = Et

162-**164**

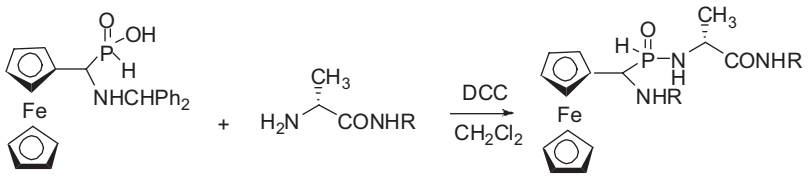
162, PG-X = Cbz-NH, R^1 = Me, R^2 = H
163, PG-X = Cbz-NH, R^1 = Me, R^2 = i-Bu
164, PG-X = Bzl-O, R^1 = Et, R^2 = H

165, X = NH, R^2 = H (racemic mixture)
166, X = NH, R^2 = i-Bu (diastereomeric mixture)
167, X = O, R^2 = H (racemic mixture)

Scheme 5.77

with several model dipeptides **169a–d** and the tripeptide D,L-alanyl- D,L-leucinyl-glycine **169e** in the presence of DCC resulted in pseudopeptides **170a–d** in 61–89% yields [296].

Recently, a new method was developed for the synthesis of phosphinamidopeptides via the Mannich reaction of aryldichlorophosphines, aldehydes, and *N*-Cbz-protected amino amides or peptide amides, and subsequent aminolysis with amino esters or peptide esters. In a typical procedure, phenyldichlorophosphine was added to a solution of *N*-Cbz-glycinamide and benzaldehyde in dry acetonitrile under a nitrogen atmosphere. The resulting mixture was stirred at room temperature for 12 h and further stirred for 24 h after addition of diisopropylethylamine and ethyl



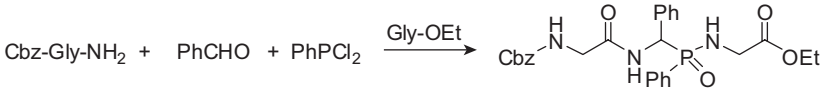
168

169a-e

170a-e

a, R = H_2C-COOH
b, R = H_3C-CH-COOH
c, R = HOOC−CH-CH_2-CH$(CH_3)_2$
d, R = HOOC-CH-CH_2-CH_2-$CONH_2$
e, R = $(H_3C)_2$CH-CH_2-CH-CO-NH-CH_2-COOH

Scheme 5.78



171

Scheme 5.79

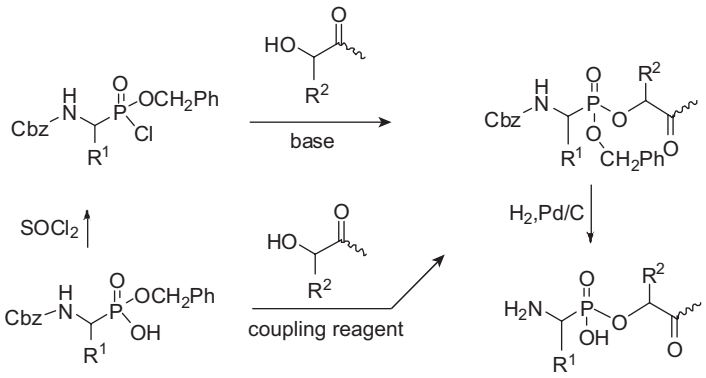
glycine. The phosphinamidopeptide **171** was obtained in a total yield of 67% (Scheme 5.79) [297].

5.6.5
Phosphonodepsipeptides Containing a Phosphonoester Moiety

Pseudopeptide in which a phosphonate ester replaces the peptide bond (phosphonodepsipeptides) are usually synthesized by the reaction of hydroxyalkane carboxylates with phosphonochloridates [292, 298–300] or direct condensation of *N*-protected aminophosphonic monoester or aminophosphinic acid with the appropriate alcohol in the presence of coupling reagents, such as DCC, DCC/base, or CCl_3CN [299, 301] (Scheme 5.80). Most frequently, these peptides are prepared in solution, but solid-phase synthesis of phosphonodepsipeptides has also been developed [302].

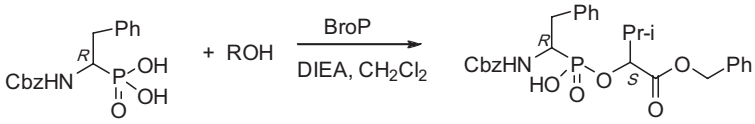
The racemization-free synthesis of α-aminophosphonate monoesters using bromotris(dimethylamino)phosphonium hexafluorophosphate (BroP) as an activating agent is quite useful. In particular, the enantiomerically pure aminophosphonates (*R*)-**172** was monoesterified with benzyl-L-hydroxy-2-isovalerate in the presence of BroP to give the corresponding phosphonodepsipeptide **173** in optically pure form (Scheme 5.81) [303].

A straightforward synthesis of phosphonodepsipeptides has been developed via a Mannich-type multicomponent condensation of benzyl carbamate, aldehydes,



R^1, R^2 = H, Me, Ph, substituted Ph

Scheme 5.80



172 **173** (72%)

ROH = (*S*)-HOCH(i-Pr)COOCH$_2$Ph

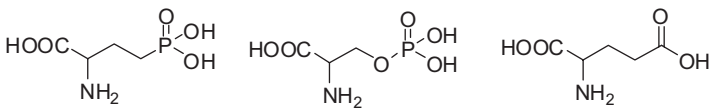
Scheme 5.81

and 1-carbethoxyalkylphosphorodichloridites (Scheme 5.82). This strategy avoids the initial preparation of aminoalkylphosphonic acid or aminoalkylphosphonous acid derivatives as starting materials [304].

5.6.6
Peptides Containing a Phosphonic or Phosphinic Acid Moiety in the Side-Chain

Pseudopeptides of this type have been discussed [252, 304–307]. Substrates for their synthesis are phosphorus analogs of acidic amino acids. For example, 2-amino-3-phosphonopropionic acid **174** may be considered as an analog of aspartic acid, while 2-amino-4-phosphonobutyris acid **175** and *O*-phosphoserine are analogs of glutamic acid. When suitably protected, amino acids **174**, **175**, and related substrates may serve as building blocks to be integrated into oligopeptides.

174 aspartic acid

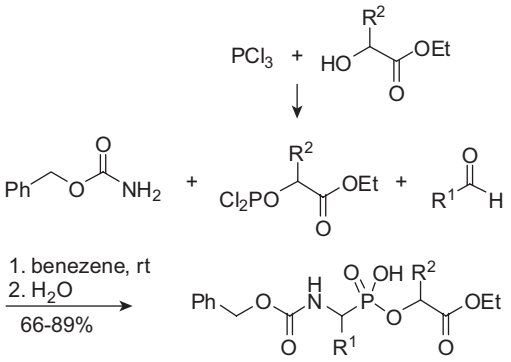


175 *O*-phosphoserine glutamic acid

As in the case of classical peptides, blocking of the amino and phosphonic (or phosphinic) acid groups is a very important aspect of the pseudopeptide synthesis. As an example, efficient synthesis of protected L-phosphonophenylalanine derivative suitable for solid-phase peptide synthesis is outlined in Scheme 5.83. Both amino acids **177** and **178** were successfully incorporated in the sequence of a potassium channel-blocking peptide derived from the sea anemone *Stichodactyla helianthus*, called ShK(L5), that selectively inhibits Kv1.3 channels [308].

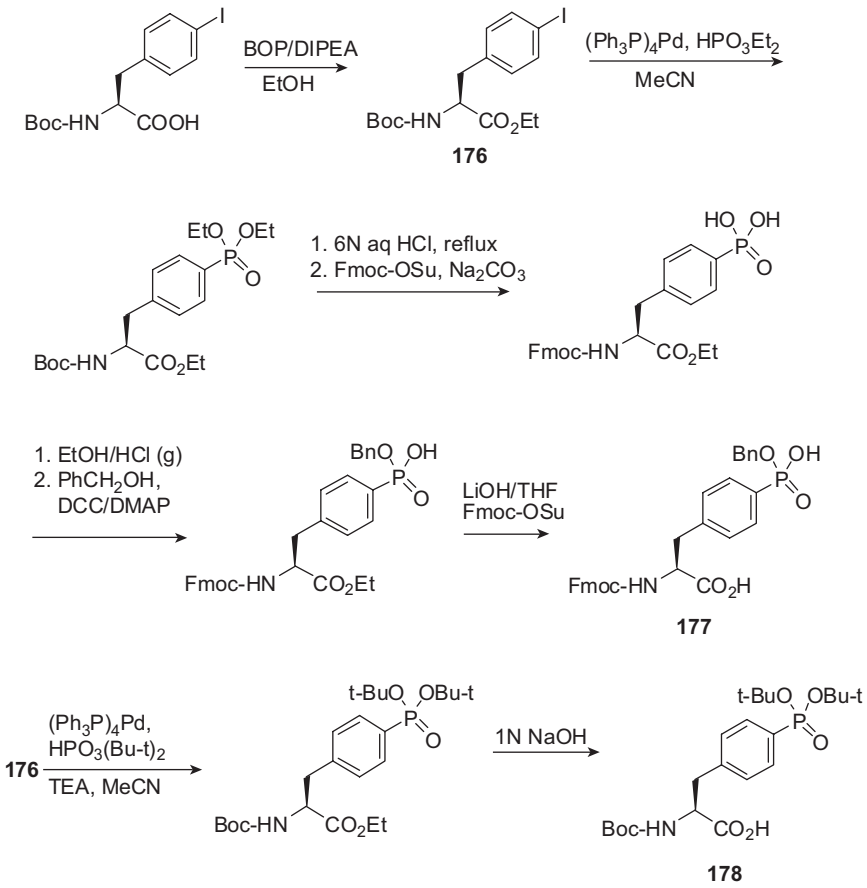
5.7
Remarks on the Practical Utility of Aminophosphonates

The natural roots of aminophosponic acids and a similarity between natural amino acids and their phosphorus analogs designate the first "circle" of practical application



R^1 = Ph, 4-ClC_6H_4, 4-BrC_6H_4, 4-$MeOC_6H_4$
R^2 = H, Me, Ph

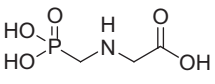
Scheme 5.82



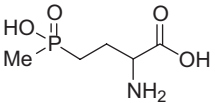
Scheme 5.83

of these compounds as biologically active agents. Bifunctionality and the presence of the phosphonic residue of aminophosphonates can also be useful in diverse areas for other practical tasks – from analytical methods of determination and extractive separation of metal ions to anticorrosion ingredients, bleaching agents, and polymers with specific properties. In a short overview, this section demonstrates some directions for practical applications of aminophosphonates.

Obviously, the herbicide glyphosate (originally marketed as Roundup) occupies the leading position in scale production and commercial use of aminophosphonates. Glyphosate was discovered by Franz in the Monsanto Company in 1971 and is a broad-spectrum, postemergence herbicide useful to control a wide variety of annual and perennial weeds. Glyphosate is formulated in the water-soluble form as an isopropylammonium salt. The systemic action of this compound enables it to be effective against deep-rooted perennial weeds. Another important peculiarity of glyphosate is its rapid metabolism by soil microorganisms to give harmless products; the major metabolite is $(HO)_2P(O)\text{-}CH_2\text{-}NH_2$ which quickly degrades further to CO_2, NH_3, and inorganic phosphate. Detailed studies of the glyphosate mode of action have shown that it is a competitive inhibitor of PEP, and forms a stable complex with the enzyme and shikimate 3-phosphate to inhibit selectively 5-enolpyruvoyl-shikimate 3-phosphate synthase – the enzyme specific for plants and microorganisms, but not in other life forms (the reader can find more details on the bioactivity of aminophosphonates elsewhere [1]).



glyphosate
(*N*-phosphonomethylglycine)



phosphinothricine

The success of glyphosate stimulated an intensive search for other analogs. A highly interesting and active compound, phosphinothricin, was discovered as component of a tripeptide from *S. hygroscopicus*. This is the first example on naturally occurring phosphinate with two P–C bonds that displays bactericidal, fungicidal, and herbicidal activity. The product has been marketed as Glufocinate (ammonium salt) by the Hoechst and as Bialaphos (Bilanafos) by Meiji Seika Kaisha. These broad-spectrum herbicides have favorable environmental characteristics to control a wide range of annual and perennial weeds. Both herbicides are potent inhibitors of glutamine synthetase – an enzyme that catalyzes the conversion of L-glutamate into L-glutamine. The generally low toxicity and environmental compatibility of the compounds are stimulating factors in the continuing search for novel active molecules of this class as low-molecular-weight bioregulators [309].

Numerous aminophosphonic and aminophosphinic acids and their derivatives have been synthesized and tested as potential agrochemicals. Many of them have shown plant-growth regulating or herbicidal activity. Some compounds have shown fungicidal and insecticidal properties. Current work in this field is directed at studies of heterocyclic aminophosphonates and aminophosphinates as well as

aminomethylenebisphosphonic acids and fluorinated derivatives in order to develop new multitarget plant enzyme inhibitors and bioregulators.

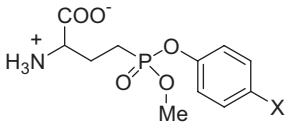
The aminophosphonic and aminophosphinic acid derivatives are classical examples of the replacement of the carboxylic function in an amino acid moiety by phosphorus residue to provide transition-state analogous inhibitors for proteolytic enzymes and phosphate mimics of phosphorylated peptides. As an example, *N*-benzyl-1-aminophosphonic acid is a potent inhibitor of human prostatic phosphatase with an IC_{50} 4 nM [310a]. Aminophosphonate derivatives, especially peptides, were studied as inhibitors of human collagenase (matrix metalloproteinase-1) and a number of serine proteinases, such as thrombin, trypsin, chymotrypsin, and human neutrophil elastase [257]. These compounds were applied in the design of HIV protease inhibitors [255]. In fact, some peptides with an aminophosphinic moiety have shown significant potency at the nanomolar level of inhibition. The substitution of a scissile amide bond with a tetrahedral phosphinic acid isostere as the central unit of symmetrical peptide substrates produces selective and tight-binding inhibitors of HIV protease. Being, in general, a nontoxic class of compounds, aminophosphonic acid derivatives are also attractive in the search for new inhibitors as potential drugs. These compounds are of interest as prodrugs especially, because some aminophosphonic acid esters possess good bioavailability.

Recently UDPGalNAc transition state analogs with an aza-sugar in place of the GalNAc moiety and an aminophosphonate residue were described as a new class of antiprotozoal agents called phosphonoxins [151]. This is the first report of such potential glycosyl transferase inhibitors. One of these compounds is a potent inhibitor of *Giardia* trophozoite growth with activity that rivals existing therapeutics and may have clinical potential as a new anti-*Giardia* drug.

Specific chemical transmitters, analogs of (*S*)-glutamic acid, a major transmitter in the mammalian central nervous system (CNS), which mediate communication between nerve cells are an important area of aminophosphonates biochemical study ([310b] and references cited therein). Numerous aspartate and glutamate analogs using substitution of the carboxyl group by a phosphonate group have been synthesized and tested in structure–activity studies on *N*-methyl-D-aspartate receptor affinity as agonists and antagonists [311]. Other targets are α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) and kainite receptors, and some aminophosphonic acids have shown potent and selective AMPA receptor antagonist activity [1]. Aminophosphonates were described in a great number of studies as agonists and antagonists of metabotropic glutamate [312, 313] and GABA receptors. Aminophosphonic and aminophosphinic acids play an important place in understanding of the roles of the receptors involved in synaptic transmission in CNS, and it is likely that aminophosphonates will be useful in the design of new drugs to treat a range of disorder of the CNS.

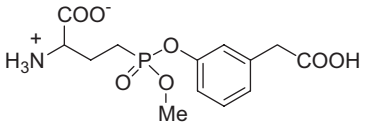
The electrophilic γ-phosphono diester analogs of glutamate **179–182** were shown to be highly potent mechanism-based inhibitors of γ-glutamyl transpeptidase (GGT, EC 2.3.2.2) [298]. In particular, one of the inhibitors (**182**) exhibited more than 6000 times higher activity toward human GGT than the classical inhibitor acivicin. GGT plays a central role in glutathione metabolism. It is involved in

a number of biological events and it is also implicated in physiological disorders, such as Parkinson's disease, neurodegenerative disease, diabetes, and cardiovascular diseases [298].

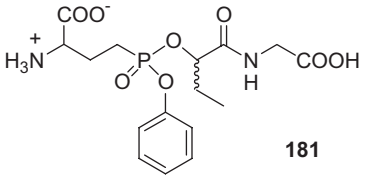


179

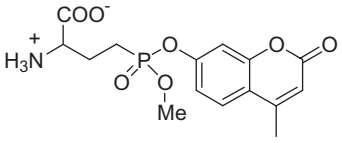
X = H, Me, MeO, CF_3, CN, Cl, NO_2, CH_2COOH



180



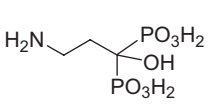
181



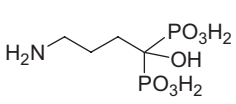
182

In the last few years fosmidomycin and its analogs have attracted significant interest in a search for effective antimalarial drugs [223–225, 227]. In recent clinical trials, fosmidomycin proved to be efficient in the treatment of patients suffering from acute, uncomplicated *Plasmodium falciparum* malaria. Fosmidomycin has the advantage of being remarkably non-toxic and exhibits activity against multiresistant parasite strains. The acetyl derivative of fosmidomycin, FR900098, was shown to be approximately twice as active against *P. falciparum in vitro* as well as in a mouse model. On this basis new efforts are directed at further modification of fosmidomycin as a lead compound. Recent study has shown that large, hydrophobic γ-aminopropylidene-bisphosphonates are inhibitors of *E. coli* cell growth. Their combination with fosmidomycin demonstrated a high synergism, opening up new possibilities for antibacterial therapy targeting the isoprenoid biosynthesis pathway [314].

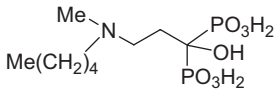
An interest in geminal bisphosphonates and their amino derivatives is associated with a development of new drugs for the treatment of tumor-induced hypercalcemia, Paget's disease of bone, osteolytic metastases, and postmenopausal osteoporosis [315, 316]. Bisphosphonates have also been found to exhibit considerable antiparasitic activity and to be immunomodulators. Some aminomethylene-bisphosphonates have been reported to be potent inhibitors of the major mevalonate–isoprene biosynthesis pathway enzyme, farnesyl diphosphate synthase from the trypanosomatids and farnesyl diphosphate synthase from *Leishmania major* – the organism responsible for cetaceous leishmaniasis [317]. The bisphosphonates pamidronate, alendronate, ibandronate, risedronate, and zoledronate are potent stimulators of γδ T cells. It has been proposed that this T cell activation may represent a new approach to the development of antibiotic compounds and potentially novel anticancer drugs.



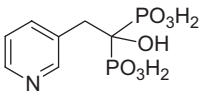
pamidronate



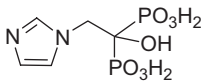
alendronate



ibandronate



risedronate



zoledronate

N-Substituted γ-aminophosphonates are potent, subtype selective agonists and antagonists of the five known sphingosine 1-phosphate receptors modulating the immune system [201, 220, 318]. *In vitro* binding assays showed that the implementation of phosphonates as phosphate mimetics provided compounds with similar receptor-binding affinities as compared to their phosphate precursors.

An ability of aminophosphonic acids to form complexes with metal ions is also useful for technical processes. Numerous patents describe using aminonophosphonates as additives for the bleaching process in paper technology to enhance paper brightness (e.g., [319]). Derivatives of aminophosphonic acids are proposed as corrosion inhibitors [320]. Quaternary aminomethyl phosphonates can be applied as scale inhibitors [321] and flame retardants for textile coatings [322].

Aminophosphonic acids are recommended to modify microporous and macroporous resins in order to remove Cu^{2+} and Ni^{2+} cations from wastewater. Aminophosphonate-chelating ion-exchange resin Duolite C467 can be used for separation of cations in waste treatment and in water purification for some metals [323–325]. The Duolite C467 resin can be recommended for Pd(II) ion removal from anodic slimes as well as Pd(II) trace analysis due to its high selectivity [326]. *N*,*N*-Di(2-ethylhexyl)amino-methylphosphonate and similar compounds are extractants with high selectivity for Pt(IV) and Pd(II) ions from aqueous chloride media into toluene [327]. A β-aminobisphosphonate receptor combined with naphthalene in the fluorophore–spacer–receptor format was proposed as a photoinduced electron transfer sensor for Cu(II) with sensitivity in the micromolar range in a wide pH range (2–10) [147].

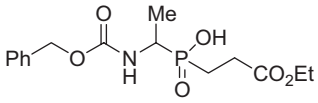
5.8 Conclusions

As can be seen from the data collected in this chapter, it is clear that the analogy between the chemistry of amino acids and aminophosphonic acids is very deep. Aminophosphonic acids efficiently mimic aminocarboxylic acids, and their phosphorus bioisosteres compete with carboxylic counterparts for the active sites of

enzymes and other cell receptors. Thus, the aminophosphonate moiety is a versatile pharmacophore that is essential for successful drug design. In particular, because of their ability to mimic transition states of hydrolysis, aminophosphonic acids have been shown to be inhibitors of various enzymes, including HIV protease. Aminophosphonic acids also received wide attention as useful tools for studying mechanisms of enzyme action. Synthetic approaches to aminophosphonic acids presented in this chapter can be applied to numerous new analogs of amino acids and peptides in which the carboxylic group was replaced by phosphonic or related function. Further studies to obtain additional details concerning the interactions of aminophosphonates substrates with enzymatic binding partners are likely to be the next challenge in aminophosphonic acid chemistry.

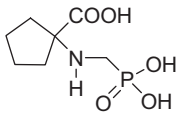
5.9
Experimental Procedures

5.9.1
Synthesis of N-Protected α-aminophosphinic Acid 10 ($R^1 = EtOCOCH_2$, $R^2 = Me$) [38]



To a stirred solution of 2-ethoxycarbonylethylphosphonous acid (0.33 mmol) and benzyl carbamate (0.34 mmol) in acetyl chloride (2 ml) was added acetaldehyde (0.34 mmol) at 0 °C. Stirring was continued at 0 °C for 30 min and then at room temperature for 6 h. Volatile components were removed under reduced pressure, and the residue was partitioned between 5% $NaHCO_3$ solution (20 ml) and ether (10 ml). The aqueous solution was separated and acidified to pH 2 with concentrated HCl and then extracted with EtOAc (4 × 20 ml). Ethyl acetate extracts were combined and dried over Na_2SO_4. The crude product, 90 mg syrup, obtained after removal of the solvent was quite pure as indicated by NMR analysis. Further purification was carried out by ion-exchange chromatography on diethylaminoethyl cellulose (eluting with 1 : 1 MeOH/0.1 M NH_4HCO_3 buffer) followed by cation exchange on Dowex AG-8 resin (H^+ form).

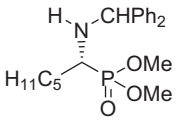
5.9.2
Synthesis of Phosphonomethylaminocyclopentane-1-carboxylic Acid (17) [85]



Paraformaldehyde (20 mmol), methanol (15 ml), and triethylamine (0.2 ml) were put into a three-necked flask equipped with a condenser, magnetic stirrer, and thermometer. The reaction mixture was heated to reflux temperature and held there for 45 min, after which it became a clear solution. To this solution was added aminocyclopentane-1-carboxylic acid (12.3 mmol) and triethylamine (12.3 mmol). This suspension was then heated at 65–70 °C and after 30 min it became a clear solution. To this solution was added diethyl phosphite (81 mmol) over a period of approximately 10 min. The reaction mixture was heated at 65–70 °C and held there for 2.5 h, after which it was cooled to room temperature, and while stirring, sodium hydroxide (3.08 g, 40% water solution) was added. The reaction mixture was heated at 105–110 °C for 5 h. Subsequently, it was treated with Dowex 50WX8-200 in order to exchange the sodium cations by hydrogen ions. The water was removed under vacuum to give **17** (1.81 g, 95%); melting point 224–225 °C.

5.9.3

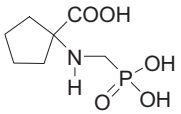
General Procedure for Catalytic Asymmetric Hydrophosphonylation. Synthesis of α-Aminophosphonate 39 ($R^1 = C_5H_{11}$, $R^2 = Ph_2CH$) [121]



To imine $C_6H_{11}CH{=}NCHPh_2$ (0.19 mmol) and dimethyl phosphite (0.95 mmol) was slowly added a solution of LnPB complex (38 mmol) in toluene/THF (7 : 1) (1.5 ml) at room temperature under argon atmosphere. After being stirred for 90 h, the mixture was quenched with water and extracted with ethyl acetate (3 × 15 ml). The combined organic layer was dried over anhydrous Na_2SO_4. Removal of the solvent and flash column chromatography (silica gel, 1 : 2, hexane/EtAc) gave the desired coupling product **39** (73.3 mg, 87%, 85% e.e.) as a colorless oil.

5.9.4

General Procedure of the Asymmetric Aminohydroxylation Reaction: Synthesis of β-Amino-α-hydroxyphosphonates 87 [159]

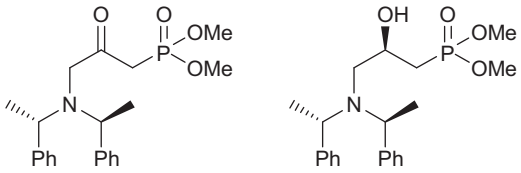


To a solution of $(DHQD)_2$-PHAL (0.05 mmol, 3.5 mol%) and chloramines-T hydrate (3.15 mmol) in *t*BuOH.H_2O 1 : 1 (20 ml) the appropriate vinylphosphonates **86** (0.72 mmol) and potassium osmate dehydrate (0.035 mmol, 2.4 mol%) were added (in that order) at room temperature under stirring. The color of the mixture changed from yellow to green in 10 min, and then turned back to yellow

in a time varying from 90 min to 2 h. The remainder of chloramines-T hydrate (1.25 mmol) and vinyl phosphonate **86** (0.72 mmol) was added, and stirring was continued at room temperature until thin-layer chromatography ($CHCl_3$/MeOH 49 : 1) indicated consumption of starting material (2–24 h). *t*BuOH was then evaporated under vacuum and the mixture was quenched with sodium sulfite (0.45 g). After stirring for 10 min, dichloromethane (10 ml) was added, the organic layers were separated, and the aqueous layer was exhaustively extracted with dichloromethane (2 × 10 ml). The combined organic extracts were dried over sodium sulfate and concentrated to yield the crude product. Purification by flash silica gel chromatography (chloroform/methanol 99 : 1 or ethyl acetate) afforded the tosylamido derivatives **87** as colorless solids.

5.9.5

Dimethyl (S,S)-(−)3-N,N-bis(α-Methylbenzyl) amino-2-oxopropylphosphonate (S,S)-100 and Dimethyl 3-[(S,S)-N,N-bis(α-methylbenzylamino)-(2R)-hydroxypropylphosphonate (R,S,S)-101 [217]

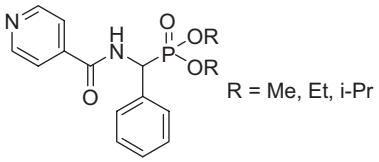


A solution of dimethyl methylphosphonate (11.56 mmol) in anhydrous THF (20 ml) was cooled at −78 °C before the slow addition of 3.61 ml (8.67 mmol) of *n*BuLi 2.4 M in hexanes. The resulting solution was stirred at −50 °C for 1.5 h, and then the solution cooled at −78 °C and slowly added to a solution of ethyl glycinate (*S,S*)-$(MePhCH)_2NCH_2COOEt$ (2.89 mmol) in anhydrous THF (20 ml). The reaction mixture was stirred at −78 °C for 4 h, quenched with aqueous NH_4Cl solution (10 ml), and extracted with ethyl acetate (3 × 20 ml). The combined organic extracts were washed with brine solution (2 × 10 ml), dried over anhydrous Na_2SO_4, filtered, and concentrated in vacuum. The crude product was purified by column chromatography (ethyl acetate-hexane 70 : 30) to afford 0.98 g, 87% yield of (*S,S*)-**100** as a colorless oil.

To a solution of β-ketophosphonate (*S,S*)-**100** (0.37 mmol), $LiClO_4$ (0.93 mmol), and dry THF (20 ml) cooled at −78 °C were added 179 mg (1.49 mmol) of catecholborane 1 M in THF. The reaction mixture was stirred at −78 °C for 4 h and at room temperature for 10 h. The reaction was quenched with aqueous NH_4Cl solution. The solvent was evaporated in vacuum, the residue was dissolved in water (10 ml), and extracted with ethyl acetate (3 × 20 ml). The combined organic extracts were dried over anhydrous Na_2SO_4, filtered, and concentrated in vacuum. The crude product was analyzed by ^{1}H-NMR at 400 MHz and ^{31}P-NMR at 200 MHz, and then purified by flash chromatography (ethyl acetate/hexane 80 : 20) to afford 128 mg, 89% yield of (*R,S,S*)-**101** as a white solid, melting point 115–118 °C, $[\alpha]_D = 23.1$ (*c* 0.59, $CHCl_3$).

5.9.6

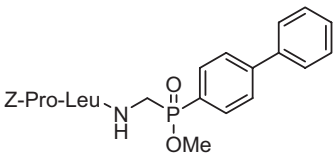
General Procedure for the Preparation of Dialkyl Phenyl(4-pyridylcarbonylamino) methyl-phosphonates 126 [261]



Triphenylphosphine (15 mmol) and hexachloroethane (15 mmol) were dissolved in 1,2-dichloroethane (20 ml) under nitrogen atmosphere for 1 h. The reacted solution was added dropwise to a mixture of the dialkyl α-aminobenzylphosphonate hydrochloride (10 mmol) and isonicotinic acid (10 mmol) in 1,2-dichloroethane (90 ml) and 4 ml of triethylamine. After 24 h the reaction was completed. The reaction mixture was acidified to pH 1 with HCl and the solution was washed with Et_2O to remove neutral materials. The aqueous phase was then adjusted to pH 11 with aqueous NaOH, and the product was extracted with CH_2Cl_2; the solvent was removed to give the crude product, which was purified by recrystallization.

5.9.7

Synthesis of 1-[(Benzyloxy) carbonyl] prolyl-N1-{[1,1′-biphenyl-4-yl-methyl)(methoxy) phosphoryl] methyl}leucinamide (159a) [281]



To a cooled (0 °C) funnel-equipped flask containing a dry THF (4 ml) solution of **158a** (0.43 mmol) and Cbz-Pro–Leu-OH (0.43 mmol), a solution of DCC (0.47 mmol) and HOBt (0.47 mmol) dissolved in 2 ml of the same solvent was added dropwise under stirring. After standing at room temperature overnight, *N,N′*-dicyclohexylurea was filtered off and the solution was evaporated under reduced pressure. The residue, dissolved in EtOAc, was washed with saturated solution of $NaHCO_3$ and brine, and organic layer dried over anhydrous Na_2SO_4. Crude oil was purified by flash chromatography (EtOAc/MeOH 95 : 5) affording 212 mg (80%) of product **159a** as 1 : 1 mixture of diastereomers.

References

1 Kukhar, V. and Hudson, H.(eds) (2000) *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity*, John Wiley & Sons, Inc., New York.

2 Horiguchi, M. and Kandatsu, M. (1959) *Nature*, **184**, 901.

3 Mastalelerz, P. and Kafarski, P. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 1–31.

4 Fields, S.C. (1999) *Tetrahedron*, **55**, 12237.

5 (a) Seidel, H.M., Freeman, S., Seto, H. and Knowles, J.R. (1988) *Nature*, **335**, 457; (b) Bowman, E.D., McQueney, M.S., Barry, R.J. and Dunaway-Mariano, D. (1988) *Journal of the American Chemical Society*, **110**, 5575; (c) Kim, J. and Dunaway-Mariano, D. (1996) *Biochemistry*, **35**, 4628; (d) Kim, A., Kim, J., Martin, B.M. and Dunaway-Mariano, D. (1998) *The Journal of Biological Chemistry*, **273**, 4443; (e) Liu, S., Lu, Z., Han, Y., Jia, Y., Howard, A., Dunaway-Mariano, D. and Herzberg, O. (2004) *Biochemistry*, **43**, 4447.

6 Kiss, T. and Lázár, I. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 285–326.

7 Saiakhov, R.D., Cherkasov, A.R., Shajmardanova, A.A., Galkin, V.I., Cherkasov, R.A., Finocchiaro, P. and Failla, S. (2002) *Phosphorus Sulfur and Silicon and the Related Elements*, **177**, 1281.

8 Garifzyanov, A.R., Nuriazdanova, G.Kh. Zakharov, S.V. and Cherkasov, R.A. (2004) *Russian Journal of Inorganic Chemistry*, **74**, 1885.

9 Choi, N. and McPartlin, M. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 327–362.

10 Hägele, G. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 217–284.

11 Quin, L.D. and Quin, G.S. (2001) *Comparative Biochemistry and Physiology*, **128**, 173.

12 Jankowski, S., Huben, K., Paneth, P., Ciesielski, W., Quin, L.D., Vokal, B. and Dolenec, T. (2005) *Polish Journal of Chemistry*, **79**, 573.

13 Hagele, G., Szakacs, Z., Ollig, J., Hermens, S. and Pfaff, C. (2000) *Heteroatom Chemistry*, **11**, 562.

14 Blazewska, K., Paneth, P. and Gajda, T. (2007) *The Journal of Organic Chemistry*, **72**, 878.

15 Rudzinska, E., Berlicki, L., Mucha, A. and Kafarski, P. (2007) *Tetrahedron – Asymmetry*, **18**, 1579.

16 Berlicki, L., Rudzinska, E. and Kafarski, P. (2003) *Tetrahedron – Asymmetry*, **14**, 1535.

17 Galushko, S.V. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York,, pp. 205–216.

18 Zarbl, E., Lammerhofer, M., Hammerschmidt, F., Wuggenig, F., Hanbauer, M., Maier, N.M., Sajovic, L. and Lindner, W. (2000) *Analytica Chimica Acta*, **404**, 169.

19 Lammerhofer, M., Zarbl, E., Elfriede, W., Linder, W., Simov, B.P. and Hammerschmidt, F. (2001) *Electrophoresis*, **22**, 1182.

20 Rudzinska, E., Dzygiel, P., Wieczorek, P. and Kafarski, P. (2000) *Journal of Chromatography. A*, **979**, 115.

21 Rudzinska, E., Wieczorek, P. and Kafarski, P. (2003) *Electrophoresis*, **24**, 2693.

22 Kudzin, Z.H. and Sochacki, M. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 363–405.

23 Engelmann, M. and Pikl, J. (1943) *Chemical Abstracts*, **37**, 20212.

24 Birum, G.H. (1974) *The Journal of Organic Chemistry*, **39**, 209.

25 Oleksyszyn, J. (1980) *Synthesis*, 722.

26 Oleksyszyn, J., Gruszecka, E., Kafarski, P. and Mastalerz, P. (1982) *Monatshefte fur Chemie*, **113**, 59.

27 Kafarski, P. and Zoń, J. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 33–74.

28 Dai, Q. and Chen, R. (1997) *Synthetic Communications*, **27**, 1653.

29 Dai, Q. and Chen, R.-Y. (1997) *Heteroatom Chemistry*, **8**, 203.

30 Dai, Q., Chen, R.Y., Zhang, C.X. and Liu, Z. (1998) *Synthesis*, 405.

31 Cai, S., Yin, W., Ma, Y., Liu, H., Zhao, Y. and Xu, J. (2007) *Phosphorus Sulfur and Silicon and the Related Elements*, **182**, 25.

32 Dai, Q., Wo, H. and Chen, R. (1998) *Heteroatom Chemistry*, **9**, 511.

33 Sienczyk, M. and Oleksyszyn, J. (2004) *Tetrahedron Letters*, **45**, 7251.

34 Van der Veken, P., El Sayed, I., Joossens, J., Stevens, C.V., Augustyns, K. and Haemers, A. (2005) *Synthesis*, 634.

35 Huang, J. and Chen, R. (2000) *Heteroatom Chemistry*, **11**, 480.

36 Miao, Z., Wang, B., Zhang, G. and Chen, R. (2006) *Bioorganic Chemistry*, **34**, 167.

37 Kumara Swamy, K.C., Kumaraswamy, S., Kumar, K.S. and Muthiah, C. (2005) *Tetrahedron Letters*, **46**, 3347.

38 Chen, S. and Coward, J.K. (1996) *Tetrahedron Letters*, **37**, 4335.

39 Matziari, M. and Yiotakis, A. (2005) *Organic Letters*, **7**, 4049.

40 Chung, S.-K. and Kang, D.-H. (1996) *Tetrahedron – Asymmetry*, **7**, 21.

41 Syamala, M. (2005) *Organic Preparations and Procedures International*, **37**, 103.

42 (a) Cherkasov, R.A. and Galkin, V.I. (1998) *Russian Chemical Reviews*, **67**, 857; (b) Cherkasov, R.A. and Galkin, V.I. (1998) *Uspekhi Khimii*, **67**, 940.

43 Lv, X.-Y., Zhang, J.-M., Xing, C.-H., Du, W.-Q. and Zhu, S.-Z. (2007) *Synthetic Communications*, **37**, 743.

44 Mu, X.-J., Lei, M.-Y., Zou, J.-P. and Zhang, W. (2006) *Tetrahedron Letters*, **47**, 1125.

45 Kabachnik, M.M., Zobnina, E.V. and Beletskaya, I.P. (2005) *Synlett*, 1393.

46 Zhan, Z.-P., Yang, R.-F. and Li, J.-P. (2005) *Chemistry Letters*, **34**, 1042.

47 Song, B., Yang, S., Hong, Y., Zhang, G., Jin, L. and Hu, D. (2005) *Journal of Fluorine Chemistry*, **126**, 1419.

48 Kaboudin, B. and As-Habei, N. (2003) *Tetrahedron Letters*, **44**, 4243.

49 Kaboudin, B. and Rahmani, A. (2003) *Synthesis*, 2705.

50 Kaboudin, B. and Saadati, F. (2004) *Synthesis*, 1249.

51 Kaboudin, B. (2001) *Chemistry Letters*, 880.

52 Kaboudin, B. and Nazari, R. (2001) *Tetrahedron Letters*, **42**, 8211.

53 Jin, L., Song, B., Zhang, G., Xu, R., Zhang, S., Gao, X., Hu, D. and Yang, S. (2006) *Bioorganic & Medicinal Chemistry Letters*, **16**, 1537.

54 Wu, J., Sun, W., Sun, X. and Xia, H.-G. (2006) *Green Chemistry*, **8**, 365.

55 Akiyama, T., Sanada, M. and Fuchibe, K. (2003) *Synlett*, 1463.

56 Bhagat, S. and Chakraborti, A.K. (2007) *The Journal of Organic Chemistry*, **72**, 1263.

57 Wu, J., Sun, W., Xia, H.-G. and Sun, X. (2006) *Organic and Biomolecular Chemistry*, **4**, 1663.

58 Azizi, N. and Saidi, M.R. (2003) *European Journal of Organic Chemistry*, 4630.

59 Heydari, A., Karimian, A. and Ipaktschi, J. (1998) *Tetrahedron Letters*, **39**, 6729.

60 Firouzabadi, H., Iranpoor, N. and Sobhani, S. (2004) *Synthesis*, 2692.

61 Ghosh, R., Maiti, S., Chakraborty, A. and Maiti, D.K. (2004) *Journal of Molecular Catalysis A – Chemical*, **210**, 53.

62 Qian, C.T. and Huang, T.S. (1998) *The Journal of Organic Chemistry*, **63**, 4125.

63 Ranu, B.C., Hajra, A. and Jana, U. (1999) *Organic Letters*, **1**, 1141.

64 Xu, F., Luo, Y., Deng, M. and Shen, Q. (2003) *European Journal of Organic Chemistry*, 4728.

65 Xu, F., Luo, Y., Wu, J., Shen, Q. and Chen, H. (2006) *Heteroatom Chemistry*, **17**, 389.

66 Chandrasekhar, S., Prakash, S.J., Jagadeshwar, V. and Narsihmulu, C. (2001) *Tetrahedron Letters*, **42**, 5561.

67 Bhattacharya, A.K. and Kaur, T. (2007) *Synlett*, 745.

68 Manabe, K. and Kobayashi, S. (2000) *Chemical Communications*, 669.

69 Yadav, J.S., Reddy, B.V.S. and Madan, C. (2001) *Synlett*, 1131.

70 Takahashi, H., Yoshioka, M., Imai, N., Onimura, K. and Kobayashi, S. (1994) *Synthesis*, 763.

71 Kaboudin, B. and Moradi, K. (2005) *Tetrahedron Letters*, **46**, 2989.

72 Sardarian, A.R. and Kaboudin, B. (1997) *Tetrahedron Letters*, **38**, 2543.

73 Kaboudin, B. and Jafari, E. (2007) *Synthesis*, 1823.

74 Yang, R., Zhao, R., Zhao, L., Yun, L. and Wang, H. (2003) *Synthesis*, 887.

75 Journet, M., Cai, D., Larsen, R.D. and Reider, P.J. (1998) *Tetrahedron Letters*, **39**, 1717.

76 Varga, T.R. (1997) *Synthetic Communications*, **27**, 2899.

77 Antipin, I.S., Stoikov, I.I., Pinkhassik, E.M., Fitseva, N.A., Stibor, I. and Konovalov, A.I. (1997) *Tetrahedron Letters*, **38**, 5865.

78 Bailly, T. and Burgada, R. (1995) *Phosphorus Sulfur and Silicon and the Related Elements*, **101**, 131.

79 Matveeva, E.D., Podrugina, T.A., Tishkovskaya, E.V., Tomilova, L.G. and Zefirov, N.S. (2003) *Synlett*, 2321.

80 Tesson, N., Dorigneux, B. and Fadel, A. (2002) *Tetrahedron – Asymmetry*, **13**, 2267.

81 Stoikov, I.I., Repejkov, S.A., Antipin, I.S. and Konovalov, A.I. (2000) *Heteroatom Chemistry*, **11**, 518.

82 Fadel, A. (1999) *The Journal of Organic Chemistry*, **64**, 4953.

83 Rabasso, N., Louaisil, N. and Fadel, A. (2006) *Tetrahedron*, **62**, 7445.

84 Maury, C., Royer, J. and Husson, H.-P. (1992) *Tetrahedron Letters*, **33**, 6127.

85 Najdenova, E., Vassilev, A., Popova, Yu. and Troev, K. (2003) *Heteroatom Chemistry*, **14**, 229.

86 Kudrimoti, S. and Bommena, V.R. (2005) *Tetrahedron Letters*, **46**, 1209.

87 Reddy, K.R., Reddy, K.S., Reddy, C.V., Mahesh, M., Raju, P.V.K. and Reddy, V.V.N. (2005) *Chemistry Letters*, **34**, 444.

88 Reddy, Y.T., Reddy, P.N., Kumar, B.S., Rajput, P., Sreenivasulu, N. and Rajitha, B. (2007) *Phosphorus Sulfur and Silicon and the Related Elements*, **182**, 161.

89 Azizi, N. and Saidi, M.R. (2003) *Tetrahedron*, **59**, 5329.

90 Azizi, N., Rajabi, F. and Saidi, M.R. (2004) *Tetrahedron Letters*, **45**, 9233.

91 Deng, S.-L. and Dong, H.-C. (2006) *Synthetic Communications*, **36**, 3411.

92 Chandrasekhar, S., Narsihmulu, C., Sultana, S.S., Saritha, B. and Prakash, S.J. (2003) *Synlett*, 505.

93 Hamilton, R., Walker, B. and Walker, B.J. (1995) *Tetrahedron Letters*, **36**, 4451.

94 Kolodiazhnyi, O.I., Grishkun, E.V., Sheiko, S., Demchuk, O., Thoennessen, H., Jones, P.G. and Schmutzler, R. (1998) *Tetrahedron – Asymmetry*, **9**, 1645.

95 Kukhar, V.P., Soloshonok, V.A. and Solodenko, V.A. (1994) *Phosphorus Sulfur and Silicon and the Related Elements*, **92**, 239.

96 Stevens, C.V., Van Meenen, E., Masschelein, K.G.R., Moonen, K., De Blieck, A. and Drabowicz, J. (2007) *Synlett*, 2549.

97 Van Meenen, E., Moonen, K., Verwée, A. and Stevens, C.V. (2006) *The Journal of Organic Chemistry*, **71**, 7903.

98 Kaboudin, B. and Moradi, K. (2006) *Synthesis*, 2339.

99 Olszewski, T.K., Boduszek, B., Sobek, S. and Kozłowski, H. (2006) *Tetrahedron*, **62**, 2183.

100 Akiyama, T., Matsuda, K. and Fuchibe, K. (2005) *Synthesis*, 2606.

101 Cristau, H.-J., Hervé, A. and Virieux, D. (2004) *Tetrahedron*, **60**, 877.

102 Szabó, A., Petneházy, I. and Jászay, Z.M. (2003) *Heteroatom Chemistry*, **14**, 235.

103 Chen, X. and Wiemer, D.F. (2003) *The Journal of Organic Chemistry*, **68**, 6108.

104 Kaboudin, B. and Jafari, E. (2006) *Synthesis*, 3063.

105 (a) Vasella, A. and Voeffray, R. (1982) *Helvetica Chimica Acta*, **65**, 1953; (b) Huber, R., Knierzinger, A., Obrecht, J.-P. and Vasella, A. (1985) *Helvetica Chimica Acta*, **68**, 1730; (c) Bernet, B., Krawczyk, E. and Vasella, A. (1985) *Helvetica Chimica Acta*, **68**, 2299; (d) Huber, R. and Vasella, A. (1990) *Tetrahedron*, **46**, 33.

106 De Risi, C., Perrone, D., Dondoni, A., Pollini, G.P. and Bertolasi, V. (2003) *European Journal of Organic Chemistry*, 1904.

107 Kaboudin, B., Haruki, T., Yamagishi, T. and Yokomatsu, T. (2007) *Tetrahedron*, **63**, 8199.

108 Déjugnat, C., Etemad-Moghadam, G. and Rico-Lattes, I. (2003) *Chemical Communications*, 1858.

109 (a) Kolodyazhnyi, O.I. and Sheiko, S.Yu. (2001) *Zhurnal Obshchei Khimii*, **71**, 1039; (b) Kolodyazhnyi, O.I. and Sheiko, S.Yu. (2001) *Russian Journal of Organic Chemistry*, **71**, 1155.

110 Louaisil, N., Rabasso, N. and Fadel, A. (2007) *Synthesis*, 289.

111 Tongcharoensirikul, P., Suarez, A.I., Voelker, T. and Thompson, C.M. (2004) *The Journal of Organic Chemistry*, **69**, 2322.

112 Szabó, A., Jászay, Z.M., Hegedüs, L., Töke, L. and Petneházy, I. (2003) *Tetrahedron Letters*, **44**, 4603.

113 Mikołajczyk, M. (2005) *Journal of Organometallic Chemistry*, **690**, 2488.

114 Szabó, A., Jászay, Z.M., Töke, L. and Petneházy, I. (2004) *Tetrahedron Letters*, **45**, 1991.

115 Davis, F.A. and Prasad, K.R. (2003) *The Journal of Organic Chemistry*, **68**, 7249.

116 Davis, F.A., Lee, S., Yan, H. and Titus, D.D. (2001) *Organic Letters*, **3**, 1757.

117 Lefebvre, I.M. and Evans, S.A. (1997) *The Journal of Organic Chemistry*, **62**, 7532.

118 Mikołajczyk, M. and Łyżwa, P. and Drabowicz, J. (1997) *Tetrahedron – Asymmetry*, **8**, 3991.

119 Mikołajczyk, M., Łyżwa, P., Drabowicz, J., Wieczorek, M.W. and Błasczyk, J. (1996) *Chemical Communications*, 1503.

120 Mikołajczyk, M. and Łyżwa, P. and Drabowicz, J. (2002) *Tetrahedron – Asymmetry*, **13**, 2571.

121 (a) Sasai, H., Arai, S., Tahara, Y. and Shibasaki, M. (1995) *The Journal of Organic Chemistry*, **60**, 6656; (b) for review see Shibasaki, M., Sasai, H. and Arai, T. (1997) *Angewandte Chemie*, **109**, 1291.

122 Gröger, H., Saida, Y., Sasai, H., Yamaguchi, K., Martens, J. and Shibasaki, M. (1998) *Journal of the American Chemical Society*, **120**, 3089.

123 Gröger, H., Saida, Y., Arai, S., Martens, J., Sasai, H. and Shibasaki, M. (1996) *Tetrahedron Letters*, **37**, 9291.

124 Akiyama, T., Morita, H., Itoh, J. and Fuchibe, K. (2005) *Organic Letters*, **7**, 2583.

125 Joly, G.D. and Jacobsen, E.N. (2004) *Journal of the American Chemical Society*, **126**, 4102.

126 Pettersen, D., Marcolini, M., Bernardi, L., Fini, F., Herrera, R.P., Sgarzani, V. and Ricci, A. (2006) *The Journal of Organic Chemistry*, **71**, 6269.

127 Schöllkopf, U. and Schuetze, R. (1987) *Liebigs Annalen der Chemie*, 45.

128 Groth, U., Richter, L. and Schöllkopf, U. (1992) *Liebigs Annalen der Chemie*, 903.

129 McNulty, J., Das, P. and Gosciniak, D. (2008) *Tetrahedron Letters*, **49**, 281.

130 Kobayashi, S., Kiyohara, H., Nakamura, Y. and Matsubara, R. (2004) *Journal of the American Chemical Society*, **126**, 6558.

131 Risch, N., Piper, S., Winter, A. and Lefarth-Risse, A. (2005) *European Journal of Organic Chemistry*, 387.

132 Dodda, R. and Zhao, C.-G. (2007) *Tetrahedron Letters*, **48**, 4339.

133 Dodda, R. and Zhao, C.-G. (2007) *Organic Letters*, **9**, 165.

134 Denmark, S.E., Chatani, N. and Pansare, S.V. (1992) *Tetrahedron*, **48**, 2191.

135 Bernardi, L., Zhuang, W. and Jørgensen, K.A. (2005) *Journal of the American Chemical Society*, **127**, 5772.

136 Aller, E., Buck, R.T., Drysdale, M.J., Ferris, L., Haigh, D., Moody, C.J., Pearson, N.D. and Sanghera, J.B. (1996) *Journal of the Chemical Society, Perkin Transactions 1*, 2879.

137 Ferris, L., Haigh, D. and Moody, C.J. (1995) *Synlett*, 921.

138 Titanyuk, I.D., Vorob'eva, D.V., Osipov, S.N. and Beletskaya, I.P. (2006) *Synlett*, 1355.

139 Yamazaki, K., Nakamura, Y. and Kondo, Y. (2003) *The Journal of Organic Chemistry*, **68**, 6011.

140 Dolence, E.K. and Roylance, J.B. (2004) *Tetrahedron – Asymmetry*, **15**, 3307.

141 Piotrowska, D.G. and Wróblewski, A.E. (2003) *Tetrahedron*, **59**, 8405.

142 Demir, A.S., Tanyeli, C., Şeşenoglu, Ö., Demic, Ş. and Evin, Ö.Ö. (1996) *Tetrahedron Letters*, **37**, 407.

143 Palacios, F., Vicario, J., Maliszewska, A. and Aparicio, D. (2007) *The Journal of Organic Chemistry*, **72**, 2682.

144 Palacios, F., Alonso, C. and de los Santos, J.M. (2005) *Chemical Reviews*, **105**, 899.

145 Gubnitskaya, E.S., Peresypkina, L.P. and Samarai, L.I. (1990) *Uspekhi Khimii*, **59**, 1386.

146 Pudovik, A.N. and Denisova, G.M. (1953) *Zhurnal Obshchei Khimii*, **23**, 263.

147 Kamila, S., Callan, J.F., Mulrooney, R.C. and Middleton, M. (2007) *Tetrahedron Letters*, **48**, 7756.

148 Cherkasov, R.A., Galkin, V.I., Khusainova, N.G., Mostovaya, O.A., Garifzyanov, A.R., Nuriazdanova, G.Kh., Krasnova, N.S. and Berdnikov, E.A. (2005) *Russian Journal of Organic Chemistry*, **41**, 1481.

149 (a) Kabachnik, M.I., Tzvetkov, E.N. and Chang, C.-Y. (1962) *Zhurnal Obshchei Khimii*, **32**, 3340; (b) Kabachnik, M.I. (1964) *Tetrahedron*, **20**, 655.

150 Cen, W. and Shen, Y. (1995) *Journal of Fluorine Chemistry*, **72**, 107.

151 (a) Suk, D.-H., Rejman, D., Dykstra, C.C., Pohl, R., Pankiewicz, K.W. and Patterson, S.E. (2007) *Bioorganic & Medicinal Chemistry Letters*, **17**, 2811; (b) Farahani, M., Antonucci, J.M. and Karam, L.R. (1998) *Journal of Applied Polymer Science*, **67**, 1545.

152 Bailly, T. and Burgada, R. (1994) *Phosphorus Sulfur and Silicon and the Related Elements*, **86**, 217.

153 (a) Alfer'ev, I.S., Kotlyarevskii, I.L., Mikhalin, N.V. and Novikova, V.M. (1983) *Izvestiya Akademii Nauk SSSR – Seriya Khimicheskaya*, 2802; (b) Matveev, S.V., Bel'skii, F.I., Matveeva, A.G., Gukasova, A.Yu., Polykarpov, Yu.M. and Kabachnik, M.I. (1998) *Izvestiya Akademii Nauk SSSR – Seriya Khimicheskaya*, 1784; (c) (1998) *Russian Chemical Bulletin*, **47**, 1736.

154 Boetzel, R., Failla, S., Finocchiaro, P. and Hagele, G. (1995) *Phosphorus Sulfur and Silicon and the Related Elements*, **104**, 71.

155 Khusainova, N.G., Mostovaya, O.A., Berdnikov, E.A., Litvinov, I.A., Krivolapov, D.B. and Cherkasov, R.A. (2005) *Russian Journal of Organic Chemistry*, **41**, 1260.

156 Swamy, K.C.K., Balaraman, E. and Kumar, N.S. (2006) *Tetrahedron*, **62**, 10152.

157 Zapata, A.J., Gu, Y. and Hammond, G.B. (2000) *The Journal of Organic Chemistry*, **65**, 227.

158 Thomas, A.A. and Sharpless, K.B. (1999) *The Journal of Organic Chemistry*, **64**, 8379.

159 Cravotto, G., Giovenzana, G.B., Pagliarin, R., Palmisano, G. and Sisti, M. (1998) *Tetrahedron – Asymmetry*, **9**, 745.

160 Qi, X., Lee, S.-H., Kwon, J.Y., Kim, Y., Kim, S.-J., Lee, Y.-S. and Yoon, J. (2003) *The Journal of Organic Chemistry*, **68**, 9140.

161 Yamauchi, K., Ohtsuki, S. and Kinoshita, M. (1984) *The Journal of Organic Chemistry*, **49**, 1158.

162 Yamauchi, K., Une, F., Tabata, S. and Kinoshita, M. (1986) *Journal of the Chemical Society, Perkin Transactions 1*, 765.

163 Tarnowski, A., Bär, T. and Schmidt, R.R. (1997) *Bioorganic & Medicinal Chemistry Letters*, **7**, 573.

164 Tarnowski, A., Retz, O., Bär, T. and Schmidt, R.R. (2005) *European Journal of Organic Chemistry*, 1129.

165 Barco, A., Benetti, S., Bergamini, P., De Risi, C., Marchetti, P., Pollini, G.P. and Zanirato, V. (1999) *Tetrahedron Letters*, **40**, 7705.

166 Duggan, M.E. and Karanewsky, D.S. (1983) *Tetrahedron Letters*, **24**, 2935.

167 Yu, P., Wang, W., Zhang, H., Yang, X., Liang, T.C. and Gao, X. (2001) *Bioorganic and Medicinal Chemistry*, **9**, 107.

168 Xu, Y., Jiang, X. and Yuan, C. (1990) *Synthesis*, 427.

169 Hays, S.J., Bigge, C.F., Novak, P.M., Drummond, J.T., Bobovski, T.P., Rice, M.J., Johnson, G., Brahce, L.J. and Coughenour, L.L. (1990) *Journal of Medicinal Chemistry*, **33**, 2916.

170 Arasappan, A. and Fuchs, P.L. (1995) *Journal of the American Chemical Society*, **117**, 177.

171 van der Klein, P.A.M., Dreef, C.E., van der Marel, G.A. and van Boom, J.H. (1989) *Tetrahedron Letters*, **30**, 5473.

172 Maier, L. (1992) *Phosphorus Sulfur and Silicon and the Related Elements*, **66**, 67.

173 (a) Stamm, H. and Gerster, G. (1980) *Tetrahedron Letters*, **21**, 1623; (b) Palacios, F., Ochoa de Retana, A.M., Martinez de Marigorta, E. and de los Santos, J.M. (2002) *Organic Preparations and Procedures International*, **34**, 219; (c) Palacios, F., Ochoa de Retana, A.M., Martinez de Marigorta, E. and de los Santos, J.M. (2001) *European Journal of Organic Chemistry*, 2401.

174 Cristau, H.-J., Pirat, J.-L., Drag, M. and Kafarski, P. (2000) *Tetrahedron Letters*, **41**, 9781.

175 Palacios, F., Aparicio, D., Ochoa de Retana, A.M., de los Santos, J.M., Gil, J.I. and de Munain, R.L. (2003) *Tetrahedron – Asymmetry*, **14**, 689.

176 Palacios, F., Ochoa de Retana, A.M. and Alonso, J.M. (2005) *The Journal of Organic Chemistry*, **70**, 8895.

177 Palacios, F., Ochoa de Retana, A.M. and Alonso, J.M. (2006) *The Journal of Organic Chemistry*, **71**, 6141.

178 Hodgson, D.M., Humphreys, P.G., Xu, Z. and Ward, J.G. (2007) *Angewandte Chemie International Edition*, **46**, 2245.

179 Wroblewski, A.E. and Drozd, J. (2007) *Tetrahedron – Asymmetry*, **18**, 1134.

180 Drag, M., Latajka, R., Gumienna-Kontecka, E., Koslowski, H. and Kafarski, P. (2003) *Tetrahedron – Asymmetry*, **14**, 1837.

181 Pousset, C. and Larchevêque, M. (2002) *Tetrahedron Letters*, **43**, 5257.

182 Yokomatsu, T., Yamagishi, T. and Shibuya, S. (1993) *Tetrahedron – Asymmetry*, **4**, 1401.

183 (a) Yamagishi, T., Suemune, K., Yokomatsu, T. and Shibuya, S. (2002) *Tetrahedron*, **58**, 2577; (b) Yamagishi, T., Kusano, T., Yokomatsu, T. and Shibuya, S. (2002) *Synlett*, 1471.

184 Dolle, R.E., Herpin, T.F. and Shimshock, Y.C. (2001) *Tetrahedron Letters*, **42**, 1855.

185 Mizrahi, D.M., Waner, T. and Segall, Y. (2001) *Phosphorus Sulfur and Silicon and the Related Elements*, **173**, 1.

186 Pachamuthu, K. and Schmidt, R.R. (2004) *Chemical Communications*, 1078.

187 (a) Blackburn, G.M., Jakeman, D.L., Ivory, A.J. and Williamson, M.P. (1994) *Bioorganic & Medicinal Chemistry Letters*, **4**, 2573; (b) Jakeman, D.L., Ivory, A.J., Williamson, M.P. and Blackburn, G.M. (1998) *Journal of Medicinal Chemistry*, **41**, 4439; (c) Prager, R.H. and Schafer, K. (1997) *Australian Journal of Chemistry*, **50**, 813.

188 (a) Davis, F.A., Ramachandar, T. and Wu, Y. (2003) *The Journal of Organic Chemistry*, **68**, 6894; (b) Hanessian, S., Bennani, Y.L. and Herve, Y. (1993) *Synlett*, 35.

189 Davis, F.A. and McCoull, W. (1999) *Tetrahedron Letters*, **40**, 249.

190 Roschenthaler, G.-V., Kukhar, V.P., Belik, M.Yu., Mazurenko, K.I. and Sorochinsky, A.E. (2006) *Tetrahedron*, **62**, 9902.

191 Gajda, A. and Gajda, T. (2008) *Tetrahedron*, **64**, 1233.

192 (a) Varlet, J.M., Collignon, N. and Savignac, P. (1981) *Tetrahedron*, **37**, 3713; (b) Maier, L. and Diel, P.J. (1995) *Phosphorus Sulfur and Silicon and the Related Elements*, **107**, 245; (c) Jang, W.B., Shin, W.S., Lee, K. and Oh, D.Y. (1997) *Synthetic Communications*, **27**, 4101.

193 (a) Wang, J., Heikkinen, L.D., Li, H., Zu, L., Jiang, W., Xie, H. and Wang, W. (2007) *Advanced Synthesis and Catalysis*, **349**, 1052; (b) Anisimova, N.A., Kuzhaeva, A.A., Berkova, G.A. and Berestovitskaya, V.M. (2006) *Russian Journal of Inorganic Chemistry*, **76**, 1937; (c) Terada, M., Ikehara, T. and Ube, H. (2007) *Journal of the American Chemical Society*, **129**, 14112.

194 Ryglowski, A. and Kafarski, P. (1996) *Tetrahedron*, **52**, 10685.

195 Yuan, C., Xu, C. and Zhang, Y. (2003) *Tetrahedron*, **59**, 6095.

196 (a) Verbruggen, C., De Craecker, S., Rajan, P. and Jiao, X.-Y. (1996) *Bioorganic & Medicinal Chemistry Letters*, **6**, 253; (b) Brigot, D., Collignon, N. and Savignac, P. (1979) *Tetrahedron*, **35**, 1345.

197 Hakogi, T., Monden, Y., Taichi, M., Iwama, S., Fujii, S., Ikeda, K. and Katsumura, S. (2002) *The Journal of Organic Chemistry*, **67**, 4839.

198 Yan, L., Hale, J.J., Lynch, C.L., Budhu, R., Gentry, A., Mills, S.G., Hajdu, R., Keohane, C.A., Rosenbach, M.J., Milligan, J.A., Shei, G.-J., Chrebet, G., Bergstrom, J., Card, D., Rosen, H. and Mandala, S.M. (2004) *Bioorganic & Medicinal Chemistry Letters*, **14**, 4861.

199 Gazizov, M.B., Khairullin, R.A., Bagauva, L.R., Sinyashin, O.G., Gaisin, L.G., Karimova, R.F. and Shaikhutdinov, R.M. (2003) *Russian Journal of Inorganic Chemistry*, **73**, 1493.

200 (a) Lu, X., Sun, J. and Zhu, J. (1992) *Heteroatom Chemistry*, **3**, 551; (b) Wroblewski, A.E. and Glowacka, I.E. (2005) *Tetrahedron*, **61**, 11930.

201 Hale, J.J., Doherty, G., Toth, L., Li, Z., Mills, S.G., Hajdu, R., Keohane, C.A., Rosenbach, M., Milligan, J., Shei, G.-J., Chrebet, G., Bergstrom, J., Card, D., Rosen, H. and Mandala, S. (2004) *Bioorganic & Medicinal Chemistry Letters*, **14**, 3495.

202 (a) Winckler, W., Pieper, T. and Keppler, B.K. (1996) *Phosphorus Sulfur and Silicon and the Related Elements*, **112**, 137; (b) Mastryukova, T.A., Lazareva, M.V. and Perekalin, V.V. (1971) *Izvestiya Akademii Nauk SSSR – Seriya Khimicheskaya*, 1353.

203 Yamagishi, T., Fujii, K., Shibuya, S. and Yokomatsu, T. (2004) *Synlett*, 2505.

204 Yamagishi, T., Fujii, K., Shibuya, S. and Yokomatsu, T. (2005) *Tetrahedron*, **62**, 54.

205 Tadeusiak, E.J. (2004) *Bioorganic Chemistry*, **32**, 473.

206 Tadeusiak, E., Krawiecka, B. and Michalski, J. (1999) *Tetrahedron Letters*, **40**, 1791.

207 Wroblewski, A.E. and Halajewska-Wosik, A. (2003) *Tetrahedron – Asymmetry*, **14**, 3359.

208 Nesterov, V.V. and Kolodiazhnyi, O.I. (2007) *Tetrahedron*, **63**, 6720.

209 Wroblewski, A.E. and Halajewska-Wosik, A. (2004) *Tetrahedron – Asymmetry*, **15**, 3201.

210 Wroblewski, A.E. and Glowacka, I.E. (2005) *Polish Journal of Chemistry*, **79**, 1895.

211 Lucet, D., Le Gall, T., Mioskowski, C., Ploux, O. and Marquet, A. (1996) *Tetrahedron – Asymmetry*, **7**, 985.

212 Verlet, J. and Collignon, N. (1979) *Canadian Journal of Chemistry*, **57**, 3216.

213 Cui, P., Tomsig, J.L., McCalmont, W.F., Lee, S., Becker, C.J., Lynch, K.R. and Macdonald, T.L. (2007) *Bioorganic & Medicinal Chemistry Letters*, **17**, 1634.

214 Chung, S.-K. and Kang, D.-H. (1997) *Tetrahedron – Asymmetry*, **8**, 3027.

215 Ordóñez, M., González-Morales, A., Ruiz, C., De La Cruz-Cordero, R. and Fernández-Zertuche, M. (2003) *Tetrahedron – Asymmetry*, **14**, 1775.

216 Ordóñez, M., De La Cruz-Cordero, R., Quinones, C. and González-Morales, A. (2004) *Chemical Communications*, 672.

217 Ordóñez, M., González-Morales, A. and Salazar-Fernández, H. (2004) *Tetrahedron – Asymmetry*, **15**, 2719.

218 De la Cruz-Cordero, R., Hernandes-Nunez, E., Fernandez-Zertuche, M., Munoz-Hernandez, M.A. and Ordonez, M. (2005) *ARKIVOC*, 277.

219 Sun, C. and Bittman, R. (2004) *The Journal of Organic Chemistry*, **69**, 7694.

220 Foss, F.W., Snyder, A.H., Davis, M.D., Rouse, M., Okusa, M.D., Lynch, K.R. and Macdonald, T.L. (2007) *Bioorganic and Medicinal Chemistry*, **15**, 663.

221 Lu, X., Byun, H.-S. and Bittman, R. (2004) *The Journal of Organic Chemistry*, **69**, 5433.

222 Ohler, E. and Kanzler, S. (1995) *Synthesis*, 539.

223 Haemers, T., Wiesner, J., Van Poecke, S., Goeman, J., Henschker, D., Beck, E., Jomaa, H. and Calenbergh, S.V. (2006) *Bioorganic & Medicinal Chemistry Letters*, **16**, 1888.

224 Kurz, T., Schlüter, K., Kaula, U., Bergmann, B., Walter, R.D. and Geffken, D. (2006) *Bioorganic and Medicinal Chemistry*, **14**, 5121.

225 Schlüter, K., Walter, R.D., Bergmann, B. and Kurz, T. (2006) *European Journal of Medicinal Chemistry*, **41**, 1385.

226 (a) Devreux, V., Wiesner, J., Jomaa, H., Rozenski, J., Van der Eycken, J. and Van Calenbergh, S. (2007) *The Journal of Organic Chemistry*, **72**, 3783; (b) Devreux, V., Wiesner, J., Jomaa, H., Van der Eycken, J. and Van Calenbergh, S. (2007) *Bioorganic & Medicinal Chemistry Letters*, **17**, 4920.

227 Devreux, V., Wisner, J., Goeman, J.L., van der Eycken, J., Jomaa, H. and Van Calenbergh, S. (2006) *Journal of Medicinal Chemistry*, **49**, 2656.

228 Fokin, A.A., Yurchenko, A.G., Rodionov, V.N., Gunchenko, P.A., Yurchenko, R.I., Reichenberg, A., Wiesner, J., Hintz, M., Jomaa, H. and Schreiner, P.R. (2007) *Organic Letters*, **9**, 4379.

229 Monbaliu, J.-C. and Marchand-Brynaert, J. (2008) *Tetrahedron Letters*, **49**, 1839.

230 Shinkai, I. and Ohta, Y. (1996) *Bioorganic and Medicinal Chemistry*, **4**, 3.

231 Kieczykowski, G.R., Jobson, R.B., Melillo, D.G., Reinhold, D.F., Grenda, V.J. and Shinkai, I. (1995) *The Journal of Organic Chemistry*, **60**, 8310.

232 Shcherbakov, B.K., Bel'skii, F.I., Gukasova, A.Yu., Polikarpov, Yu.M. and Kabachnik, M.I. (1998) *Russian Chemical Bulletin*, **47**, 1732.

233 Vachal, P., Hale, J.J., Lu, Z., Streckfuss, E.C., Mills, S.G., MacCoss, M., Yin, D.H., Algayer, K., Manser, K., Kesisoglou, F., Ghosh, S. and Alani, L.L. (2006) *Journal of Medicinal Chemistry*, **49**, 3060.

234 Turhanen, P.A. and Vepsäläinen, J.J. (2006) *Beilstein Journal of Organic Chemistry*, **2**, 2.

235 Xie, Y., Ding, H., Qian, L., Yan, X., Yang, C. and Xie, Y. (2005) *Bioorganic & Medicinal Chemistry Letters*, **15**, 3267.

236 Chaleix, V. and Lecouvey, M. (2007) *Tetrahedron Letters*, **48**, 703.

237 Maalouf, M.A., Wiemer, A.J., Kuder, C.H., Hohl, R.J. and Wiemer, D.F. (2007) *Bioorganic and Medicinal Chemistry*, **15**, 1959.

238 Moonen, K., Laureyn, I. and Stevens, C.V. (2004) *Chemical Reviews*, **104**, 6177.

239 Murata, Y., Woodward, R.M., Miledi, R. and Overman, L.E. (1996) *Bioorganic & Medicinal Chemistry Letters*, **6**, 2073.

240 Kehler, J., Ebert, B., Dahl, O. and Kroggsgaard-Larsen, P. (1998) *Journal of the Chemical Society, Perkin Transactions 1*, 3241.

241 Orain, D. and Mattes, H. (2005) *Synlett*, 3008.

242 Orain, D. and Mattes, H. (2006) *Tetrahedron Letters*, **47**, 1253.

243 Casero, F., Cipolla, L., Lay, L., Nicotra, F., Panza, L. and Russo, G. (1996) *The Journal of Organic Chemistry*, **61**, 3428.

244 Wen, X. and Hultin, P.G. (2004) *Tetrahedron Letters*, **45**, 1773.
245 Chan, T.-H., Xin, Y.-C. and von Itzstein, M. (1997) *The Journal of Organic Chemistry*, **62**, 3500.
246 Streicher, H. and Bohner, C. (2002) *Tetrahedron*, **58**, 7573.
247 Robiette, R., Defacqz, N., Stofferis, J. and Marchand-Brynaert, J. (2003) *Tetrahedron*, **59**, 4167.
248 Anisimova, N.A., Kuzhaeva, A.A., Deiko, L.I., Berkova, G.A. and Berestovitskaya, V.M. (2003) *Russian Journal of Inorganic Chemistry*, **73**, 1489.
249 Al Quntar, A.A.A., Srivastava, H.K., Srebnik, M., Melman, A., Ta-Shma, R. and Shurki, A. (2007) *The Journal of Organic Chemistry*, **72**, 4932.
250 Johnson, R.L. and Rao, K.S.S.P. (2005) *Bioorganic & Medicinal Chemistry Letters*, **15**, 57.
251 Filosa, R., Marinozzi, M., Costantino, G., Hermit, M.B., Thomsen, C. and Pellicciari, R. (2006) *Bioorganic and Medicinal Chemistry*, **14**, 3811.
252 Kafarski, P. and Lejczak, B. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 173–204.
253 Kafarski, P. and Lejczak, B. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 407–442.
254 Oleksyszyn, J. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 537–558.
255 Peyman, A. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 559–578.
256 Green, D.St.C., Skordalakes, E., Scully, M.F. and Deadman, J.J. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 579–596.
257 Markwell, R. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York,, pp. 597–619.
258 Palacios, F., Alonso, C. and de los Santos, J.M. (2004) *Current Organic Chemistry*, **8**, 1481.
259 Ntai, I., Manier, M.L., Hachey, D.L. and Bachmann, B.O. (2005) *Organic Letters*, **7**, 2763.
260 Ntai, I., Phelan, V.V. and Bachmann, B.O. (2006) *Chemical Communications*, 4518.
261 Hua, F., Meijuan, F., Xiaoxia, L., Guo, T. and Yufen, Z. (2007) *Heteroatom Chemistry*, **18**, 9.
262 Ravaschino, E.L., Docampo, R. and Rodriguez, J.B. (2006) *Journal of Medicinal Chemistry*, **49**, 426.
263 Rinnová, M., Nefzi, A. and Houghten, R.A. (2002) *Tetrahedron Letters*, **43**, 4103.
264 Hatam, M., Tehranfar, D. and Martens, J. (1995) *Synthetic Communications*, **25**, 1677.
265 Solodenko, V.A. and Kukhar, V.P. (1989) *Tetrahedron Letters*, **30**, 6917.
266 Solodenko, V.A., Kasheva, T.N., Mironenko, D.A., Kozlova, E.V., Shvyadas, V.P. and Kukhar, V.P. (1992) *Zhurnal Obshchei Khimii*, **62**, 1472.
267 Natchev, I.A. (1988) *Tetrahedron*, **44**, 1511.
268 Baylis, E.K. (1995) *Tetrahedron Letters*, **36**, 9385.
269 Baylis, E.K. (1995) *Tetrahedron Letters*, **36**, 9389.
270 Jiao, X.Y., Verbruggen, C., Borloo, M., Bollaert, W., De Groot, A., Dommisse, R. and Haemers, A. (1994) *Synthesis*, 23.
271 Grobelny, D. (1989) *Synthetic Communications*, **19**, 1177.
272 Boyd, E.A. and Regan, A.C. (1984) *Phosphorus Sulfur and Silicon and the Related Elements*, **21**, 184.

273 McCleery, P.P. and Tuck, B. (1989) *Journal of the Chemical Society, Perkin Transactions 1*, 1319.

274 (a) Khomutov, A.R., Osipova, T.I., Khurs, E.N., Alferov, K.V. and Khomutov, R.M. (1996) *Russian Chemical Bulletin*, **45**, 1961; (b) Khomutov, A.R., Osipova, T.I., Khurs, E.N., Alferov, K.V. and Khomutov, R.M. (1996) *Izvestiya Akademii Nauk SSSR – Seriya Khimicheskaya*, 2066.

275 Cristau, H.-J., Coulombeau, A., Genevois-Borella, A., Sanchez, F. and Pirat, J.-L. (2002) *Journal of Organometallic Chemistry*, **643/644**, 381.

276 Vassiliou, S., Xeilari, M., Yiotakis, A., Grembecka, J., Pawełczak, M., Kafarski, P. and Mucha, A. (2007) *Bioorganic and Medicinal Chemistry*, **15**, 3187.

277 Matziari, M., Nasopoulou, M. and Yiotakis, A. (2006) *Organic Letters*, **8**, 2317.

278 Matziari, M., Beau, F., Cuniasse, P., Dive, V. and Yiotakis, A. (2004) *Journal of Medicinal Chemistry*, **47**, 325.

279 Berlicki, Ł., Obojska, A., Forlani, G. and Kafarski, P. (2005) *Journal of Medicinal Chemistry*, **48**, 6340.

280 Bartley, D.M. and Coward, J.K. (2005) *The Journal of Organic Chemistry*, **70**, 6757.

281 Bianchini, G., Aschi, M., Cavicchio, G., Crucianelli, M., Preziuso, S., Gallina, C., Nastari, A., Gavuzzo, E. and Mazza, F. (2005) *Bioorganic and Medicinal Chemistry*, **13**, 4740.

282 Hirschmann, R., Yager, K.M., Taylor, C.M., Moore, W., Sprengeler, P.A., Witherington, J., Phillips, B.W. and Smith, A.B. III (1995) *Journal of the American Chemical Society*, **117**, 6370.

283 Hirschmann, R., Yager, K.M., Taylor, C.M., Witherington, J., Sprengeler, P.A., Phillips, B.W., Moore, W. and Smith, A.B. III (1997) *Journal of the American Chemical Society*, **119**, 8177.

284 Chen, R.-Y. and Dan, S.-C. (1990) *Phosphorus Sulfur and Silicon and the Related Elements*, **51/52**, 11.

285 Natchev, I.A. (1991) *Tetrahedron Letters*, **47**, 1239.

286 Yamauchi, K., Kinoshita, M. and Imoto, M. (1972) *Bulletin of the Chemical Society of Japan*, **45**, 2531.

287 Yamauchi, K., Mitsuda, Y. and Kinoshita, M. (1975) *Bulletin of the Chemical Society of Japan*, **48**, 3285.

288 Hariharan, M., Chaberek, S. and Martell, A.E. (1973) *Synthetic Communications*, **3**, 375.

289 McLeod, D.A., Brinkworth, R.I., Ashley, J.A., Janda, K.D. and Wirsching, P. (1991) *Bioorganic & Medicinal Chemistry Letters*, **1**, 653.

290 Bartlett, P.A. and Lamden, L.A. (1986) *Bioorganic Chemistry*, **14**, 356.

291 Malachowski, W.P. and Coward, J.K. (1994) *The Journal of Organic Chemistry*, **59**, 7616.

292 Malachowski, W.P. and Coward, J.K. (1994) *The Journal of Organic Chemistry*, **59**, 7625.

293 Mucha, A., Kafarski, P., Plenat, F. and Cristau, H.-J. (1994) *Tetrahedron*, **50**, 12743.

294 Sampson, N.S. and Bartlett, P.A. (1988) *The Journal of Organic Chemistry*, **53**, 4500.

295 Mucha, A., Grembecka, J., Cierpicki, T. and Kafarski, P. (2003) *European Journal of Organic Chemistry*, 4797.

296 Lewkowski, J., Skowroński, R., Krasowska, D. and Karpowicz, R. (2006) *Tetrahedron Letters*, **47**, 1589.

297 Li, B., Cai, S., Du, D.-M. and Xu, J. (2007) *Organic Letters*, **9**, 2257.

298 Han, L., Hiratake, J., Kamiyama, A. and Sakata, K. (2007) *Biochemistry*, **46**, 1432.

299 Li, T. and Janda, K.D. (1995) *Bioorganic & Medicinal Chemistry Letters*, **5**, 2001.

300 Bartlett, P.A., Hanson, J.E. and Giannousis, P.P. (1990) *The Journal of Organic Chemistry*, **55**, 6268.

301 Campagne, J.-M., Coste, J. and Jouin, P. (1995) *The Journal of Organic Chemistry*, **60**, 5214.

302 Maffre-Lafon, D., Escale, R., Dumy, P., Vidal, J.-P. and Girard, J.-P. (1994) *Tetrahedron Letters*, **35**, 4097.

303 Galéotti, N., Coste, J., Bedos, P. and Jouin, P. (1996) *Tetrahedron Letters*, **37**, 3997.

304 Xu, J. and Gao, Y. (2006) *Synthesis*, 783.
305 Kafarski, P. and Lejczak, B. (2001) *Current Medicinal Chemistry*, **1**, 301.
306 van Zutphen, S., Margarit, V.J., Mora, G. and Le Floch, P. (2007) *Tetrahedron Letters*, **48**, 2857.
307 Han, L., Hiratake, J., Tachi, N., Suzuki, H., Kumagai, H. and Sakata, K. (2006) *Bioorganic and Medicinal Chemistry*, **14**, 6043.
308 Chauhan, S.S., Varshney, A., Verma, B. and Pennington, M.W. (2007) *Tetrahedron Letters*, **48**, 4051.
309 Berlicki, L. and Kafarski, P. (2005) *Current Organic Chemistry*, **9**, 1829.
310 Lane, D.E. (2000) In *Aminophosphonic and Aminophosphinic Acids: Chemistry and Biological Activity* (eds V. Kukhar and H. Hudson), John Wiley & Sons, Inc., New York, pp. 483–535.
311 Conti, P., Pinto, A., Tamborini, L., Grazioso, G., De Sarro, G., Bräuner-Osborne, H., Szabo, G., Hársing, L.G. and De Micheli, C. (2007) *ChemMedChem*, **2**, 1639.
312 Sibille, P., Lopez, S., Brabet, I., Valenti, O., Oueslati, N., Gaven, F., Goudet, C., Bertrand, H.-O., Neyton, J., Marino, M.J., Amalric, M., Pin, J.-P. and Acher, F.C. (2007) *Journal of Medicinal Chemistry*, **50**, 3585.
313 Selvam, C., Goudet, C., Oueslati, N., Pin, J.-P. and Acher, F.C. (2007) *Journal of Medicinal Chemistry*, **50**, 4656.
314 Leon, A., Liu, L., Yang, Y., Hudock, M.P., Hall, P., Yin, F., Studer, D., Puan, K.-J., Morita, C.T. and Oldfield, E. (2006) *Journal of Medicinal Chemistry*, **49**, 7331.
315 Widler, L., Jaeggi, K.A., Glatt, M., Muller, K., Bachmann, R., Bisping, M., Born, A.-R., Cortesi, R., Guiglia, G., Jeker, H., Klein, R., Ramseier, U., Schmid, J., Schreiber, G., Seltenmeyer, Y. and Green, J.R. (2002) *Journal of Medicinal Chemistry*, **45**, 3721.
316 Kotsikorou, E. and Oldfield, E. (2003) *Journal of Medicinal Chemistry*, **46**, 2932.
317 Sanders, J.M., Gomez, A.O., Mao, J., Meints, G.A., Van Brussel, E.M., Burzynska, A., Kafarski, P. and Gonzalez-Pacanowska, D. and Oldfield, E. (2003) *Journal of Medicinal Chemistry*, **46**, 5171.
318 Yan, L., Hale, J.J., Lynch, C.L., Budhu, R., Gentry, A., Mills, S.G., Hajdu, R., Keohane, C.A., Rosenbach, M.J., Milligan, J.A., Shei, G.-J., Chrebet, G., Bergstrom, J., Card, D., Rosen, H. and Mandala, S.M. (2004) *Bioorganic & Medicinal Chemistry Letters*, **14**, 4861.
319 Duggirala, P.Y. and Shevchenko, S.M. (2007) *Chemical Abstracts*, **147**, 54348.
320 Lee, K.-W., Hwang, I.-S., Ko, Y.-S., Yoon, B.-M., Kim, K.-H., Kim, K.-S., Song, S.-Y., Lee, H.-J. and Kim, B.-M. (2005) *Chemical Abstracts*, **143**, 175184.
321 Redmore, D. (1983) *Chemical Abstracts*, **100**, 215268.
322 Leitner, G.J. (1983) *Chemical Abstracts*, **99**, 23948.
323 Wismer, M.W., Woodling, R. and Day, J.C. (2007) *Chemical Abstracts*, **146**, 189457.
324 Vaaramaa, K. and Lehto, J. (1997) *Reactive & Functional Polymers*, **33**, 19.
325 McKenna, C.E., Kashemirov, B.A., Favier, I. and Wolfe, T.D. (2006) *Chemical Abstracts*, **144**, 356565.
326 Hubicki, Z., Leszczynska, M., Lodyga, B. and Lodyga, A. (2006) *Minerals Engineering*, **19**, 1341.
327 Ohto, K., Nagata, J., Honda, S., Yoshizuka, K., Inoue, K. and Baba, Y. (1997) *Solvent Extraction and Ion Exchange*, **15**, 115.

6
Chemistry of Silicon-Containing Amino Acids

Yingmei Qi and Scott McN. Sieburth

6.1
Introduction

Among the design possibilities for unnatural amino acids, the substitution of silicon for carbon can be the most subtle [1–4]. As the element most similar to carbon, replacement of carbon with silicon leads to small changes in overall size and electronic character, particularly when compared with other elements or other substitution options. The nontoxic nature of silicon and its organic derivatives make it an appealing option for novel bioactive substances. Moreover, the anthropogenic nature of all C−Si bonds enhances the excitement of bioactive organosilanes.

A broad selection of amino acids with silicon substitution have been prepared and studied (Figure 6.1). Like carbon, silicon carries four substituents. The simplest with three methyl groups is the most commonly found for amino acid substitution, but many other variations are possible.

6.1.1
Stability of Organosilanes

Any discussion of the use of silicon in a biological milieu must include the caveat that some structures will not be suitable. The strength of the C−Si bond (375 kJ/mol for H_3Si-CH_3) is comparable to the C−C bond (377 kJ/mol for H_3C-CH_3) [5], making simple alkyl and aryl substituents well suited for biological applications; however, allylic substitution can be labile to acid and other electrophiles (Figure 6.2). Silicon α to a carbonyl has, in all cases, a level of instability (see also Section 6.3.1). Another effect of the lower electronegativity of silicon, relative to carbon, is the polarity of the Si−H bond. A single hydrogen on silicon is stable but easily oxidized *in vivo* to a silanol [6]. Additional hydrogens on silicon elevate the instability. The exceptional strength of the Si−O bond (512 kJ/mol for $Me_3Si-OMe$) makes silanols and silyl ethers intrinsically stable [5], but these are also very easily exchanged with solvent. Enantiomercially pure trialkylsilanols racemize rapidly in aqueous solvents [7].

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

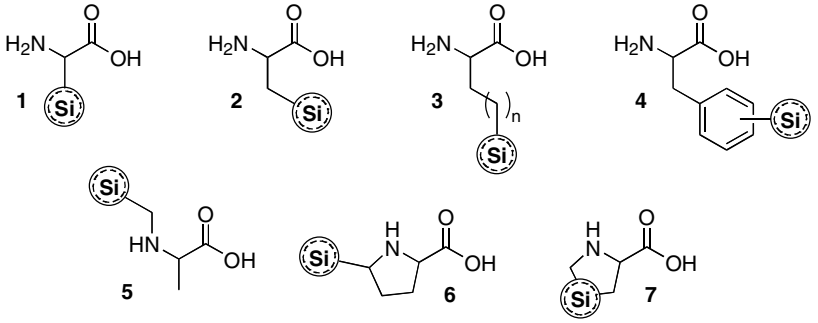


Figure 6.1 Placement of silicon in amino acids.

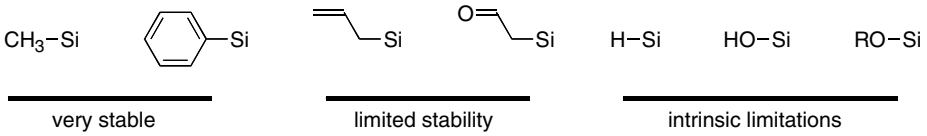


Figure 6.2 Consequences of silicon substitution.

Silanols are also very well known for their tendency to self-condense, forming siloxanes (silicones) [8, 9].

6.1.2 Sterics and Electronics

Substitution of silicon for a quaternary carbon results in an increase in both size and lipophilicity of a molecule (Figure 6.3). The C–Si bond is 20% longer than a C–C bond and the volume of a trimethylsilyl group is 20% larger than a *tert*-butyl group [10]. The enhanced lipophilicity of organosilanes can be seen in the comparison of *tert*-butylbenzene and trimethylsilylbenzene [11].

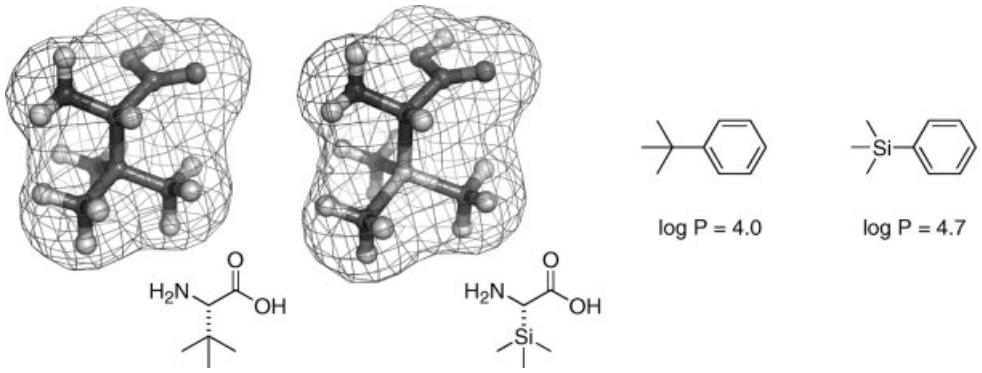


Figure 6.3 Steric and lipophilic effect of silicon substitution.

6.2
Synthesis of Silicon-Containing Amino Acids

6.2.1
Synthesis of α-Silyl Amino Acids and Derivatives

Amino acids with silicon attached directly to the α-carbon were among the last silicon-substituted amino acids to be reported, in part because of the stability of these structures (see Section 6.3.1). Three fundamental approaches to these are exemplified in Scheme 6.1. A very general and practical approach to these structures involves disconnection **A**, pathway (*1*), in which a carbene inserts into a carbamate N−H bond [12]. Both Boc and Cbz derivatives are readily accessible by this approach. Unfortunately, this method is not enantioselective.

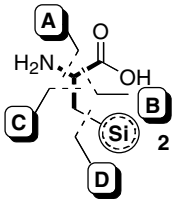
The earliest example of an α-silyl-α-amino acid was prepared using disconnection **B**, pathway (*2*), in which a Boc group was used to direct metallation [13]. This approach suffers from a lack of selectivity, with a competitive metallation of the benzyl position of **11**, not shown.

A second and stereoselective method for α-silyl-α-amino acid synthesis, pathway (*3*), utilizes the zirconium imine complex **15** and its reaction with an optically active carbonate [14]. The stereochemistry of the carbonate controls formation of the new chiral center of the α-silyl-α-amino acid. Examples without aryl nitrogen substitution have not been reported.

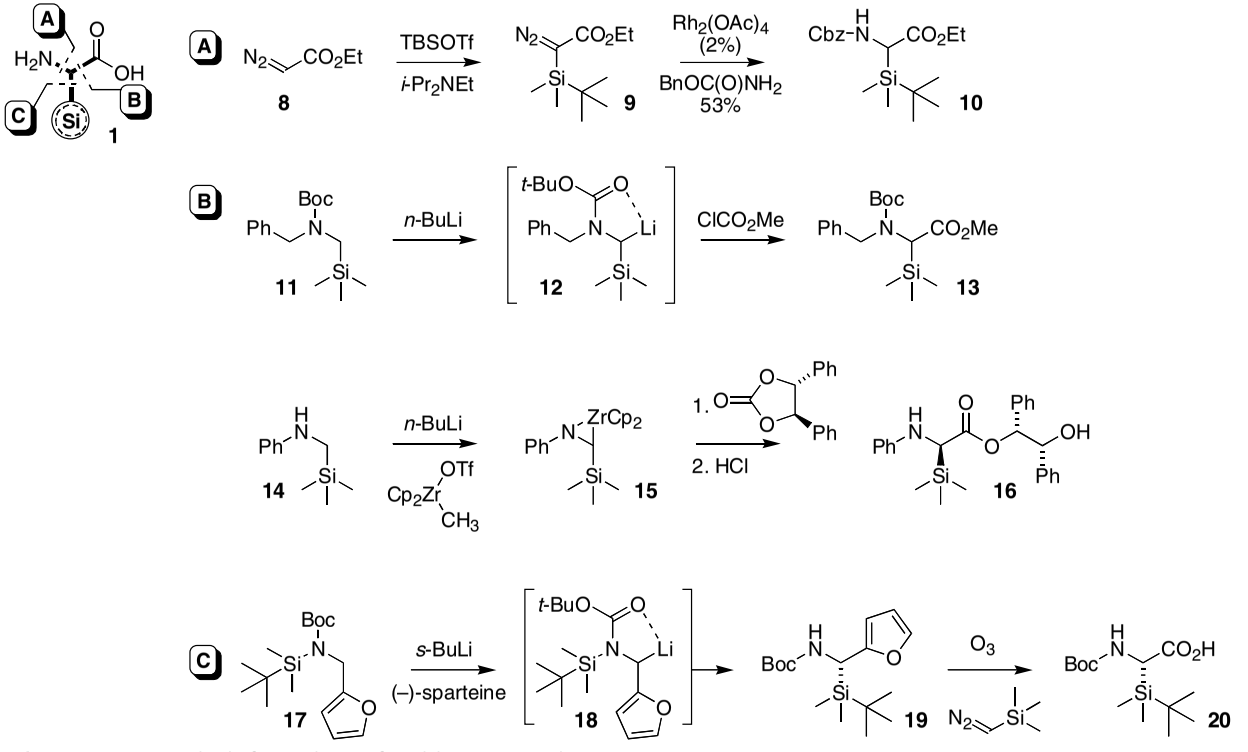
A disconnection **C** method that is stereoselective involves reverse-aza-Brook rearrangement of a furfuryl amine derivative **17**, pathway (*4*) [10, 15]. Asymmetric (−)-sparteine-mediated metallation α to the amine, followed by migration of the silicon, sets the stereogenic center with high enantioselectivity. Ozonolysis followed by trimethylsilyl diazomethane then gives the Boc-protected amino acid methyl ester.

6.2.2
Synthesis of β-Silylalanine and Derivatives

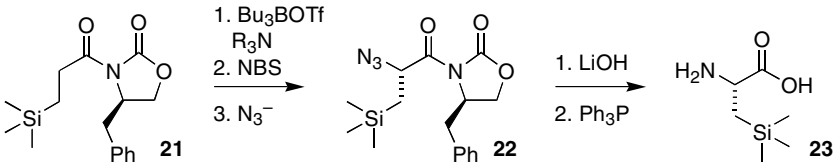
The majority of silicon-substituted amino acid research, and some of the earliest, involved β-trimethylsilylalanine **23**. Three of the four fundamental bond disconnections of **2** have been used to assemble this target, with only examples of **B** missing.



Asymmetric introduction of the C−N bond **A** was shown by Walkup *et al.* to work well with the Evans chiral auxiliary (Scheme 6.2) [16]. Beginning with β-trimethylsilylpropionyl derivative **21**, bromination of the boron enolate and azide inversion of

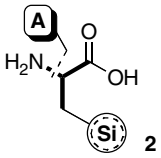


Scheme 6.1 Four methods for synthesis of α-silyl-α-amino acids.



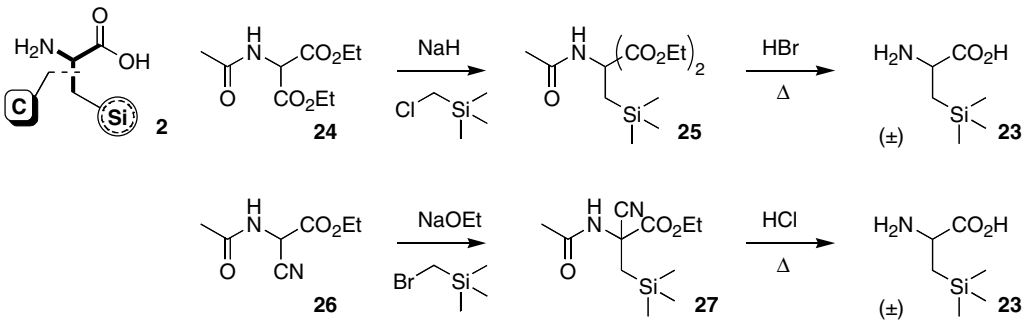
Scheme 6.2 Asymmetric synthesis of β-trimethylsilylalanine.

this product sets the stereochemistry of **22**. Hydrolysis and reduction of the azide provides β-trimethylsilylalanine **23**.

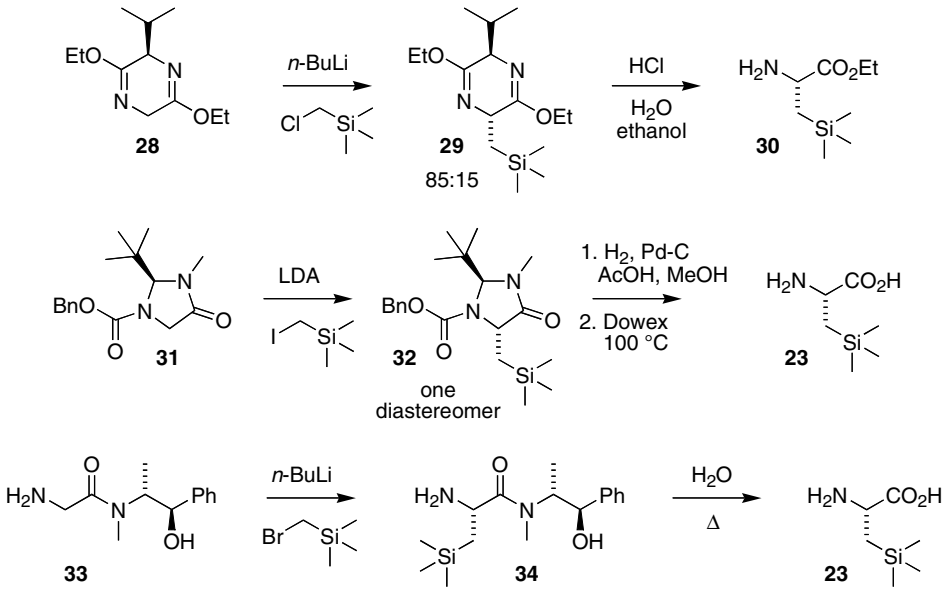


The first synthesis of β-trimethylsilylalanine utilized a malonate approach and disconnection **C**. Alkylation of diethyl α-acetamido malonate **24** or cyanoacetate **26** with chloromethyltrimethylsilane or the corresponding bromide, introduced the silyl group in good yield. Hydrolysis, deprotection of the amine, and decarboxylation gave the parent β-trimethylsilylalanine **23**, demonstrating the robustness of this silicon substitution [17, 18]. Naturally, the product **23** is obtained as a racemate (Scheme 6.3).

Three enantioselective methods using the **C** disconnection have been described (Scheme 6.4) [19–23]. Alkylation of Schöllkopf's reagent **28** with chloromethyltrimethylsilane gives an 85: 15 mixture of diastereoisomers **29**. Hydrolysis of the heterocycle with acidic ethanol yields the amino acid ester **30** in good yield and stereochemical purity [19, 20]. The five-membered Seebach imidazolone **31** can be alkylated with iodomethyltrimethylsilane, producing a single diastereomeric product **32**. A two-step hydrolysis then yields the amino acid **23** [21]. Alternatively, alkylation of the dianion of the Myers pseudoephedrine amide of glycine **33** with



Scheme 6.3 A traditional approach to racemic β-trimethylsilylalanine.



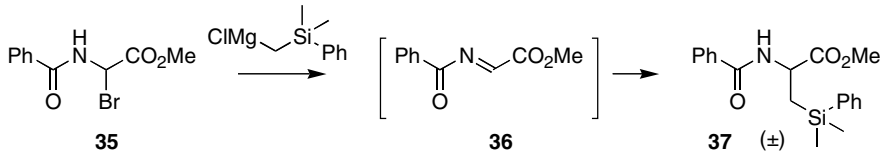
Scheme 6.4 Three enantioselective methods for β-trimethylsilylalanine synthesis.

bromomethyltrimethylsilane yields the product **34** with greater than 99% d.e. Hydrolysis of this amide by simply heating in water gives the (*S*) amino acid **23** [22].

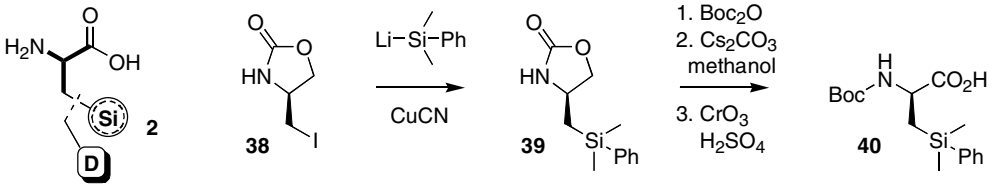
A **C** approach to silylalanine reported by Trost and Lee uses an inverted nucleophilicity. Treatment of α-bromo glycine derivative **35** with an excess of (phenyldimethylsilyl)methyl magnesium chloride initiates dehydrohalogenation and then traps the imine **36** to give the racemic, protected amino acid derivative **37** (Scheme 6.5) [23].

Disconnection **D** has been described by Sibi *et al.* [24], using a silyl cuprate. The optically active iodomethyloxazolidinone **38** undergoes nucleophilic displacement of the iodide with dimethylphenylsilyl lithium and Cu(I) cyanide to install the silicon. Boc-derivatization of the nitrogen, hydrolysis of the heterocycle, and oxidation of the primary alcohol gave the dimethylphenylsilylalanine in Boc-protected form **40** (Scheme 6.6) [24].

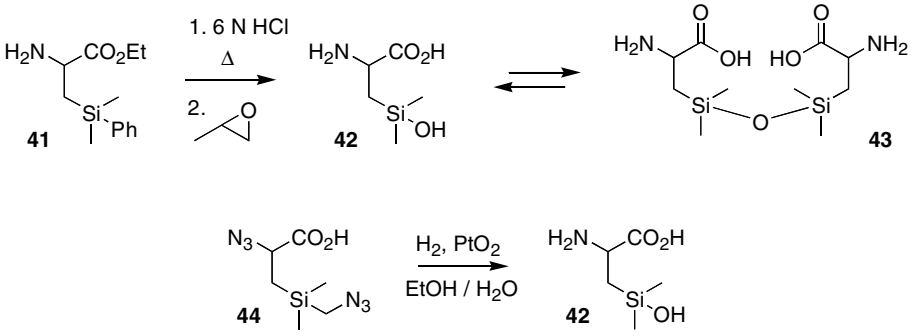
The unusual β-hydroxydimethylsilylalanine **42** was prepared by Tacke *et al.* when the ester of dimethylphenylsilylalanine was hydrolyzed in hydrochloric acid and then neutralized with propylene oxide (Scheme 6.7). Under these strongly acidic



Scheme 6.5 Trost and Lee's umpolung **C** approach to β-silylalanine.



Scheme 6.6 Asymmetric synthesis of β-dimethylphenylsilylalanine.



Scheme 6.7 Two preparations of β-phenyldimethylsilylalanine.

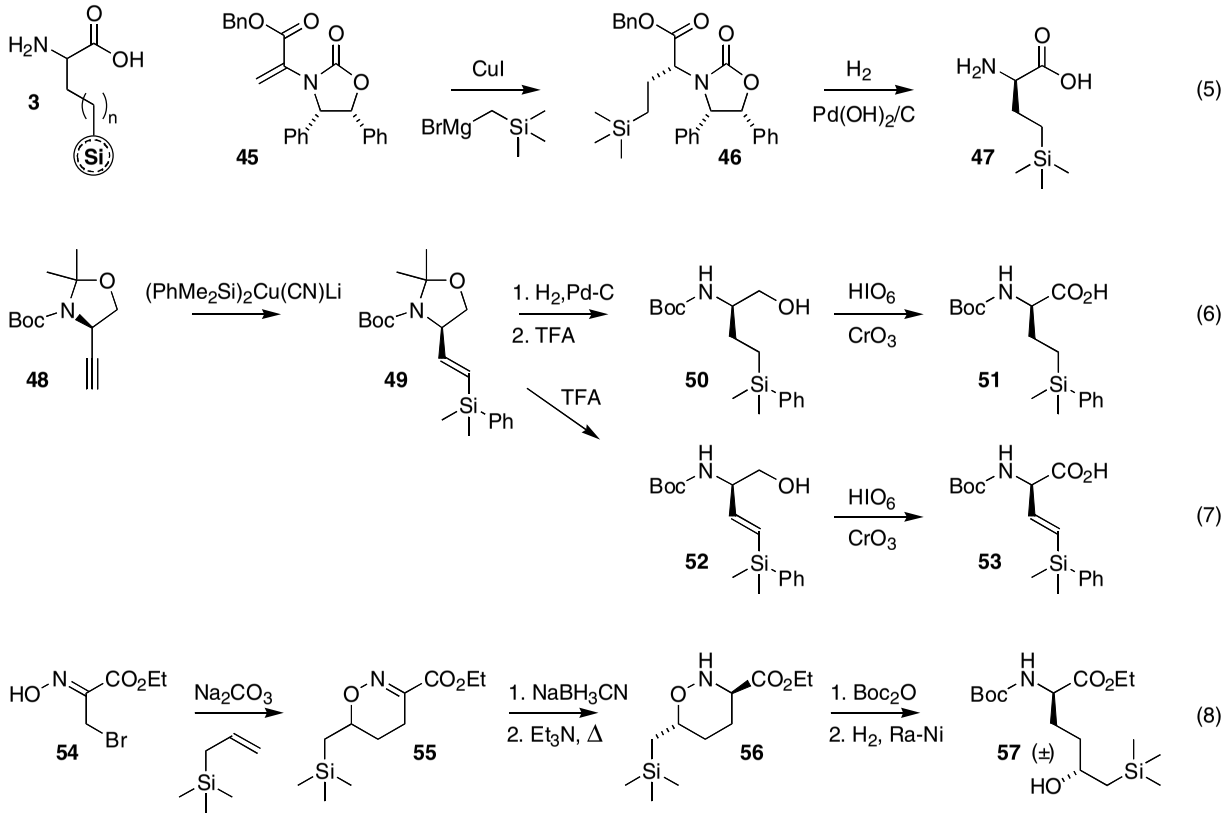
conditions the phenylsilane was also hydrolyzed. This phenylalanine derivative was found to be monomeric when dissolved in water, but exclusively dimeric **43** when isolated [25]. An unexpected preparation of silanol **42** occurred when diazido amino acid **44** was reduced catalytically [26]. Most reductions of azidomethylsilanes to the corresponding amines have been reported with a phosphine [27] or with lithium aluminum hydride [28, 29]; however, the latter method would be incompatible with the carboxylic acid.

6.2.3
Synthesis of ω-Silyl Amino Acids and Derivatives

A number of examples of amino acids with silicon at the end of a carbon side-chain two carbons or longer have been reported, generally with control of absolute stereochemistry.

Hegedus added the trimethylsilylmethyl cuprate to the unsaturated ester-carbamate **45** (Scheme 6.8, pathway *5*). Protonation of the resulting enolate gave the protected amino acid **46** with 12 : 1 diastereoselectivity. Deprotection of the amine and acid gave the amino acid **47** [30].

The *cis*-cuprate addition to acetylene **48**, using the bis(dimethylphenylsilyl)copper lithium reagent, produces the vinyl silane **49**. Hydrogenation, hydrolysis, and oxidation of the primary alcohol, pathway (*6*), yields γ-silyl amino acid **51**. Without the hydrogenation the corresponding vinylsilane **53** is produced, via pathway (*7*) [31].



Scheme 6.8 Three approaches to ω-silyl amino acids.

A bifunctional amino acid side-chain terminating in a trimethylsilyl group is formed by hetero-Diels–Alder reaction of an unsaturated nitroso ester intermediate, formed by dehydrohalogenation of bromooxime **54**, via pathway (*8*). Allyl trimethylsilane undergoes a regiospecific cycloaddition. Reduction of the resulting oxazine and epimerization gave the thermodynamically favored trans isomer **56**. Nitrogen protection and reductive cleavage of the N—O bond gave the protected amino acid **57** [32].

6.2.4
Synthesis of Silyl-Substituted Phenylalanines

A number of silicon-substituted phenylalanines **4** have been prepared using standard approaches (Scheme 6.9) [33–36]. Alkylation of formaminomalonate with *p*-silylbenzyl bromide **58**, followed by basic deprotection, acidification, and decarboxylation gave the racemic phenylalanine **60** [33, 34]. The unique double phenylalanine linked through a dimethylsilyl group **62**, pathway (*10*), was prepared similarly [33, 34]. Another traditional synthesis of the *p*-trimethylsilylphenylalanine **66** employed condensation of hydantoin **63** with *p*-trimethylsilyl benzaldehyde and hydrogenation of product **65**, via pathway (*11*). Enzymatic hydrolysis of the substituted hydantoin gave the amino acid in 98% e.e. [35]. As part of a solid-phase peptide synthesis program, Silverman and Lee developed the asymmetric synthesis outlined in pathway (*12*) [36]. The intermediate unsaturated amino acid derivative **69** was reduced with a chiral rhodium catalyst to set the stereogenic center in **70**. Addition of a Boc group to the amide and hydrazinolysis of the acetyl group gave the protected amino acid **70** [36]. Naphthalene **71** and thiophene **72** were prepared with similar chemistry.

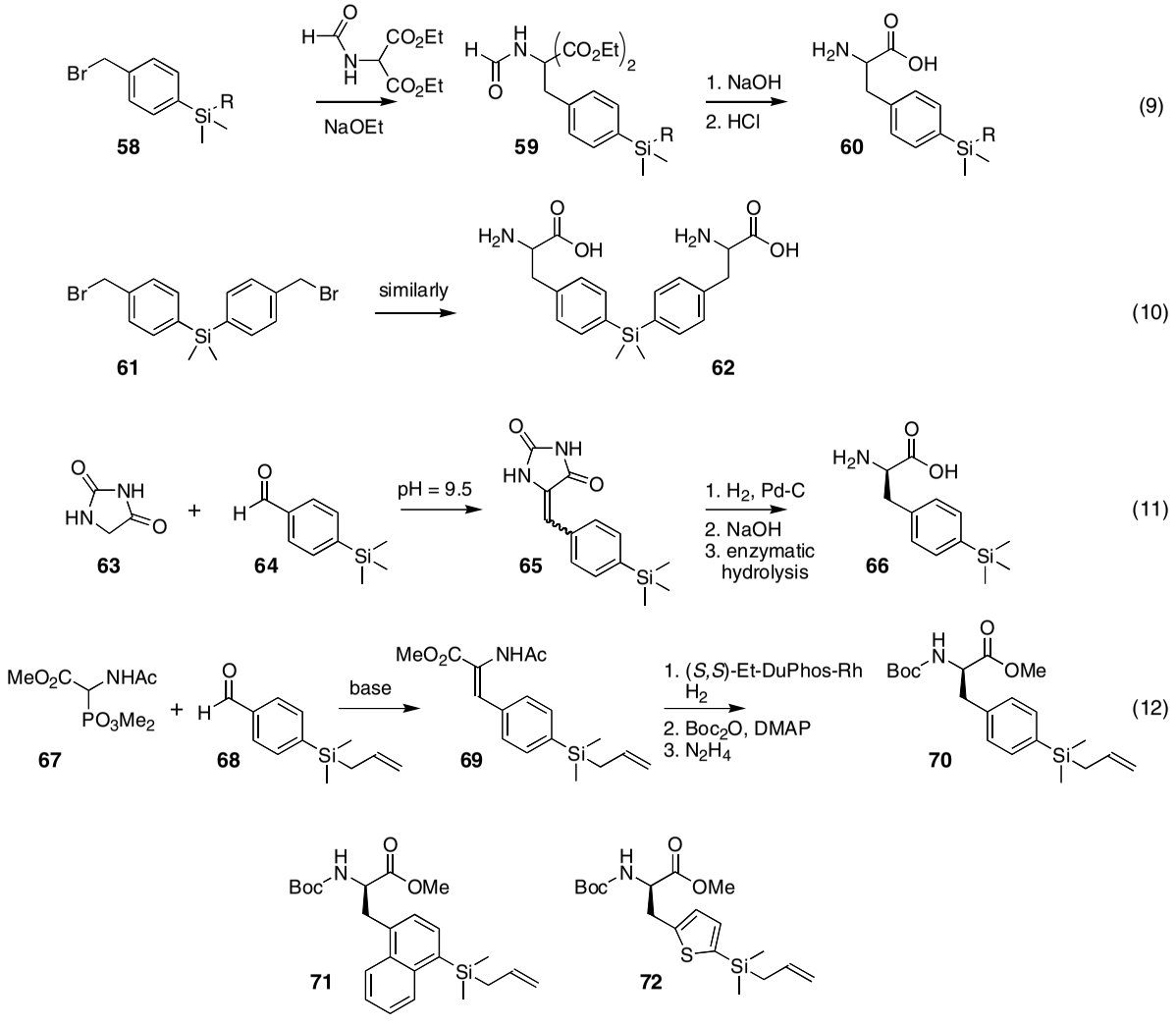
6.2.5
Synthesis of Amino Acids with Silicon α to Nitrogen

Moeller has described the preparation of several amino acids bearing silicon α to nitrogen (Scheme 6.10) (**5** and **6**, Figure 6.1). Proline with a 5-dimethylphenylsilyl group **76** was prepared in three steps from a protected proline: oxidation adjacent to nitrogen gave methoxy substitution which was then replaced with phenylsulfone to give **75**. Displacement of the sulfone with dimethylphenylsilyl cuprate then gave the 5-(*R*)-silylproline **76** [37].

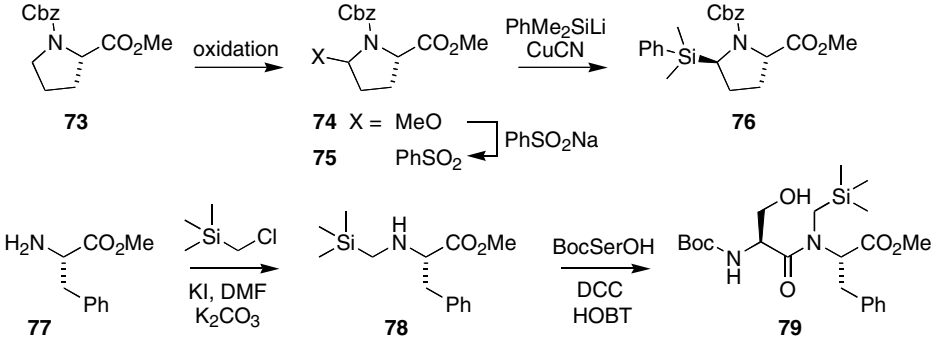
Silanes can also be attached as a nitrogen substituent, prepared by direct alkylation of the amino acid amine by the readily available chloromethylsilanes to give **78** [37, 38]. In addition to their novelty as amino acids, the silicon groups α to nitrogen are latent carbocations, revealed through chemical or electrochemical oxidation (see Section 6.4.2).

6.2.6
Synthesis of Proline Analogs with Silicon in the Ring

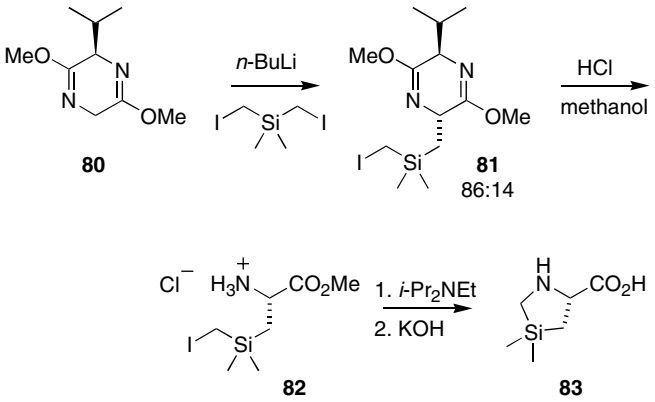
Replacement of the C-4 methylene group of proline with a dimethylsilyl group, prepared in three steps with good enantioselectivity, has been achieved by alkylating



Scheme 6.9 Synthesis of silicon-substituted phenylalanines and analogs.



Scheme 6.10 Synthesis of 5-silylproline and *N*-(silylmethyl)-phenylalanine.

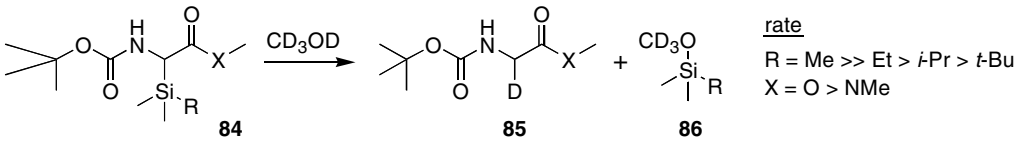


Scheme 6.11 Synthesis of 4-dimethylsilaproline.

Schöllkopf's reagent **80** with bis(iodomethyl)dimethylsilane to give **81**, setting the desired stereochemistry with 72% d.e. (Scheme 6.11). Hydrolysis gave β-(iodomethy) dimethylsilylalanine **82**. Neutralization of the amine salt results in intramolecular alkylation and formation of the proline analog **83** (**7**, Figure 6.1) [39].

6.3 Reactions of Silicon-Containing Amino Acids

In general, amino acids carrying a silicon group undergo standard peptide coupling chemistry without incident. Silicon groups are larger than their carbon counterparts, however, they can present a *lower* steric impediment in some cases [10]. This counterintuitive relationship is a consequence of the longer Si–C bond [40–42].



84 85 86

Scheme 6.12 Instability of α-silyl amino acid derivatives.

6.3.1
Stability of the Si–C Bond

There are many examples of standard protection, coupling, and deprotection steps used with organosilane amino acids without complications (see Sections 6.3.2 and 6.3.3) [43]. The only exceptions are the α-silyl amino acids **1**, exemplified by **84** (Scheme 6.12). The two substructures of this amino acid, an α-silyl amine and an α-silyl acid, have substantial stability individually, but when the amine and carbonyl group both flank the silane the resulting structure is substantially less stable, leading to hydrolysis or methanolysis of the Si–C bond. The relative rate of methanolysis of the Boc-protected amino acid **84** is substantially slower for the amide than for the ester and more sterically hindered silanes are more stable than less hindered silanes [10].

6.3.2
Functional Group Protection

Acidic esterification of trimethylsilylalanine **2** has been reported (but see also Scheme 6.7) [44], as has dicyclohexylcarbodiimide esterification [45], and protection of the amine with Boc, Cbz [44, 46], and Fmoc groups. Protection of silylphenylalanine **4** with Boc and Fmoc proceeds without problem [47]. Protection of silylproline **7** with Boc and Fmoc has been reported [39].

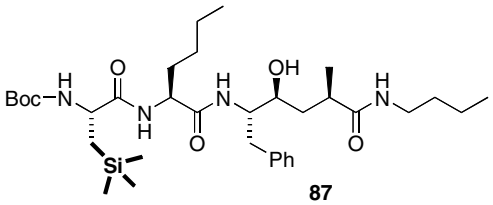
6.3.3
Functional Group Deprotection

The following are examples of acidic removal of Boc groups from trimethylsilylalanines **2** [16], silylphenylalanine **4** [48, 49], and silylprolines **6** and **7** [43, 50]. Hydrogenolysis of Cbz groups on silicon-substituted prolines **6** has been reported [37]. Basic removal of Fmoc groups from trimethylsilylalanine **2** is firmly established [16, 20, 43].

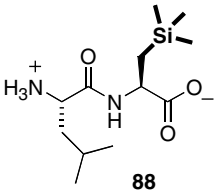
6.4
Bioactive Peptides Incorporating Silicon-Substituted Amino Acids

6.4.1
Use of β-Silylalanine

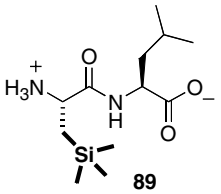
Trimethylsilylalanine (**2**, Figure 6.1) is the most studied of the silicon amino acids in biological applications. Replacement of phenylalanine and leucine with this amino



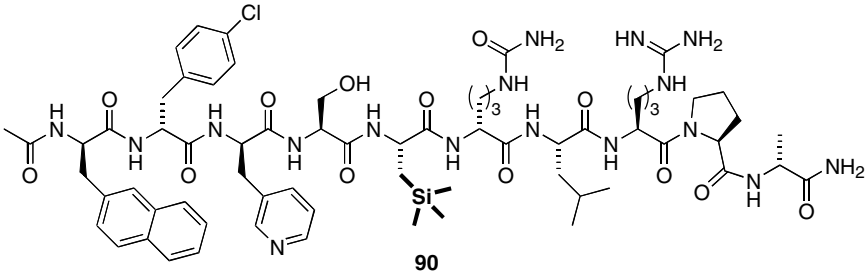
87



88

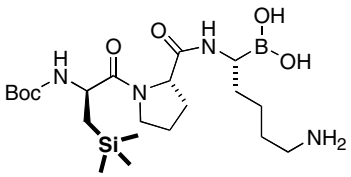


89



90

GnRH receptor agonist



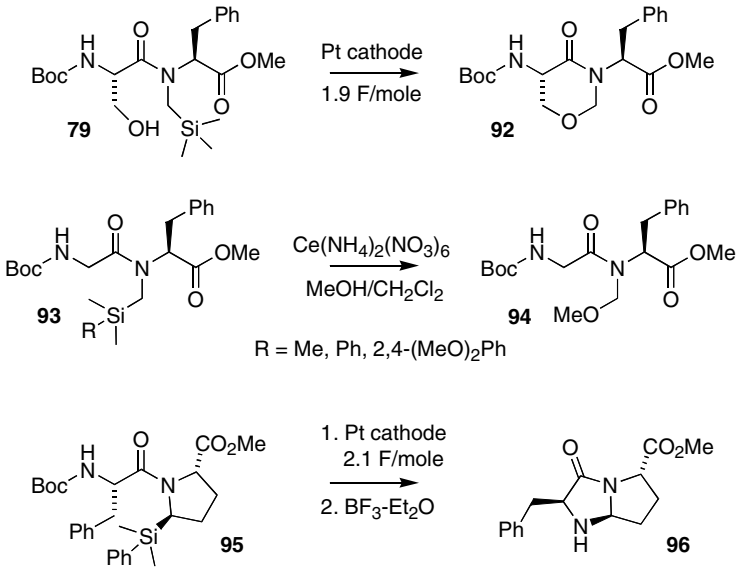
91

thrombin-selective inhibitor

Scheme 6.13 Bioactive peptides incorporating β-trimethylsilylalanine.

acid to give products with comparable activity has suggested an isosteric relationship (Scheme 6.13). Silicon-containing renin inhibitor **87** was nearly as potent as the phenylalanine analog [19]. Similarly, when each leucine of a Leu–Leu thermolysin inhibitor was replaced with trimethylsilylalanine to give compounds **88** and **89**, inhibition was enhanced, but replacing both leucines with trimethylsilylalanine gave a substantial loss in potency [44].

Incorporation of trimethylsilylalanine into a gonadotropin-releasing hormone receptor agonist to give compound **90** found a similar level of activity compared to the carbon and germanium analogs [20]. Use of the D-trimethylsilylalanine to



Scheme 6.14 Silylmethyl-substituted peptides and their reactivity under electrochemical and chemically oxidative conditions.

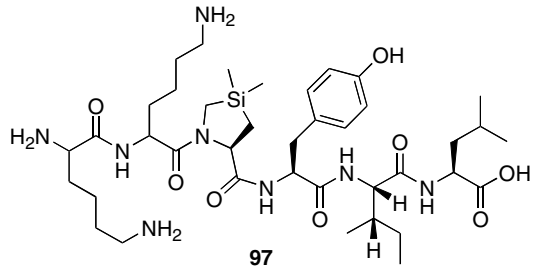
prepare thrombin inhibitor **91** resulted in a high level of potency and selectivity (Scheme 6.14) [46]. See also Scheme 6.15.

6.4.2
Use of *N*-Silylalkyl Amino Acids

A silylalkyl group on nitrogen of a peptide is a latent reactive group that can be activated by reagents not typically used in peptide chemistry, enabling modification of a peptide after its construction (Scheme 6.14). Oxidation of an *N*-trimethylsilylmethyl or *N*-aryldimethylsilylmethyl amide electrochemically or using ceric ammonium nitrate yields an iminium ion intermediate that can be trapped by oxygen nucleophiles intra- or intermolecularly (e.g., to give **92** and **94**). An aryl substituent on the silicon allows for fine tuning of the oxidation reactivity. Similarly, 5-silylproline **95** can be oxidized and trapped with methanol. When the resulting product is treated with a Lewis acid, the dipeptide cyclizes to give the bicyclic product **96** [37, 38].

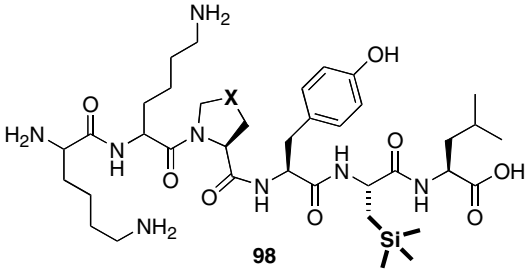
6.4.3
Use of Silaproline

The effect of the dimethylsilyl group in silaproline **83** (**7**, Figure 6.1) on the conformation of peptides containing it was found to be small. When this amino



97

degradation-resistant neurotensin analog



greatly reduced binding (**X** = CH_2 or $SiMe_2$)

Scheme 6.15 Silaproline neurotensin analog.

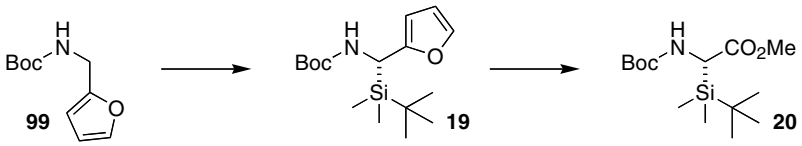
acid was incorporated into the neurotensin fragment **97**, replacing a proline, the resulting peptide had a binding affinity similar to the neurotensin fragment itself, but was endowed with greatly enhanced resistance to proteolytic degradation *in vivo*. In contrast, replacing the isoleucine with trimethylsilylalanine, with and without the silaproline substitution, greatly reduced receptor binding [50].

6.5 Conclusions

Beginning more than 50 years ago, preparation of amino acids carrying or incorporating silicon has largely explored analogs of alanine, phenylalanine, and proline. The high strength of the Si–C bond makes these analogs suitable for incorporation in peptides without difficulty. The only exception to this generalization is the α-silyl amino acids (**1**, Figure 6.1), which are prone to hydrolytic loss of the silicon group. Incorporation of silicon-containing amino acids into biologically active peptides has demonstrated that bioactivity can be maintained while simultaneously minimizing their inactivation by proteolytic enzymes. Clearly, this area is still in its infancy despite its substantial history.

6.6
Experimental Procedures

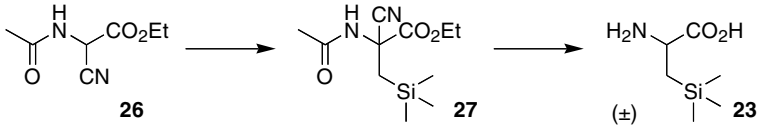
6.6.1
L-β-Trimethylsilylalanine 23



To a −78 °C, 0.05 M solution of Boc-protected furfuryl amine **99** and *tert*-butyldimethylsilyl chloride (1.1 equiv.) in tetrahydrofuran (THF) was added dropwise *n*BuLi (1 equiv.). After 1 h, the mixture was warmed to room temperature for 4 h, recooled to −78 °C, and a second portion of *n*BuLi (1 equiv.) was slowly added. After stirring for 3 h, the mixture was diluted with aqueous ammonium chloride and the aqueous phase was extracted with ether. Concentration and silica gel chromatography (20 : 1 hexanes/ether) gave **19** (97%).

Into a −78 °C, 0.5 M solution of **19** in 5 : 1 methylene chloride/methanol containing a small quantity of Sudan III was bubbled a stream of ozone until the red color dissipated. Argon was bubbled through the solution for 1 h and then the mixture was warmed to 0 °C. Trimethylsilyl diazomethane (1.3 equiv.) was added and after 1 h the mixture was diluted with water. The aqueous phase was extracted with ethyl acetate, and the combined organics were dried over $MgSO_4$ and concentrated. Flash chromatography (10 : 1 hexanes/ethyl acetate) gave **20** as a colorless solid (80%).

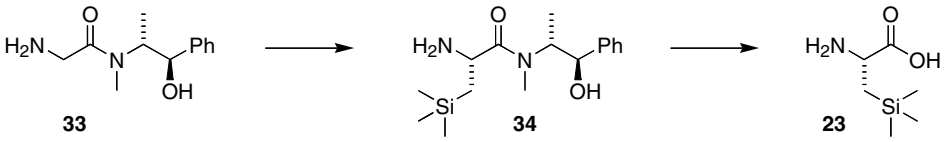
6.6.2
(±)-β-Trimethylsilylalanine 23



Ethyl acetamidocyanoacetate **26** was added to a 1.2 M suspension of sodium ethoxide (1.03 equiv.) in dimethylsulfoxide, followed by addition of bromomethyltrimethylsilane (1 equiv.), and the resulting mixture was stirred overnight. After concentration *in vacuo* by 50%, the mixture was diluted with four volumes of water and extracted with ether. After concentration the product was recrystallized from ethanol/water to give **27**.

A 0.25 M solution of compound **28** in 10% sodium hydroxide was heated to 100 °C for 3 h, then cooled and acidified to pH 5 with concentrated HCl. The resulting precipitate was recrystallized from water to give **23** (36%).

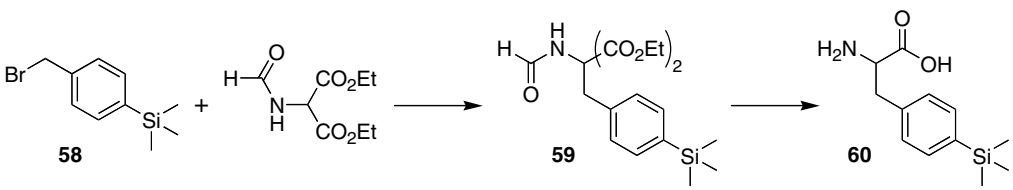
6.6.3
L-β-Trimethylsilylalanine 23



A solution of lithium diisopropylamide (1.95 equiv.) in THF was added to a 0 °C suspension of (*R,R*)-pseudoephedine glycine amide (1 equv.) and LiCl (6 equiv.) in THF (total volume to give about 0.2 M in **33**). After 20 min bromomethyltrimethylsilane (1.25 equivalents) was added and the resulting mixture was stirred for 1 day. After addition of 1 M HCl, the aqueous phase was extracted twice with ethyl acetate. The aqueous phase was taken to pH 14 by addition of 50% NaOH and then extracted with methylene chloride. The combined extracts were dried over K_2CO_3, filtered, and concentrated. Silica gel chromatography (23 : 1 : 1 methylene chloride/methanol/triethylamine) gave **34**, which was recrystallized from 1 : 1 hexanes/butyl acetate in 57% overall yield.

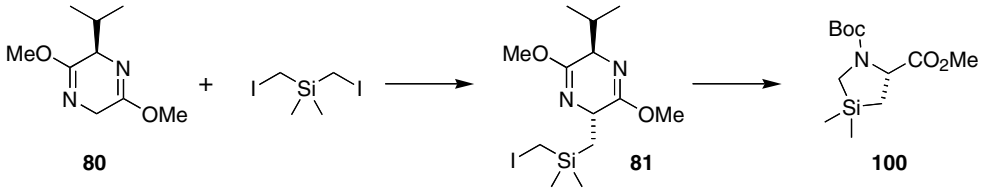
A 0.25 M suspension of compound **34** in water was heated to reflux for 20 h. On cooling and dilution with an additional five portions of water, washing with methylene chloride and concentration of the aqueous phase gave a solid which was triturated with ethanol to give an 82% yield of **23**.

6.6.4
(±)-*p*-Trimethylsilylphenylalanine 60



Diethyl formamidomalonate was added to 0.5 M sodium ethoxide (1 equiv.) in ethanol, followed by *p*-trimethylsilylbenzyl bromide (1 equiv.), and the resulting mixture was stirred at room temperature overnight. After concentration the residue was taken up in 1 : 2 ether/water. The aqueous phase was extracted with ether, the combined organics were dried over $MgSO_4$, and recrystallized from aqueous ethanol to give **59** (93%). Malonate **59** was dissolved in ethanolic sodium ethoxide. Water was added and the resulting mixture was heated to reflux for 5 min. On cooling, the precipitated disodium salt was collected and washed with ethanol. This salt was treated with the following procedure 3 times: taken up in 1.5 M HCl in methanol and stirred for 20 h; the mixture was evaporated, dissolved in methanol, and evaporated again. The resulting product was triturated with water and then washed with cold water. After digestion with hot ethanol, amino acid **60** was recrystallized from 15% acetic acid.

6.6.5
L-4-Dimethylsilaproline 100



To a −70 °C solution of Schöllkopf's reagent **80** in THF (0.3 M in **80**) was added *n*BuLi in hexane (1.03 equiv.). After 15 min a THF solution of bis(iodomethyl) dimethylsilane (1.8 equiv.) was added. This mixture was stirred at −70 °C for 3 h and then warmed to room temperature. The mixture was diluted with ether, washed with water, dried over magnesium sulfate, concentrated, and then purified by silica gel chromatography (99 : 1 ether/hexanes) to give **81** (57%).

A 0.3 M of **81** in 3 : 1 methanol/10% HCl was stirred at room temperature for 2 h and then concentrated. The residue was taken up in methanol and reconcentrated 3 times, then dissolved in 2 : 3 methylene chloride/ether. After addition of diisopropylethylamine (2.2 equiv.) and stirring for 4 h, di-*tert*-butyl dicarbonate (2.1 equiv.) was added and the resulting solution was stirred overnight. The solution was concentrated, taken up in ethyl acetate, washed with aqueous potassium hydrogen sulfate and water, and then dried over magnesium sulfate. Concentration and silica gel chromatography (1 : 9 ethyl acetate/hexanes) gave **100** (70%).

References

1 Tacke, R. and Linoh, H. (1989) Bioorganosilicon chemistry, in *The Chemistry of Organic Silicon Compounds* (eds S. Patai and Z. Rappoport), John Wiley & Sons, Inc., New York, pp. 1143–1206.

2 Franz, A.K. (2007) *Current Opinion in Drug Discovery & Development*, **10**, 654–671.

3 Showell, G.A. and Mills, J.S. (2003) *Drug Discovery Today*, **8**, 551–556.

4 Bains, W. and Tacke, R. (2003) *Current Opinion in Drug Discovery & Development*, **6**, 526–543.

5 Luo, Y.R. (2003) *Handbook of Bond Dissociation Energies in Organic Compounds*, CRC Press, New York.

6 Fessenden, R.J. and Hartman, R.A. (1970) *Journal of Medicinal Chemistry*, **13**, 52–54.

7 Tacke, R., Linoh, H., Ernst, L., Moser, U., Mutschler, E., Sarge, S., Cammenga, H.K. and Lambrecht, G. (1987) *Chemische Berichte*, **120**, 1229–1237.

8 Stark, F.O., Falender, J.R. and Wright, A.P. (1982) Silicones, in *Comprehensive Organometallic Chemistry* (eds G. Wilkinson, F.G.A. Stone and E.W. Abel), Pergamon, New York, pp. 305–363.

9 Brown, S.S., Kendrick, T.C., McVie, J. and Thomas, D.R. (1995) Silicones, in *Comprehensive Organometallic Chemistry II* (eds E.W. Abel, F.G.A. Stone and G. Wilkinson), Pergamon, New York, pp. 111–135.

10 Liu, G. and Sieburth, S.McN. (2005) *Organic Letters*, **7**, 665–668.

11 Fessenden, R.J. and Fessenden, J.S. (1980) *Advances in Organometallic Chemistry*, **18**, 275–299.

12 Bolm, C., Kasyan, A., Drauz, K., Günther, K. and Raabe, G. (2000) *Angewandte Chemie International Edition*, **39**, 2288–2290.

13 Sieburth, S.McN., Somers, J.J. and O'Hare, H.K. (1996) *Tetrahedron*, **52**, 5669–5682.

14 Chen, J.X., Tunge, J.A. and Norton, J.R. (2002) *The Journal of Organic Chemistry*, **67**, 4366–4369.

15 Liu, G. and Sieburth, S.McN. (2003) *Organic Letters*, **5**, 4677–4679.

16 Walkup, R.D., Cole, D.C. and Whittlesey, B.R. (1995) *The Journal of Organic Chemistry*, **60**, 2630–2634.

17 Birkofer, L. and Ritter, A. (1958) *Liebigs Annalen der Chemie*, 22–33.

18 Porter, T.H. and Shive, W. (1968) *Journal of Medicinal Chemistry*, **11**, 402–403.

19 Weidmann, B. (1992) *Chimia*, **46**, 312–313.

20 Tacke, R., Merget, M., Bertermann, R., Bernd, M., Beckers, T. and Reissmann, T. (2000) *Organometallics*, **19**, 3486–3497.

21 Fitzi, R. and Seebach, D. (1988) *Tetrahedron*, **44**, 5277–5292.

22 Myers, A.G., Gleason, J.L., Yoon, T. and Kung, D.W. (1997) *Journal of the American Chemical Society*, **119**, 656–673.

23 Trost, B.M. and Lee, C. (2001) *Journal of the American Chemical Society*, **123**, 12191–12201.

24 Sibi, M.P., Harris, B.J., Shay, J.J. and Hajra, S. (1998) *Tetrahedron*, **54**, 7221–7228.

25 Tacke, R., Schmid, T. and Merget, M. (2005) *Organometallics*, **24**, 1780–1783.

26 Smith, R.J. and Bienz, S. (2004) *Helvetica Chimica Acta*, **87**, 1681–1696.

27 Daiss, J.O., Burschka, C. and Tacke, R. (2005) *Journal of Organometallic Chemistry*, **690**, 678–684.

28 Tomoda, S., Matsumoto, Y., Takeuchi, Y. and Nomura, Y. (1986) *Chemistry Letters*, 1193–1196.

29 Kim, J. and Sieburth, S.McN. (2004) *The Journal of Organic Chemistry*, **69**, 3008–3014.

30 Lander, P.A. and Hegedus, L.S. (1994) *Journal of the American Chemical Society*, **116**, 8126–8132.

31 Reginato, G., Mordini, A., Valacchi, M. and Grandini, E. (1999) *The Journal of Organic Chemistry*, **64**, 9211–9216.

32 Gallos, J.K., Sarli, V.C., Massen, Z.S., Varvogli, A.C., Papadoyanni, C.Z., Papaspyrou, S.D. and Argyropoulos, N.G. (2005) *Tetrahedron*, **61**, 565–574.

33 Frankel, M., Gertner, D., Shenhar, A. and Zilkha, A. (1963) *Journal of the Chemical Society*, 5049–5051.

34 Rotman, A., Gertner, D. and Zilkha, A. (1967) *Canadian Journal of Chemistry*, **45**, 2469–2471.

35 Tsuji, Y., Yamanaka, H., Fukui, T., Kawamoto, T. and Tanaka, A. (1997) *Applied Microbiology and Biotechnology*, **47**, 114–119.

36 Lee, Y. and Silverman, R.B. (2001) *Tetrahedron*, **57**, 5339–5352.

37 Sun, H. and Moeller, K.D. (2002) *Organic Letters*, **4**, 1547–1550.

38 Sun, H. and Moeller, K.D. (2003) *Organic Letters*, **5**, 3189–3192.

39 Vivet, B., Cavelier, F. and Martinez, J. (2000) *Journal of Organic Chemistry*, **5**, 807–811.

40 Kitching, W., Olszowy, H.A., Drew, G.M. and Adcock, W. (1982) *The Journal of Organic Chemistry*, **47**, 5153–5156.

41 Penman, K.G., Kitching, W. and Adcock, W. (1989) *The Journal of Organic Chemistry*, **54**, 5390–5391.

42 Arnason, I., Kvaran, A., Jonsdottir, S., Gudnason, P.I. and Oberhammer, H. (2002) *The Journal of Organic Chemistry*, **67**, 3827–3831.

43 Cavelier, F., Marchand, D., Martinez, J. and Sagan, S. (2004) *The Journal of Peptide Research*, **63**, 290–296.

44 Ishikawa, H., Yamanaka, H., Kawamoto, T. and Tanaka, A. (1999) *Applied Microbiology and Biotechnology*, **53**, 19–22.

45 Jacob, L.A., Chen, B.L. and Stec, D. (1993) *Synthesis*, 611–614.

46 von Matt, A., Ehrhardt, C., Burkhard, P., Metternich, R., Walkinshaw, M. and Tapparelli, C. (2000) *Bioorganic and Medicinal Chemistry*, **8**, 2291–2303.

47 Lee, Y. and Silverman, R.B. (1999) *Journal of the American Chemical Society*, **121**, 8407–8408.

48 Gu, W., Liu, S. and Silverman, R.B. (2002) *Organic Letters*, **4**, 4171–4174.

49 Gu, W. and Silverman, R.B. (2003) *The Journal of Organic Chemistry*, **68**, 8774–8779.

50 Cavelier, F., Vivet, B., Martinez, J., Aubry, A., Didierjean, C., Vicherat, A. and Marraud, M. (2002) *Journal of the American Chemical Society*, **124**, 2917–2923.

Part Two
Amino Acid Organocatalysis

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

7
Catalysis of Reactions by Amino Acids

Haibo Xie, Thomas Hayes, and Nicholas Gathergood

7.1
Introduction

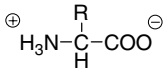
Since the major breakthrough in asymmetric catalysis made by William Knowles and his colleagues in 1970s, the demand for chiral compounds, usually as single enantiomers, has sharply escalated. This was driven not only by the demands of the pharmaceutical industry but also by other applications, including agricultural chemicals, flavors, fragrances, and materials [1, 2]. Several strategies for synthesizing such compounds to meet this increasing demand for chiral compounds, such as transition metal catalysis [3], biocatalysis [4], and organocatalysis [5, 6], have been developed. Interest in organocatalysis spectacularly increased in the past decade as a result of both the novelty of the concept and, more importantly, the fact that efficiency and selectivity of many organocatalytic reactions without the assistance of transition metal catalysts meet the standard of established organic reactions combined with a green chemistry concept.

Amino acids play central roles both as building blocks of proteins and as intermediates in metabolism. An α-amino acid is a molecule that contains both amine and carboxyl functional groups (Figure 7.1). Their particular structural characteristics determine their key role in the formation of proteins and as bifunctional asymmetric catalysts for stereoselective synthesis. The two functional groups can act as both acid and base and can promote chemical transformations in concert, similar to enzymatic catalysis [7].

While all of these criteria apply for amino acids, proline, a secondary amino acid with a pyrrolidine-based structure, dominated early investigations. Its pyrrolidine ring determines its higher pK_a value, and unique nucleophilic reactivity, compared to other primary amino acids. This facilitates the formation of iminium ions and enamines with carbonyl compounds more readily than most other amines, including cyclic examples such as piperidine. The cooperating contribution offered by the carboxylate group enhanced and expanded its applications in organocatalytic

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1



R: H or side chain

Figure 7.1 Structure of an α-amino acid.

processes (Figure 7.2) [8]. As a result, proline has been studied as an effective catalyst and investigated for several powerful asymmetric transformations, such as the Aldol, Mannich, and Michael reactions. In addition, to provide opportunities for a better understanding of the origin of homochirality [9–11], amino acids as fundamental building blocks of organisms need to be investigated for their possible asymmetric transformation and amplification into prebiotic units such as sugars.

Although amino acids, especially (*S*)-proline, have shown high efficiency in many valuable organic transformations affording corresponding products with high regio-, diastereo-, and enantioselectivities, sometimes such transformations suffered drawbacks. These include high catalyst loading, excess of nucleophile, long reaction times, and low solubility of catalyst in organic reaction media. To overcome these drawbacks and improve the catalytic performance of proline, some modifications to the structure of (*S*)-proline have been reported, thus allowing the fine-tuning of catalytic properties and improving the activities and conditions. Illustrated examples of modified amino acids are included in this chapter; however, several excellent recent reviews cover these derivatives in more detail [9–11].

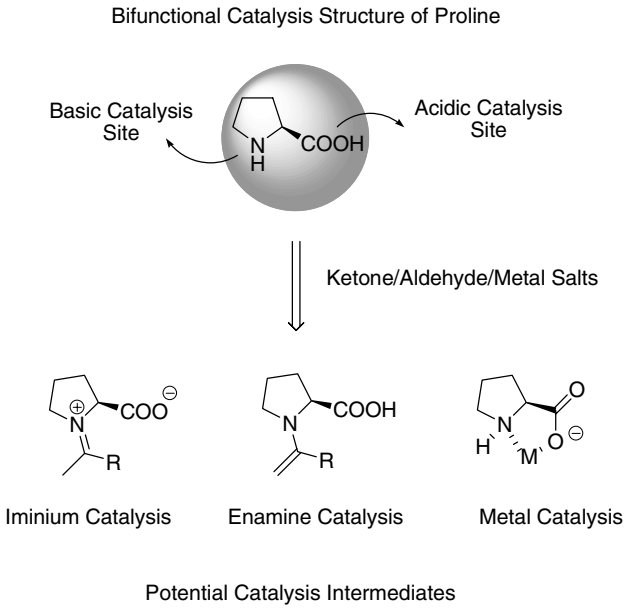
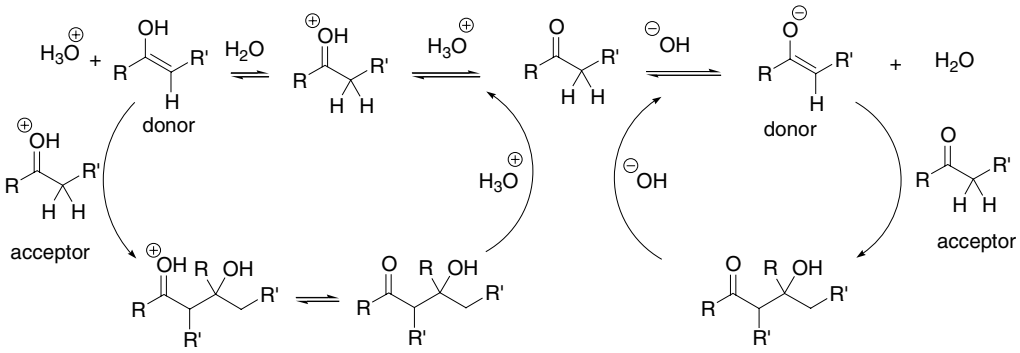


Figure 7.2 Potential catalysis intermediates of bifunctional proline-catalyzed transformations.



Scheme 7.1 Traditional acidic and basic catalytic process of Aldol reaction.

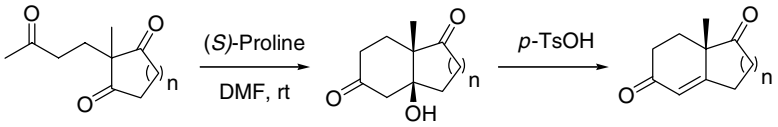
7.2
Aldol Reaction

The Aldol reaction is a useful carbon–carbon bond-forming reaction, also providing high "atom economy." It is a typical example of electrophilic substitution at the α-carbon in enols or enolate anions. The fundamental transformation usually catalyzed by base or equivalent acid is dimerization of an aldehyde or ketone to form a β-hydroxyl aldehyde or ketone by α-carbanionic addition of one reactant molecule to the carbonyl group of a second reactant molecule [12] (Scheme 7.1).

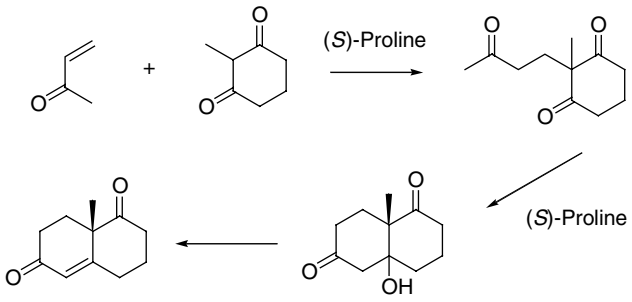
7.2.1
Intramolecular Aldol Reaction and Mechanisms

7.2.1.1 Intramolecular Aldol Reaction

The amino acid-catalyzed asymmetric intramolecular Aldol reaction (Robinson-type annulation) was discovered in the 1970s by Hajos and Parrish [13] and Wiechert *et al.* [14], who first demonstrated the potential of enantioselective reactions for the synthesis of natural products, even at a larger scale. The procedure involves the use of (*S*)-proline as a catalyst and perchloric acid as a cocatalyst, which leads to the final Aldol condensation, by an *in situ* acid-catalyzed dehydration reaction. Under these conditions, the corresponding yields and enantioselectivity vary with the structures of the starting materials, amount of catalyst, and the solvents used in the reaction (Scheme 7.2).



Scheme 7.2 Hajos–Parrish–Eder–Sauer–Wiechert reactions.



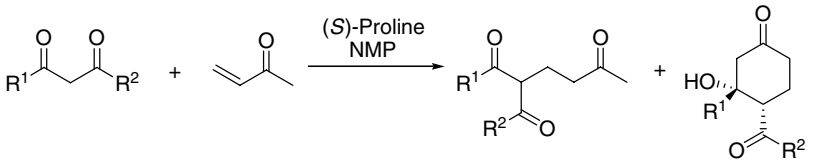
49% yield, 76% ee in 89 h for a single-step Robinson annulation

Scheme 7.3 (*S*)-Proline-catalyzed Robinson annulation sequence.

Instead of using the triketone derivative as the starting reagent, acyclic diketone derivatives and methyl vinyl ketone have been investigated [15]. These studies showed that (*S*)-proline can also catalyze the entire annulation sequence providing the Wieland–Miescher ketone more rapidly and with an overall yield (~49%) and enantiomeric purity (76% e.e.) similar to that of the two-step reaction sequence (Scheme 7.3). This tandem process involves an initial Michael-type addition to form the corresponding triketone intermediate [16], which finally undergoes the asymmetric Robinson reaction (Scheme 7.4).

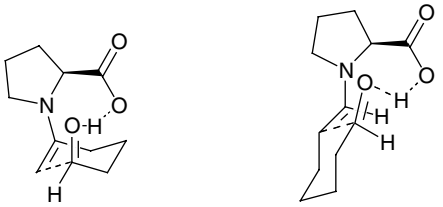
According to the Baldwin rules, 6-enol*endo* aldolizations are very common and favored, but 6-enol*exo* aldolizations are also permitted. While the usefulness of the proline-catalyzed Hajos–Parrish–Eder–Sauer–Wiechert reaction has been illustrated in a broad context, only 6-enol*endo* aldolizations have been described so far. By inspecting possible transition-state models, List *et al.* [17] realized that in addition to the established 6-enol*endo* aldolizations via transition state A (Scheme 7.5), proline should also catalyze corresponding 6-enol*exo* aldolizations via the chair-like assembly B. In 2003, they first reported the highly enantioselective (*S*)-proline-catalyzed enol*exo* aldolization of dicarbonyl compounds, such as dialdehydes and ketoaldehydes. Both the diastereoselectivity and the enantioselectivity are sensitive to the aldehyde substituents.

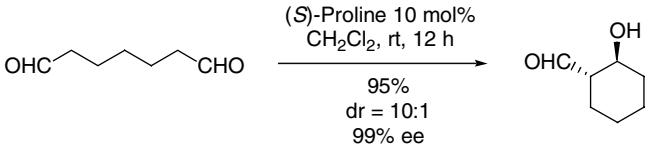
The above methodology has been successfully applied to an efficient asymmetric organocatalytic intramolecular 5-enol*exo* Aldol reaction affording *cis*-selectively 3-hydroxy-2,3-dihydrobenzofurans and derivatives [18]. The authors found that the substituent on the aromatic ring or R^3 had a negligible effect on enantioselectivity,



$R^1 = R^2$ = thiophenyl	52% yield	47% yield (67% ee)
$R^1 = R^2$ = pyridinyl	0% yield	81% yield (80% ee)
R^1 = pyridinyl, R^2 = 4-MeOPh	0% yield	93% yield (50% ee)

Scheme 7.4 Organocatalytic transformation of 1,3-diketones into optically active cyclohexanones.



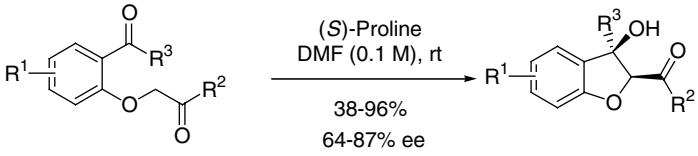


Scheme 7.5 Direct catalytic asymmetric enol*exo* aldolization catalyzed by (*S*)-proline.

and the R^2 substituent had a pronounced effect due to increased hindrance. Enantiomeric excess ranging from 64 to 87% was obtained (Scheme 7.6).

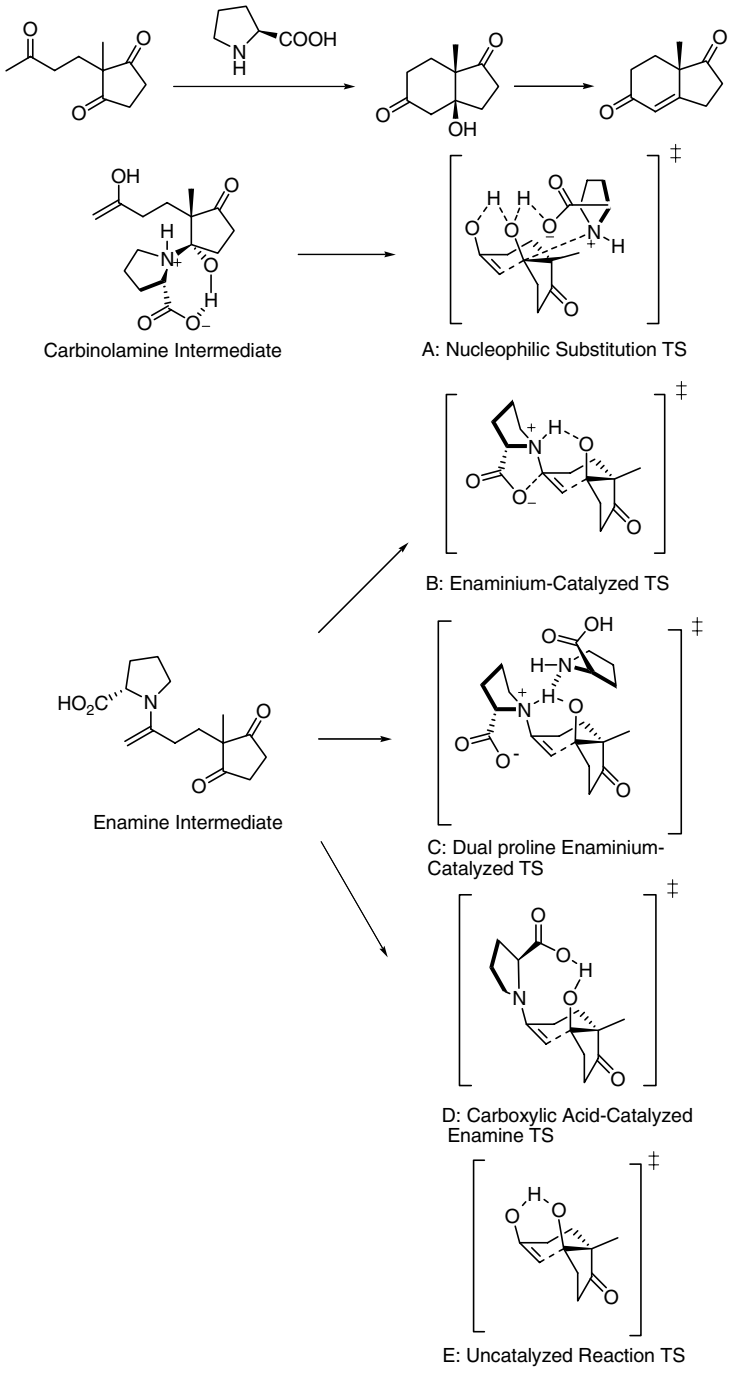
7.2.1.2 **Mechanisms**

Since Hajos and Parrish introduced the first intramolecular Aldol cyclization of triketones, extensive work has been performed to understand the mechanism of this reaction. Initially, Hajos and Parrish proposed two possible mechanisms for the intramolecular Aldol reactions based on carbinolamine and enamine intermediates potentially formed between the proline catalyst and triketones [19]. One mechanism is via a nucleophilic substitution route (Scheme 7.7, A), and the other is an enaminium-catalyzed process (B). The nucleophilic substitution process involves the attack of proline on one of the cyclic carbonyl groups to form a carbinolamine intermediate; the subsequent C−C bond-forming step consists of the displacement of the proline moiety by nucleophilic attack by the side-chain enol. The enaminium path involves the attachment of proline to the acyclic carbonyl group to form an enaminium intermediate that acts as the nucleophile in the subsequent C−C bond formation with concomitant N−H···O hydrogen transfer. In the 1980s,



R^1= H, 2-Methyl, 2,3-Methylenedioxo, 2,4-Dibromo
R^2= Me, Et
R^3= H, Me

Scheme 7.6 Asymmetric organocatalytic synthesis of *cis*-substituted dihydrobenzofuranols via intramolecular Aldol reactions.



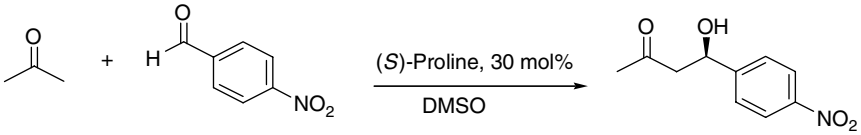
Scheme 7.7 Mechanisms proposed for the proline-catalyzed intramolecular Aldol reaction.

Agami *et al.* [20–22] proposed a modification of the original enaminium-catalyzed mechanism proposed by Hajos and Parrish based on experimental evidence from Molines and Wakselman [23]. The modified mechanism involves the presence of a second proline molecule assisting in the N−H · · · O hydrogen transfer, thus enabling conjugation of the nitrogen lone pair with the enamine system. A new mechanism D was introduced by Jung [19], which involves attack of an enamine intermediate accompanied by proton transfer from the proline carboxylic acid group to the developing alkoxide, in which one proline molecule plays both a basic and an acidic bifunctional role in the process. Yet, the theoretical calculation studies of Houk and Clemente indicated that the carboxylic acid-catalyzed model (D) is lowest in energy and therefore the mechanism of the proline-catalyzed Aldol cyclization is more favorable [24]. For example, the energy of pathway D is around 10 kcal/mol lower than that of the uncatalyzed process, and is 31 kcal/mol lower than that of pathway B. In addition, the energy of carbinolamine intermediate is around 12 kcal/mol higher than the carboxylic acid-catalyzed transition state. Therefore, the transition structure leading to the Aldol product via A must be even higher in energy. This illustrates the difficulty of nucleophilic attack of an enol on a tertiary carbon as well as the lower nucleophilicity of enols with respect to enamines. The stereoselectivities of intramolecular Aldol cyclizations catalyzed by both primary acids and proline have also been investigated based on theoretical calculations by Houk *et al.* [24]. The results reproduced the observed enantioselectivities during the process. The conformation flexibility of the primary amino acids allows a good alignment for the O−H · · · O hydrogen bonding in the transition structures in either *anti* or *syn* approaches, which explains the lower enantioselectivity observed when primary amino acids are used as catalysts in the cyclization. The enhanced enantioselectivity observed with primary amino acids in the substituted case arose from the differential steric repulsion between the substituents on the enamine nitrogen, hydrogen, and methylene, with substituents of the *Z*-enamine. These transition structures are lower in energy than their *E* counterparts because, in the latter, the enamine substituent is located in a more crowded area, so the all-*cis* bicyclic products are predicted to be the only diastereomers. In the (*S*)-proline-catalyzed reaction, the steric repulsion with the R^1 substituent significantly destabilizes the *Z*-enamine, so the diastereomeric outcome is expected to be opposite to that observed with primary amino acids. This destabilization is responsible for the significant increase in the activation barrier for the C−C bond-forming step compared to the unsubstituted case.

7.2.2 Intermolecular Aldol Reaction and Mechanisms

7.2.2.1 Intermolecular Aldol Reaction

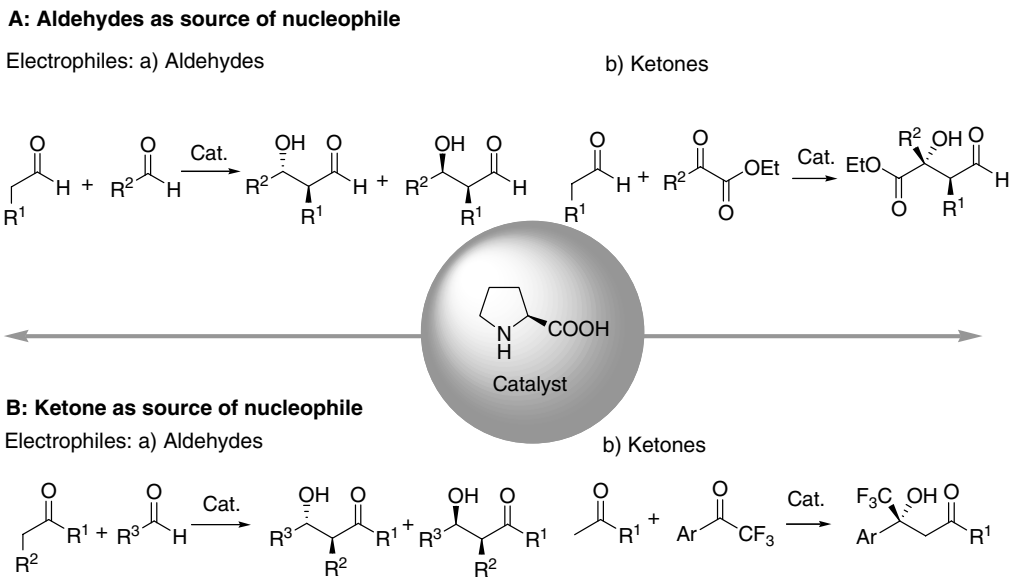
Renewed interest in this reaction was awakened by the observation that proline is able to catalyze not only intramolecular but also intermolecular reactions with high selectivity and high yield. In 2000, with the inspiration of the action of class I aldolases, List *et al.* [25] were first to demonstrate that proline and its derivatives have significant ability to mediate intermolecular asymmetric Aldol reactions between unmodified



Scheme 7.8 Proline-catalyzed intermolecular Aldol reaction.

acetone and a variety of aldehydes with moderate to good enantioselectivities. In these processes (Scheme 7.8), the ketones acted as a source of nucleophiles and aldehydes as electrophiles. They found that none of the commercially available proline derivatives showed significantly improved enantioselectivity compared to proline. Modest enantiomeric excess (60–77% e.e.) has been obtained for the synthesis of aromatic aldols. Unbranched aldehydes did not yield any significant amount of the desired cross-Aldol product. However, the reaction of acetone with isobutyraldehyde gave Aldol product in 97% yield and 96% e.e. Instead of using acetone, other dialkyl ketones, such as 3-pentanone, acetylcyclohexanes, and 3-methyl-2-butanone, were used as a source of the nucleophile. However, they were not reactive toward the Aldol reaction with *p*-nitrobenzaldehyde under comparative reaction conditions.

This method has been extended recently. The current progress has demonstrated that, for the intermolecular Aldol reaction, both ketones and aldehydes can act as the source of the nucleophile, as well as the source of the electrophile (Scheme 7.9). Without doubt, the successful use of proline as catalyst in the intermolecular Aldol reaction will expand the research field of organocatalysis and lead to efficiency in new methodology for constructing chiral compounds [11].



Scheme 7.9 (*S*)-Proline-catalyzed intermolecular Aldol reactions.

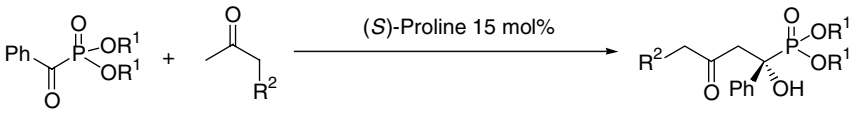
Scheme 7.10 Acyclic amino acid-catalyzed direct asymmetric Aldol reactions.

In stark contrast to proline, usually linear amino acids are considered to be poor catalysts for the intermolecular enantioselective Aldol reaction between unmodified ketones and aldehydes. In 2005, Cordova *et al.* [26] reported that acyclic amino acids and their derivatives are able to catalyze asymmetric intermolecular Aldol reactions with high stereoselectivity. For example, simple natural and unnatural primary amino acid derivatives catalyzed the reaction between 2,2-dimethyl-1,3-dioxan-5-one and an aldehyde with high yield and up to >99% e.e. (Scheme 7.10).

Intrigued by the enzyme-like enantioselectivity of the acyclic amino acid-catalyzed asymmetric Aldol reaction, the authors investigated the origins of its stereoselectivity. Their theoretical and experimental data showed that only one amino acid molecule is involved in the transition state; for alanine, the carboxylic acid-catalyzed enamine mechanism is more favorable than other mechanisms because it requires the lowest activation energy. Furthermore, the calculations demonstrate that such a mechanism accurately predicts the correct stereochemistry of the product [27].

Although ketones are poor electrophiles and therefore are considered unsuitable candidates for the Aldol reaction, this transformation has been achieved in certain cases using activated nonenolizable ketones, such as 1-aryl-2,2,2,-trifluoroethanones [28], and α-keto carbonyl compounds [29, 30].

α-Hydroxy phosphonic acid derivatives have been shown to be very important enzyme inhibitors, such as human immunodeficiency virus protease and polymerase inhibitors. A novel cross-Aldol reaction of α-keto phosphonates and ketones was used to synthesize α-hydroxy phosphonates with excellent enantioselectivities [31]. In the presence of 20 mol% (*S*)-proline as catalyst, a good enantiomeric excess of 71% of the desired Aldol product was obtained in a yield of 85% at room temperature. Alkyl-substituted ketophosphonates are also good substrates for this reaction and the best results are obtained for isopropyl derivatives (Scheme 7.11).



R^1= Me, R^2= Me, 87% yield, 69% ee
R^1= Me, R^2= Et, 83% yield, 69% ee
R^1= Me, R^2= *i*-Pr, 82% yield, 74% ee
R^1= *i*-Pr, R^2= OMe, 93% yield, 85% ee

Scheme 7.11 Organocatalytic enantioselective synthesis of α-hydroxy phosphonates.

For the intermolecular Aldol reaction, aldehydes can also act as a source of the nucleophile. In 2002, MacMillan and Northrup [32] reported the first direct and enantioselective cross-Aldol reaction of aldehydes (Scheme 7.12). The capacity of proline to catalyze asymmetric cross-Aldol reactions between nonequivalent aldehydes was investigated. Syringe pump addition of propionaldehyde to a series of aldehyde acceptors in the presence of amino acid catalyst effectively suppressed homodimerization of the donor aldehyde while providing good yields (75–87%) of the desired β-hydroxyaldehyde as a mixture of *anti*- and *syn*-diastereoisomers (up to 21 : 1 dr) with excellent enantioselectivity (91–99%).

The great success obtained in the cross-Aldol reaction between aldehydes has been further expanded to other electrophilic aldehydes, such as 1,3-dithianyl-2-carboxaldehyde [33], aqueous formaldehyde [34], and α-alkoxyacetaldehydes [35]. Especially, the use of α-alkoxyacetaldehydes allowed access to the polyol architecture. (*S*)-Proline was also able to catalyze the auto-Aldol reaction of aldehydes to obtain their dimers and trimeric products, directly giving polyketide derivatives (see Section 7.2.3).

The intermolecular Aldol reaction between aldehydes has been proven to show the potential for the synthesis of α-amino acids. The reaction of a glycine aldehyde derivative with an excess of different aldehydes gave an *anti*-β-hydroxy-α-amino aldehyde as the main product [36].

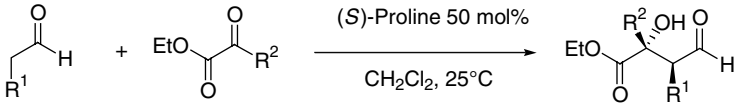
The intermolecular Aldol reaction, in which aldehydes act as nucleophiles and ketones as electrophiles, has also been investigated [37]. (*S*)-Proline is able to catalyze the reaction of different aldehydes with highly electrophilic ketones, such as trifluoropyruvate, or diethyl ketomalonate, giving expected products with good yields and enantioselectivities. However, poor diastereoselectivities were observed with trifluoropyruvate (Scheme 7.13).

7.2.2.2 **Mechanisms**

The originally proposed mechanism of the proline-catalyzed intermolecular Aldol reaction via an enamine mechanism was introduced by List *et al.* based on the class I aldolase mechanism [25]. In the proposed mechanism, the proline functions as a

R^1= Me, R^2= Et, 80% yield, 4:1 *anti*/*syn*, 99% ee
R^1= Me, R^2= *cy*-C_6H_{11}, 87% yield, 14:1 *anti*/*syn*, 99% ee
R^1= Me, R^2= Ph, 81% yield, 3:1 *anti*/*syn*, 99% ee
R^1= Me, R^2= *i*-Pr, 82% yield, 24:1 *anti*/*syn*, > 99% ee
R^1= Bu, R^2= *i*-Pr, 80% yield, 24:1 *anti*/*syn*, 98% ee
R^1= Bn, R^2= *i*-Pr, 75% yield, 19:1 *anti*/*syn*, 91% ee

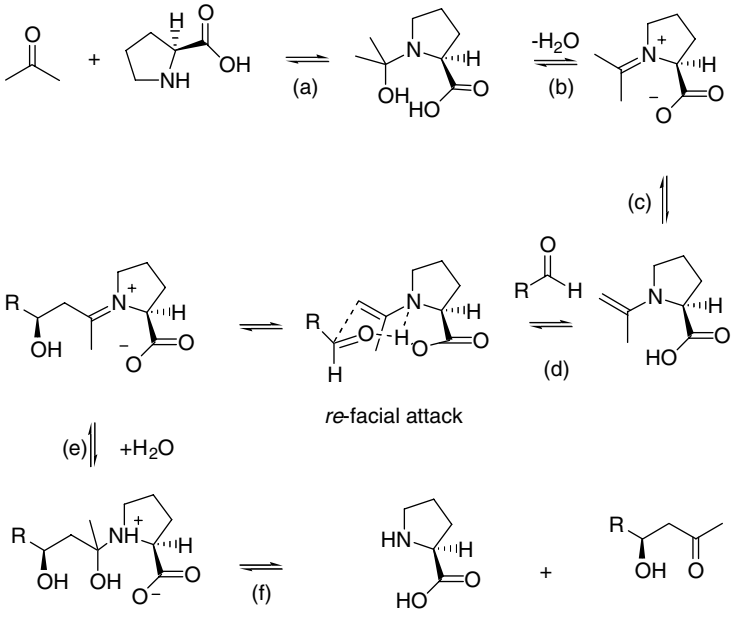
Scheme 7.12 Direct and enantioselective cross-Aldol reaction of aldehydes.



R^1= Me, R^2= CO_2Et, 90% yield, 90% ee
R^1= Et, R^2= CO_2Et, 91% yield, 85% ee
R^1= *i*-Pr, R^2= CO_2Et, 88% yield, 85% ee
R^1= allyl, R^2= CO_2Et, 94% yield, 88% ee
R^1= Hx, R^2= CO_2Et, 91% yield, 84% ee
R^1= Me, R^2= CF_3, 98% yield, 3:2 dr, 67/81% ee

Scheme 7.13 Aldol reaction between aldehydes as nucleophiles and ketones as electrophiles.

"microaldolase" that provides both the nucleophilic amino group and an acid/base cocatalyst in the form of carboxylate (Scheme 7.14). This cocatalyst may facilitate each individual step (a–f) of the mechanism. The enantioselectivity can be explained with a metal-free version of a Zimmerman–Traxler-type transition state. The tricyclic



(a) nucleophilic attack of the amino group
(b) the dehydration of the carbinol amine intermediate
(c) the deprotonation of the iminium species
(d) the carbon-carbon bond-forming step
(e and f) hydrolysis of the iminium-aldol intermediate

Scheme 7.14 Proline-catalyzed intermolecular Aldol reaction.

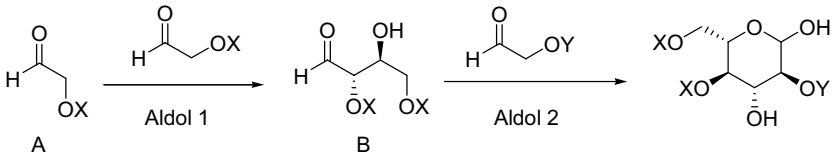
hydrogen-bonded framework provides for enantiofacial selectivity. The proposed multistep reaction mechanism has been confirmed using density functional theory calculations. The stereoselectivities of proline-catalyzed asymmetric intermolecular Aldol reaction can be predicted through quantum mechanical calculation based on this proposed mechanism [38].

7.2.3 Carbohydrate Synthesis

7.2.3.1 Carbohydrate Synthesis

Carbohydrates are an important class of natural products. They are involved in life-essential metabolic processes, signal transduction, and immune response. They are also the building blocks of oligo- and polysaccharides. Therefore, they are interesting as chiral pool materials and as building blocks for the synthesis of natural products and pharmaceuticals [10]. Although the synthesis of carbohydrates from simple achiral precursors is a fundamental process catalyzed by enzymes, it is still a challenging issue with the increasing development of glycobiology and carbohydrate-based pharmaceuticals. Most conventional synthetic approaches usually have drawbacks, such as multiple steps, requiring protective groups, and sometimes, a subsequent reduction or oxidation is necessary [39–41]. The demands and challenges of this area inspired the increased development of reaction design and methodological advancement for selective construction of natural and unnatural carbohydrates. Since List and Notz introduced the highly regioselective and diastereoselective intermolecular Aldol reaction catalyzed by amino acids, the similarity between the Aldol reaction product and carbohydrates has initiated a new strategy to construct carbohydrate-type structures [42]. The practical execution of this carbohydrate synthesis strategy would require the invention of two new Aldol technologies: (a) an enantioselective Aldol union of α-oxyaldehyde substrates (Aldol step 1) and (b) a diastereoselective Aldol coupling between trioxy-substituted butanals and an α-oxyaldehyde enolate (Aldol step 2) [39] (Scheme 7.15).

MacMillan *et al*. recently reported the two-step synthesis of carbohydrates based on this concept [35]. They showed that *O*-protected glycol aldehydes could be converted in the presence of (*S*)-proline into the corresponding Aldol products with high enantiomeric excess; however, the electronic nature of the oxyaldehyde

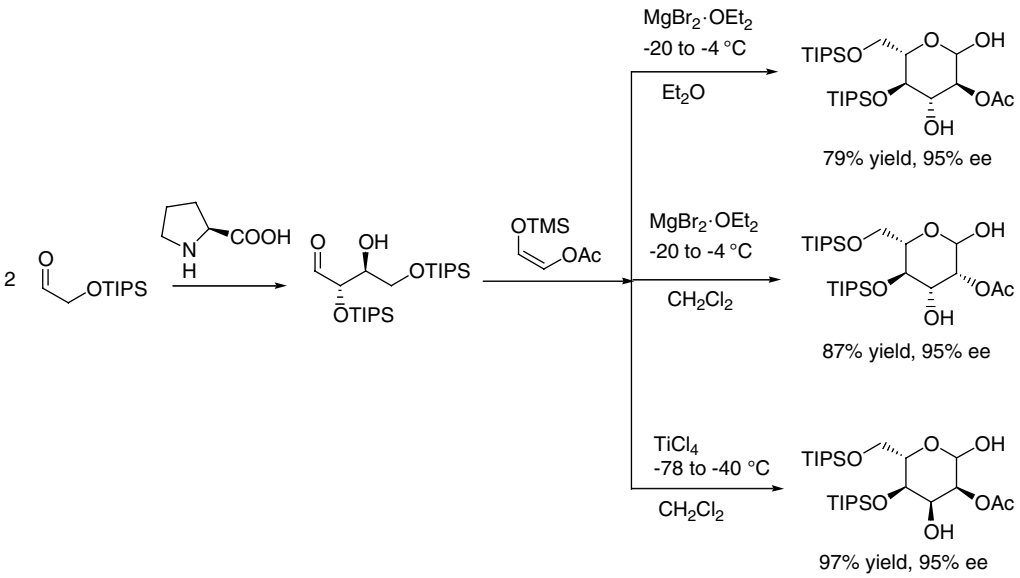


Aldol 1 requires α-oxyaldehyde A (reagent) is reactive in aldol union
Aldol 2 also requires α-oxyaldehyde B (product) is nonreactive in aldol union

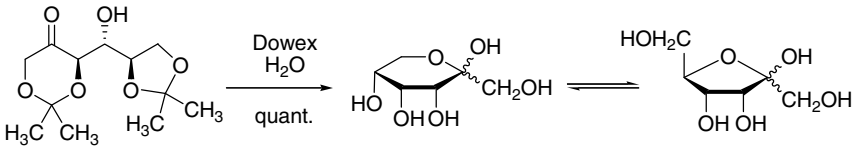
Scheme 7.15 Two-step carbohydrate synthesis: iterative aldehyde aldol.

substituent has a pronounced effect on the overall efficacy of the process. For example, electron-rich substituted derivatives could be converted with good results, whereas electron-withdrawing substituted derivatives do not participate in this transformation. Interestingly, the reaction of protected glycoaldehydes can be extended to α-alkyl-substituted aldehydes. The principal issue in this reaction is that the nonequivalent aldehydes must selectively partition into two discrete components, a nucleophilic donor and an electrophilic acceptor. More important, the α-oxyaldehyde products of this new Aldol protocol are apparently inert to further proline-catalyzed enolization or the enamine addition reaction, which could be subjected to a second Aldol reaction catalyzed by Lewis acids, such as $MgBr_2$ and $TiCl_4$. The stereochemical outcome of the reaction can be effectively controlled by altering the reaction conditions. The usage of $MgBr_2$ resulted in the formation of glucose derivatives in diethyl ether, whereas mannose was formed in dichloromethane. In the presence of $TiCl_4$, the corresponding allose derivative was obtained [43] (Scheme 7.16).

In 2001, Enders and Grondal reported the direct synthesis of selectively protected ketoses in one step through the diastereo- and enantioselective organocatalytic Aldol reaction between 2,2-dimethyl-1,3-dioxan-5-one and suitable aldehyde carbonyl components [44]. The optimal reaction conditions reported so far use (*S*)-proline as the catalyst and DMF as the solvent at a temperature of 2 °C. High enantioselectivities and excellent diastereoselectivities were obtained; however, the yields were sensitive to the structure of the aldehydes used. Usually, good yields, excellent



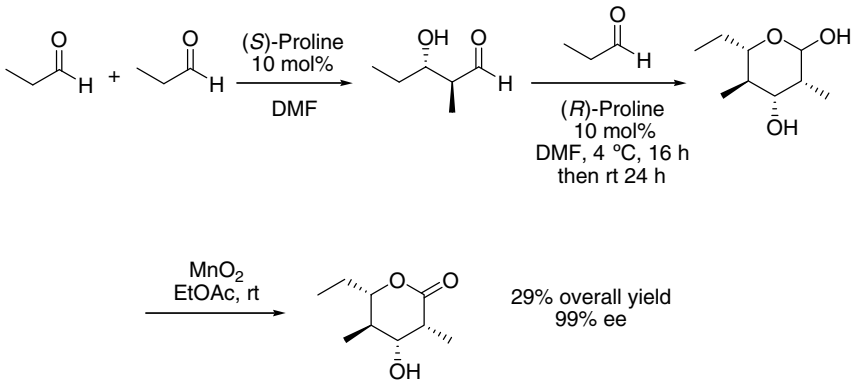
Scheme 7.16 Proline-catalyzed carbohydrate synthesis according to MacMillan *et al.* [43]. TIPS = triisopropylsilyl; TMS = trimethylsilyl.



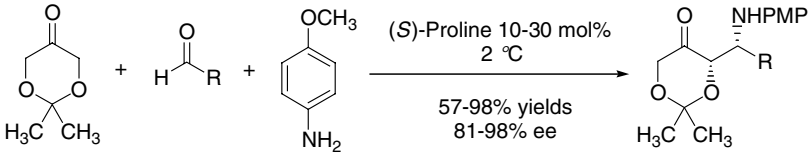
Scheme 7.17 Synthesis of D-piscose from Aldol products.

anti/*syn* ratios, and enantiomeric excesses were observed with α-branched aldehydes, and lower yields and selectivity were observed with linear aldehydes. This may be due to the linear aldehydes undergoing self-Aldol condensation, which is in direct competition with the cross-Aldol reaction. Interestingly, when *S*-configured, enantiomerically pure *N*-Boc- and *N*-Cbz-protected Garner aldehydes were used, (*S*)-proline proved to be the appropriate catalyst. However, (*R*)-proline was the catalyst of choice for reactions of glyceraldehyde. The double acetonide protected D-piscose was quantitatively deprotected with an acidic ion exchange resin to give the parent D-piscose from 2,2-dimethyl-1,3-dioxan-5-one in 76% overall yield (Scheme 7.17). This new protocol offers a viable simple alternative for the asymmetric synthesis of the rare ketosugar in practically one step.

At the same time, Cordova *et al*. also realized the two-step pathway to construct sugar-type skeletons with proline catalysis [45]. They showed that propionaldehyde dimerized utilizing (*S*)-proline catalysis, and then the corresponding Aldol product was treated with propionaldehyde in the presence of a catalytic amount of (*R*)-proline, to yield the hexose. They disclosed the direct amino acid-catalyzed asymmetric *de novo* synthesis of hexoses with excellent chemo-, diastereo-, and enantioselectivity. The employment of a two-step direct catalytic synthetic protocol furnished either L- or D-sugars in most cases with >99% e.e. This two-step Aldol strategy opens up a novel route to enantiomerically pure δ-lactones from simple aldehydes. For example, the hexoses obtained can be quantitatively converted into δ-lactones by oxidation with MnO_2 (Scheme 7.18).



Scheme 7.18 Two-step direct amino acid-catalyzed enantioselective synthesis of hexoses.



R = $CH(OCH_3)_2$, CO_2Et, CH_2OBn

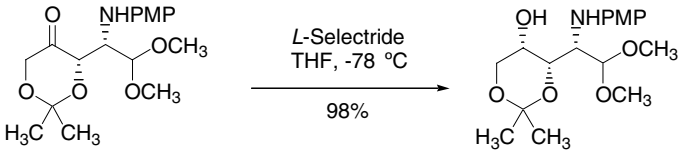
Scheme 7.19 Asymmetric synthesis of selectively protected amino sugars and derivatives.

7.2.3.2 **Synthesis of Amino Sugars**

Amino sugars are a class of carbohydrates in which one or more hydroxyl functions are replaced with amino groups. They are found as parts of glycoproteins, glycolipids, and aminoglycosides, and many biologically active secondary metabolites contain free, methylated, or acetylated amino groups. Amino sugars are usually prepared in multistep synthetic sequences. Amino acids, particularly proline, have been regarded as an effective catalyst for the direct asymmetric three-component Mannich reaction of ketones, aldehydes, and amines to give β-amino ketones in high yields and enantioselectivities (Scheme 7.19). Enders *et al.* [46] and Cordova *et al.* ([47] and references therein), on the basis of their organocatalytic [$C_3 + C_x$] concept for the direct synthesis of carbohydrates, reported the successful development of a diastereo- and enantioselective Mannich variant that paves the way for selectively protected amino sugars and their derivatives.

They found that the optimal reaction temperature is between 2 °C and ambient temperature. Lower temperatures resulted in a decrease in both the reaction rate and the diastereo- and enantioselectivity. Independent of the solvent employed, the addition of 1–10 equiv. of water led to an increase in the stereoselectivity. Analogously, *N*-Boc- or *N*-Cbz-protected (*S*)-configured enantiomerically pure Garner aldehydes were investigated for their Mannich reactions. The authors found that when (*S*)-proline was used, only traces of the desired products could be isolated, whereas the (*R*)-proline-catalyzed reaction provided the Mannich products in good yields (57–67%) and excellent stereoselectivity (>96% e.e.) [46].

The Mannich product obtained could be transformed into the corresponding β-amino alcohol. Reduction of the keto function with L-selectride proceeded at −78 °C with high stereocontrol (Scheme 7.20).



Scheme 7.20 Reduction of β-aminoketone to β-amino alcohol with L-selectride.

7.3
Mannich Reaction

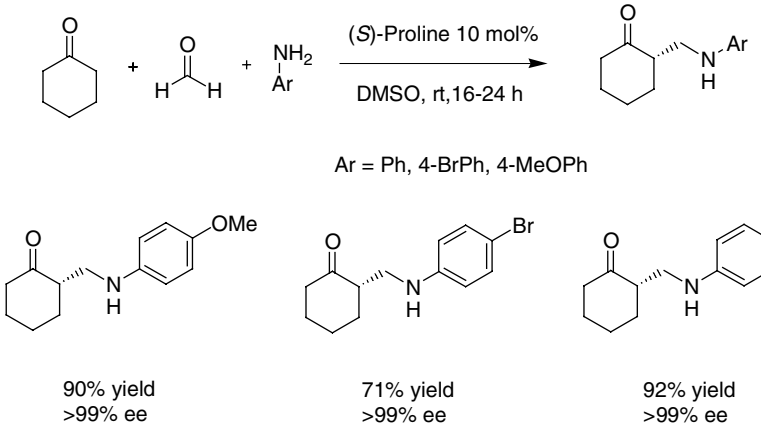
The potential use of amino acids for the asymmetric Mannich reaction has been realized. Amino acids have been reported to catalyze both indirect and direct variants of the Mannich reaction. The Mannich reaction is a powerful synthetic method for the construction of C−C bonds for molecules containing nitrogen [5, 9, 48]. Chiral nitrogenous molecules are found as pharmaceutical drugs and natural products, with a great research effort being expended to find asymmetric organocatalyzed Mannich reactions. The classical Mannich reaction [49], whereby an amino methyl group is placed α to a carbonyl group, has recently been extended to generate highly functionalized nitrogen-containing molecules. The Mannich reaction has also been developed to provide access to β-amino carbonyls, β-amino alcohols, 1,2-amino alcohols, 1,2-diamines and 1,4-diamines, and α-functionalized amino acids [48].

7.3.1
α-Aminomethylation

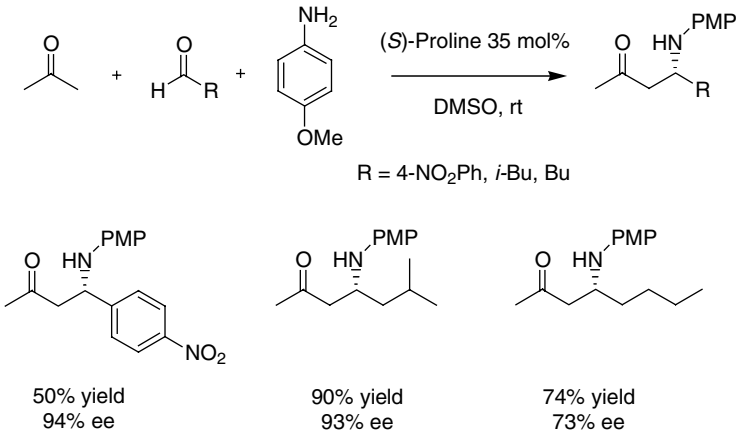
Cordova *et al*. have published a stereoselective version of the classic Mannich reaction [50]. They reported a direct catalytic enantioselective α-aminomethylation of ketones using (*S*)-proline as the catalyst (Scheme 7.21). The α-aminomethylated ketones were prepared with yields up to 92% and enantioselectivities >99%.

7.3.2
Direct Mannich Reaction

List has extended the potential of the Mannich reaction by developing a direct asymmetric Mannich reaction catalyzed by (*S*)-proline [51]. The major advantage was that no preformed activated enolates and imines or modified donors were required.



Scheme 7.21 Direct catalytic one-pot three-component α-aminomethylation reaction with anilines.



Scheme 7.22 Direct Mannich reaction with acetone as nucleophile catalyzed by (*S*)-proline.

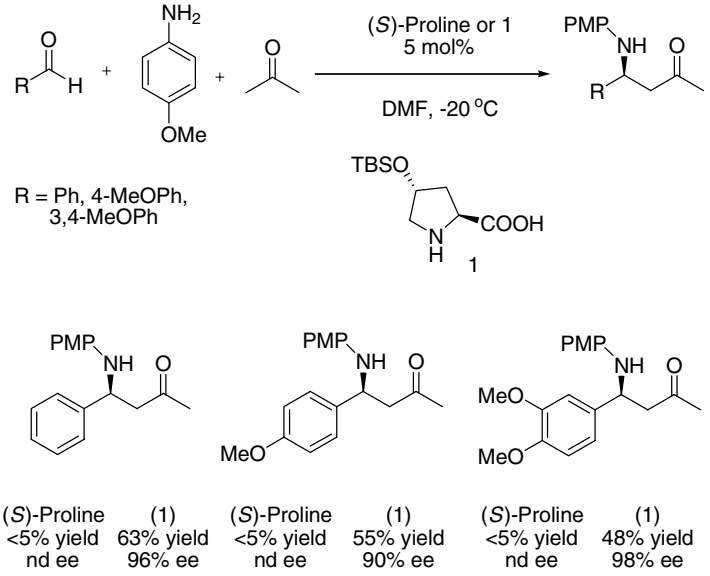
Acetone was reacted with a wide range of substituted and unsubstituted aliphatic aldehydes (Scheme 7.22). The one-pot reaction furnished the aminocarbonyl products in good yields and high enantioselectivity.

Initially, there were several limitations to the usage of (*S*)-proline for the direct asymmetric three-component Mannich reaction: (i) a high catalyst loading is required and (ii) proline is not an effective catalyst for electron-rich aldehyde substrates. Many research groups have optimized the Mannich reaction to remove these limitations associated with (*S*)-proline.

Hayashi *et al.* have, however, found that under high pressure (*S*)-proline can catalyze reactions with electron-rich aldehydes and the reaction products are obtained with higher yields and stereoselectivities compared to the reaction performed under normal pressure [52]. A highly active 4-siloxyproline catalyst described by Hayashi requires a lower catalyst loading compared to (*S*)-proline and extends the scope of the Mannich reaction to include electron-rich aldehydes [53]. Good yields and high enantioselectivities were obtained using 4-siloxyproline compared to (*S*)-proline where the yields were less than 5% and enantioselectivities were not determined (Scheme 7.23).

List *et al.* selected the *para*-methoxyphenylamine (PMP) as a protecting group, originally used by Bravo *et al.* [54], because it could be easily removed under oxidative conditions. However, the Boc-protected amine has been found to be effective for the proline-catalyzed Mannich reaction of aldehydes with *N*-Boc-imines (Scheme 7.24) [55]. The Boc group can be removed under mild conditions compared to the PMP group, although the *N*-Boc-protected imine is preformed compared to the three-component method reported by List [51].

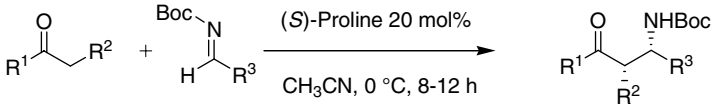
A direct organocatalyzed asymmetric three-component Mannich reaction using acyclic amino acids has been reported by Cordova. Initially, acyclic amino acids were found to be effective catalysts for the Aldol reaction [26]. They extended the scope of the Mannich reaction by screening amino acids such as (*S*)-valine, (*S*)-alanine, (*S*)-serine, (*S*)-leucine, (*S*)-isoleucine, and (*S*)-alanine tetrazole for catalysis of the three-



Scheme 7.23 A *trans*-4-siloxyproline-catalyzed asymmetric three-component Mannich reaction.

component Mannich reaction (Scheme 7.25) [56]. The amino acids generated the Mannich products in good yields and enantioselectivities (>99%). Reactions proceeded with excellent chemoselectivity, and the primary amino acid did not act as an amine component under their reactions conditions. Addition of 2 or 3 equiv. of water was found to slightly improve yields. Also, addition of 1 equiv. of dicyclohexylamine led to increased yields; however, longer reaction times resulted in decreased enantioselectivity. The modification of alanine to the tetrazole derivative improved yields albeit with lower enantioselectivity for the asymmetric Mannich reaction.

The Mannich reaction has been extended for the enantioselective synthesis of 1,2-amino alcohols [57]. (*S*)-Proline-catalyzed Mannich reaction using hydroxyacetone as



R^1= H, R^2= Me, R^3= Ph, 91%, >99:1 dr, >99% ee

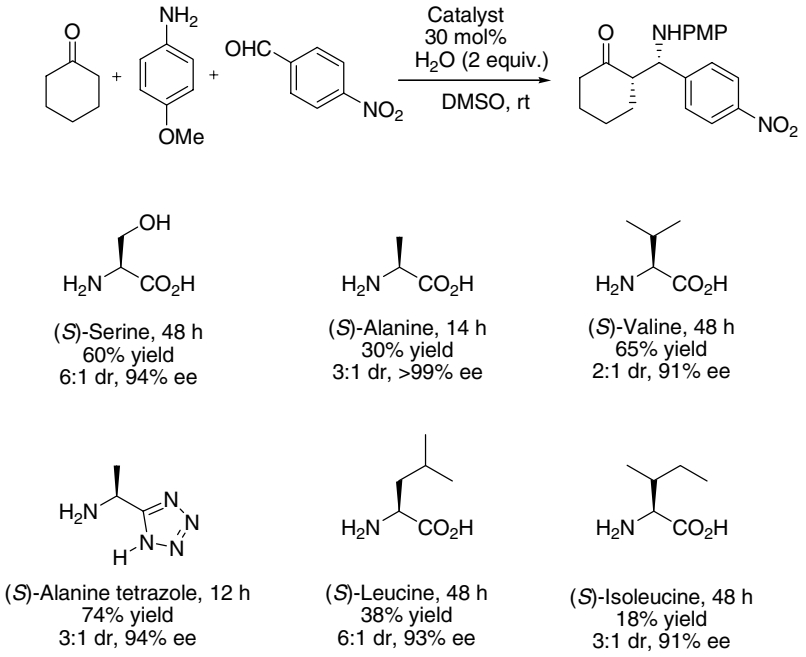
R^1= H, R^2= *i*-Pr, R^3= Ph, 88%, >99:1 dr, >99% ee

R^1= H, R^2= *i*-Pr, R^3= 4-MeOPh, 80%, >99:1 dr, >99% ee

R^1= H, R^2= *i*-Pr, R^3= Naphth, 82%, >99:1 dr, >99% ee

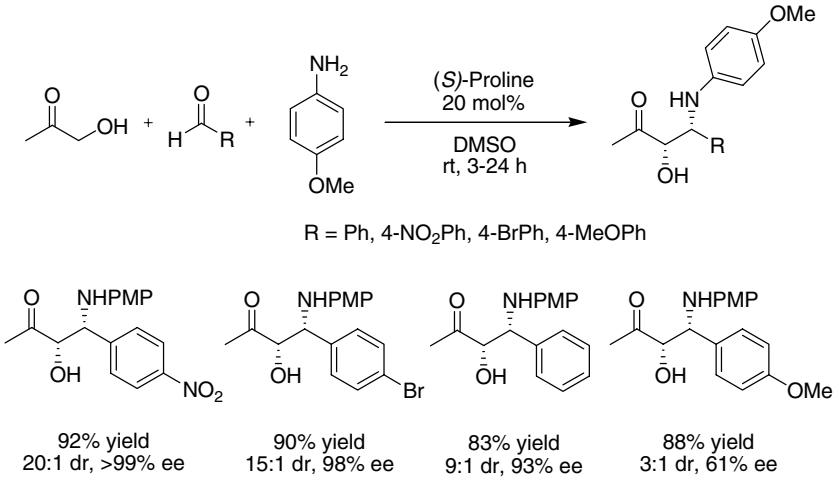
R^1= Me, R^2= H, R^3= Ph, 73%, >99% ee

Scheme 7.24 Asymmetric Mannich reaction of aldehydes with *N*-Boc-imines.

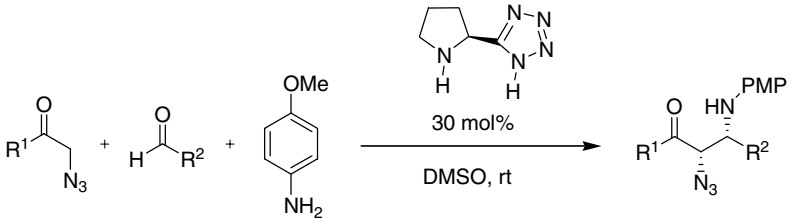


Scheme 7.25 Direct asymmetric Mannich reaction catalyzed by acyclic amino acids.

the ketone donor yielded *syn*-1,2-amino alcohols with an excellent regioselectivity, high stereoselectivities, and good yields (Scheme 7.26). A lower yield (57%) was obtained when aliphatic aldehydes were used compared to aromatic aldehydes (up to 92%).



Scheme 7.26 Direct Mannich reaction for *syn*-1,2-amino alcohols.



R^1= Me, R^2= CO_2Et, 96%, 91:9 *syn*/*anti*, 99/99% ee

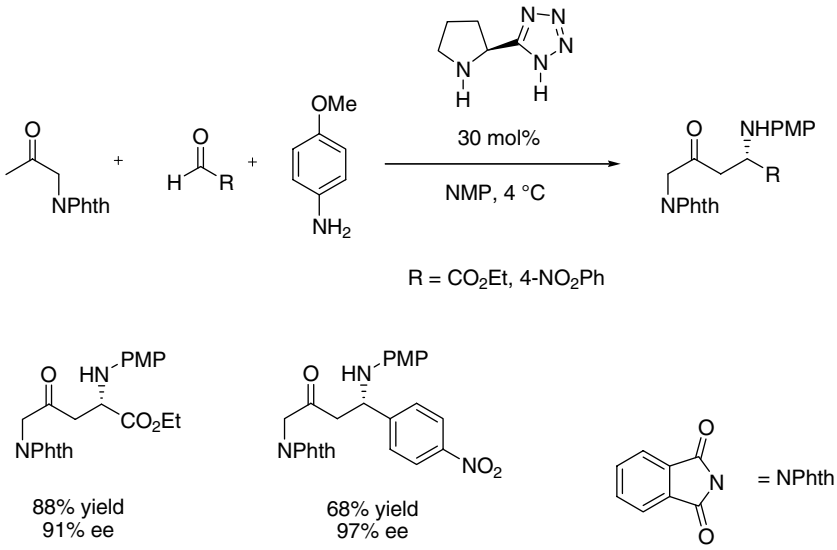
R^1= Ph, R^2= CO_2Et, 87%, 88:12 *syn*/*anti*, 99/64% ee

R^1= Me, R^2= CH_2OBn, 83%, 80:20 *syn*/*anti*, 85/29% ee

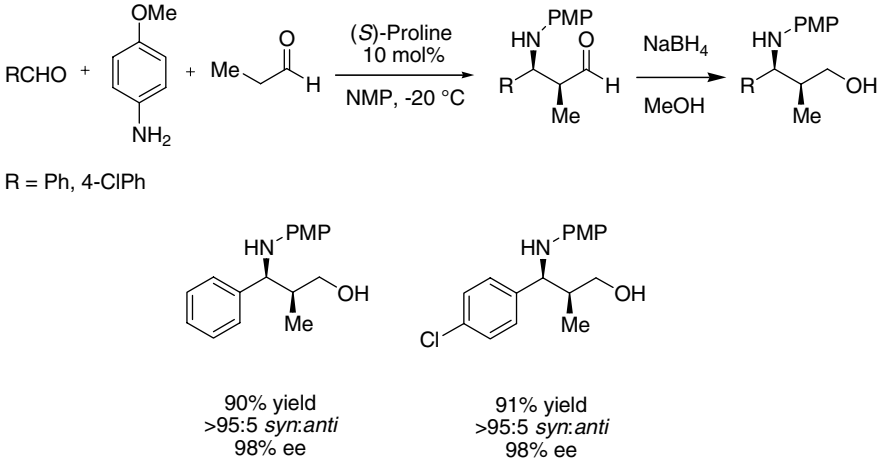
R^1= Et, R^2= CH_2OBn, 80%, 85:15 *syn*/*anti*, 82/79% ee

Scheme 7.27 Direct Mannich reaction for 1,2-azidoamines.

Barbas *et al*. extended the scope of the Mannich reaction for the generation of 1,2- and 1,4-diamines [58]. The diamines were prepared by an organocatalytic asymmetric Mannich reaction of protected amino ketones with imines using an (*S*)-proline-derived tetrazole catalyst. The regioselectivity of the reaction was controlled by protecting the amine functionality. Azido ketones were used as donors and reacted with various imines to yield chiral 1,2-azidoamines (Scheme 7.27), and α-phthalimido ketones were reacted with imines to yield chiral 1,4-diamines (Scheme 7.28). Both 1,2-azidoamines and 1,4-diamines were furnished with excellent yields and enantioselectivities.



Scheme 7.28 Direct Mannich reaction for protected 1,4-diamines.



Scheme 7.29 Direct three-component Mannich reaction using aldehydes as nucleophiles.

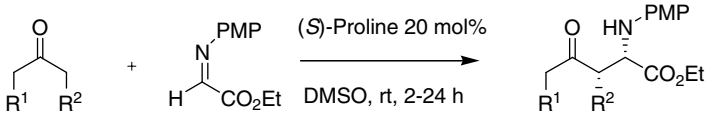
The use of aldehydes as nucleophiles for the Mannich reaction was demonstrated by Hayashi *et al.* [59]. Two aldehydes were reacted together where one was used as the Mannich donor and the other was a Mannich acceptor to furnish β-aminoaldehydes (Scheme 7.29). The reaction was performed using NMP as solvent, at $-20\,^{\circ}C$, to suppress cross-Aldol and homo-Mannich adducts, affording high yields and stereoselectivities. The β-aminoaldehydes were reduced with $NaBH_4$ as the aldehyde products easily racemize, to give β-amino alcohols with high yields.

7.3.3
Indirect Mannich Reaction Using Ketone Donors

α-Functionalized amino acids have been prepared using the asymmetric Mannich reaction. Enormous interest has been focused on the generation of α-functionalized amino acids due to their use in drugs and peptidomimetics. Barbas *et al.* have developed a highly enantioselective route via the Mannich reaction for the synthesis of α-functionalized amino acids [60]. Various unmodified ketone donors were reacted with an α-iminoglyoxylate acceptor molecule in a stereoselective manner catalyzed by (*S*)-proline (Scheme 7.30). The Mannich adducts were furnished in good yields, diastereoselectivities, and excellent enantioselectivities. Removal of the PMP protecting group under oxidative conditions followed by ester hydrolysis can yield α-functionalized amino acids.

7.3.4
anti-Mannich Reactions

The Mannich reactions using (*S*)-proline as catalyst are predominantly *syn*-selective. *anti*-Selective Mannich reactions catalyzed by amino acids present a significant challenge. Barbas and Cordova first reported an asymmetric organocatalytic *anti*-



R^1= H, R^2= H, 82%, 95% ee

R^1= H, R^2= Me, 72%, >19:1 dr, >99% ee

R^1= H, R^2= OH, 62%, >19:1 dr, 99% ee

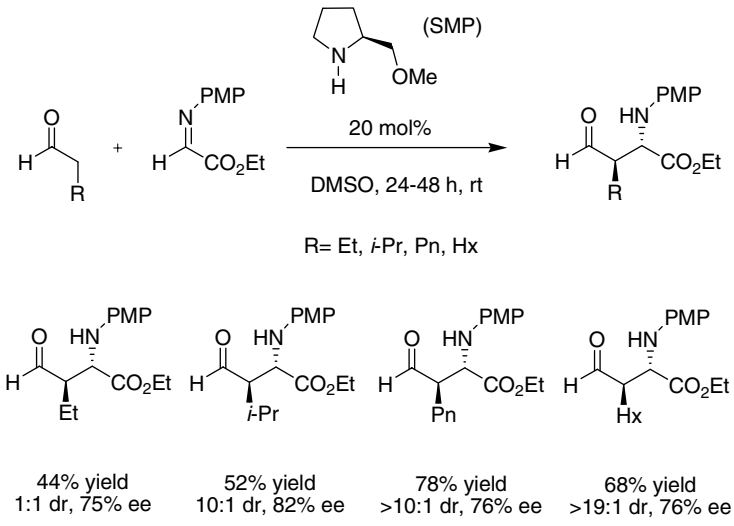
R^1,R^2= -$(CH_2)_3$-, 81%, >19:1 dr, >99% ee

Scheme 7.30 (*S*)-Proline-catalyzed Mannich reaction of ketone donors with *N*-PMP-protected α-imino ethyl glyoxylate.

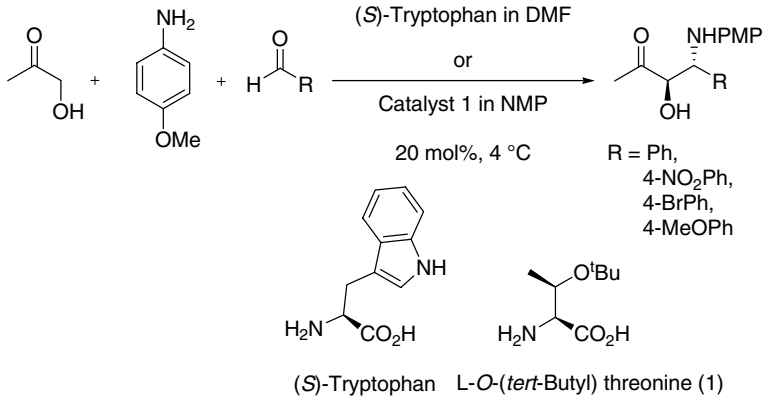
selective Mannich reaction using (*S*)-2-methoxymethylpyrrolidine (SMP) as catalyst [61]. They reacted various α-substituted aliphatic aldehydes with *N*-PMP-protected α-imino ethyl glyoxylate in DMSO using SMP as organocatalyst at room temperature (Scheme 7.31).

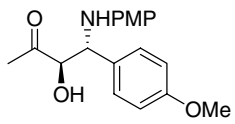
The *anti*-Mannich adducts were obtained with moderate yields, good diastereoselectivities, and good enantioselectivities. SMP gave functionalized amino acid derivatives with *anti* stereochemistry.

(*S*)-Tryptophan- and L-threonine-derived catalysts have been found to be effective for the *anti*-selective Mannich reaction [62]. Hydroxyacetone was reacted with a variety of imines in DMF for (*S*)-tryptophan and *N*-methylpyrrolidone (NMP) for L-*O*-(*tert*-butyl) threonine at 4 °C (Scheme 7.32). Reaction times using (*S*)-tryptophan were faster (16–20 h) than those using L-*O*-(*tert*-butyl) threonine (48 h). *anti*-Amino

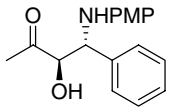


Scheme 7.31 SMP-catalyzed Mannich reaction of unmodified aldehydes with *N*-PMP-protected α-imino ethyl glyoxylate.

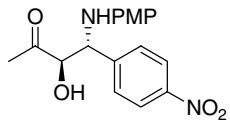




(*S*)-Tryptophan
72% yield, 1.3:1 dr
53% ee

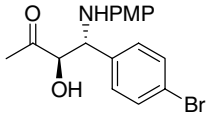


(*S*)-Tryptophan
75% yield, 4:1 dr
77% ee



(*S*)-Tryptophan
95% yield, 12:1 dr
95% ee

Catalyst 1
85% yield, >15:1 dr
98% ee



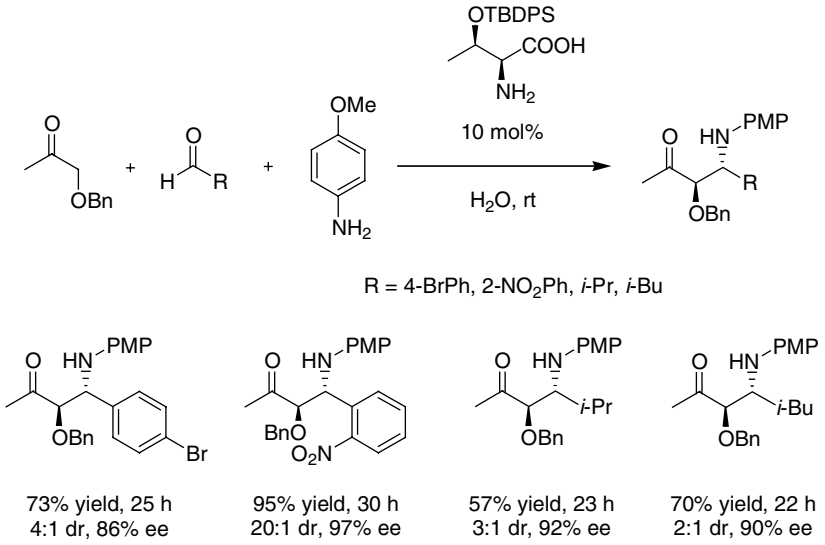
(*S*)-Tryptophan
89% yield, >10:1 dr
93% ee

Catalyst 1
71% yield, >10:1 dr
94% ee

Scheme 7.32 *anti*-Selective Mannich reaction catalyzed by acyclic amino acids.

alcohols were prepared in good yields with excellent diastereoselectivities and enantioselectivities.

Lu *et al.* reported the first direct three-component *anti*-selective Mannich reaction to be catalyzed by a modified primary amino acid in water [63]. Threonine-derived organocatalysts were found to be the most effective. Benzyl-protected hydroxyacetone as the donor molecule was reacted with various aliphatic and aromatic aldehydes using *O*-(*tert*-butyldiphenylsilyl) threonine as an organocatalyst in water (Scheme 7.33). Mannich products were furnished in good yields, diastereoselectivities, and enantioselectivities. Higher yields and diastereoselectivities were obtained using aromatic aldehydes compared to aliphatic aldehydes. Ionic liquids have also been considered as an alternative to traditional organic solvents as a medium for the Mannich reaction. Barbas and his group have found that imidazolium ionic liquids facilitate the proline-catalyzed asymmetric Mannich reaction of *N*-PMP-



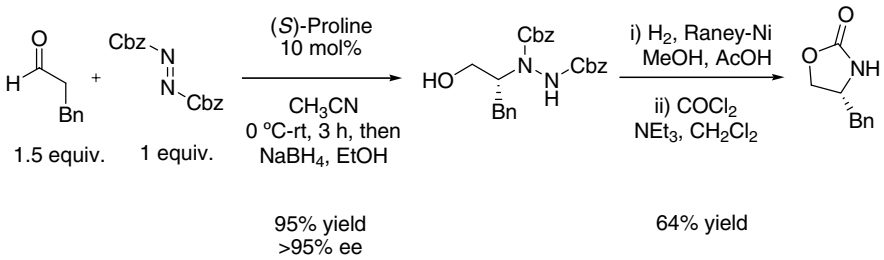
Scheme 7.33 *anti*-Selective Mannich reaction catalyzed by protected L-threonine in water.

protected α-imino ethyl glyoxylate with aldehydes and ketones to furnish the amino acids with good yields and enantioselectivities [64].

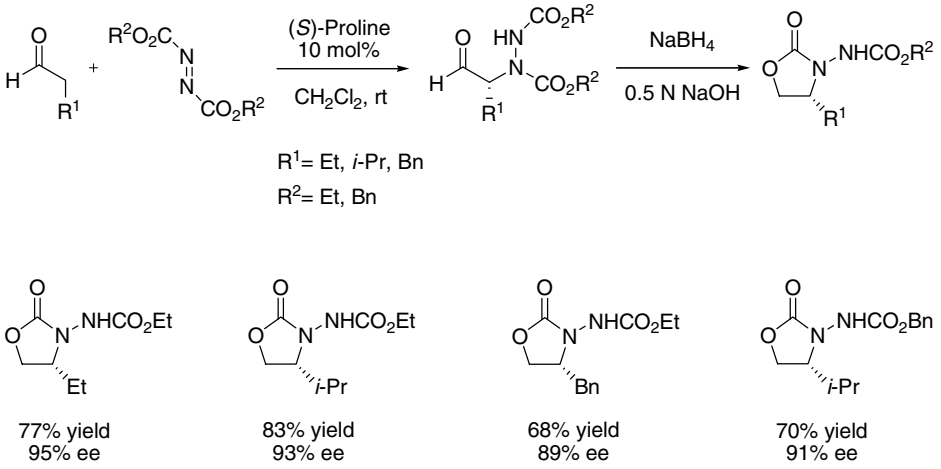
7.4 α-Amination Reaction

The development of a direct asymmetric organocatalytic α-amination reaction for aldehydes was independently reported by List [65] and Jørgensen [66]. Interest in this reaction stems from the versatile products that can be produced, such as amino alcohols, α-amino acids, and 4-substituted oxazolidinone compounds.

List reported a direct catalytic asymmetric α-amination of aldehydes using (*S*)-proline as catalyst and a dialkylazodicarboxylate as the nitrogen electrophile with a slight excess of aldehyde in acetonitrile (Scheme 7.34) [65]. The resulting α-hydrazino aldehydes were reduced with $NaBH_4$ *in situ* as the product is prone to racemization. List isolated the amination products as their corresponding configurationally stable



Scheme 7.34 Direct catalytic asymmetric α-amination of aldehydes using (*S*)-proline by List [65].



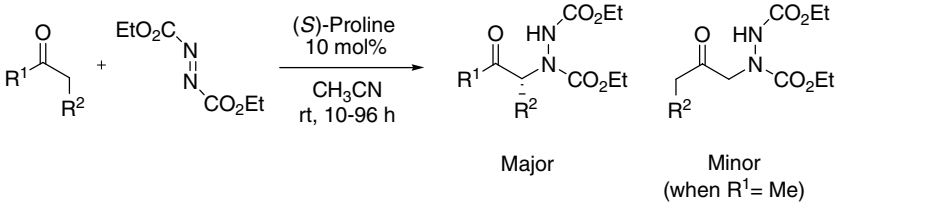
Scheme 7.35 Direct catalytic asymmetric α-amination of aldehydes using (*S*)-proline by Jørgensen *et al.* [66].

crystalline 2-hydrazino alcohols with high yields and excellent enantioselectivities. The 2-hydrazino alcohols were transformed into 4-substituted oxazolidinone compounds by hydrogenation over Raney nickel and subsequent condensation with phosgene.

Jørgensen *et al.* also obtained α-aminated products of aldehydes in good yields and with excellent enantioselectivities (Scheme 7.35). Reduction and subsequent cyclization of the α-amino aldehyde products directly gave the *N*-amino-oxazolidinones [66].

Jørgensen *et al.* extended the α-amination reaction to ketone substrates (Scheme 7.36) [67]. They found the reaction to be highly regioselective with amination occurring at the most substituted carbon. The α-hydrazino ketones were obtained in good yields and excellent enantioselectivities.

Brase *et al.* have reported the α-amination of α,α-disubstituted aldehydes using (*S*)-proline that allows for the formation of α,α-disubstituted amino aldehydes, acids, and alcohols [68]. Various α,α-disubstituted aldehydes were reacted with diethyl azodi-

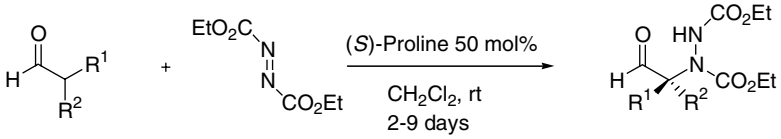


R1= Me, R2= Bn, 92% yield, major/minor 82:18, 98% ee

R1= Me, R2= *i*-Pr, 69% yield, major/minor 76:24, 99% ee

R1= Et, R2= Me, 79% yield, 94% ee

Scheme 7.36 Direct catalytic asymmetric α-amination of ketones using (*S*)-proline.



R^1= Me, R^2= Et, 52% yield, 28% ee

R^1= Me, R^2= Ph, 62% yield, 80% ee

R^1= Me, R^2= 2-Naph, 54% yield, 86% ee

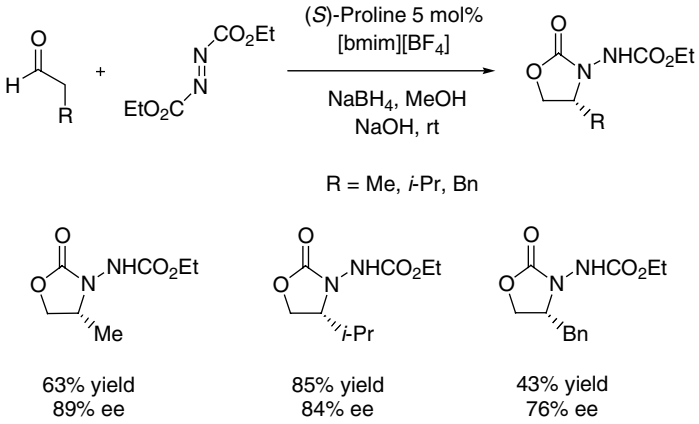
Scheme 7.37 Direct catalytic asymmetric α-amination of α,α-disubstituted aldehydes using (*S*)-proline.

carboxylate (DEAD) using 50 mol% (*S*)-proline in dichloromethane at room temperature (Scheme 7.37). The α-disubstituted amino aldehydes were obtained in moderate yields and with good enantioselectivities. However, long reaction times were required and low enantioselectivities were obtained for dialkyl-substituted aldehydes.

Ionic liquids as an alternative reaction media have been reported for the α-amination of aldehydes and ketones using (*S*)-proline (Scheme 7.38). The ionic liquid [bmim][BF_4] gave the highest yield (85%) and enantioselectivity (84% e.e.). For the α-amination of ketones, the authors obtained both mono- and bis-aminated products [69].

7.5 Michael Reaction

The Michael reaction is a key C–C bond-forming reaction that involves various Michael donor and acceptor molecules. Recent interest in the direct asymmetric organocatalyzed Michael reaction is due to the diverse array of products that can be obtained, such as γ-nitrocarbonyl compounds, functionalized pyrrolidines, and



Scheme 7.38 Direct catalytic asymmetric α-amination of aldehydes using (*S*)-proline in an ionic liquid.

1,4-Michael adducts [6, 70]. There are many modes of activation for the catalyzed Michael reaction, including catalysis via iminium ion-based activation of the Michael acceptor and enamine-based activation of Michael donor molecules. The potential usage of natural amino acids as organocatalysts for the Michael reaction has been investigated using a wide range of Michael donors and acceptors. It has been found that (*S*)-proline is not a highly efficient enantioselective catalyst for the Michael reaction as products are obtained with low enantioselectivities [71]. However, various chiral amine derivatives of (*S*)-proline have been employed for the Michael reaction to induce higher enantioselectivities.

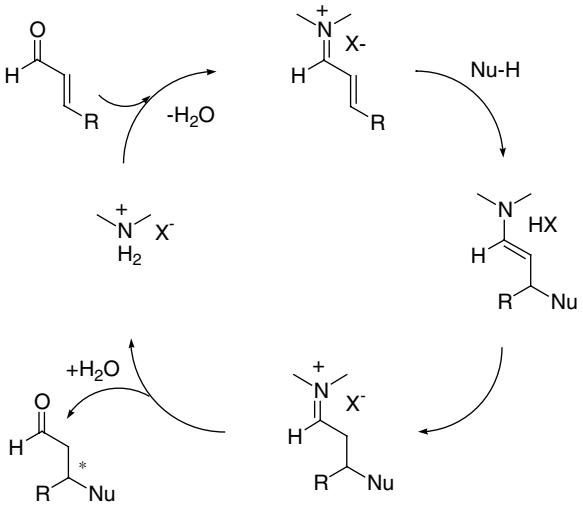
7.5.1
Mechanism for Iminium Ion-Catalyzed Michael Reaction

In the iminium ion-catalyzed Michael reaction, the Michael acceptor is activated by reversible condensation into a chiral iminium ion. The reversible condensation of an unsaturated carbonyl compound with a chiral secondary amine gives a chiral unsaturated iminium ion. A face-selective reaction with the nucleophile gives an enamine intermediate, which can then be hydrolyzed to a chiral carbonyl compound and regenerates the chiral secondary amine (Scheme 7.39).

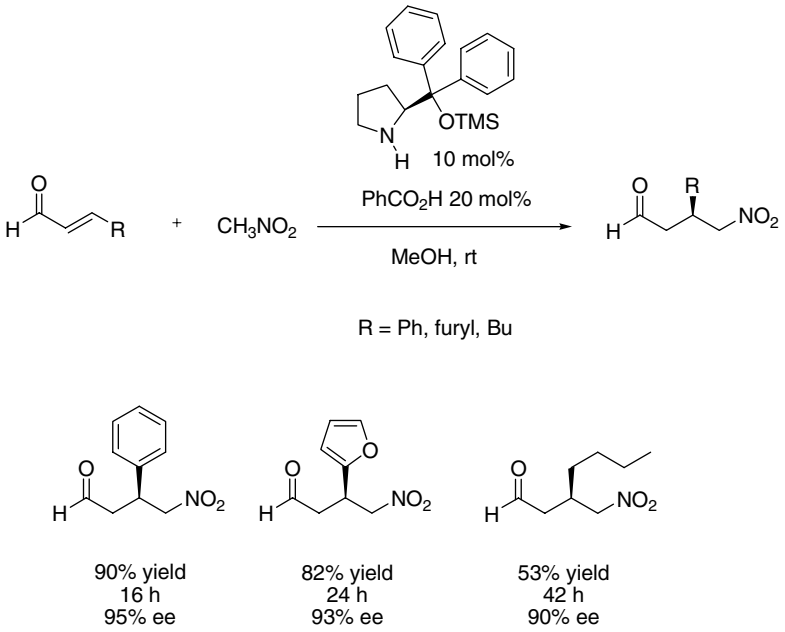
The substrates reported by various groups for an iminium-based Michael reaction involve α,β-unsaturated carbonyl compounds as Michael acceptors and Michael donors such as nitroalkanes and malonates that can form stabilized carbanions.

7.5.1.1 Iminium Ion-Catalyzed Intermolecular Michael Reactions

A diphenylprolinol silyl ether catalyst has been reported to be effective for addition of nitroalkanes to α,β-unsaturated aldehydes via chiral iminium ion catalysis



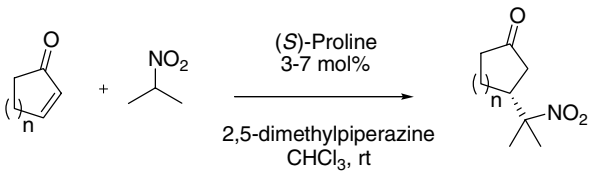
Scheme 7.39 Mechanism for iminium-catalyzed Michael reaction.



Scheme 7.40 Direct organocatalyzed intermolecular Michael reaction between nitroalkanes and α,β-unsaturated aldehydes.

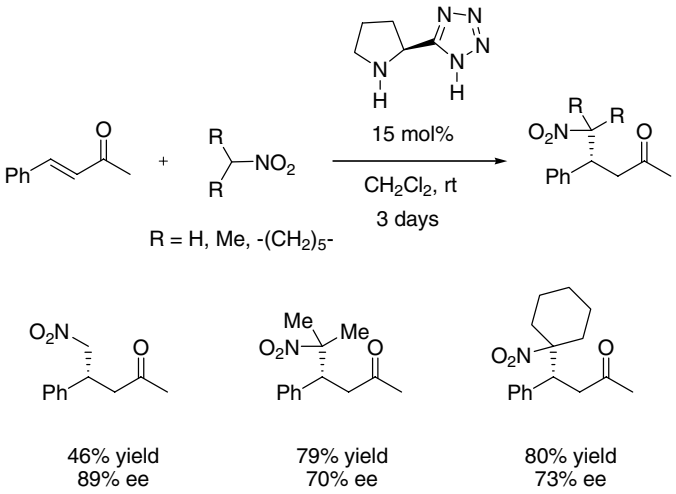
(Scheme 7.40). Addition to α,β-unsaturated aldehydes is more challenging because of the competitive 1,2-addition reaction [72]. The diphenylprolinol silyl ether generates Michael products using methanol as solvent in moderate yields and enantioselectivities greater than 90%. Aromatic α,β-unsaturated aldehydes generated products with higher yields compared to aliphatic analogs. Similar enantioselectivities were obtained using aliphatic and aromatic α,β-unsaturated aldehydes. However, longer reaction times were required for aliphatic α,β-unsaturated aldehydes (42–120 h) compared to aromatic α,β-unsaturated aldehydes (16–48 h).

The conjugate addition of nitroalkanes with cyclic enones catalyzed by (*S*)-proline, using *trans*-2,5-dimethylpiperazine, as an additive has been reported (Scheme 7.41) [73]. The effect of the additive is to increase the enantioselectivity



n=1, 66% yield, 75% ee
n=2, 88% yield, 93% ee

Scheme 7.41 Direct organocatalyzed intermolecular Michael reaction between cyclic enones and nitroalkanes.



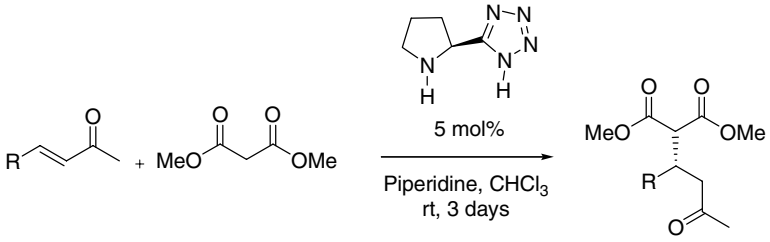
Scheme 7.42 Direct organocatalyzed intermolecular Michael reaction of acyclic enones with nitroalkanes using oven-dried glassware.

of the reaction as (*S*)-proline gave moderate selectivities. The reaction generated higher yields for six-membered rings compared to five- and seven-membered rings, with higher enantioselectivities for six- and seven-membered rings compared to five-membered rings. The effect of nitrocyclopentane and nitrocycloheptane as nucleophiles was also reported, but yields and enantioselectivities remained similar to nitrocyclohexane apart from the addition of nitrocycloheptane to cycloheptanone where lower yields were observed compared to other nitroalkanes.

Ley *et al.* have reported a proline tetrazole catalyst that facilitates the addition of nitroalkanes to acyclic enones (Scheme 7.42) [74]. 4-Phenyl-3-buten-2-one was reacted with various nitroalkanes. The yields and enantioselectivities for this particular reaction were moderate. The reaction was performed under different conditions using oven-dried glassware (conditions used for Scheme 7.42) and nonoven-dried glassware. Significantly lower yields were obtained using nonoven-dried glassware, but enantioselectivities increased slightly compared to oven-dried glassware conditions.

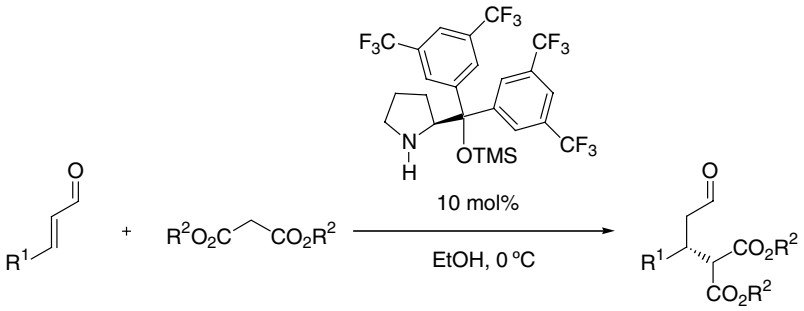
Ley *et al.* have also used proline tetrazole to catalyze the addition of malonates to enones (Scheme 7.43) [75]. Addition of malonates to enones proceeds via the iminium ion mechanism. Ley investigated the scope of the addition of malonates to enones. Under optimized conditions, dimethyl malonate was added to a series of enones such as 4-phenyl-3-buten-2-ones, with high yields and enantioselectivities. When an electron-withdrawing CF_3 group is present at the *para* position of the phenyl ring, higher yields and enantioselectivities are obtained compared to the electron-donating hydroxy substituent. Thiophene- and furan-substituted enones also lead to adducts with high yields.

Jørgensen *et al.* reported the first organocatalytic enantioselective conjugate addition of malonates to aromatic α,β-unsaturated aldehydes (Scheme 7.44) [76]. The Michael adducts were obtained with good yields and good to high



Scheme 7.43 Direct organocatalyzed intermolecular Michael addition of acyclic enones with malonates.

enantio-selectivities. Reaction scope was investigated and only protic solvents such as ethanol were found to be suitable for this reaction. Jørgensen *et al.* reacted benzyl and methyl malonates with various aromatic α,β-unsaturated aldehydes using ethanol as solvent and obtained adducts in high yields.



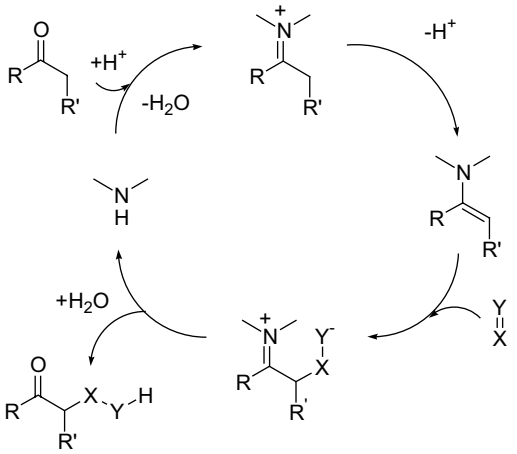
R^1= Ph, R^2= Bn, 80% yield, 91% ee

R^1= Ph, R^2= Me, 85% yield, 94% ee

R^1= 4-BrPh, R^2= Bn, 84% yield, 90% ee

R^1= 4-MeOPh, R^2= Bn, 93% yield, 92% ee

Scheme 7.44 Direct organocatalyzed intermolecular Michael addition of α,β-unsaturated aldehydes with malonates.



Scheme 7.45 Mechanism for enamine-catalyzed Michael reaction.

7.5.2 Mechanism for the Enamine-Catalyzed Michael Reaction

In the enamine-catalyzed Michael reaction, the Michael donor becomes activated by a reversible condensation of a carbonyl compound with a chiral secondary amine that generates a chiral enamine intermediate (Scheme 7.45). The generated chiral enamine intermediate can react with an electrophile to generate an iminium ion intermediate that becomes hydrolyzed to liberate the Michael adduct and regenerate the chiral secondary amine.

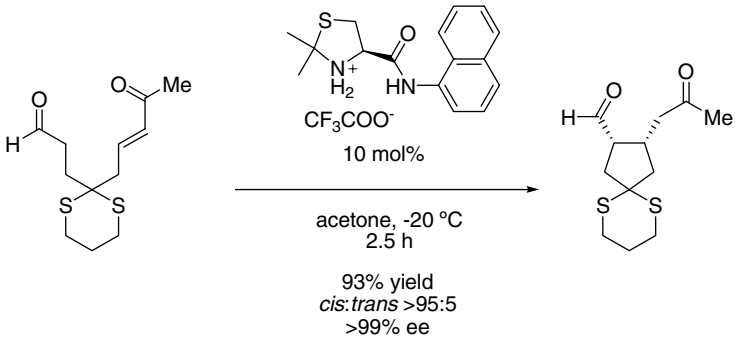
7.5.2.1 Enamine-Catalyzed Intramolecular Michael Reactions

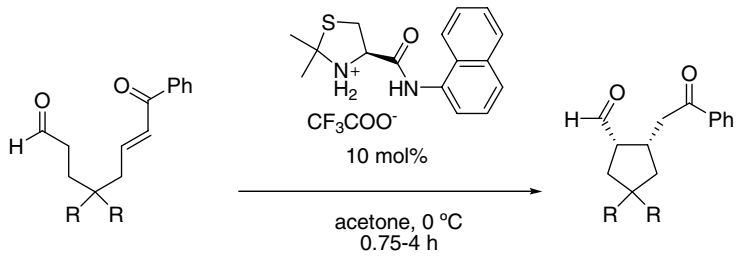
A cysteine-derived organocatalyst for an intramolecular Michael reaction has been designed by Hayashi *et al.* [77]. The cysteine-derived organocatalyst was applied to the reaction of formyl enones for the synthesis of chiral disubstituted cyclopentane derivatives (Scheme 7.46). Both aromatic and aliphatic enones are suitable substrates for this reaction, furnishing products in good yields and with high enantioselectivities. Substituents α to the aldehyde lead to an increase in diastereoselectivity but with a slight loss in yield.

7.5.2.2 Enamine-Catalyzed Intermolecular Michael Reactions

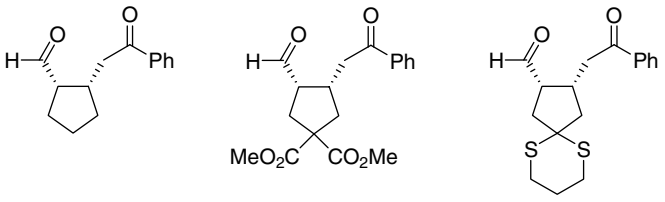
Various Michael acceptor molecules have been investigated for the intermolecular Michael reaction, such as β-nitrostyrene, α,β-unsaturated carbonyl compounds, and vinyl sulfones. The nucleophilic enamines generated from aldehydes and ketones attack Michael acceptors in a stereoselective manner depending on the chiral organocatalyst, generating Michael adducts with high enantioselectivities and yields.

List *et al.* screened (*S*)-proline for the reaction of various unmodified ketones and β-nitrostyrene (Scheme 7.47) [78]. The yields obtained were good, but the reaction gave products with very low enantioselectivities.





R = H, CO_2Me, $-S(CH_2)_3S-$

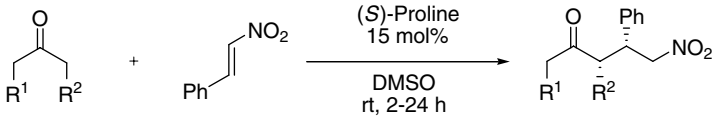


100% yield
cis:trans 89:11
99% ee

95% yield
cis:trans 91:9
97% ee

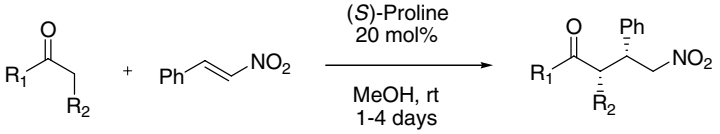
93% yield
cis:trans >95:5
99% ee

Scheme 7.46 Direct organocatalyzed Michael reaction using a cysteine-derived organocatalyst.



R^1= H, R^2= H, 97% yield, 7% ee
R^1= H, R^2= Me, 85% yield, 3:1 dr, 10% ee
R^1, R^2= $-(CH_2)_3-$, >94% yield, >20:1 dr, 23% ee

Scheme 7.47 Proline-catalyzed intermolecular Michael reaction between unmodified ketones and β-nitrostyrene.



R^1= Me, R^2= H, 93% yield, 12% ee

R^1= Et, R^2= Me, 74% yield,16:1 dr, 76% ee

R^1, R^2= -$(CH_2)_4$-, 79% yield, >20:1 dr, 57% ee

Scheme 7.48 Proline-catalyzed intermolecular Michael reaction between unmodified ketones and β-nitrostyrene in methanol.

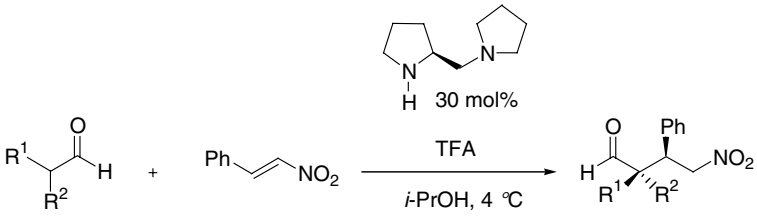
Enders and Seki found that Michael products can be furnished with higher enantioselectivities by using (*S*)-proline in methanol (Scheme 7.48) [79]. γ-Nitroketones were afforded with good to excellent yields. Enantioselectivities were found to be higher for the more substituted ketone donors, and excellent diastereoselectivities were obtained.

Furthermore, alternative reaction media such as ionic liquids have been reported for the intermolecular Michael reaction. Ionic liquids have been found to improve the enantioselectivity for the asymmetric Michael addition of ketones to β-nitrostyrene [80].

Due to the low enantioselectivities of (*S*)-proline-catalyzed intermolecular Michael additions, alternative chiral amines have been explored for their potential as organocatalysts for the intermolecular Michael addition reaction.

(*S*)-(+)-1-(2-Pyrrolidinylmethyl)pyrrolidine was found to be efficient for the direct asymmetric Michael reaction of α,α-disubstituted aldehydes with β-nitrostyrene for the synthesis of quaternary carbon-containing products (Scheme 7.49) [81]. Products were obtained with high yields and good enantioselectivities albeit with poor diastereoselectivities.

Alexakis *et al.* reported the first asymmetric organocatalyzed Michael addition of α-hydroxyketones to nitroolefins using a chiral diamine, (*S,S*)-*N*-isopropyl-2,2′-bipyrrolidine (*i*PBP) (Scheme 7.50) [82]. The reaction scope was investigated

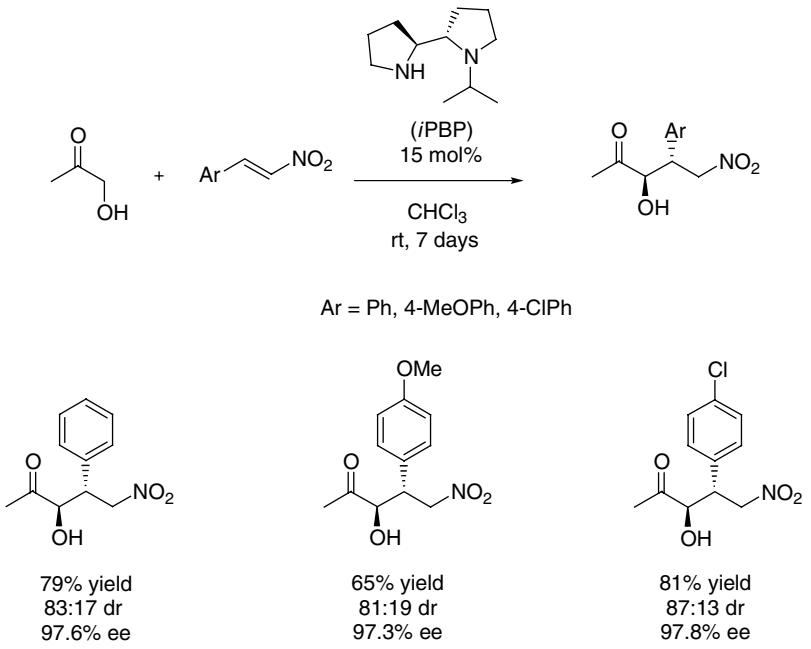


R^1= Me, R^2= Me, 87% yield, 80% ee

R^1= Et, R^2= Me, 94% yield, 3:1 dr, 81/75% ee

R^1= Pr, R^2= Me, 95% yield, 3:1 dr, 86/67% ee

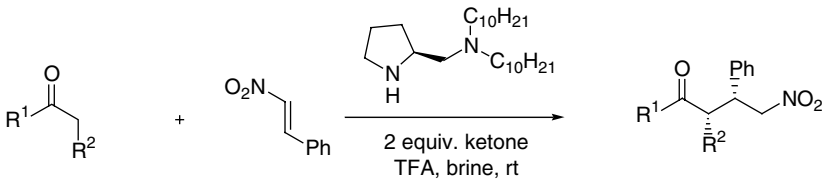
Scheme 7.49 Direct organocatalyzed Michael addition of α,α-disubstituted aldehydes to β-nitrostyrene.



Scheme 7.50 Direct organocatalyzed Michael addition of hydroxyketones with β-nitrostyrene.

and it was found that polar aprotic solvents such as chloroform were preferable. α-Hydroxyacetone was reacted with different arylnitroolefins. Extended reaction times were needed to generate products with high yields and enantioselectivities. Higher yields and stereoselectivities were obtained when an electron-withdrawing group was present on the aryl moiety.

Barbas *et al*. have reported a direct organocatalytic Michael reaction of ketones and aldehydes with β-nitrostyrene in brine using a hydrophobic organocatalyst (Scheme 7.51) [83]. In previous studies, Barbas used a hydrophobic organocatalyst



R^1= Me, R^2= H, 87% yield, 32% ee

R^1, R^2= -$(CH_2)_4$-, 93% yield, 19:1 dr, 89% ee

R^1, R^2= -$(CH_2)_3$-, 75% yield, 3:1 dr, 80% ee

R^1= H, R^2= Et, 99% yield, 2:1 dr, 38% ee

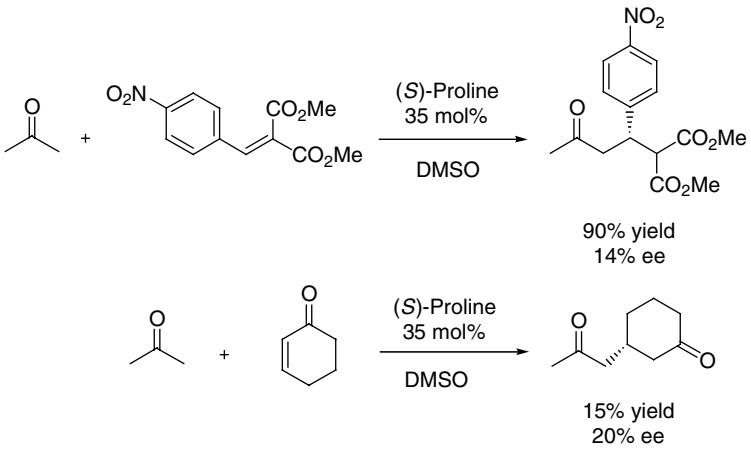
Scheme 7.51 Direct organocatalyzed Michael addition of ketones and aldehydes with β-nitrostyrene in brine.

for the direct Aldol reaction in water [84]. For the Michael reaction, this organocatalyst is employed with TFA in brine using various ketones and aldehydes as Michael donors for β-nitrostyrene. In general, an excess of ketone donor (10–20 equiv.) is employed in the Michael reaction using organic solvents. However, when performing the reaction in brine, a lower amount of ketone donor (2 equiv.) is tolerated (Scheme 7.51). Michael adducts were generated in high yields, but enantioselectivities varied for different donors. Cyclic ketones such as cyclohexanone (89% e.e.) and cyclopentanone (80% e.e.) produced adducts with the highest enantioselectivities while acetone (32% e.e.) and butanal (38% e.e.) yielded products with low enantioselectivities.

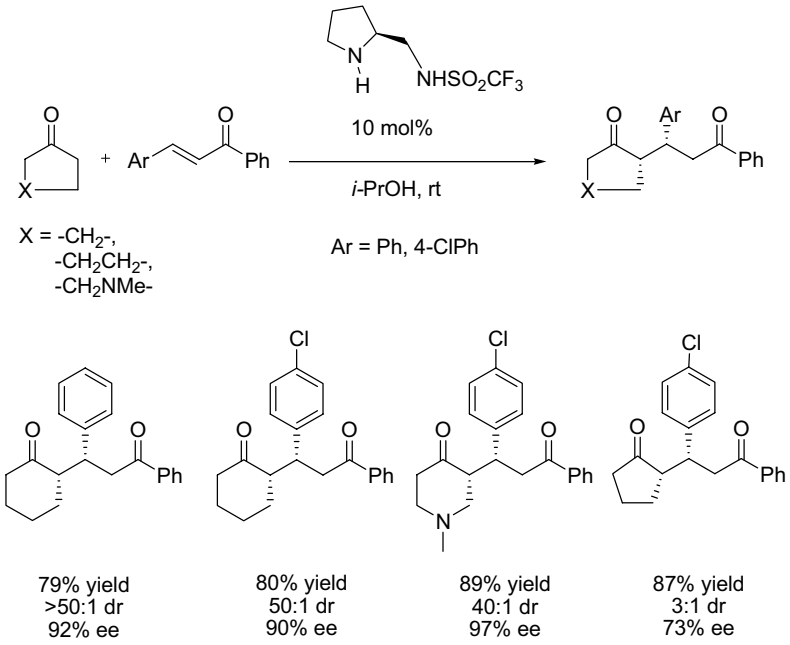
List reported a Michael reaction between acetone and cyclohexenone or aromatic alkylidene malonates as acceptors in the presence of (*S*)-proline (35 mol%) in DMSO (Scheme 7.52) [8]. Higher yields were observed using aromatic alkylidene malonates (90%) compared to cyclohexenone; however, enantioselectivities were poor.

An (*S*)-pyrrolidine sulfonamide has been reported by Wang *et al.* for the intermolecular Michael addition reaction of cyclic ketones with chalcones (Scheme 7.53) [85]. Michael adducts were formed in high yields, enantioselectivities, and diastereoselectivities. In general, six-membered rings efficiently generated Michael adducts (73–89%) with high enantioselectivities and diastereoselectivities. Cyclopentanone reacted at a faster rate with high yields, although diastereoselectivity was reduced.

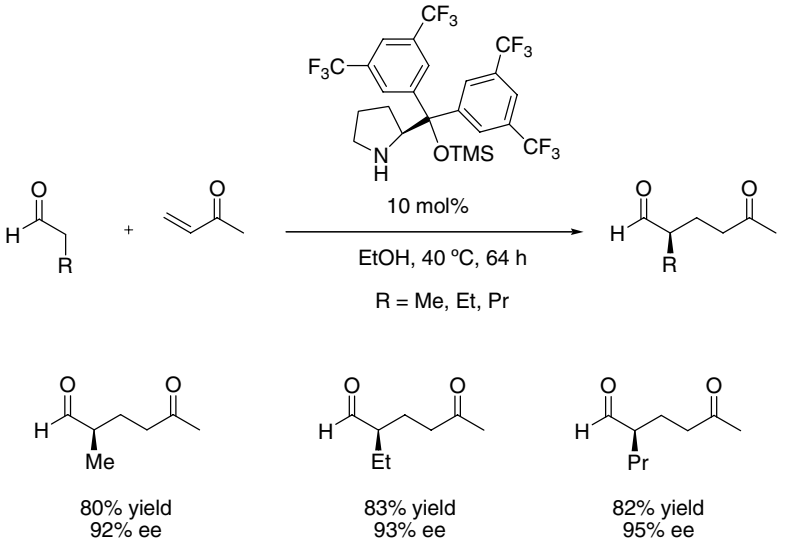
Jørgensen and his group have reported the Michael addition of aldehydes to methyl vinyl ketone using the organocatalyst bis-trifluoromethylprolinol (Scheme 7.54). Michael adducts were furnished in good yields and with high enantioselectivities [86].



Scheme 7.52 (*S*)-Proline-catalyzed Michael addition of acetone with cyclohexenone and aromatic alkylidene malonates.



Scheme 7.53 Direct organocatalyzed Michael addition of cyclic ketones with chalcones.



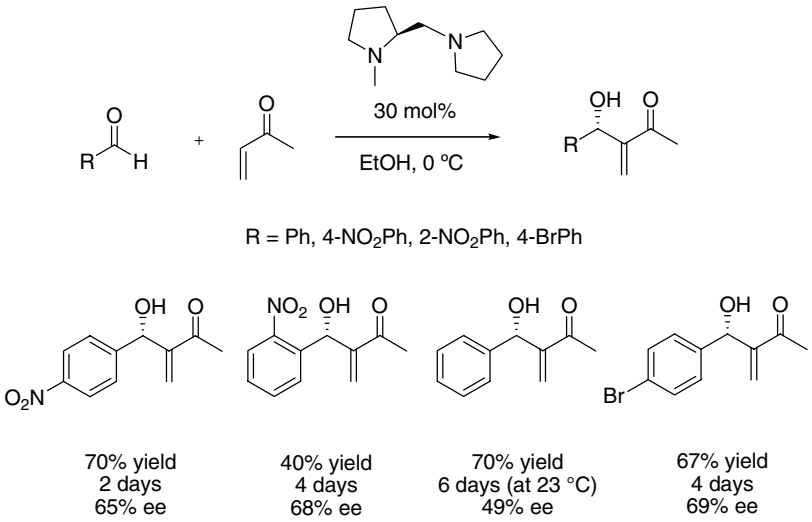
Scheme 7.54 Direct organocatalyzed Michael addition of unmodified aldehydes to methyl vinyl ketone.

7.6
Morita–Baylis–Hillman Reaction and Its Aza-Counterpart

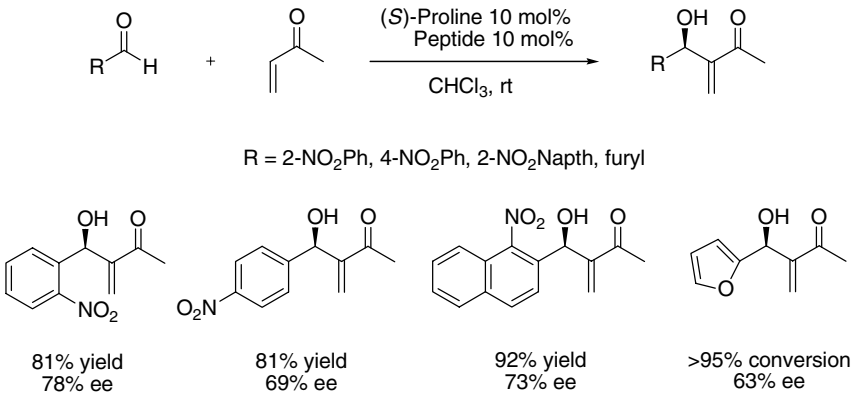
7.6.1
Morita–Baylis–Hillman Reactions

The Morita–Baylis–Hillman (MBH) reaction involves the addition of electrophilic alkenes to carbonyl electrophiles using nucleophiles such as tertiary amines or tertiary phosphines as an organic catalyst [87]. Enantiomerically enriched β-hydroxy-α-methylene carbonyl compounds are useful building blocks for natural product synthesis [88]. The use of amino acids to promote the MBH reaction has been investigated. However, amino acids cannot catalyze the MBH reaction on their own. Amino acids such as (*S*)-proline have been employed as cocatalysts for MBH reactions, where a nucleophile such as imidazole or DABCO is used. A chiral diamine (*S*)-1-methyl-2-(pyrrolidinylmethyl)pyrrolidine has been found to be effective for MBH reactions using various aldehydes and methyl vinyl ketone as substrates (Scheme 7.55) [89]. Ethanol was the optimal solvent, and the MBH adducts are formed with moderate yields and enantioselectivities. Methyl vinyl ketone was used in excess and reaction times ranged from 2 to 6 days.

The application of (*S*)-proline as a cocatalyst with imidazole as a nucleophilic source was reported by Shi *et al.* [90] to produce MBH adducts in high yields. Miller *et al.* extended the idea of a dual catalyst system by developing peptides that act as nucleophiles with the aid of (*S*)-proline to yield MBH products in good yields and higher enantioselectivities (Scheme 7.56) [91]. Methyl vinyl ketone and various aldehydes were used as a model system when designing their peptides. Heptapeptides and octapeptides were found to increase the enantioselectivity of the reaction. The



Scheme 7.55 Direct organocatalyzed Morita–Baylis–Hillman reaction between unmodified aldehydes and methyl vinyl ketone.



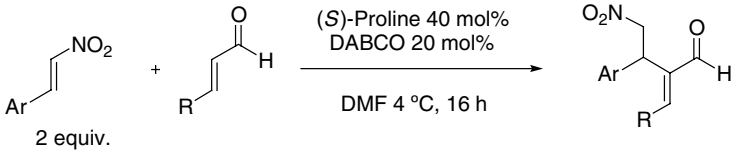
Scheme 7.56 Direct organocatalyzed Morita–Baylis–Hillman reaction between unmodified aldehydes and methyl vinyl ketone.

MBH products were obtained with good yields and enantioselectivities. Aliphatic, nonactivated aromatic, and branched aliphatic aldehydes were also screened as substrates for the MBH reaction with methyl vinyl ketone, but no reaction occurred.

Cordova *et al.* have extended the MBH reaction to electrophilic β-nitrostyrene acceptors for the generation of MBH adducts (Scheme 7.57) [92]. They used a cocatalyst system containing 1,4-diazabicyclo[2.2.2]octane (DABCO) as nucleophile and (*S*)-proline to catalyze the reaction of different nitrostyrenes with various α,β-unsaturated aldehydes. The MBH adducts were prepared with moderate yields, excellent *E*-selectivity, but poor enantioselectivity.

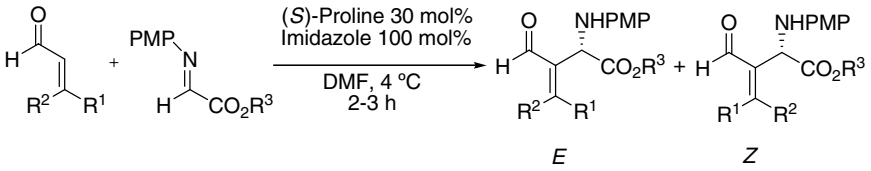
7.6.2
Aza-Morita–Baylis–Hillman Reactions

β-Amino carbonyl compounds bearing a α-alkylidene group can be made using the aza-Morita–Baylis–Hillman (aza-MBH). A catalyst system using (*S*)-proline and imidazole has been reported by Barbas *et al.* to be effective for the formation of aza-MBH adducts using various β-substituted α,β-unsaturated aldehydes and α-imino



Ar = Ph, R = Et, 57% yield, >25:1 *E*/*Z*, <10% ee
Ar = Ph, R = Bu, 62% yield, >25:1 *E*/*Z*, <10% ee
Ar = 2-BrPh, R = Et, 41% yield, >25:1 *E*/*Z*, <10% ee
Ar = 3-ClPh, R = Et, 55% yield, >25:1 *E*/*Z*, <10% ee

Scheme 7.57 (*S*)-Proline-catalyzed Morita–Baylis–Hillman reaction between nitrostyrene and unsaturated aldehydes.



R^1= Me, R^2= H, R^3= *i*-Pr, 68% yield, 16:1 *E*/*Z*, 99% ee

R^1= Et, R^2= H, R^3= Et, 58% yield, 10:1 *E*/*Z*, 99% ee

R^1= Pr, R^2= H, R^3= Et, 50% yield, 8:1 *E*/*Z*, 97% ee

R^1= Me, R^2= Me, R^3= Et, 39% yield, 92% ee

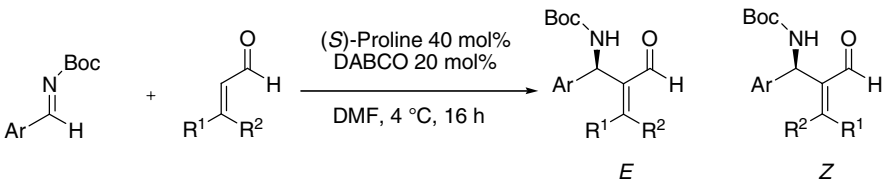
Scheme 7.58 A catalyzed aza-Morita–Baylis–Hillman reaction using β-substituted α,β-unsaturated aldehydes and α-imino esters protected with PMP group.

esters protected with a PMP group as substrates (Scheme 7.58) [93]. Aza-MBH adducts were isolated with moderate yields and excellent enantioselectivities. Nonenolizable aldehydes did not furnish the aza-MBH adducts.

Cordova *et al*. have reported the use of Boc-imines as acceptors for the aza-MBH reaction using (*S*)-proline and DABCO as cocatalyst system (Scheme 7.59) [94]. They reacted aryl *N*-Boc-imines with unmodified α,β-unsaturated aldehydes. Aza-MBH adducts were prepared with moderate yields and excellent enantioselectivities.

7.7
Miscellaneous Amino Acid-Catalyzed Reactions

In addition to Aldol, Mannich, Michael, and Morita–Baylis–Hillman reactions, amino acids have also been proved to be versatile catalysts or cocatalysts for other important chemical transformations such as Diels–Alder, Knoevenagel, and oxidation reactions.



R^1 = Me, R^2 = H, Ar = Ph, 45% yield, 4:1 *E*/*Z*, 99% ee

R^1 = Et, R^2 = H, Ar = Ph, 51% yield, 9:1 *E*/*Z*, 99% ee

R^1 = Et, R^2= H, Ar = 4-ClPh, 53% yield, 9:1 *E*/*Z*, 99% ee

R^1 = Et, R^2 = H, Ar = 4-MeOPh, 56% yield, 9:1 *E*/*Z*, 99% ee

Scheme 7.59 A dual catalyzed aza-Morita–Baylis–Hillman reaction using aryl *N*-Boc-imines and unmodified unsaturated aldehydes.

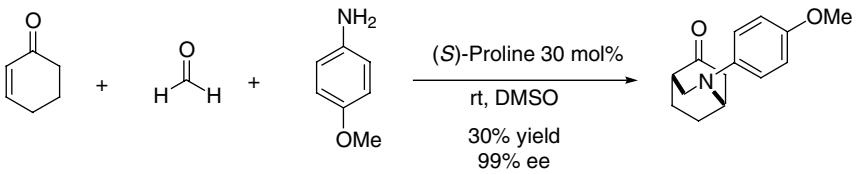
7.7.1
Diels–Alder Reaction

Early research on the Diels–Alder reaction catalyzed by amino acids by Asato *et al.* [95] reported that, in the presence of (*S*)-proline in ethanol, cyclic optically active products can be obtained through the dimerization of α,β-unsaturated aldehydes. The formed enamine and/or iminium ion between proline and the aldehyde may undergo a Diels–Alder reaction to give the corresponding products. Recently, Cordova *et al.* reported the first one-pot three-component direct catalytic enantioselective aza-Diels–Alder reaction catalyzed by (*S*)-proline with excellent chemo-, regio-, and stereoselectivity (Scheme 7.60) [96]. For example, the amino acid-catalyzed *trans*-formation of formaldehyde, α,β-unsaturated cyclic ketones, and aromatic amines furnished the desired azabicyclic ketones with up to >99% e.e. and satisfactory yields in most cases. On the basis of the absolute stereochemistry of the aza-Diels–Alder adducts, the authors proposed a tandem one-pot three-component Mannich/Michael reaction pathway, which first involved the formation of chiral enamine between the catalyst and α,β-unsaturated ketone (Scheme 7.61). Next, the *in situ* generated imine attacks the *si*-face of the chiral diene via transition state I, and an activated iminium salt is formed. The secondary amine of the chiral iminium salt performs a subsequent selective 6-*endo*-trig cyclization to furnish the corresponding chiral azabicycle. Finally, the amino acid derivative is released and the desired aza-Diels–Alder adducts are obtained by hydrolysis.

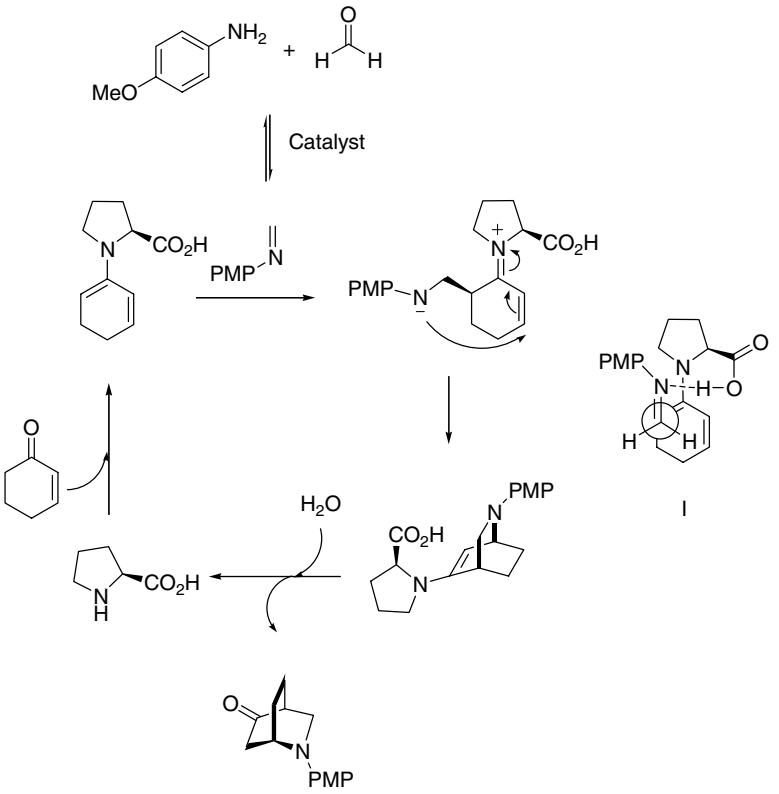
In 2006, Aznar *et al.* reported a proline-catalyzed imino-Diels–Alder reaction to construct 4-piperidine ring systems (Scheme 7.62) [97]. They found that the reaction has broad scope and the *meso*-2,6-diaryl-4-piperidones were obtained with excellent diastereoselectivities and good yields independent of the aryl substituents. This (*S*)-proline-catalyzed cycloaddition methodology has the advantage that neither the use of a metal catalyst nor the restriction of having to isolate the diene intermediate prior to the cycloaddition is required, and the reaction can easily be run on a multigram scale. Formation of an iminium zwitterion is postulated during the proposed reaction pathway (Scheme 7.62).

7.7.2
Knoevenagel Condensation

Knoevenagel condensation of carbonyl compounds on molecules containing an active methylene group is one of the most important methods for the preparation of



Scheme 7.60 One-pot three-component asymmetric aza-Diels–Alder reaction.



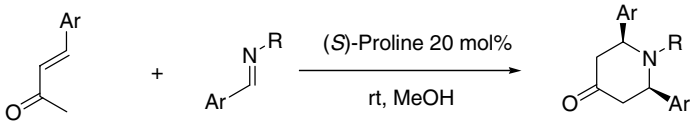
Scheme 7.61 The plausible reaction pathway and transition-state one-pot three-component asymmetric aza-Diels–Alder reaction.

substituted alkenes. Cardillo *et al.* reported a straightforward method for the synthesis of alkylinene/arylinene malonates and arylidene cyanoacetates, utilizing proline as an alternative to traditional catalysts (Scheme 7.63) [98].

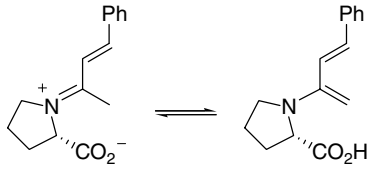
The generated α,β-unsaturated ester can act as one of the three components in a Knoevenagel/Diels–Alder sequence. This domino reaction affords highly substituted spiro[5,5]undecane-1,5,9-triones with high enantio- and diastereoselectivity (Scheme 7.64) [99].

7.7.3
Reduction and Oxidation

Several reduction reactions involving proline have been developed, such as hydrogenation of α,β-unsaturated ketones and ketones [100, 101] and reduction of racemic epoxides [102–104] to give enantioenriched alcohols. A suitable metal cocatalyst and reducing reagents are usually involved, and proline plays a role as the source of asymmetric induction.



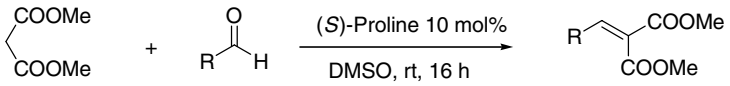
Ar = Ph, R = allyl, de >93%, 77% yield
Ar = Ph, R = homoallyl, de >97%, 59% yield
Ar = *p*-F-Ph, R = allyl, de >95%, 66% yield
Ar = *p*-Cl-Ph, R = allyl, de >97%, 64% yield
Ar = *o*-Br-Ph, R = allyl, de >96%, 66% yield
Ar = *o*-I-Ph, R = allyl, de >96%, 67% yield



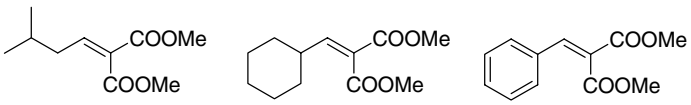
Iminium zwitterion

Scheme 7.62 Preparation of *meso*-2,6-disubstituted-4-piperidones.

Oxidation reactions assisted by amino acids have also been investigated. Levina and Muzart reported the synthesis of optically active allylic esters with enantiomeric excesses up to 54% by acyloxylation of the alkenes with *t*-BuOOC(O)Ph in the presence of catalytic amounts of copper compounds and chiral amino acids [105]. Asymmetric oxidations with singlet molecular oxygen have previously been considered to be in the domains of enzymes and chiral transition metal complexes. In 2004, Cordova *et al.* reported a direct amino acid-catalyzed asymmetric α-oxidation of aldehydes and ketones with molecular oxygen affording α-hydroxy aldehydes and

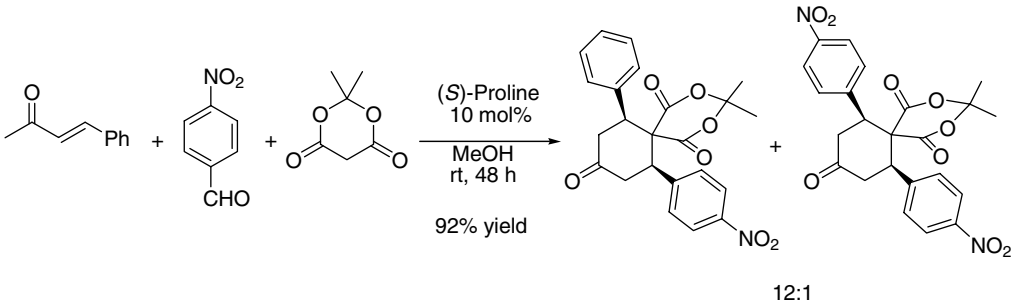


R = *i*-Bu, Cy, Ph



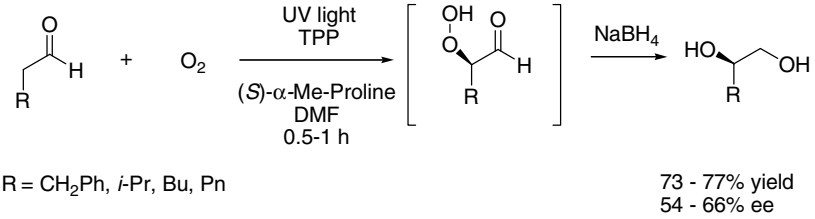
92% yield 97% yield 100% yield

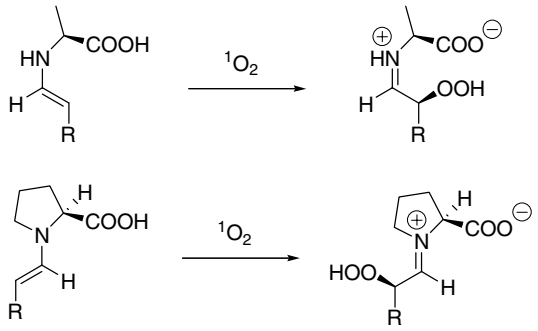
Scheme 7.63 Proline-catalyzed Knoevenagel condensation between dimethyl malonate and aldehydes.



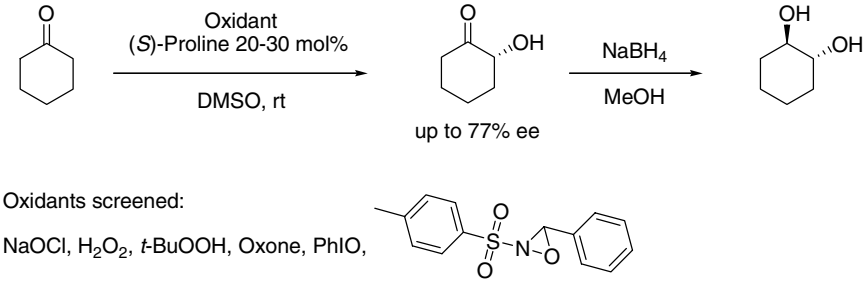
Scheme 7.64 Proline-catalyzed asymmetric three-component domino Knoevenagel/Diels–Alder reaction.

terminal diols [106] and α-hydroxy ketones [107]. They suggest that this may warrant investigation for a prebiotic entry for the creation of α-hydroxy aldehydes. The catalyst screening study revealed that a methyl group at the α-position of proline significantly increased the stereoselectivity of the terminal diol synthesis (Scheme 7.65) [106]. Furthermore, they found that the reactions with molecular $^3O_2$ did not provide the diols. Thus, molecular $^1O_2$ was the fastest reacting electrophile and not $^3O_2$. Moreover, no diol product was formed without addition of the amino acid catalyst. They proposed that a plausible reaction proceeded via a catalytic enamine mechanism. Hence, the (2*R*)-α-hydrogen peroxide aldehyde intermediate was formed via





Scheme 7.65 Direct catalytic asymmetric incorporation of $^1O_2$ to aldehydes and plausible mechanism. TPP: tetraphenylporphyrin.



Scheme 7.66 Direct organocatalytic α-oxidation of ketones.

molecular $^1O_2$ proton abstraction of the (*S*)-amino acid carboxyl group, which induced the stereochemical information, together with addition to the *si*-face of the amino acid-derived enamine (Scheme 7.65).

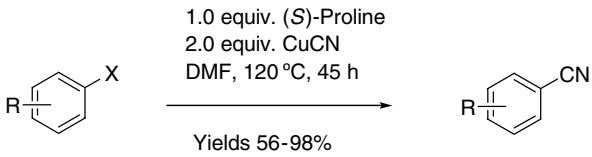
Cordova *et al.* [108] reported the direct organocatalytic α-oxidation of ketones yielding α-hydroxylated ketones with up to 77% e.e., using a range of oxidants, and then reduction to chiral diols (Scheme 7.66).

7.7.4
Rosenmund–von Braun Reaction

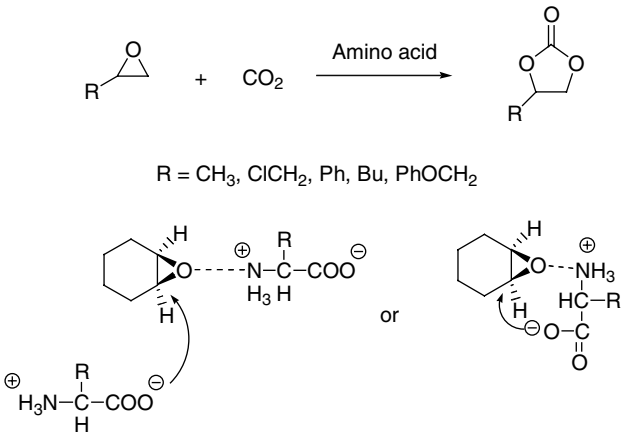
Aryl nitriles are important intermediates for the preparation of dyes, pesticides, agrochemicals, and bioactive compounds. Ding *et al.* reported that (*S*)-proline is an effective additive to promote the Rosenmund–von Braun cyanation of aryl bromides and aryl iodides with CuCN at relatively lower temperatures. The results show the modified process exhibits excellent functional group compatibility with over 80% yields in most cases (Scheme 7.67) [109]. *N*-Methylglycine, *N*,*N*-dimethylglycine, and (*S*)-tryptophan were also screened as additives; however, no effect on the Rosenmund–von Braun reaction was observed.

7.7.5
Activation of Epoxides

Chemical fixation of carbon dioxide into valuable products through coupling reaction of carbon dioxide with epoxides has received great interest from synthetic and



Scheme 7.67 (*S*)-Proline-promoted CuCN-catalyzed cyanation of aryl bromides and iodides.

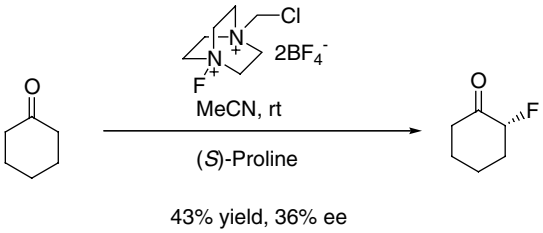


Scheme 7.68 Coupling of epoxides with CO_2 catalyzed by amino acids.

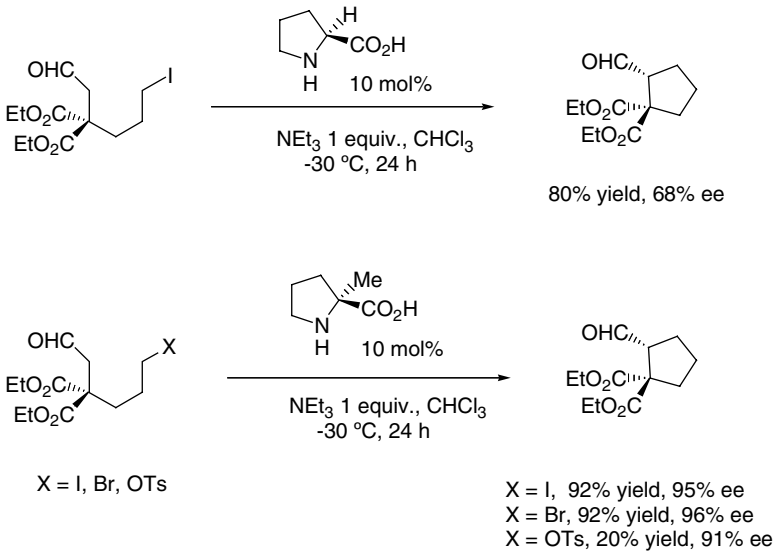
sustainable chemistry research groups. Many catalyst systems for this transformation have been developed and are based on a bifunctional concept. Taking into account the special bifunctional structure of amino acids, Jiang *et al.* reported that naturally occurring α-amino acids were a class of efficient catalysts for coupling carbon dioxide with epoxides to give the corresponding cyclic carbonates with high selectivity [110]. They proposed that the cooperation of the ammonium group of one amino acid and the carboxylate ion of another played a key activation factor for this transformation. The simple catalytic system, low catalyst loading, and the natural properties of the catalyst offer significant "greenness" for this process (Scheme 7.68).

7.7.6
α-Fluorination of Aldehydes and Ketones

There has been a growing interest in the selective introduction of one or more fluorine atoms to organic molecules due to the unique properties of fluorinated compounds. Enders and Huttl first reported the organocatalytic direct α-fluorination of aldehydes and ketones employing Selectfluor as the fluorine source (Scheme 7.69) [111]. Although corresponding α-fluoroaldehydes and α-fluoroketones can be prepared in moderate to good yields (47–78%), low asymmetric inductions (36% e.e.) have been observed.



Scheme 7.69 α-Fluorination of cyclohexanone.



Scheme 7.70 α-Alkylation of carbonyl compounds.

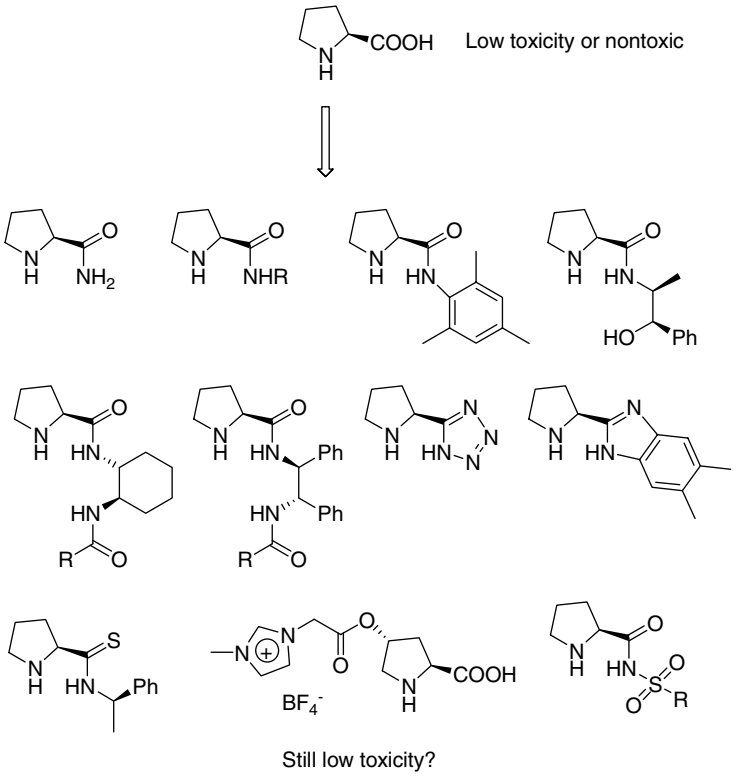
7.7.7
S_N2 Alkylation

α-Alkylation of carbonyl compounds consists of fundamental carbon–carbon σ-bond-forming reactions in organic synthesis. In 2004, List and Vignola reported the catalytic asymmetric intramolecular α-alkylation of aldehydes catalyzed by (*S*)-proline and (*S*)-α-methylproline derivative [112]. Higher reaction rates and higher enantioselectivities were observed with (*S*)-α-methylproline (Scheme 7.70). To explore the scope of this reaction, various leaving groups (such as -I, -Br, and -OTs) were investigated, with halides affording increased enantioselectivities and yields.

7.8
Sustainability of Amino Acid Catalysis

7.8.1
Toxicity and Ecotoxicity of Amino Acid Catalysis

Organocatalysis has been the subject of considerable interest as a synthetic methodology alongside metal-catalyzed chemical synthesis. In many publications, one of the prominent advantages of organocatalysts over several metal-catalyzed processes is their low toxicity and absence of heavy metal ions contaminating final products or accumulating in the environment. However, there has not been a detailed study of structure–toxicity relationship of organocatalysts. For naturally occurring (*S*)-proline, a low toxicity and environmental impact would be expected; however, the development



Scheme 7.71 Proline and proline derivatives.

of closely related amino acid derivatives raises questions over the toxicity of novel compounds (Scheme 7.71) [6]. The potential toxicity of organocatalysts to humans and other organisms is significant, while ecotoxicity has the potential to limit biodegradation. There is need to investigate the structure–toxicity relationship of organocatalysts. This will assist the design of sustainable, efficient catalysts for synthesis and the pharmaceutical industry. In medicinal chemistry, the field of peptidomimetics investigates the effects modified amino acids and peptides have on biological activity. Lessons learnt there may help to avoid designing novel toxic amino acid organocatalysts.

7.8.2
Amino Acid Catalysis and Green Chemistry

Green chemistry aims to develop and design chemical products and processes that reduce or eliminate the use and generation of hazardous substances [113, 114]. The catalysis of reactions by amino acids provides advantages including their lack of sensitivity to moisture and oxygen, their ready availability, low cost, catalyzing atom economic transformations, which confers considerable direct benefit on the

production of pharmaceutical intermediates compared to many (transition) metal catalysts. Initially, to achieve satisfactory yields and stereoselectivities, polar solvents with high dielectric constants, such as DMSO and DMF, were often used as solvents for reactions catalyzed by amino acids. Recently, increased attention on the sustainability of chemical industry has motivated the use of environment-friendly reaction media, such as ionic liquids [115–123], water [124, 125], and polyethyleneglycol [126–128] for amino acid-catalyzed reactions. To facilitate their recycling and reuse, immobilized amino acid-catalyzed chemical transformations have been investigated ([129, 130] and references therein), although the heterogeneous immobilization catalysis still encountered problems, with lower catalytic activities and enantioselectivity. However, recent amino acid/ionic liquid homogeneous immobilization technology has made great progress in those transformations and paved the way for a more sustainable process of amino acid catalysis [131–134].

7.9 Conclusions and Expectations

Enantioselective organocatalysis, as a new asymmetric catalytic methodology, has had a major impact on organic synthesis over the past 10 years and has bridged the fields of stereochemistry and pharmaceutically active molecules. The development in organocatalysis initiated new methodologies for synthetic chemistry and has provided new strategies to construct bioactive molecules. Amino acids, as versatile catalysts, have been one of the most important classes of organocatalysis, and it is anticipated that the scale of amino acid catalysis will expand rapidly and future amino acid derivative catalysis will become a core asymmetric synthetic methodology. In addition, compared to several transition metal-catalyzed processes of chemical synthesis, the most prominent advantages of proline organocatalysis include low toxicity and lack of sensitivity to moisture and oxygen. Although amino acids, as basic units of protein, have low toxicity, there is no detailed study of structure–toxicity relationships of amino acid derivatives. This data will enable a better understanding and stimulate design of sustainable catalysts for the pharmaceutical industry. In summary, enantioselective organocatalysis is an essential tool for the synthetic chemist.

7.10 Typical Procedures for Preferred Catalysis of Reactions by Amino Acids

Direct catalytic asymmetric eno*lexo* aldolizations (Scheme 7.5) Typical aldolization procedure: Dicarbonyl (1 mmol) was dissolved in dry dichloromethane (10 ml) and treated with (*S*)- or (*R*)-proline (12 mg, 0.1 mmol, 10 mol%). The mixture was stirred at room temperature until the starting material had disappeared (8–16 h). Designed Aldol product can be isolated after standard aqueous workup, but is unstable over extended periods at room temperature.

Proline-catalyzed direct asymmetric Aldol reactions (Scheme 7.8) The amino acid catalyst (0.03–0.04 mmol) was stirred in 1 ml of DMSO/acetone (4 : 1) for 15 min. 4-Nitrobenzaldehyde (0.1 mmol) was added and the mixture was stirred for 4–24 h. The mixture was treated with saturated aqueous ammonium chloride solution (1 ml) and extracted with EtOAc. The organic layer was dried ($MgSO_4$), filtered, and concentrated to give after column chromatography (hexanes/EtOAc (3 : 1)) pure designed Aldol product.

Acyclic amino acid-catalyzed direct asymmetric Aldol reactions (alanine) (Scheme 7.10) A catalytic amount of (*S*)-amino acid (0.15 mmol, 30 mol%) was added to a vial containing acceptor aldehyde (0.5 mmol), donor ketone (1.5 mmol), and H_2O (5 mmol, 90 µl) in DMSO (2 ml). After 3–4 days of vigorously stirring at room temperature, the reaction mixture was poured into an extraction funnel that contained brine (5.0 ml), which was diluted with distilled H_2O (5.0 ml) and EtOAc (15 ml). The reaction vial was also washed with EtOAc (2 ml), which was poured into the extraction funnel. The aqueous phase was extracted with EtOAc (2×15 ml). The combined organic phases were dried with Na_2SO_4 and the solvent removed under reduced pressure. The reaction can also be quenched by directly putting the reaction mixture on a silica-gel column. The crude Aldol product was purified by silica-gel column chromatography (EtOAc : pentane mixtures) to furnish the desired Aldol product.

Direct amino acid-catalyzed asymmetric synthesis of sugars (Scheme 7.18) A suspension of propionaldehyde (144 µl, 4.0 mmol) in DMF (2.5 ml) was added slowly over 16–24 h to a stirring suspension of the corresponding aldehyde (2 mmol) and (*S*)-proline [or (*R*)-proline] (11.5 mg, 0.10 mmol) in DMF at 4 °C. After 16 h, the resulting solution was diluted with diethyl ether and washed successively with water and brine. The combined aqueous layers were back-extracted with three portions of EtOAc. The combined organic layers were dried over anhydrous $MgSO_4$, which was subsequently removed by filtration. Next, the solvent was removed under reduced pressure following purification of the crude product mixture by silica-gel column chromatography (EtOAc : pentane mixtures) to afford the corresponding Aldol adduct. This product was dissolved in DMF (1 ml), with 10 mol% of (*R*)-proline [or (*S*)-amino acid]. Then, a suspension of propionaldehyde (2 equiv.) in DMF (2 ml) was added slowly over the course of 16 h to the reaction mixture at 4 °C. Next, the solution was allowed to react at room temperature and stirred for 24 h. The reaction was quenched by extraction. The combined aqueous layers were back-extracted with three portions of EtOAc. The combined organic layers were dried over anhydrous $MgSO_4$, which was subsequently removed by filtration. Finally, the solvent was removed under reduced pressure following purification of the crude product mixture by silica-gel column chromatography (EtOAc : pentane mixtures) to afford the desired compounds.

Direct Mannich reaction with acetone as nucleophile catalyzed by (S)-proline (Scheme 7.22) A suspension of (*R*)- or (*S*)-proline (40 mg, 0.35 mmol), *p*-anisidine (135 mg, 1.1 mmol), and an aldehyde (1 mmol) in DMSO/acetone (4 : 1, 10 ml) was stirred at

room temperature for 12–48 h. Reaction was worked up by adding phosphate-buffered saline (PBS) solution (pH 7.4), extracting with ethyl acetate, drying the organic layer with $MgSO_4$, and purifying the crude product by column chromatography with hexane/ethyl acetate mixtures.

anti-Selective Mannich reaction catalyzed by acyclic amino acids (Scheme 7.32) General procedure for three-component Mannich reactions using (*S*)-tryptophan as catalyst. A mixture of DMF (2 ml), *p*-anisidine (0.45 mmol, 1 equiv.), aldehyde (0.5 mmol, 1.1 equiv.), hydroxyacetone (5 mmol, 11 equiv.), and (*S*)-tryptophan (0.1 mmol, 0.2 equiv.) was vigorously stirred at 4 °C for 16–20 h. Then, the mixture was diluted with AcOEt and half-saturated ammonium chloride solution was added. The mixture was extracted with AcOEt (four times). The combined organic layers were washed with brine, dried over $MgSO_4$, concentrated *in vacuo*, and purified by flash column chromatography (hexanes/ethyl acetate) to afford the desired Mannich addition product.

General procedure for three-component Mannich reactions using O-t-Bu-L-Thr (Scheme 7.32) A mixture of 1-methyl-2-pyrrolidinone (NMP, 1 ml), *p*-anisidine (0.45 mmol, 1 equiv.), aldehyde (0.5 mmol, 1.1 equiv.), hydroxyacetone (2.5 mmol, 5.5 equiv.), and *O*-*t*-Bu-L-Thr (0.1 mmol, 0.2 equiv.) was vigorously stirred at 4 °C for 48 h. Then, the mixture was diluted with AcOEt and half-saturated ammonium chloride solution was added. The mixture was extracted with AcOEt (four times). The combined organic layers were washed with brine, dried over $MgSO_4$, concentrated *in vacuo*, and purified by flash column chromatography (hexanes/ethyl acetate) to afford the desired Mannich addition product.

Direct catalytic asymmetric α-amination of aldehydes using (S)-proline in an ionic liquid (Scheme 7.38) (*S*)-Proline (5 mol%) and the aldehyde (1.1 equiv.) were added to the degassed ionic liquid (1 ml). After stirring the resulting mixture for 15 min at room temperature, DEAD (1 mmol, 40% solution in toluene) was added and vigorous stirring was continued. The product was isolated by extraction with several portions of diethyl ether followed by solvent evaporation using a rotary evaporator. Crude product was dissolved in MeOH (3 ml). The mixture was cooled to 3 °C and $NaBH_4$ (1.1 equiv.) was added. After stirring for 15 min at room temperature, a 0.5 M solution of NaOH was added and stirring was continued for 25 min before water (7 ml) was added. The mixture was extracted with EtOAc (3 × 5 ml), and the combined organic layers were dried with Na_2SO_4. After evaporating the solvent *in vacuo*, the crude product was purified by column chromatography on SiO_2 with hexane/EtOAc (4 : 1) or hexane/dichloromethane (2 : 1) as eluent.

Direct organocatalyzed Michael addition of unmodified aldehydes to methyl vinyl ketone (Scheme 7.54) Methyl vinyl ketone (82 µl, 1.0 mmol) was added to a solution of aldehyde (2.0 mmol) and catalyst (60 mg, 0.1 mmol) in EtOH (0.5 ml) and stirred at 40 °C for 64 h. The crude reaction mixture was purified by flash chromatography, without prior evaporation of the solvent, to give 2-substituted 5-oxo-hexanal.

Direct organocatalyzed BMH reaction between unmodified aldehydes and methyl vinyl ketone (Scheme 7.55) To an ethanol solution (0.45 ml) of aldehyde (0.5 mmol) and methyl vinyl ketone (2.5 mmol) was added (*S*)-1-methyl-2-(1-pyrrolidinylmethyl) pyrrolidine (25.2 mg, 0.15 mmol) at 0 °C, and the reaction mixture was stirred for 2 days at this temperature. After removing the volatile materials under reduced pressure, the residue was purified by thin-layer chromatography to afford the BMH adducts.

Catalytic asymmetric intramolecular α-alkylation of aldehydes (Scheme 7.70) A solution of the corresponding aldehyde (0.5 mmol) in dry chloroform (2.5 ml) was added to the catalyst (*S*)- or (*R*)-α-methylproline (5–20 mol%) and the mixture was cooled to −30 °C. After the addition of triethylamine (0.5 mmol, 70 μl), the mixture was stirred for 24–48 h. The mixture was filtered and concentrated. The residue was purified on a silica-gel column (20% EtOAc/hexane) to give the pure aldehyde cyclized product.

References

1 Hawkins, J.M. and Watson, T.J.N. (2004) *Angewandte Chemie – International Edition*, **43**, 3224–3228.

2 Halpern, J. and Trost, B.M. (2004) *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 5347.

3 Indolese, A.F., Wasescha, D., Eisele, T. and Bucher, C. (2005) *Chimia*, **59**, 22–24.

4 Pollard, D.J. and Woodley, J.M. (2007) *Trends in Biotechnology*, **25**, 66–73.

5 (a) Dalko, P.I. and Moisan, L. (2004) *Angewandte Chemie – International Edition*, **43**, 5138–5175; (b) Gaunt, M.J., Johansson, C.C.C., McNally, A. and Vo, N.T. (2007) *Drug Discovery Today*, **12**, 8–27.

6 Pellissier, H. (2007) *Tetrahedron*, **63**, 9267–9331.

7 Steinhagen, H. and Helmchen, G. (1996) *Angewandte Chemie – International Edition in English*, **35**, 2339–2342.

8 List, B. (2001) *Synlett*, 1675–1686.

9 List, B. (2002) *Tetrahedron*, **58**, 5573–5590.

10 Kazmaier, U. (2005) *Angewandte Chemie – International Edition*, **44**, 2186–2188.

11 Guillena, G., Najera, C. and Ramon, D.J. (2007) *Tetrahedron – Asymmetry*, **18**, 2249–2293.

12 Mahrwald, R. (ed.) (2004) *Modern Aldol Reactions*, 2 Vols, Wiley-VCH Verlag GmbH, Weinheim, pp. 1218–1223.

13 Hajos, Z.G. and Parrish, D.R. (1974) *The Journal of Organic Chemistry*, **39**, 1615–1621.

14 Eder, U., Sauer, G. and Wiechert, R. (1971) *Angewandte Chemie – International Edition in English*, **10**, 496–497.

15 Bui, T. and Barbas, C.F. (2000) *Tetrahedron Letters*, **41**, 6951–6954.

16 Gryko, D. (2005) *Tetrahedron – Asymmetry*, **16**, 1377–1383.

17 Pidathala, C., Hoang, L., Vignola, N. and List, B. (2003) *Angewandte Chemie – International Edition*, **42**, 2785–2788.

18 Enders, D., Niemeier, O. and Straver, L. (2006) *Synlett*, 3399–3402.

19 Jung, M.E. (1976) *Tetrahedron*, **32**, 3–31.

20 Agami, C., Meynier, F., Puchot, C., Guilhem, J. and Pascard, C. (1984) *Tetrahedron*, **40**, 1031–1038.

21 Agami, C., Puchot, C. and Sevestre, H. (1986) *Tetrahedron Letters*, **27**, 1501–1504.

22 Agami, C. (1988) *Bulletin de la Societe Chimique de France*, 499–507.

23 Molines, H. and Wakselman, C. (1976) *Tetrahedron*, **32**, 2099–2103.

24 Clemente, F.R. and Houk, K.N. (2005) *Journal of the American Chemical Society*, **127**, 11294–11302.

25 List, B., Lerner, R.A. and Barbas, C.F. (2000) *Journal of the American Chemical Society*, **122**, 2395–2396.

26 Cordova, A., Zou, W., Ibrahem, I., Reyes, E., Engqvist, M. and Liao, W.-W. (2005) *Chemical Communications*, 3586–3588.

27 Bassan, A., Zou, W., Reyes, E., Himo, F. and Cordova, A. (2005) *Angewandte Chemie – International Edition*, **44**, 7028–7032.

28 Qiu, L.-H., Shen, Z.-X., Shi, Q.-C., Liu, Y.-H. and Zhang, Y.-W. (2005) *Chinese Journal of Chemistry*, **23**, 584–588.

29 Wang, Y.-J., Shen, Z.-X., Li, B. and Zhang, Y.-W. (2006) *Chinese Journal of Chemistry*, **24**, 1196–1199.

30 Samanta, S. and Zhao, C.-G. (2006) *Tetrahedron Letters*, **47**, 3383–3386.

31 Samanta, S. and Zhao, C.-G. (2006) *Journal of the American Chemical Society*, **128**, 7442–7443.

32 Northrup, A.B. and MacMillan, D.W.C. (2002) *Journal of the American Chemical Society*, **124**, 6798–6799.

33 Storer, R.I. and MacMillan, D.W.C. (2004) *Tetrahedron*, **60**, 7705–7714.

34 Casas, J., Sunden, H. and Cordova, A. (2004) *Tetrahedron Letters*, **45**, 6117–6119.

35 Northrup, A.B., Mangion, I.K., Hettche, F. and MacMillan, D.W.C. (2004) *Angewandte Chemie – International Edition*, **43**, 2152–2154.

36 Thayumanavan, R., Tanaka, F. and Barbas, C.F. (2004) *Organic Letters*, **6**, 3541–3544.

37 Bogevig, A., Kumaragurubaran, N. and Jørgensen, K.A. (2002) *Chemical Communications*, 620–621.

38 Bahmanyar, S., Houk, K.N., Martin, H.J. and List, B. (2003) *Journal of the American Chemical Society*, **125**, 2475–2479.

39 Ogawa, T. (1994) *Chemical Society Reviews*, **23**, 397–407.

40 Danishefsky, S.J. and Bilodeau, M.T. (1996) *Angewandte Chemie – International Edition in English*, **35**, 1380–1419.

41 Hudlicky, T., Entwistle, D.A., Pitzer, K.K. and Thorpe, A.J. (1996) *Chemical Reviews*, **96**, 1195–1220.

42 Notz, W. and List, B. (2000) *Journal of the American Chemical Society*, **122**, 7386–7387.

43 Northrup, A.B. and MacMillan, D.W.C. (2004) *Science*, **305**, 1752–1755.

44 Enders, D. and Grondal, C. (2005) *Angewandte Chemie – International Edition*, **44**, 1210–1212.

45 Casas, J., Engqvist, M., Ibrahem, I., Kaynak, B. and Cordova, A. (2005) *Angewandte Chemie – International Edition*, **44**, 1343–1345.

46 Enders, D., Grondal, C., Vrettou, M. and Raabe, G. (2005) *Angewandte Chemie – International Edition*, **44**, 4079–4083.

47 Ibrahem, I., Zou, W., Xu, Y. and Cordova, A. (2006) *Advanced Synthesis and Catalysis*, **348**, 211–222.

48 (a) Notz, W., Tanaka, F. and Barbas, C.F. (2004) *Accounts of Chemical Research*, **37**, 580–591; (b) Jarvo, E.R. and Miller, S.J. (2002) *Tetrahedron*, **58**, 2481–2495.

49 Mannich, C. and Kroesche, W. (1912) *Archiv der Pharmazie*, **250**, 647–667.

50 Ibrahem, I., Casas, J. and Cordova, A. (2004) *Angewandte Chemie – International Edition*, **43**, 6528–6531.

51 List, B. (2000) *Journal of the American Chemical Society*, **122**, 9336–9337.

52 Hayashi, Y., Tsuboi, W., Shoji, M. and Suzuki, N. (2003) *Journal of the American Chemical Society*, **125**, 11208–11209.

53 Hayashi, Y., Yamaguchi, J., Hibino, K., Sumiya, T., Urushima, T., Shoji, M., Hashizume, D. and Koshino, H. (2004) *Advanced Synthesis and Catalysis*, **346**, 1435–1439.

54 Bravo, P., Guidetti, M., Viani, F., Zanda, M., Markovsky, A.L., Sorochinsky, A., Soloshonok, I.V. and Soloshonok, V.A. (1998) *Tetrahedron*, **54**, 12789–12806.

55 Yang, J.W., Stadler, M. and List, B. (2007) *Angewandte Chemie – International Edition*, **46**, 609–611.

56 Ibrahem, I., Zou, W., Engqvist, M., Xu, Y.M. and Cordova, A. (2005) *Chemistry – A European Journal*, **11**, 7024–7029.

57 List, B., Pojarliev, P., Biller, W.T. and Martin, H.J. (2002) *Journal of the American Chemical Society*, **124**, 827–833.

58 Chowdari, N.S., Ahmad, M., Albertshofer, K., Tanaka, F. and Barbas, C.F. (2006) *Organic Letters*, **8**, 2839–2842.

59 Hayashi, Y., Tsuboi, W., Ashimine, I., Urushima, T., Shoji, M. and Sakai, K. (2003) *Angewandte Chemie – International Edition*, **42**, 3677–3680.

60 Cordova, A., Notz, W., Zhong, G., Betancort, J.M. and Barbas, C.F. (2002) *Journal of the American Chemical Society*, **124**, 1842–1843.

61 Córdova, A. and Barbas, C.F. (2002) *Tetrahedron Letters*, **43**, 7749–7752.

62 Ramasastry, S.S.V., Zhang, H., Tanaka, F. and Barbas, C.F. (2007) *Journal of the American Chemical Society*, **129**, 288–289.

63 Cheng, L., Wu, X. and Lu, Y. (2007) *Organic and Biomolecular Chemistry*, **5**, 1018–1020.

64 Chowdari, N.S., Ramachary, D.B. and Barbas, C.F. (2003) *Synlett*, 1906–1909.

65 List, B. (2002) *Journal of the American Chemical Society*, **124**, 5656–5657.

66 Bogevig, A., Juhl, K., Kumaragurubaran, N., Zhuang, W. and Jørgensen, K.A. (2002) *Angewandte Chemie – International Edition*, **41**, 1790–1793.

67 Kumaragurubaran, N., Juhl, K., Zhuang, W., Bogevig, A. and Jørgensen, K.A. (2002) *Journal of the American Chemical Society*, **124**, 6254–6255.

68 Vogt, H., Vanderheiden, S. and Brase, S. (2003) *Chemical Communications*, 2448–2449.

69 Kotrusz, P., Alemayehu, S., Toma, S., Schmalz, H. and Adler, A. (2005) *European Journal of Organic Chemistry*, 4904–4911.

70 (a) Tsogoeva, S.B. (2007) *European Journal of Organic Chemistry*, 1701–1716; (b) Ballini, R., Bosica, G., Fiorini, D., Palmieri, A. and Petrini, M. (2005) *Chemical Reviews*, **105**, 933–972; (c) Sulzer-Mosse, S. and Alexakis, A. (2007) *Chemical Communications*, 3123–3135.

71 Mukherjee, S., Yang, J.W., Hoffmann, S. and List, B. (2007) *Chemical Reviews*, **107**, 5471–5569.

72 Gotoh, H., Ishikawa, H. and Hayashi, Y. (2007) *Organic Letters*, **9**, 5307–5309.

73 Hanessian, S. and Pham, V. (2000) *Organic Letters*, **2**, 2975–2978.

74 (a) Mitchell, C.E.T., Brenner, S.E. and Ley, S.V. (2006) *Chemical Communications*, 5346–5348; (b) Mitchell, C.E.T., Brenner, S.E., Garcia-Fortanet, J. and Ley, S.V. (2005) *Organic and Biomolecular Chemistry*, **4**, 2039–2049.

75 Knudsen, K.R., Mitchell, C.E.T. and Ley, S.V. (2006) *Chemical Communications*, 66–68.

76 Brandau, S., Landa, A., Franzen, J., Marigo, M. and Jørgensen, K.A. (2006) *Angewandte Chemie – International Edition*, **45**, 4305–4309.

77 Hayashi, Y., Gotoh, H., Tamura, T., Yamaguchi, H., Masui, R. and Shoji, M. (2005) *Journal of the American Chemical Society*, **127**, 16028–16029.

78 List, B., Pojarliev, P. and Martin, H.J. (2001) *Organic Letters*, **3**, 2423–2425.

79 Enders, D. and Seki, A. (2002) *Synlett*, 26–28.

80 (a) Kotrusz, P., Toma, S., Schmalz, H. and Adler, A. (2004) *European Journal of Organic Chemistry*, 1577–1583; (b) Rasalkar, M.S., Potdar, M.K., Mohile, S.S. and Salunkhe, M.M. (2005) *Journal of Molecular Catalysis A – Chemical*, **235**, 267–270.

81 Mase, N., Thayumanavan, R., Tanaka, F. and Barbas, C.F. (2004) *Organic Letters*, **6**, 2527–2530.

82 Andrey, O., Alexakis, A. and Bernardinelli, G. (2003) *Organic Letters*, **5**, 2559–2561.

83 Mase, N., Watanabe, K., Yoda, H., Takabe, K., Tanaka, F. and Barbas, C.F. (2006) *Journal of the American Chemical Society*, **128**, 4966–4967.

84 Mase, N., Nakai, Y., Ohara, N., Yoda, H., Takabe, K., Tanaka, F. and Barbas, C.F. (2006) *Journal of the American Chemical Society*, **128**, 734–735.

85 Wang, J., Li, H., Zu, L. and Wang, W. (2006) *Advanced Synthesis and Catalysis*, **348**, 425–428.

86 Franzen, J., Marigo, M., Fielenbach, D., Wabnitz, T.C., Kjaersgaard, A. and Jørgensen, K.A. (2005) *Journal of the American Chemical Society*, **127**, 18296–18304.

87 Morita, K., Suzuki, Z. and Hirose, H. (1968) *Bulletin of the Chemical Society of Japan*, **41**, 2815.

88 Masson, G., Housseman, C. and Zhu, J. (2007) *Angewandte Chemie – International Edition*, **46**, 4614–4628.

89 Hayashi, Y., Tamura, T. and Shoji, M. (2004) *Advanced Synthesis and Catalysis*, **346**, 1106–1110.

90 Shi, M., Jiang, J.-K. and Li, C.-Q. (2002) *Tetrahedron Letters*, **43**, 127–130.

91 (a) Imbriglio, J.E., Vasbinder, M.M. and Miller, S.J. (2003) *Organic Letters*, **5**, 3741–3743; (b) Vasbinder, M.M., Imbriglio, J.E. and Miller, S.J. (2006) *Tetrahedron*, **62**, 11450–11459.

92 Vesely, J., Rios, R. and Córdova, A. (2008) *Tetrahedron Letters*, **49**, 1137–1140.

93 Utsumi, N., Zhang, H., Tanaka, F. and Barbas, C.F. (2007) *Angewandte Chemie – International Edition*, **46**, 1878–1880.

94 Vesely, J., Dziedzic, P. and Córdova, A. (2007) *Tetrahedron Letters*, **48**, 6900–6904.

95 Asato, A.E., Watanabe, C., Li, X.Y. and Liu, R.S.H. (1992) *Tetrahedron Letters*, **33**, 3105–3108.

96 Sunden, H., Ibrahem, I., Eriksson, L. and Cordova, A. (2005) *Angewandte Chemie – International Edition*, **44**, 4877–4880.

97 Aznar, F., Garcia, A.-B. and Cabal, M.-P. (2006) *Advanced Synthesis and Catalysis*, **348**, 2443–2448.

98 Cardillo, G., Fabbroni, S., Gentilucci, L., Gianotti, M. and Tolomelli, A. (2003) *Synthetic Communications*, **33**, 1587–1594.

99 Ramachary, D.B., Chowdari, N.S. and Barbas, C.F. (2003) *Angewandte Chemie – International Edition*, **42**, 4233–4237.

100 (a) Klabunovskii, E.I., Vedenyapin, A.A., Airapetov, Y.S. and Fridman, Y.D. (1978) *Reaction Kinetics and Catalysis Letters*, **9**, 73–77; (b) Ninomiya, T. (1972) *Bulletin of the Chemical Society of Japan*, **45**, 2548–2550.

101 Tungler, A. and Fogassy, G. (2001) *Journal of Molecular Catalysis A – Chemical*, **173**, 231–247.

102 Wallbaum, S. and Martens, J. (1992) *Tetrahedron – Asymmetry*, **3**, 1475–1504.

103 Mehler, T., Behnen, W., Wilken, J. and Martens, J. (1994) *Tetrahedron – Asymmetry*, **5**, 185–188.

104 Brunel, J.M., Maffei, M. and Buono, G. (1993) *Tetrahedron – Asymmetry*, **4**, 2255–2260.

105 Levina, A. and Muzart, J. (1995) *Tetrahedron – Asymmetry*, **6**, 147–156.

106 Cordova, A., Sunden, H., Engqvist, M., Ibrahem, I. and Casas, J. (2004) *Journal of the American Chemical Society*, **126**, 8914–8915.

107 Sunden, H., Engqvist, M., Casas, J., Ibrahem, I. and Cordova, A. (2004) *Angewandte Chemie – International Edition*, **43**, 6532–6535.

108 Engqvist, M., Casas, J., Sunden, H., Ibrahem, I. and Cordova, A. (2005) *Tetrahedron Letters*, **46**, 2053–2057.

109 Wang, D., Kuang, L., Li, Z. and Ding, K. (2008) *Synlett*, 69–72.

110 Qi, C., Jiang, H., Wang, Z., Zou, B. and Yang, S.R. (2007) *Synlett*, 255–258.

111 Enders, D. and Huttl, M.R.M. (2005) *Synlett*, 991–993.

112 Vignola, N. and List, B. (2004) *Journal of the American Chemical Society*, **126**, 450–451.

113 Tang, S.L.Y., Smith, R.L. and Poliakoff, M. (2005) *Green Chemistry*, **7**, 761–762.

114 Constable, D.J.C., Dunn, P.J., Hayler, J.D., Humphrey, G.R., Leazer, J.L., Linderman, R.J., Lorenz, K., Manley, J., Pearlman, B.A., Wells, A., Zaks, A. and Zhang, T.Y. (2007) *Green Chemistry*, **9**, 411–420.

115 Loh, T.-P., Feng, L.-C., Yang, H.-Y. and Yang, J.-Y. (2000) *Tetrahedron Letters*, **43**, 8741–8743.

116 Kotrusz, P., Kmentová, I., Gotov, B., Toma, Š. and Solčániová, E. (2000) *Chemical Communications*, 2510–2511.

117 Gruttadauria, M., Riela, S., Meo, P.L., D'Anna, F. and Noto, R. (2004) *Tetrahedron Letters*, **45**, 6113–6116.

118 Gruttadauria, M., Riela, S., Aprile, C., Meo, P.L., D'Anna, F. and Noto, R. (2006) *Advanced Synthesis and Catalysis*, **348**, 82–92.

119 Kitazume, T., Jiang, Z., Kasai, K., Mihara, Y. and Suzuki, M. (2003) *Journal of Fluorine Chemistry*, **121**, 205–212.

120 Lombardo, M., Pasi, F., Easwar, S. and Trombini, C. (2007) *Advanced Synthesis and Catalysis*, **349**, 2061–2065.

121 Cordova, A. (2004) *Tetrahedron Letters*, **45**, 3949–3952.

122 Aprile, C., Giacalone, F., Gruttadauria, M., Marculescu, A.M., Noto, R., Revell, J.D. and Wennemers, H. (2007) *Green Chemistry*, **9**, 1328–1334.

123 Wu, L.-Y., Yan, Z.-Y., Xie, Y.-X., Niu, Y.-N. and Liang, Y.-M. (2007) *Tetrahedron – Asymmetry*, **18**, 2086–2090.

124 Chimni, S.S. and Mahajan, D. (2006) *Tetrahedron – Asymmetry*, **17**, 2108–2119.

125 Aratake, S., Itoh, T., Okano, T., Usui, T., Shoji, M. and Hayashi, Y. (2007) *Chemical Communications*, 2524–2526.

126 Benaglia, M., Cinquini, M., Cozzi, F., Puglisi, A. and Celentano, G. (2002) *Advanced Synthesis and Catalysis*, **344**, 533–542.

127 Gu, L., Wu, Y., Zhang, Y. and Zhao, G. (2007) *Journal of Molecular Catalysis A – Chemical*, **263**, 186–194.

128 Benaglia, M., Cinquini, M., Cozzi, F., Puglisi, A. and Celentano, G. (2003) *Journal of Molecular Catalysis A – Chemical*, **204**, 157–163.

129 Gruttadauria, M., Giacalone, F., Marculescu, A.M., Meo, P.L., Riela, S. and Noto, R. (2007) *European Journal of Organic Chemistry*, 4688–4698.

130 Doyaguez, E.G., Calderon, F., Sanchez, F. and Fernandez-Mayoralas, A. (2007) *The Journal of Organic Chemistry*, **72**, 9353–9356.

131 Luo, S., Mi, X., Zhang, L., Liu, S., Xu, H. and Cheng, J.-P. (2007) *Tetrahedron*, **63**, 1923–1930.

132 Luo, S., Zhang, L., Mi, X., Qiao, Y. and Cheng, J.-P. (2007) *The Journal of Organic Chemistry*, **72**, 9350–9352.

133 Miao, W. and Chan, T.H. (2006) *Advanced Synthesis and Catalysis*, **348**, 1711–1718.

134 Ni, B., Zhang, Q. and Headley, A.D. (2007) *Green Chemistry*, **9**, 737–739.

Part Three
Enzymes

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

8
Proteases as Powerful Catalysts for Organic Synthesis

Andrés Illanes, Fanny Guzmán, and Sonia Barberis

8.1
Enzyme Biocatalysis

Enzymes are the catalysts of life. In fact, each biochemical reaction of cell metabolism requires catalysis by a specific enzyme to proceed at the pace required to sustain life. Enzymes are protein molecules that have evolved to perform efficiently under the mild conditions required to preserve the functionality and integrity of biological systems, and so can be considered as catalysts that have been optimized through evolution to perform their physiological task. Enzymes are capable of catalyzing a wide range of chemical reactions and it is not presumptuous to state that any organic chemical reaction already described might have an enzyme able to catalyze it. In fact, nature is an unlimited source of enzymes; to get the right enzyme for a certain chemical reaction is then a matter of searching, and this is certainly challenging and exciting if one realizes that only a very small fraction of all living forms have already been isolated. Moreover, it is now possible to obtain DNA metagenomic pools from the environment without requiring to know the organism from which it comes and then express it into a suitable host organism [1, 2], and also to perform genetic remodeling of structural genes by site-directed mutagenesis [3].

Enzymes have been naturally tailored to perform under physiological conditions, but biocatalysis refers to the use of enzymes as process catalysts (biocatalysts) under artificial conditions (*in vitro*) so that a major challenge in biocatalysis is then to transform these physiological catalysts into process catalysts able to withstand the harsh reaction conditions of an industrial process. A biocatalyst, as any catalyst, acts by reducing the energy barrier of a chemical reaction, without being altered as a consequence of it. However, because of their complex molecular structure, enzymes display quite distinct properties when compared with chemical catalysts. Specificity and high activity under moderate conditions are outstanding properties of enzyme biocatalysts; in addition, they have a high turnover number, are highly biodegradable, and are usually considered as natural products. These are technologically appealing properties [4, 5]. However, enzymes are complex molecules that are intrinsically labile and costly to produce, which are definite disadvantages in comparison with

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

chemical catalysts [6]. Actually, enzyme stabilization under process conditions is a major issue in biocatalysis and represents the main constraint for its technological use [7]. Within the last decades, much effort has been devoted to enzyme stabilization and several strategies of enzyme stabilization have been developed. Most important among them are the chemical modification of the enzyme structure [8, 9], immobilization on solid matrices [10–13], auto-immobilization in its own matrix by protein crystallization [14, 15], and aggregation [16–19], and the different strategies of protein engineering, including site-directed mutagenesis [20–23], directed evolution by tandem mutagenesis [24–26], and gene-shuffling based on polymerase-assisted [27, 28] and ligase-assisted recombination [29]. Screening for intrinsically stable enzymes is also a prominent area of research in biocatalysis. Organisms able to survive and thrive in extreme environmental conditions are a promising source for highly stable enzymes [30–32], and genes from such extremophiles have been cloned into suitable hosts to develop productive biological systems [33].

Most industrial enzymatic processes to date refer to reactions conducted by hydrolases in aqueous media for the degradation of complex molecules (often polymers) into simpler molecules in conventional processes with modest added value [34]. Most conventional uses of enzymes refer to the use of hydrolases (mainly proteases and glycosidases) as process catalysts or additives for the food, feed, detergent, leather, and textile industries; despite the impressive advances in biocatalysis, they still represent the major share of the enzyme market. This is because hydrolases are robust, usually extracellular, and have no coenzyme requirements, which make them ideal process biocatalysts. Most important applications of those enzymes are summarized in Table 8.1. Food enzymes are the most widely used and detailed information on them can be found in books devoted to the subject (e.g., [35]). A review on the subject can be found elsewhere [36].

Enzyme immobilization widened the scope of application allowing less stable, intracellular, and nonhydrolytic enzymes to be developed as process biocatalysts [37, 38], as illustrated by the case of glucose isomerase for the production of high fructose syrup and the production of acrylamide from acrylonitrile by nitrile hydratase. The former is the most spectacular success of enzyme technology at large scale [39], with an estimated annual output of over 10 million tons of high fructose syrup, representing more than 50% of the sweeteners market and still growing at a yearly rate of 3–4% [40]. The latter is the second largest enzymatic process, with production of acrylamide in Japan exceeding 30 000 tons/year, representing 40% of its total world market [41, 42]; the enzyme process has several advantages over the chemical process, associated with the high conversion efficiency obtained under moderate conditions, and is now conducted with nitrile hydratase-rich cells of *Rhodococcus rhodocrous* with no requirement of organic coenzymes [43, 44].

Enzymes are by no means ideal process catalysts, but their high specificity and activity under moderate conditions are prominent characteristics that are being increasingly appreciated by different production sectors, among which the pharmaceutical and fine-chemical industries stand out [45–47]. In fact, the trend is now to develop enzymatic processes for organic synthesis where the potential added value is quite high. There is an impressive number of organic reactions of technological

Table 8.1 Enzymes of commercial relevance.

Enzyme	Source	Application
Carbohydrases		
α-amylase	mold	baking, confectionery
α-amylase	bacteria	starch thinning; detergents; fabric desizing
glucoamylase	mold	glucose syrup
β-amylase and pullulanase	plant, bacteria	glucose syrup
pectinase	mold	fruit juice extraction and clarification; wine making
cellulose	mold	fruit juice extraction and clarification; textile stone-washing and bio-polishing, detergents; digestive aid
hemicellulase	mold, bacteria	baking, bleaching of wood pulp
lactase	yeast, mold	delactosed milk and dairy products; whey treatment and upgrading
invertase	yeast, mold	confectionery
phytase	bacteria	animal nutrition
β-glucanase	mold	animal nutrition; brewing
naringinase	mold	juice debittering
Proteases		
papain	papaya	yeast and meat extracts; beer chill-proofing; protein hydrolyzates; meat tenderization; leather bating; animal nutrition; digestive aid; anti-inflammatory
bromelain	pineapple	pharmaceutical: anti-inflammatory; burn debridement; enhancement of drug absorption
pepsin	animal	cheese making
rennin	animal, recombinant yeasts and molds	cheese making
neutral protease	mold, bacteria	baking; protein hydrolyzates
alkaline protease	bacteria	detergents; stickwater recovery (a byproduct of the fish processing industry)
aminopeptidase	mold, bacteria	debittering of protein hydrolyzates
Other hydrolases		
pancreatin	animal	digestive aid; leather bating
lipase	animal, mold, yeast, bacteria	flavor development in milk and meat products; detergents
aminoacylase	mold	food and feed supplementation
penicillin acylase	bacteria, mold	β-lactam precursors for semisynthetic β-lactam antibiotics
urease	bacteria	removal of urea in alcoholic beverages
Nonhydrolytic enzymes		
glucose isomerase	bacteria, actinomycetes	production of high-fructose syrups
glucose oxidase	mold	food and beverage preservation
catalase	bacteria	food preservation, peroxide removal in milk
nitrile hydratase	bacteria	acrylamide
lactamase–α-amino-ε-caprolactam racemase	bacteria	production of L-lysine
aspartate ammonia lyase	bacteria	production of L-aspartic acid

relevance that have been studied using enzymes or enzyme-containing cell biocatalysts [48, 49]. However, the use of enzyme catalysts for organic synthesis, despite its enormous potential, has been confronted with several drawbacks, biocatalysts having been considered too expensive, not easily available, too unstable, only acting properly in their natural aqueous habitat and on their natural substrates with very narrow substrate specificity, and requiring complex cofactors [50]. Some of these appreciations may be argued against, while others are being solved by technological developments. Enzyme prices have dropped consistently over recent decades, and screening of novel enzyme sources and advances in genetic/protein engineering have contributed significantly to this decrease. Enzymes are indeed unstable catalysts, but enzyme stabilization has been a major concern in biocatalysis and advances in the field are impressive [7, 51] so that now enzymes stable enough even under stringent conditions are available. Despite their drawbacks, enzymes are attractive catalysts for performing organic synthesis since they match the fundamental principles of environmentally benign manufacturing, sustainable development, and green chemistry [50]. Selectivity is maybe the most attractive property of an enzyme biocatalyst for performing organic synthesis: enzymes are highly regio- and enantioselective, which are very valuable attributes for the synthesis of pharmaceuticals and other biologically active compounds [52, 53]. Beyond health and safety considerations, selectivity has a profound influence in process economics by reducing the number of steps and protective reactions, and reducing downstream operations for product purification. Increasing enantioselectivity of enzymes has been attained by directed evolution [54] and even inversion of the natural enantioselectivity has been attained by site-directed mutagenesis [55]. Enzymes are active at mild environmental conditions, which is also a valuable attribute especially for the production of labile compounds, representing an advantage in terms of energy consumption and reactor design.

Enzymes whose physiological role is the catalysis of reactions of synthesis are, by contrast to most industrial enzymes, labile intracellular proteins requiring coenzymes; therefore, their application as process catalysts is confronted with technological difficulties, since enzymes have to be stabilized and coenzymes retained and recycled [56–58]. More technologically promising is the use of hydrolases that can, under certain conditions, catalyze the formation of the type of bond they usually hydrolyze. In this way proteases and acylases can catalyze the formation of peptide bonds [59–61], carbohydrases can catalyze the formation of glycosidic linkages [62–65], and lipases can catalyze esterification, interesterification, and transesterification reactions [66–70]. Hydrolases have tremendous potential as catalysts in organic synthesis [71], but nonconventional (nonaqueous) media are frequently required to fully exploit their synthetic capacity and depress their hydrolytic potential. The use of nonconventional media is now well established for biocatalysis [72–75], and impressive advances have been obtained in conducting synthetic reactions both in organic solvent [76–81] and ionic liquid media [82–85]. Most work has been done in organic media, but the use of organic solvents contradicts to some extent the principle of green chemistry so that the use of ionic liquids has been thoroughly studied even though their use still presents several challenges: difficulty in purification and

controlling water activity and pH, high viscosity, and problematic product recovery [82]. Complete and updated information on ionic liquids can be obtained from the *Ionic Liquids Today* newsletter (www.iolitec.de).

Biocatalysis has evolved from traditional processes in aqueous media, where the substrates and usually the enzyme as well are dissolved, to synthetic processes with different forms of biocatalysts performing in different kinds of nonconventional media. As a consequence, enzymes are now considered as valuable tools in organic chemistry. In 2000, it was claimed that about 100 synthetic processes were being conducted by biocatalysis at varying levels of industrial production, mainly for the synthesis of pharmaceutical and agrochemical precursors [86]. It is foreseeable that, pushed by the impressive advances both in biocatalyst and medium engineering, many new products of biocatalysis will reach the market within the next decade and a large portfolio of novel applications will develop.

8.2
Proteolytic Enzymes: Mechanisms and Characteristics

Proteolytic enzymes comprise a group of hydrolases referred to as peptidases (EC 3.4. X.X), which share the common feature of attacking peptide bonds. These enzymes are usually termed proteases. Proteases are among the best studied enzymes in terms of structure–function relationships [49, 87].

There are six groups of proteases termed serine, threonine, cysteine, aspartic, glutamic, and metallo, according to the groups that play a primary catalytic role. Serine, cysteine, and threonine proteases are quite different from aspartic, glutamic, and metalloproteases. In the first three groups, the nucleophile in the catalytic center is part of an amino acid residue, while in the others the nucleophile is an activated water molecule. In cysteine proteases the nucleophile is a sulfhydryl group and the catalytic mechanism is similar to the serine proteases in which the proton donor is a histidine residue. A schematic representation of the catalytic mechanism of serine and cysteine proteases is presented in Figures 8.1 and 8.2.

The catalytic mechanism of serine proteases has been thoroughly studied. The catalytic triad formed by three amino acid residues, His57, Ser195 (hence the name "serine protease"), and Asp102, is located in the active site of all serine proteases. Each amino acid residue of the triad plays a role in catalysis: the OH group in Ser195 acts as a nucleophile and attacks the carbonyl carbon in the peptide bond of the substrate; the electron pair in the nitrogen atom of His57 accepts the hydrogen atom from the OH group of Ser195, coordinating the attack of the peptide bond, while the carboxyl group in Asp102 establishes a hydrogen bond with His 57, enhancing the electronegative character of the above-mentioned electron pair. Catalysis of peptide bond hydrolysis has been described by a "ping-pong" mechanism [88] in which the substrate (protein or peptide) binds to the active site, the N-terminal portion of the substrate is released, then water (acting as a second substrate) binds to the active site, and the C-terminal portion of the substrate is released.

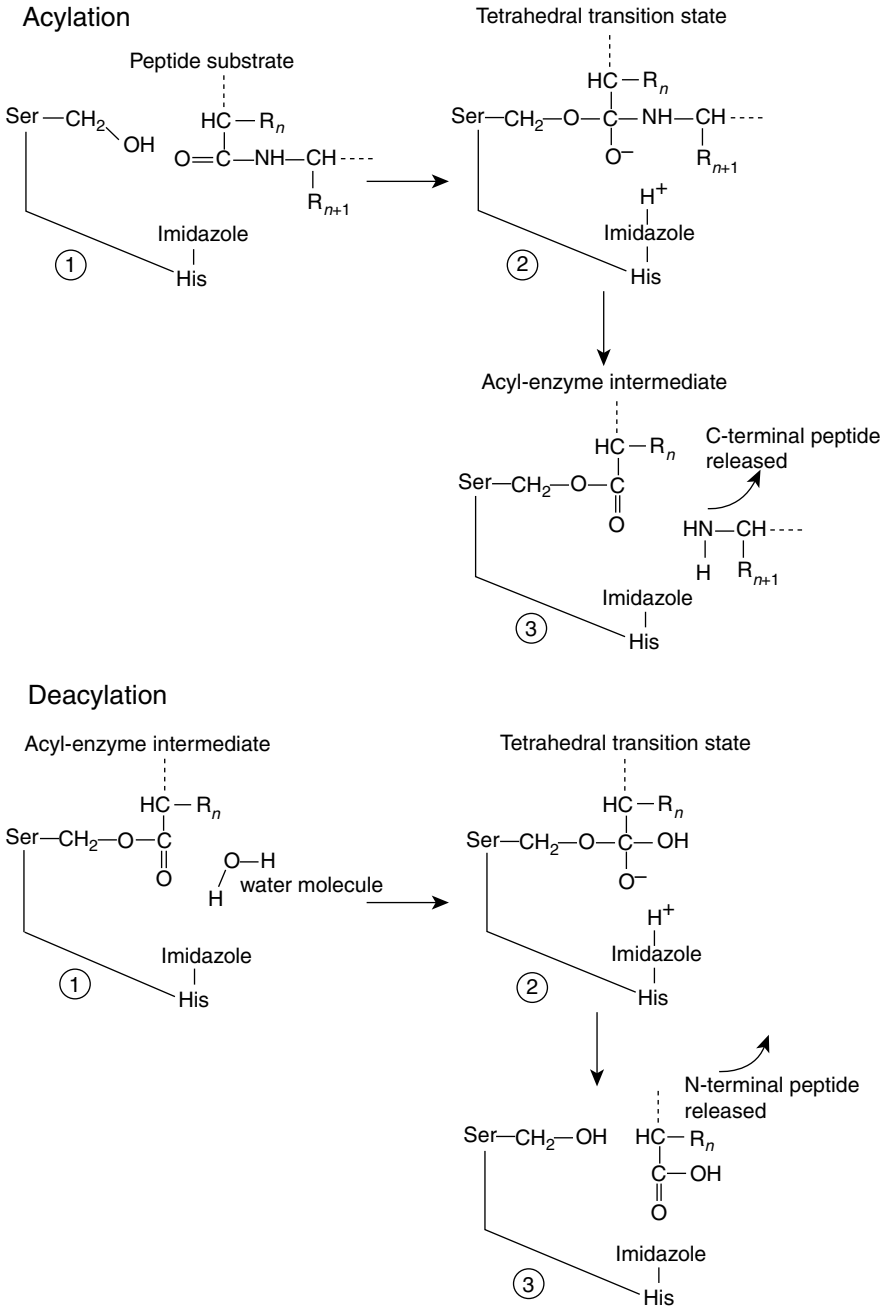


Figure 8.1 Schematic representation of the catalytic mechanism of serine proteases.

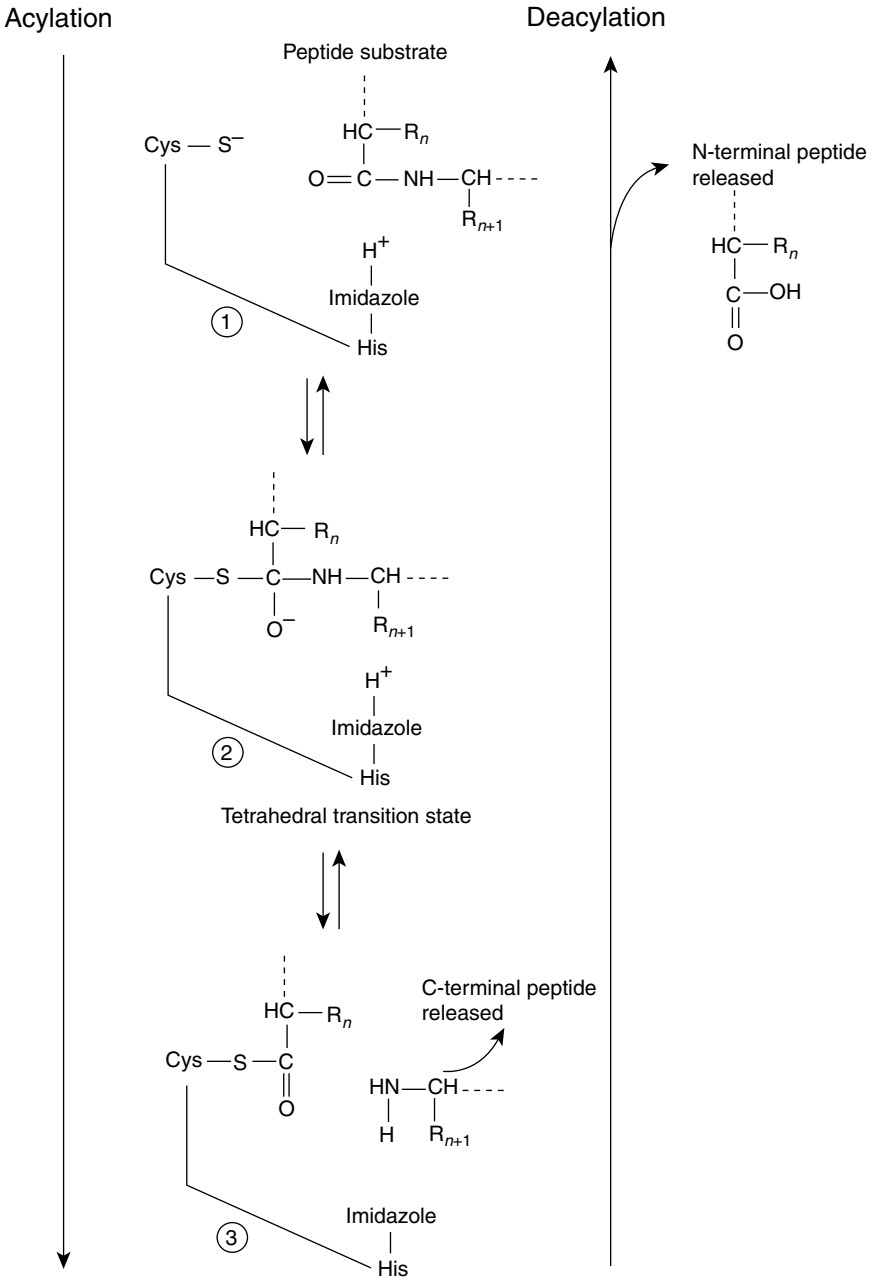


Figure 8.2 Schematic representation of the catalytic mechanism of cysteine proteases.

Proteases that cleave off single amino acids from either end of the protein or peptide chain are termed exoproteases, while endoproteases are those that attack peptide bonds in the interior of the protein or peptide chain, so that quite few free amino acids are produced as end products. The hydrolysis products are smaller polypeptides and peptides. Proteases can also be classified as alkaline [89], neutral [90], and acid proteases [91], according to the range of pH at which they exhibit maximum activity. This pH dependence has profound technological consequences and dictates to a great extent the kind of applications of a protease. According to its temperature profile, proteases are classified as thermophilic [92], mesophilic [93], or psychrophilic [94]. Thermophilic proteases are interesting for having unusually high thermal stability, while psychrophilic proteases are interesting for being very active at low temperatures, properties that can be exploited. Proteases can be extracellular or intracellular. Extracellular proteases are highly desirable because they are much simpler to recover and purify and are usually more robust [95, 96].

According to the principle of microscopic reversibility, both the hydrolysis and the formation of the peptide bond proceed by the same mechanism and through the same intermediate. In this way proteases are able not only to catalyze the hydrolysis of a peptide bond but also its formation, so that under proper conditions proteases can be used for peptide synthesis, as will be analyzed in Section 8.5.

8.3 Proteases as Process Catalysts

Proteases constitute the largest category of industrial enzymes, their application covering industrial sectors like food and beverages, detergents, leather, and pharmaceuticals [97–99]. The world market for proteases has been estimated at US$3 billion, representing roughly half of the world market of enzymes [100]. Proteases have been outstanding since the onset of enzyme technology: a normalized pancreatin preparation used as a bating agent in leather manufacture (to make the leather pliable before tanning) and later introduced for detergent formulations was patented by Röhm in 1908, and plant proteases were patented by Wallerstein for their use in brewing in 1911 [34]. Microbial proteases are now the most important in terms of market share because of the advantages of their intensive production. However, plant and some animal proteases are still prelevant for certain medical and industrial applications.

Acid and neutral proteases are relevant to the food industry: those from plant and animal origin are still important, especially in pharmaceutical products and some food applications. They represent roughly 15% of the whole protease market, but microbial proteases are now preferred for the production of protein hydrolyzates [101, 102] and other applications in the food sector. Plant-derived enzymes, such as papain and bromelain, are still industrially prevalent [103, 104] and they represent roughly 5% of the total market. Papain (EC 3.4.22.2), a cysteine protease obtained from the latex of papaya (*Carica papaya* and *Carica candamarcensis*), is the most prominent plant enzyme marketed today. It is widely used in meat

tenderization [105], beer clarification [106], yeast extract production, stain removal [107], and, in highly purified form, in several cosmetic and medical applications [108]. Bromelain, a complex of cysteine proteases extracted from pineapple stems [109], is another plant protease with different applications, mainly in the medical area [110, 111] as a wound-healing, anti-inflammatory, digestive aid, and appetite-suppressant agent. Plant and animal proteases have been usual components of digestive aids [112]; recently, a crude protease extract from kiwifruit has had considerable market success as a digestive aid (http://www.vitalfoods.co.nz/ourproducts/zylax/). Protein and genetic engineering of neutral proteases have been devoted to produce more potent and stable enzyme preparations [113]. A striking example is the case of chymosin, which is a very specific aspartic protease that hydrolyzes the Phe105–Met106 bond of κ-casein and, to a lesser extent, four other peptide bonds [114]; it is the hydrolysis of the Phe105–Met106 peptide bond that triggers the clotting of casein in the presence of calcium ions to yield curd [115]. No other enzyme is so specific for milk clotting and chymosin is therefore the best choice for cheese making, because of high yields of curding and proper flavor development [116]. The traditional source of chymosin is calf rennet, obtained as a byproduct of veal production. Shortage of rennet as a source of chymosin for cheese making became critical [117] and the first approach to solve the crisis was the screening of chymosin activity in microbial strains [118]. Microbial rennets were introduced to the market in the 1960s and still represent a significant share of the chymosin market [116]. However, they are not as acceptable as calf chymosin because curd yields are lower and the flavor developed upon maturation is not so pleasant. A further step was made by genetic engineering, and much information accumulated on the cloning, expression, and production of chymosin in microbial hosts [119]. The best hosts for chymosin expression were fungal and yeast strains, and the recombinant enzymes from *Aspergillus oryzae* and *Kluyveromyces lactis* have been in the market for more than 20 years now [120–122] with great commercial success – they are probably the most widely used enzymes in cheese making. Protein engineering has also been applied to improve chymosin performance by conferring the enzyme with increased specificity and better pH profile [123, 124].

Alkaline proteases are of paramount importance in detergent formulation and considerable progress has been made since their introduction in the 1960s [125]. The case of alkaline proteases is one of the best examples of successful application of genetic and protein engineering techniques for industrial enzyme production [89, 126, 127]. Site-directed mutagenesis has been successful in obtaining proteases more stable under harsh laundry conditions [128]. Gene shuffling [129] and phage display approaches [130] have been tried to obtain more resistant variants, but apparently no significant improvements have been obtained over the already genetically improved proteases. Apart from detergents, alkaline proteases are used in tanning [131–133] and fish-meal processing [134, 135].

New sources of proteases are continuously being reported, especially endogenous proteases from plant species [136] and exotic organisms that thrive in extreme environments – their proteases being abnormally stable and/or active in extreme conditions [100, 137].

8.4
Proteases in Organic Synthesis

Proteases are robust and stable enzymes, active at mild conditions, with pH optima in the range of 6–8 that do not require stoichiometric cofactors, being highly stereo- and regioselective [138]. These properties are important to their use as catalysts in organic synthesis. This is possible because proteases can catalyze not only the cleavage of peptide bonds, but also their formation [139, 140]. Subtilisin, chymotrypsin, trypsin, and papain have been widely used proteases in the chemical synthesis of peptides. However, the broad specificity of proteases restricts their application in peptide synthesis, since the peptide product that accumulates during the reaction can be attacked by the proteases simultaneously with the synthesis reaction [138]. Even so, the synthesis of peptides is the most outstanding application of proteases in organic synthesis.

Apart from the hydrolysis and formation of peptide bonds between amino acids, proteases can catalyze several types of reactions of synthesis. Proteases have been used for the synthesis of peptide aldehydes [141] and the acylation of a broad variety of nucleophiles, including noncoded amino acids [142, 143], non-amino acid-derived amines, amino alcohols, symmetric and asymmetric diamines [138], and *meso* and *pro* chiral diols [144–146]. Proteases have also been used in the esterification and transesterification of alcohols and carboxylic acids [147, 148]; in the case of transesterification reactions an active-site nucleophile is required so that serine and cysteine proteases are to be used, while in the former case all kinds of proteases can be used [138]. The use of proteases has also been reported in the regiospecific hydrolysis of esters and the kinetic resolution of racemic mixtures [75, 138]. In this latter case, lipases and esterases are more useful than proteases for non-amino acid derivatives [71, 149, 150]. Proteases have been used in the synthesis of glycoconjugates [138, 151], where they can be utilized for the formation of peptide bonds of glycopeptides or act as selective acylating agents [152]. They have been also used in a strategy for the synthesis of glycoproteins based on a combination of enzymatic glycopeptide coupling and glycosyltransferase reactions [153].

8.5
Peptide Synthesis

Peptides are heteropolymers composed of amino acid residues linked by peptidic bonds between the carboxyl group of one amino acid residue and the α-amino group of the next one. The size of the molecule determines the technology most suitable for its synthesis. Recombinant DNA technology is particularly suitable for the synthesis of large peptides and proteins, as illustrated by the case of insulin and other hormones [154]. Chemical synthesis is a viable technology for the production of small and medium size peptides ranging from about 5 to 80 residues [155]. Enzymatic synthesis is more restricted and has been applied mainly for the synthesis of di- and tripeptides [156]. The conclusion is that the technologies for peptide production are not competitive with each other for most of the cases.

8.5.1
Chemical Synthesis of Peptides

The chemical route is often a better technological option than the biotechnological methods of recombinant DNA and biocatalysis for the synthesis of medium-size peptides that comprise most therapeutic molecules. The synthesis of peptides was originally performed in solution. However, solid-phase peptide synthesis (SPPS) introduced by Merrifield [157] has many advantages over the classical system in solution: the reaction can be automated and the problem of solubilization of the peptide no longer exists since it remains attached to a solid support [158]. A general scheme for stepwise SPPS is presented in Figure 8.3.

There are two main systems for SPPS: *t*Boc/Bzl and Fmoc/*t*Bu. In the former, the *t*Boc group is used for the protection of the N^{α}-amino group and a benzyl ester or cyclohexyl ester for the side-chains of several amino acids; in the latter, the Fmoc group is used for the protection of the N^{α}-amino group and the *tert*-butyl group for the side-chains of several amino acids [159, 160]). The first step is the coupling of the C-terminal amino acid to the solid matrix. The N^{α} (A) group is then removed by treatment with trifluoroacetic acid (TFA) in the *t*Boc strategy and with piperidine in the Fmoc strategy. The next (N^{α}-protected) amino acid is coupled to the peptide chain

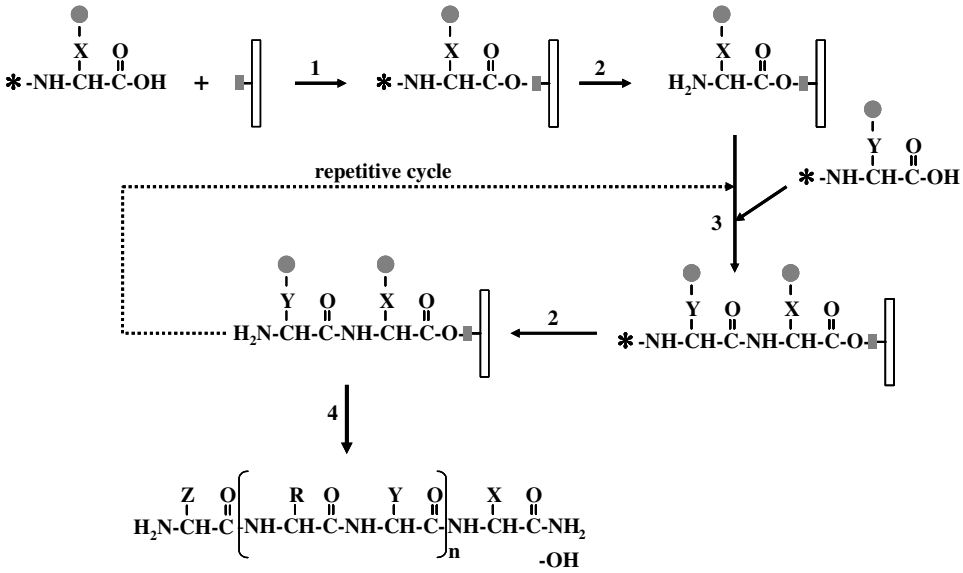


Figure 8.3 Scheme for the stepwise solid-phase peptide synthesis: shaded circles = side-chain protecting group; asterisks = N^{α}-amino protecting group; small shaded bars = functional group in the resin (Cl or NH_2); large open bars = resin. (*1*) Coupling of first residue, (*2*) N^{α} deprotection, (*3*) coupling of following residues (repetitive cycle), and (*4*) cleavage, side-chain deprotection.

bound to the polymeric matrix and, once coupled, its N^{α}-amino group is deprotected. This coupling–deprotection cycle is repeated until the desired amino acid sequence has been synthesized. Finally, the peptide–matrix complex is cleaved and side-chain protecting groups are removed to yield the peptide with either a free acid or amide depending on the chemical nature of the functional group in the solid matrix. Solid matrices should meet several requirements: particles should be of conventional and uniform size, mechanically robust, easily filterable, chemically inert and stable under the conditions of synthesis, and highly accessible to the solvents allowing the penetration of the reagents and the enlargement of the peptide chain within its microstructure. Several polymeric supports are now available that can be derivatized with functional groups to produce a highly stable linkage to the peptide being synthesized [161].

The main SPPS strategies are sequential synthesis, convergent synthesis, and chemical ligation. Sequential synthesis involves the stepwise addition of amino acids until the desired sequence is synthesized. This strategy is used mainly for the synthesis of small to medium size peptides up to 50 amino acid residues in length. In convergent synthesis, peptides (up to 50 amino acid residues) are separately produced by sequential synthesis and then linked in solution or in solid phase to obtain the desired higher-molecular-weight peptide or protein [162]. The advantage of convergent synthesis is that each peptide fragment is purified and characterized before being linked, but reaction rates for the coupling of fragments are substantially lower than for the activated amino acid species in the conventional stepwise synthesis and the C-terminal of each peptide fragment may be racemized during coupling. Some of these problems have been circumvented by using mixtures of solvents to increase solubility, by using prolonged reaction times to increase the efficiency of coupling and by using glycine or proline at the C-terminal to avoid the problem of racemization [163]. Convergent synthesis represents the best option for the chemical synthesis of large peptides and proteins and a variety of large peptides have been successfully produced by convergent synthesis, such as the human cathepsin inhibitor P41icf [164], the antiviral peptide T-20 [165], and new-generation polyproline dendrimers, which are structures for the delivery of peptide drugs [166, 167].

Chemical ligation is particularly useful for the chemical synthesis of large peptides and proteins [168, 169], and is based on the chemical linkage of short unprotected peptides that are functionalized with groups that react chemoselectively with only one group on the acceptor peptide, thus preserving the integrity of the unprotected side-chains. Many proteins and peptides of biological interest, like synthetic vaccines, have been synthesized by chemical ligation using a variety of ligands with the formation of thioester [170], oxime [171], disulfide, and thiazolidine bonds [172].

Manual and automated systems are available for small- and large-scale synthesis of only one peptide or several peptides at a time (multiple peptide synthesis). Manual synthesis of individual peptides can be performed in syringes of different sizes provided with a bottom sintered-glass or plastic filter. Multiple peptide synthesis at the micromolar level can be conducted in functionalized cellulose, polypropylene, or

polyvinylidene difluoride membranes, according to the spot-synthesis methodology [173]. Fmoc/*t*Bu and *t*Boc/Bzl multiple peptide synthesis at the millimolar level can be performed using the tea-bag methodology [174], in which up to 400 peptides of 20 residues or less can be synthesized at the same time. Several systems for automated *t*Boc/Bzl and Fmoc/*t*Bu SPPS up to 5-kg scale are available from several suppliers [160].

The precise quantification of the amount of peptide synthesized is determined through amino acid analysis by high-performance liquid chromatography, which can be performed before or after cleavage of the peptide from the resin [175]. The determination of peptide structure can be done by circular dichroism and nuclear magnetic resonance. The analyses determine if the peptide has a secondary structure and if that structure is an α-helix or β-turn [176].

Most peptides produced by SPPS are intended for pharmacological use, and may have some restrictions because of the degradation by endogenous proteases, hepatic clearance, interaction with different receptors, and low membrane permeability. Changes in the side-chains of single amino acids and the modification of the peptide chain backbone have been used to solve the problem [177]. These products are quite interesting due to their wide range of biological properties such as immunosuppressant, antibiotic, antifungal, anti-inflammatory, and antitumoral activities [178, 179]. Better recognition by antibodies can be obtained when using constrained structure peptidomimetics in which the amino acids comprising a region of a defined structure in the natural peptide are synthesized and bound by its N- and C-termini through a nonprotein ligand [180, 181]. Peptidomimetics have also been produced to inhibit protein–protein interaction [182] and increase antimicrobial action [183]. Production of peptides on a large scale presents several challenges. Peptides of less than 30 amino acid residues are mainly produced by sequential SPPS [163], while larger peptides (up to 60 residues) must be produced by convergent synthesis in which protected peptide fragments are synthesized by SPPS and then linked by liquid-phase synthesis [184]. Larger-size peptides and proteins are preferably produced by chemoselective ligation [185], in which all the unprotected linked fragments have been previously synthesized by SPPS. In terms of large-scale production, the high cost of materials and the release of environmentally aggressive or even toxic residues are major constraints. Costs of reagents are usually high; therefore, the use of large excess, which is frequent at the laboratory scale and in the early stages of development, is inadequate for large-scale production since large amounts of reagents cannot be wasted, not only for cost but also for environmental considerations [186–188]. Process validation should consider reproducibility in terms of yields of intermediate and final products, and consistency in the profile of impurities in the product [189]. This is a complex task in the case of the synthesis of peptides because of the complexity and the number of operations involved in the production process. SPPS is at present the most mature technology for producing therapeutic peptides; more than 40 of those products are now in the market and a much higher number are in different phases of approval. Several of those peptides are being produced at large scale [188].

8.5.2 Enzymatic Synthesis of Peptides

Peptide synthesis represents the most important application of proteases in organic synthesis. Nonconventional (nonaqueous) media are required to perform peptide synthesis, so selectively depressing the hydrolytic potential of the enzyme [190, 191]. In order to do so, organic solvents [77, 192], supercritical fluids [193, 194], solid state [195–198], and ionic liquids [82, 148] have been used as media for enzymatic synthesis of peptides. As compared to chemical synthesis, biocatalysis offers the advantages of higher selectivity (regio- and stereo-), low or no requirement for group protection, negligible byproduct formation, and no risk of product racemization [52]. However, the range of amino acids serving as substrates is narrow, productivities are lower, mainly because of the low activity and stability of proteases in nonconventional media, and no general protocols for synthesis have been developed as opposite to the case of SPPS [60, 188]. Proteases used for peptide synthesis are selected on the basis of their specificity for amino acid residues on either side of the splitting point, and include the majority of the commercially available endo- and exoproteases [52, 156]. The broad specificity of proteases restricts their application in peptide synthesis, since the peptide product that accumulates during the reaction can be attacked hydrolytically by the proteases simultaneously with the synthesis reaction [199].

Enzymatic peptide synthesis can proceed by two distinct mechanisms: thermodynamic and kinetic control [200].

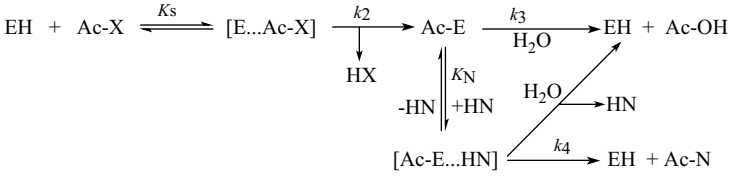
The thermodynamically controlled synthesis (TCS) of peptides with proteases represents the reverse reaction of peptide bond cleavage, as shown in the scheme [201]:

$$\mathrm{R1COO^- + H_3^+ NR2} \overset{K_{ion}}{\rightleftharpoons} \mathrm{R1CO_2H + H_2NR2} \overset{K_{con}}{\rightleftharpoons} \mathrm{R1CO\text{-}NHR2 + H_2C}$$

where K_{ion} is the equilibrium constant of ionization and K_{con} is the equilibrium constant of conversion. According to the principle of microscopic reversibility, both the formation and the hydrolysis of the peptide bond proceed by the same mechanism and through the same intermediate. The formation of the acyl intermediate from a carboxylic acid is a slow reaction and represents the rate-limiting step in TCS [138]. Distinctive advantages of TCS are the use of an acyl donor with the free carboxylic group and the possibility of using any type of protease, independent of their catalytic mechanism. Its main drawbacks are the low reaction rates and product yield attainable – the latter being determined by the equilibrium constant of the reaction, the high amount of enzyme biocatalyst often required, and the need for precise reaction conditions to displace the equilibrium towards synthesis. Conditions that favor equilibrium displacement towards synthesis can be in severe conflict with enzyme activity and stability, so it is not an easy task to obtain experimental conditions where high yields are attainable. Manipulation of the equilibrium of ionization (i.e., by pH change) and the equilibrium of the reaction (i.e., by product precipitation or by modification of medium composition) are valuable strategies to displace the equilibrium towards peptide bond formation

[202, 203]. The use of nonconventional (nonaqueous) media is by far the most effective strategy as will be analyzed ahead.

The kinetically controlled synthesis (KCS) of peptides with proteases is represented by the following scheme [138]:



where EH is the free enzyme, Ac-X is the acyl donor substrate, [E · · · Ac-X] is the Michaelis–Menten acyl-enzyme complex, HX is the released group, Ac-E is the acyl-enzyme intermediate, HN is the acceptor substrate (nucleophile), Ac-N is the product of synthesis (peptide), and Ac-OH is the product of hydrolysis of the acyl donor. As shown in the above scheme, the acyl donor, which requires being activated in the form of an ester or an amide, first binds to the enzyme to form a tetrahedral enzyme–substrate complex [E · · · Ac-X] that then collapses to form the covalent acyl-enzyme intermediate [Ac-E]. This intermediate can be nucleophilically attacked by water and by the nucleophile (HN), which can be an amine, an alcohol, or a thiol that competes with water for the deacylation reaction. The success of the synthetic reaction depends on the kinetics of these nucleophilic reactions. Contrary to TCS, only serine or cysteine proteases can be used to perform KCS, because the enzyme acts in this case as a transferase and catalyzes the transference of an acyl group from the acyl donor to the amino acid nucleophile through the formation of an acyl-enzyme intermediate. As depicted in Figure 8.4, KCS proceeds faster and with significantly higher conversion yields than TCS. It also requires lower enzyme: substrate ratios

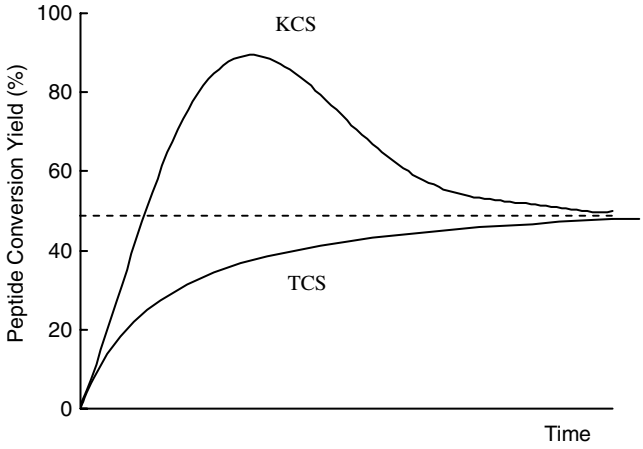


Figure 8.4 Enzymatic peptide synthesis under thermodynamic and kinetic control strategies. Dashed line = equilibrium conversion yield.

because the acyl donor is now in the form of an activated carboxylic acid [138]. Different from TCS, a maximum conversion yield is obtained in KCS at a certain point, after which it may significantly decrease as product hydrolysis outweighs synthesis. Similar behavior is observed in the case of the synthesis of β-lactam antibiotics with acylases [204, 205].

Despite their good catalytic properties, proteases are not ideal catalysts for the synthesis of peptides. As already mentioned, their specificity and selectivity limit their potential, particularly in the case of large peptides, where unwanted hydrolytic reactions may occur over the peptide product and the substrates, and the conditions of medium composition, temperature, and pH required for synthesis can be detrimental both for protease activity and stability [138, 206, 207]. However, there are different strategies to overcome such problems, which comprise the engineering of the reaction medium, the biocatalyst and the substrate [156, 198].

Medium engineering refers to the rational manipulation of the reaction medium to positively influence the properties of the enzyme with respect to the synthetic reaction [208]. This frequently implies the substitution of the usual aqueous medium for a nonconventional medium where the activity of remaining water is low enough to displace the equilibrium towards synthesis (in the case of TCS) or to selectively depress the competing hydrolysis reactions (in the case of KCS). The addition of organic cosolvents reduces the activity of water in the reaction medium, which favors synthesis, and also reduces the dielectric constant of the medium, which in turn reduces the acidity of the carboxyl group of the acyl donor and so increases the equilibrium constant K_{ion}, also promoting the synthetic reaction. The use of cosolvents is a rather simple strategy, but high concentrations are usually detrimental for enzyme activity [209, 210], even though microbial strains, highly tolerant to organic cosolvents, have been reported. Increased enzyme stability is associated with amino acid residues located in the surface region of the enzyme [211]. The decrease of water activity by organic cosolvents also favors peptide synthesis in KCS by reducing the hydrolysis of the acyl-enzyme intermediate and the final product, but again the reaction medium can be harmful to the enzyme [212–214]. There are, however, notable exceptions, like polyols and glymes in which proteases and other peptide bond forming enzymes have been successfully employed for synthesis [215–218]. A biphasic reaction medium is another option for performing peptide synthesis since the partition of the peptide products from the aqueous phase that contains the enzyme to the organic phase drives the equilibrium towards synthesis, with the additional benefit that the product is no longer subjected to hydrolysis [213, 214, 219–221]. It has been extensively used for peptide synthesis and represents a flexible strategy that can accommodate the properties of substrates and products [222, 223]. However, reaction rates in biphasic systems are low because of the limitation of substrate diffusion and because proteases tend to denature at the water/solvent interface [202, 224]. Biphasic systems are not adequate to perform KCS since the neutral esters commonly used as activated acyl donors partition poorly into the aqueous phase and therefore the concentration in that phase, where the enzymatic reaction occurs, is low. An alternative to the classic biphasic system has been proposed by Clapés *et al.* [225] as reaction medium for enzymatic peptide synthesis using water-in-oil emulsion with a high water content (95%) which differs from conventional reverse micelles where

the water phase containing the enzyme represents a small percentage of the volume which corresponds mostly to the oil phase [226–228]. Solubilization of proteases in hydrophobic solvents by ion-pairing with anionic surfactants at submicellar concentrations has been successfully applied to the synthesis of peptides. Enzymes display very high synthetic activities in those systems, leading to much higher reaction rates than those obtained with insoluble particle biocatalysts suspended in organic media. In addition, water control in such a reaction medium was not critical since no significant hydrolysis of the product was promoted [229, 230]. Such systems have been occasionally used in organic synthesis [231], but its application in biocatalysis is still in development. Suspension of nearly anhydrous enzymes in hydrophobic solvents can be considered as a promising strategy for enzymatic synthesis since it exploits to the highest extent the advantages of working in a nonconventional medium [77]: high thermostability [76, 232], potentially favorable changes in substrate specificity [77, 233–235], and ease of biocatalyst and product recovery [236]. The enzyme biocatalyst is simply a lyophilized or acetone-precipitated enzyme powder that is suspended in the organic medium in which it is completely insoluble [237, 238]. Immobilization is unnecessary since the enzyme is already insolubilized in the reaction medium, even though it might help by offering an increased contact surface with the substrate and provide additional stabilization [215, 239]. The use of neat hydrophobic solvents with very low water content can, in principle, be effective for peptide synthesis by reducing the backward hydrolytic reaction [200]. However, proteases may exhibit very low activity and substrates and products may be poorly soluble in this kind of medium. In fact, despite that in some cases no significant differences in molecular conformation have been found in organic media and in water [240], activities are much lower than in water or water/cosolvent media [241]. Notwithstanding this, proteases can be quite stable in organic media and usually much more stable than in water/cosolvent media [232, 242]. Water activity is a critical variable and needs to be carefully controlled to optimize enzyme performance [239, 243, 244]. The main drawback of neat hydrophobic solvents is the dramatic decrease in enzyme activity, which severely limits its application for large-scale synthesis making the cost of the biocatalyst a critical factor [245]. Among the many reasons for the poor expression of enzyme activity in these media are diffusional restrictions which, is reasonable since the enzyme is completely insoluble in such media being in the form of a particle. This very much resembles what happens with enzymes immobilized onto solid supports. In fact, diffusional restrictions in such media have been correlated with size and specific activity of the biocatalyst particle [246]. Molecular stiffness of the enzyme molecule in low-water organic media is also a plausible cause of the reduction in activity [247]. However, improvements in activity have been obtained by the addition of water mimics, such as formamide or ethylene glycol, to the reaction medium or the addition of crown ethers during the preparation of the lyophilized enzyme [248–251].

Biocatalyst engineering refers to the strategies for improving biocatalysts under synthetic conditions. Such strategies go from screening of novel enzyme sources to physical and chemical modification of the catalyst and to genetic and protein engineering [20, 138, 252]. Screening of novel protease sources has been a very active area of research and several novel proteases, mainly from extremophiles, have

been isolated and characterized as amenable for performing organic synthesis [31, 198, 253]. Also, metagenomics has been successfully applied to isolate novel proteases from uncultured microbiota in the environment [89, 254, 255]. Insolubilization by immobilization to a solid carrier or by protein aggregation are biocatalyst engineering strategies for producing robust enzyme catalysts well suited to withstand the harsh conditions required for performing synthetic reactions. Immobilization has been extensively used to produce protease biocatalysts for peptide synthesis [192, 215, 256–261]. Insolubilization by protein cross-linking is a promising technology to produce biocatalysts for synthesis. Cross-linked enzyme aggregates (CLEAs) are robust novel biocatalysts produced by precipitation of the enzyme protein followed by cross-linking with bifunctional reagents. Their specific activity is very high since there is no inert matrix and the whole mass of biocatalyst is essentially pure enzyme protein. Production is simple and inexpensive since the enzyme does not require to be purified to any extent [16, 262]. CLEAs prepared by precipitation with salts, organic solvents, and polymers, and cross-linked with glutaraldehyde have been quite effective for peptide bond formation in the synthesis of β-lactam antibiotics in nonaqueous environments [18, 263, 264]. However, in the specific case of peptide synthesis with proteases, autoproteolysis can play a role because of the close proximity and flexibility of the enzyme molecules. Despite this, CLEAs of trypsin [17] and alcalase [19] have been reported. Cross-linked enzyme crystals obtained by cross-linking of previously crystallized enzyme protein are very stable and robust biocatalysts [14] that have been successfully applied to the synthesis of peptides [81, 265], even though in this case the high cost of the biocatalyst is a drawback for its large-scale application. Proteases have been preferred models for biocatalyst engineering strategies that consider the manipulation of the genes encoding the enzyme protein. In fact, subtilisin has been extensively studied [128], and considerable progress has been made in engineering this protease and its substrates for peptide bond formation in aqueous media [266], and for increased thermal stability and stability in nonconventional media [267–269]. Site-directed mutagenesis has also been employed to improve the properties of trypsin for performing peptide synthesis [198], and directed evolution based on tandem random mutagenesis has been successfully applied to improve the thermal stability of subtilisin [20, 270] and increase the activity and selectivity of an endoprotease from *Escherichia coli* [271].

Substrate engineering refers to the modification of the substrate rather than the enzyme to improve reaction. Even though most proteases recognize more than one amino acid, not all are able to couple any amino acid sequence. Due to the specificity of proteases to a particular amino acid, only those acyl donors that have a specific amino acid in the C-terminal position can be coupled without side-reactions. In this way, the manipulation of the leaving group may be useful to increase the specificity of the protease for a previously less specific amino acid, so increasing reaction rate [272, 273]. The manipulation of the leaving group affects the aminolysis/hydrolysis ratio of an acyl donor and therefore the conversion yield, since the acyl-enzyme intermediate formed is the same regardless of the change produced in the leaving group. The influence of the structure of the acyl donor in the selectivity of α-chymotrypsin and the efficiency of nucleophiles on peptide synthesis in organic

solvents was reported by Fischer *et al.* [274]. Another approach is the use of mimetic substrates where, contrary to the classic manipulation of the leaving group, a mimetic substrate is designed to bind to the active site of the enzyme. In this way, serine and cysteine proteases can react with non specific amino acids or peptide sequences without altering the enzyme or the reaction medium [275]. The main advantage of this strategy is that mimetic substrates allow the formation of the acyl-enzyme intermediate and the nucleophile reaction to establish a peptide bond that cannot be further hydrolyzed because it does not correspond to the protease specificity [198, 276, 277]. However, this approach is limited to reactions with nonspecific amino acid containing peptides, whereas the coupling of specific ones leads to unwanted cleavages due to the native proteolytic activity of the biocatalyst [278].

Small protease synthesized peptides are being used successfully for human and animal nutrition and also as pharmaceuticals and agrochemicals. The most striking example is the synthesis of the leading noncaloric sweetener aspartame, where the former chemical route is being progressively replaced by enzymatic synthesis with the protease thermolysin in an organic medium [45, 222, 279, 280] and more recently in a fully aqueous medium at high substrate concentrations within the framework of sustainable technology [281, 282]. Other small-size peptides are also being produced by protease synthesis, like the lysine sweet dipeptide [156, 283], the neuroactive pain regulator dipeptide kyotorphin [284–286], the nociception regulator pentapeptide enkephalin [197, 228, 287–289], the C-terminal octapeptide of cholecystokinin (CCK-8) [290, 291], a 33-amino-acid hormone responsible for stimulating the digestion of fat and protein in the small intestine [292], vasopressin, a tetrapeptide involved in blood pressure regulation [293], the neuro-octapeptide oxytocin [245], and the octapeptide fragment of the opioid peptide dynorphin [294, 295]. Several other examples of enzymatically synthesized biologically active peptides have been reported [156, 213, 218, 296, 297]. Process development for enzymatic peptide synthesis requires the definition of a synthetic strategy, the proper selection of the proteases to be used for each coupling step according to their specificity, the sequencing of peptide bond formation (one pot or consecutive reactions), the formulation of the reaction medium (aqueous, organic, or biphasic), and the selection of the amino acid and peptide derivatives acting as acyl donor and nucleophile [156, 245, 284]. Although there are some rules for conducting the synthesis, the high number of critical variables in an enzymatic process makes its optimization cumbersome and, in this sense, enzymatic synthesis of peptides is a less mature technology than chemical synthesis since no general protocols of synthesis are available. Each situation is a particular case that has to be extensively studied and optimized. Despite this, many of the limitations of chemical synthesis can be overcome by protease synthesis. The high specificity and high reactivity under mild operation conditions of enzyme biocatalysts are properties that may have a strong impact on process economics, since these will reduce the number of operations required for the synthesis, will have a less stringent requirement for equipment, will reduce the energy input required, and will produce less environmental impact [298]. The size of the peptide is a major constraint for the enzymatic synthesis since no established and automated protocols have yet been developed.

As the peptide chain grows, requirements of protection and modification of several operational variables at each step makes the process cumbersome, so that in practice only peptides of less than 10 residues have been synthesized, as illustrated by the examples given above. Convergent enzymatic synthesis may be used, as in chemical synthesis, and has been elegantly applied to the synthesis of the CCK-8 using a combination of proteases and reaction media for each step [188, 290, 299, 300].

The use of organic solvents, usually required for performing enzymatic peptide synthesis [301], imposes additional restrictions on process design and engineering, and contradicts the clean technology concept [245, 302], so major efforts are being directed toward their replacement by greener systems, among which ionic liquids [82, 198] and solid or semisolid reaction media are prominent [196, 303–305]. Recently, Meng *et al.* [291] reported the enzymatic synthesis of the CCK pentapeptide (CCK-5) where the coupling was conducted in aqueous buffer and no hazardous chemicals were necessary. However, these systems are still at an early stage of technological development. Despite the technological advances in peptide synthesis with proteases, a general approach for the process remains to be formulated, and low productivity, low conversion yield, and high cost of enzymes are problems yet to be solved to make it competitive in a broad spectrum of cases. More specific, active, and stable enzymes are being pursued, and the trend should be to improve existing proteases by the use of the modern technologies of genetic/protein engineering and automated screening of proteases from novel microbial strains and also from the environmental metagenome.

8.6 Conclusions

Chemical synthesis, especially solid phase, can be considered now as the most mature technology for peptide synthesis, being especially suited for medium-size peptides, which comprise the most relevant pharmaceutical and healthcare products. However, a lack of specificity and environmental burden are drawbacks of chemical processes that can be successfully overcome by enzyme biocatalysis, since many of the present constraints of enzymatic processes for peptide synthesis are being solved through research and development in different areas of enzyme biocatalysis. Proteases are readily available enzymes that have been used traditionally as industrial commodities for a long time exploiting their hydrolytic potential over proteinaceous substrates. However, they are also well suited as biocatalysts for performing peptide synthesis in a proper reaction medium, where they may be used as such or as starting point for developing modified enzymes especially suited to match the requirements for peptide synthesis. Site-directed mutagenesis and directed evolution of protease-producing genes will be major strategies for developing improved proteases for peptide synthesis. The combination of chemical and enzymatic synthesis has proven to be a valuable technological option since the good properties of each technology can be synergistically used in the context of one process objective as recently shown for the case of the synthesis of the tripeptide RGD (Arg–Gly–Asp) amide [306].

References

1 Nield, B.S., Willows, R.D., Torda, A.E., Gillings, M.R., Holmes, A.J., Helena Nevalainen, K.M., Stokes, H.W. and Mabbutt, B.C. (2002) New enzymes from environmental cassette arrays: functional attributes of a phosphotransferase and an RNA-methyltransferase. *Protein Science*, **13**, 1651–1659.

2 Bertram, M., Hildebrandt, P., Weiner, D.P., Patel, J.S., Bartnek, F., Hitchman, T.S. and Bornscheuer, U.T. (2008) Characterization of lipases and esterases from metagenomes for lipid modification. *Journal of the American Oil Chemists Society*, **85**, 47–53.

3 Abián, O., Grazú, V., Hermoso, J., Gonzalez, R., Garcia, J.L., Fernandez-Lafuente, R. and Guisan, J.M. (2004) Stabilization of penicillin G acylase from *Escherichia coli*: site-directed mutagenesis of the protein surface to increase multipoint covalent attachment. *Applied and Environmental Microbiology*, **70**, 1249–1251.

4 Benkovic, S.J. and Ballesteros, A. (1997) Biocatalysts – the next generation. *Trends in Biotechnology*, **15**, 385–386.

5 Wegman, M.A., Janssen, M.H.A., van Rantwijk, F. and Sheldon, R.A. (2001) Towards biocatalytic synthesis of β-lactam antibiotics. *Advanced Synthesis and Catalysis*, **343**, 559–576.

6 Bommarius, A.S. and Broering, J.M. (2005) Established and novel tools to investigate biocatalyst stability. *Biocatalysis and Biotransformation*, **23**, 125–139.

7 Illanes, A. (1999) Stability of biocatalysts. *Electronic Journal of Biotechnology*, **2**, 1–9.

8 Özturk, D.C., Kazan, D. and Erarslan, A. (2002) Stabilization and functional properties of *Escherichia coli* penicillin G acylase by covalent conjugation of anionic polysaccharide carboxymethyl cellulose. *World Journal of Microbiology & Biotechnology*, **18**, 881–888.

9 Mislovičová, D., Masarová, J., Bučo, M. and Gemeiner, P. (2006) Stability of penicillin G acylase modified with various polysaccharides. *Enzyme and Microbial Technology*, **39**, 579–585.

10 Kim, J., Grate, J.W. and Wang, P. (2006) Nanostructures for enzyme stabilization. *Chemical Engineering Science*, **61**, 1017–1026.

11 Wilson, L., Fernández-Lorente, G., Fernández-Lafuente, R., Illanes, A., Guisán, J.M. and Palomo, J.M. (2006) CLEAs of lipases and poly-ionic polymers: a simple way of preparing stable biocatalysts with improved properties. *Enzyme and Microbial Technology*, **39**, 750–755.

12 Mateo, C., Palomo, J.M., Fernandez-Lorente, G., Guisán, J.M. and Fernandez-Lafuente, R. (2007) Improvement of enzyme activity, stability and selectivity via immobilization techniques. *Enzyme and Microbial Technology*, **40**, 1451–1463.

13 Rodrigues, D.S., Mendes, A.A., Adriano, W.S., Gonçalves, L.R.B. and Giordano, R.L.C. (2008) Multipoint covalent immobilization of microbial lipases on chitosan and agarose activated by different methods. *Journal of Molecular Catalysis B – Enzymatic*, **51**, 100–109.

14 Häring, D. and Schreier, P. (1999) Cross-linked enzyme crystals. *Current Opinion in Chemical Biology*, **3**, 35–38.

15 Roy, J.J. and Abraham, T.E. (2006) Preparation and characterization of cross-linked enzyme crystals of laccase. *Journal of Molecular Catalysis B – Enzymatic*, **38**, 31–36.

16 Cao, L., van Langen, L. and Sheldon, R. (2003) Immobilised enzymes: carrier-bound or carrier free? *Current Opinion in Biotechnology*, **14**, 387–394.

17 Schoevaart, R., Wolbers, M.W., Golubovic, M., Ottens, M., Kieboom, A.P.G., van Rantwijk, F., van der Wielen,

L.A.M. and Sheldon, R.A. (2004) Preparation, optimization, and structures of cross-linked enzyme aggregates (CLEA). *Biotechnology and Bioengineering*, **87**, 754–762.

18 Illanes, A., Wilson, L., Caballero, E., Fernández-Lafuente, R. and Guisán, J.M. (2006) Crosslinked penicillin acylase aggregates for synthesis of β-lactam antibiotics in organic medium. *Applied Biochemistry and Biotechnology*, **133**, 189–202.

19 Sheldon, R.A., Sorgedrager, M. and Janssen, M.H.A. (2007) Use of cross-linked enzyme aggregates (CLEAs) for performing biotransformations. *Chemistry Today*, **25**, 62–67.

20 Adamczak, M. and Hari Krishna, S. (2004) Strategies for improving enzymes for efficient biocatalysis. *Food Technology and Biotechnology*, **42**, 251–264.

21 Bardy, S.L., Ng, S.Y.M., Carnegie, D.S. and Jarrell, K.F. (2005) Site-directed mutagenesis analysis of amino acids critical for activity of the type I signal peptidase of the archaeon *Methanococcus voltae*. *Journal of Bacteriology*, **187**, 1188–1191.

22 Morley, K.L. and Kazlauskas, R.J. (2005) Improving enzyme properties: when are closer mutations better? *Trends in Biotechnology*, **23**, 231–237.

23 Leisola, M. and Turunen, O. (2007) Protein engineering: opportunities and challenges. *Applied Microbiology and Biotechnology*, **75**, 1225–1232.

24 Arnold, F. (2001) Combinatorial and computational challenges for biocatalyst design. *Nature*, **409**, 253–257.

25 Sylvestre, J., Chautard, H., Cedrone, F. and Delcourt, Marc. (2006) Directed evolution of biocatalysts. *Organic Process Research & Development*, **10**, 562–571.

26 Boersma, Y.L., Dröge, M.J. and Quax, W.J. (2007) Selection strategies for improved biocatalysts. *FEBS Journal*, **274**, 2181–2195.

27 Shibuya, H., Kaneko, S. and Hayashi, K. (2000) Enhancement of the thermostability and hydrolytic activity of xylanase by random gene shuffling. *The Biochemical Journal*, **349**, 651–656.

28 Kaur, J. and Sharma, R. (2006) Directed evolution: an approach to engineer enzymes. *Critical Reviews in Biotechnology*, **26**, 165–199.

29 Chodorge, M., Fourage, L., Ullmann, C., Duvivier, V., Masson, J.M. and Lefevre, F. (2005) Rational strategies for directed evolution of biocatalysts – application to *Candida antarctica* lipase B (CALB). *Advanced Synthesis and Catalysis*, **347**, 1022–1026.

30 Adams, M. and Kelly, R. (1998) Finding and using hyperthermophilic ezymes. *Trends in Biotechnology*, **16**, 329–332.

31 van den Burg, B. (2003) Extremophiles as a source for novel enzymes. *Current Opinion in Microbiology*, **6**, 213–218.

32 Gomes, J. and Steiner, W. (2004) The biocatalytic potential of extremophiles and extremozymes. *Food Technology and Biotechnology*, **42**, 223–235.

33 Haki, G.D. and Rakshit, S.K. (2003) Developments in industrially important thermostable enzymes: a review. *Bioresource Technology*, **89**, 17–34.

34 Neidleman, S.L. (1991) Historical perspective on the industrial uses of biocatalysts, in *Biocatalysts for Industry* (ed. J.S. Dordick), Plenum Press, New York, pp. 21–33.

35 Whitehurst, R.J. and Law, B.A. (2002) In *Enzymes in Food Technology*, CRC Press, Boca Raton, FL, p. 255.

36 Illanes, A. (2000) Biocatalysts for the food industry, in *Trends in Food Engineering* (eds J.E. Lozano, C. Añón, E. Parada-Arias and G.V. Barbosa-Cánovas), Technomic, Lancaster, pp. 287–298.

37 Katchalsky-Katzir, E. (1993) Immobilized enzymes. Learning from past successes and failures. *Trends in Biotechnology*, **11**, 471–478.

38 D' Souza, S. (1999) Immobilized enzymes in bioprocess. *Current Science*, **77**, 69–78.

39 Carasik, W. and Carroll, O.J. (1983) Development of immobilized enzymes

for production of high-fructose corn syrup. *Food Technology*, **37**, 85–91.

40 Bhosale, S.H., Rao, M.B. and Deshpande, V.V. (1996) Industrial aspects of glucose isomerase. *Microbiological Reviews*, **60**, 280–300.

41 Yamada, H. and Kobayashi, M. (1996) Nitrile hydratase and its application to industrial production of acrylamide. *Bioscience, Biotechnology, and Biochemistry*, **60**, 1391–1400.

42 Miller, J. and Nagarajan, V. (2000) The impact of biotechnology on the chemical industry in the 21st century. *Trends in Biotechnology*, **18**, 190–196.

43 Buchholz, K., Kasche, V. and Bornscheuer, U.T. (2005) In *Biocatalysts and Enzyme Technology*, Wiley-VCH Verlag GmbH, Weinheim, p. 448.

44 Liese, A., Seelbach, K. and Buchholz, A. (2006) Lyases, in *Industrial Biotransformations*, 2nd edn (eds A. Liese, K. Seelbach and C. Wandrey), Wiley-VCH Verlag GmbH, Weinheim, pp. 447–503.

45 Schmid, A., Dordick, J.S., Hauer, B., Kiener, A., Wubbolts, M. and Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature*, **409**, 258–268.

46 Thomas, S.M., DiCosimo, R. and Nagarajan, V. (2002) Biocatalysis: applications and potentials for the chemical industry. *Trends in Biotechnology*, **20**, 238–242.

47 Bruggink, A., Straathof, A.J.J. and van der Wielen, L.A.M. (2003) Fine chemical industry for life science products: green solutions to chemical challenges. *Advances in Biochemical Engineering/ Biotechnology*, **80**, 69–113.

48 Koeller, K.M. and Wong, C.H. (2001) Enzymes for chemical synthesis. *Nature*, **409**, 232–240.

49 Drauz, K. and Waldmann, H. (2002) In *Enzyme Catalysis in Organic Synthesis*, vol. **II**, Wiley-VCH Verlag GmbH, Weinheim, p. 998.

50 Bommarius, A.S. and Riebel, B.R. (2004) In *Biocatalysis: Fundamentals and Applications*, Wiley-VCH Verlag GmbH, Weinheim, p. 611.

51 O'Fágáin, C. (2003) Enzyme stabilization – recent experimental progress. *Enzyme and Microbial Technology*, **33**, 137–149.

52 Faber, K. (1997) In *Biotransformations in Organic Chemistry*, Springer, Berlin, p. 402.

53 Vakhlu, J., Johri, S., Verma, V., Koul, S., Parshad, R., Taneja, S.C. and Qazi, G.N. (2005) Purification and properties of enantioselective ester hydrolase from a strain of *Trichosporon* species (DSMZ 11829). *Enzyme and Microbial Technology*, **37**, 330–339.

54 Reetz, M.T. (2006) Directed evolution of enantioselective enzymes as catalysts for organic synthesis. *Advances in Catalysis*, **49**, 1–69.

55 Ijima, Y., Matoishi, K., Terao, Y., Doi, N., Yanagawa, H. and Ohta, H. (2005) Inversion of enantioselectivity of asymmetric biocatalytic decarboxylation by site-directed mutagenesis based on the reaction mechanism. *Chemical Communications*, 877–879.

56 Adlercreutz, P., Iborra, J.L., Schmid, E. and Pedersen, S. (1994) Applications, in *Applied Biocatalysis* (eds J.M.S. Cabral, D.B.L. Boros and J. Tramper), Harwood, Chur, pp. 108–156.

57 Wandrey, C. (2004) Biochemical reaction engineering for redox reactions. *The Chemical Record*, **4**, 254–265.

58 Liu, W. and Wang, P. (2007) Cofactor regeneration for sustainable enzymatic biosynthesis. *Biotechnology Advances*, **25**, 369–384.

59 Illanes, A., Wilson, L., Altamirano, C., Cabrera, Z., Alvarez, L. and Aguirre, C. (2007) Production of cephalexin in organic medium at high substrates concentrations with CLEA of penicillin acylase and PGA-450. *Enzyme and Microbial Technology*, **40**, 195–203.

60 Meng, L.P., Joshi, R. and Eckstein, H. (2007) Protease-catalyzed synthesis of the tripeptide CCK26-28, a fragment of CCK-8. *Amino Acids*, **33**, 91–96.

61 Terreni, M., Ubiali, D., Bavaro, T., Pregnolato, M., Fernández-Lafuente, R. and Guisán, J.M. (2007) Enzymatic synthesis of cephalosporins. The immobilized acylase from *Arthrobacter viscosus*: a new useful biocatalyst. *Applied Microbiology and Biotechnology*, **77**, 579–587.

62 Bucke, C. (1996) Oligosaccharide synthesis using glycosidases. *Journal of Chemical Technology and Biotechnology*, **67**, 217–220.

63 Cheng, C.C., Yu, M.C., Cheng, T.C., Sheu, D.C., Duan, K.J. and Tai, W.L. (2006) Production of high-content galacto-oligosaccharide by enzyme catalysis and fermentation with *Kluyveromyces marxianus*. *Biotechnology Letters*, **28**, 793–797.

64 Panesar, P.S., Panesar, R., Singh, R.S., Kennedy, J.F. and Kumar, H. (2006) Microbial production, immobilization and applications of β-galacosidase. *Journal of Chemical Technology and Biotechnology*, **81**, 530–543.

65 Nakkharat, P. and Haltrich, D. (2007) β-galactosidase from *Talaromyces termophilus* immobilized onto Eupergit C for production of galacto-oligosaccharides during lactose hydrolysis in batch and packed-bed reactor. *World Journal of Microbiology & Biotechnology*, **23**, 759–764.

66 Cernia, E., Palocci, C. and Soro, S. (1998) The role of reaction medium in lipase-catalyzed esterifications and tansesterifications. *Chemistry and Physics of Lipids*, **93**, 157–168.

67 Undurraga, D., Markovits, A. and Erazo, S. (2001) Cocoa butter equivalent through enzymic interesterification of palm oil midfraction. *Process Biochemistry*, **36**, 933–939.

68 Osório, N.M., Gusmão, J.H., da Fonseca, M.M. and Ferreira-Dias, S. (2005) Lipase-catalysed interesterification of palm stearin with soybean oil in a continuous fluidised-bed reactor. *European Journal of Lipid Science and Technology*, **107**, 455–463.

69 Li, L., Du, W., Liu, D., Wang, L. and Li, Z. (2006) Lipase-catalyzed transesterification of rapeseed oils for biodiesel production with a novel organic solvent medium. *Journal of Molecular Catalysis B – Enzymatic*, **43**, 58–62.

70 Kumari, V., Shah, S. and Gupta, M.N. (2007) Preparation of biodiesel by lipase-catalyzed transesterification of high free fatty acid containing oil from *Madhuca indica*. *Energy Fuels*, **21**, 368–372.

71 Bornscheuer, U.T. and Kazlauskas, R.J. (1999) In *Hydrolases in Organic Synthesis*, Wiley-VCH Verlag GmbH, Weinheim, p. 336.

72 Ballesteros, A., Bornscheuer, U., Capewell, A., Combes, D., Condoret, J.S., Koenig, K., Kolisis, F.N., Marty, A., Menge, U., Scheper, T., Stamatis, H. and Xenakis, A. (1995) Enzymes in non-conventional phases. *Biocatalysis and Biotransformation*, **13**, 1–42.

73 Hari Krishna, S. (2002) Developments and trends in enzyme catalysis in non-conventional media. *Biotechnology Advances*, **20**, 239–267.

74 Halling, P.J. (2004) What can we learn by studying enzymes in non-aqueous media? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 1287–1296.

75 Carrea, G. and Riva, S. (2008) In *Organic Synthesis with Enzymes in Non-Aqueous Media*, Wiley-VCH Verlag GmbH, Weinheim, p. 310.

76 Klibanov, A. (2001) Improving enzymes by using them in organic solvents. *Nature*, **409**, 241–246.

77 Clark, D. (2004) Characteristics of nearly dry enzymes in organic solvents: implications for biocatalysis in the absence of water. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 1299–1307.

78 Gupta, M.N. and Roy, I. (2004) Enzymes in organic media: forms, functions and applications. *European Journal of Biochemistry*, **271**, 2573–2583.

79 Krieger, N., Bhatnagar, T., Baratti, J., Baron, A.M., de Lima, V.M. and Mitchell, D. (2004) Non-aqueous biocatalysis in heterogeneous solvent systems. *Food Technology and Biotechnology*, **42**, 279–286.

80 Torres, S. and Castro, G. (2004) Non-aqueous biocatalysis in homogeneous solvent systems. *Food Technology and Biotechnology*, **42**, 271–277.

81 Amorim Fernandes, J., McAlphine, M. and Halling, P. (2005) Operational stability of subtilisin CLECs in organic solvents in repeated batch and continuous operation. *Biochemical Engineering Journal*, **24**, 11–15.

82 Park, S. and Kazlauskas, R. (2003) Biocatalysis in ionic liquids – advantages beyond green technology. *Current Opinion in Biotechnology*, **14**, 432–437.

83 van Rantwijk, F., Lau, R.M. and Sheldon, R.A. (2003) Biocatalytic transformations in ionic liquids. *Trends in Biotechnology*, **21**, 131–138.

84 Durand, J., Teuma, E. and Gómez, M. (2007) Ionic liquids as a medium for enantioselective catalysis. *Comptes Rendus Chimie*, **10**, 152–177.

85 Ha, S.H., Lan, M.N., Lee, S.H., Hwang, S.M. and Koo, Y.M. (2007) Lipase-catalyzed biodiesel production from soybean oil in ionic liquids. *Enzyme and Microbial Technology*, **41**, 480–483.

86 Wandrey, C., Liese, H. and Kihumbu, D. (2000) Industrial biocatalysis: past, present, and future. *Organic Process Research & Development*, **4**, 286–290.

87 Krowarsch, D., Zakrzewska, M., Smalas, A.O. and Otlewski, J. (2005) Structure–function relationships in serine protease–bovine pancreatic trypsin inhibitor interaction. *Protein and Peptide Letters*, **12**, 403–407.

88 Gaertner, H. and Puigserver, A. (1989) Kinetics and specificity of serine proteases in peptide synthesis catalyzed in organic solvents. *European Journal of Biochemistry*, **181**, 207–213.

89 Gupta, R., Beg, Q.K. and Lorenz, P. (2002) Bacterial alkaline proteases: molecular approaches and industrial applications. *Applied Microbiology and Biotechnology*, **59**, 15–32.

90 Uehara, H., Yamane, K. and Maruo, B. (1979) Thermosensitive, extracellular neutral proteases in *Bacillus subtilis*: isolation, characterization, and genetics. *Journal of Bacteriology*, **139**, 583–590.

91 Tremacoldi, C.R., Watanabe, N.K. and Carmona, E.C. (2004) Production of extracellular acid proteases by *Aspergillus clavatus*. *World Journal of Microbiology & Biotechnology*, **20**, 639–642.

92 De Azeredo, L.A.I., De Lima, M.B., Coelho, R.R.R. and Freire, D.M.G. (2006) A low-cost fermentation medium for thermophilic protease production by *Streptomyces* sp. 594 using feather meal and corn steep liquor. *Current Microbiology*, **53**, 335–339.

93 Wintrode, P.L., Miyazaki, K. and Arnold, F.H. (2000) Cold adaptation of a mesophilic subtilisin-like protease by laboratory evolution. *The Journal of Biological Chemistry*, **275**, 31635–31640.

94 Villeret, V., Chessa, J.P., Gerday, C. and van Beeumen, J. (1997) Preliminary crystal structure determination of the alkaline protease from the Antarctic psychrophile *Pseudomonas aeruginosa*. *Protein Science*, **6**, 2462–2464.

95 Setyorini, E., Kim, Y.J., Takenaka, S., Murakami, S. and Aoki, K. (2006) Purification and characterization of a halotolerant intracellular protease from *Bacillus subtilis* strain FP-133. *Journal of Basic Microbiology*, **46**, 294–304.

96 Karbalaei-Heidari, H.R., Ziaee, A.A., Schaller, J. and Amoozegar, M.A. (2007) Purification and characterization of an extracellular haloalkaline protease produced by the moderately halophilic bacterium, *Salinivibrio* sp. strain AF-2004. *Enzyme and Microbial Technology*, **40**, 266–272.

97 Hultin, H.O. (1983) Current and potential uses of immobilized enzymes. *Food Technology*, **37**, 66–82.

98 Klein, M.D. and Langer, R. (1986) Immobilized enzymes in clinical medicine: an emerging approach to new drug therapies. *Trends in Biotechnology*, **4**, 179–186.

99 van Brunt, J. (1987) Biosensors for bioprocesses. *Biotechnology*, **5**, 437–440.

100 Chellapan, S., Jasmin, C., Basheer, S., Elyas, K., Bhat, S. and Chandrasekaran, M. (2006) Production, purification and partial characterization of a novel protease from *Engyodontium album* BTMFS 10 under solid state fermentation. *Process Biochemistry*, **41**, 956–961.

101 Barzana, E. and García-Garibay, M. (1994) Production of fish protein concentrates, in *Fisheries Processing* (ed. A.M. Martin), Chapman & Hall, London, pp. 206–222.

102 Nielsen, P.M. and Olsen, H.S. (2002) Enzymic modification of food protein, in *Enzymes in Food Technology* (eds R.J. Whitehurst and B.A. Law), CRC Press, Boca Raton, FL, pp. 109–143.

103 Tucker, G.A. and Woods, L.F.J. (1995) In *Enzymes in Food Processing*, Blackie, London, p. 336.

104 Uhlig, H. (1998) In *Industrial Enzymes and Their Applications*, Wiley-IEEE, New York, p. 454.

105 Ashie, I.N.A., Sorensen, T.L. and Nielsen, P.M. (2002) Effects of papain and a microbial enzyme on meat proteins and beef tenderness. *Journal of Food Science*, **67**, 2138–2142.

106 Monsan, P., Duteurtre, B., Moll, M. and Durand, G. (1978) Use of papain immobilized on spherosil for beer chillproofing. *Journal of Food Science*, **43**, 424–427.

107 Gebreselassie, P., Hoic, D., Masters, J.G. and Prencipe, M. (2002) Oral composition providing enhanced tooth stain removal. US Patent 6379654. 30 April 2002.

108 Pendzhiev, A.M. (2002) Proteolytic enzymes of papaya: medicinal applications. *Pharmaceutical Chemistry Journal*, **36**, 315–317.

109 Rowan, A.D., Butte, D.J. and Barrett, A.J. (1990) The cysteine proteinases of the pineapple plant. *The Biochemical Journal*, **266**, 869–875.

110 Taussig, S.J. and Batkin, S. (1988) Bromelain, the enzyme complex of pineapple (*Ananas comosus*) and its clinical application. An update. *Journal of Ethnopharmacology*, **22**, 191–203.

111 Gregory, S. and Kelly, N.D. (1996) Bromelain: a literature review and discussion of its therapeutic applications. *Alternative Medicine Review*, **1**, 243–257.

112 Bilton, G.L. (1984) Enzyme-containing digestive aid compositions. US Patent 4447412. 5 August 1984.

113 Imanaka, T., Shibazaki, M. and Takagi, M. (1986) A new way of enhancing the thermostability of proteases. *Nature*, **324**, 695–697.

114 Kim, S.Y., Gunasekaran, S. and Olson, N.F. (2004) Combined use of chymosin and protease from *Cryphonectria parasitica* for control of meltability and firmness of Cheddar cheese. *Journal of Dairy Science*, **87**, 274–283.

115 Budtz, P. (1994) Microbial rennets for cheese making. Novo-Nordisk Technical Report A-6045a.

116 van den Berg, G. (1992) Fermentation-produced chymosin; technological aspects of its use for cheesemaking. *Bulletin of the International Dairy federation*, **296**, 13–17.

117 Green, M.L. (1977) Review of the progress of dairy science: milk coagulants. *The Journal of Dairy Research*, **44**, 159–188.

118 Neelakantan, S., Mohanty, A.K. and Kaushik, J.K. (1999) Production and use of microbial enzymes for dairy processing. *Current Science*, **77**, 143–149.

119 Johnson, M.E. and Lucey, J.A. (2006) Major technological advances and trends in cheese production. *Journal of Dairy Science*, **89**, 1174–1178.

120 Dunn-Coleman, N.S., Bloebaum, P., Berka, R.M., Bodie, E., Robinson, N., Armstrong, G., Ward, M., Przetak, M., Carter, G.L. and LaCost, R. (1991) Commercial levels of chymosin production by *Aspergillus*. *Bio/Technology*, **9**, 976–981.

121 Tramper, J. (1994) Applied biocatalysis: from product request to idea to product, in *Applied Biocatalysis* (eds J.M.S. Cabral, D.B.L. Boros and J. Tramper), Harwood, Chur, pp. 1–45.

122 van Dijck, P.W.M. (1999) Chymosyn and phytase made by genetic engineering. *Journal of Biotechnology*, **67**, 77–80.

123 Mantafounis, M. and Pitts, J. (1990) Protein engineering of chymosin; modification of the optimum pH of enzyme catalysis. *Protein Engineering*, **3**, 605–609.

124 Mohanty, A.K., Mukhopadhyay, U.K., Grover, S. and Batish, V.K. (1999) Bovine chymosin production by rDNA technology and application in cheese manufacture. *Biotechnology Advances*, **17**, 205–217.

125 Maurer, K. (2004) Detergent proteases. *Current Opinion in Biotechnology*, **15**, 330–334.

126 Teplyakov, A.V., van der Laan, J.M., Lammers, A.A., Kelders, H., Kalk, K.H., Misset, O., Mulleners, L.J.S.M. and Dijkstra, B.W. (1992) Protein engineering of the high-alkaline serine protease PB92 from *Bacillus alcalophilus*: functional and structural consequences of mutation at the S4 substrate binding pocket. *Protein Engineering*, **5**, 413–420.

127 Tjalsma, H., Koetje, E.J., Kiewiet, R., Kuipers, O.P., Kolkman, M., Laan, J., Daskin, R., Ferrari, E. and Bron, S. (2004) Engineering of quorum-sensing systems for improved production of alkaline protease by *Bacillus subtilis*. *Journal of Applied Microbiology*, **96**, 569–578.

128 Bryan, P.N. (2000) Protein engineering of subtilisin. *Biochimica et Biophysica Acta*, **1543**, 223–238.

129 Ness, J.E., Kim, S., Gottman, A., Pak, R., Krebber, A., Borchert, T.V., Govindarajan, S., Mundorff, E.C. and Minshull, J. (2002) Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nature Biotechnology*, **20**, 1251–1255.

130 Legendre, D., Laraki, N., Grässlund, T., Bjørnvad, M.E., Bouchet, M., Nygren, P.Å., Borchert, T.V. and Fastrez, J. (2000) Display of active subtilisin 309 on phage: analysis of parameters influencing the selection of subtilisin variants with charged substrate specificity from libraries using phosphonylating inhibitors. *Journal of Molecular Biology*, **296**, 87–102.

131 Varela, H., Ferrari, M.D., Belobradjic, L., Vázquez, A. and Loperena, M.L. (1997) Skin unhairing proteases of *Bacillus subtilis*: production and partial characterization. *Biotechnology Letters*, **19**, 755–758.

132 Tang, X.M., Lakay, F.M., Shen, W., Shao, W.L., Fang, H.Y., Prior, B.A., Wang, Z.X. and Zhuge, J. (2004) Purification and characterisation of an alkaline protease used in tannery industry from *Bacillus licheniformis*. *Biotechnology Letters*, **26**, 1421–1424.

133 Thanikaivelan, P., Rao, J.R., Nair, B.U. and Ramasami, T. (2004) Progress and recent trends in biotechnological methods for leather processing. *Trends in Biotechnology*, **22**, 181–188.

134 Schaffeld, G., Bruzzone, P., Illanes, A., Curotto, E. and Aguirre, C. (1989) Enzymatic treatment of stickwater from fishmeal industry with the protease from *Cucurbita ficifolia*. *Biotechnology Letters*, **11**, 521–522.

135 Aguilera, J.M. (1994) Identifying opportunities for engineering applications in agriculture, food and nutrition biochemistry in Latin America, in *Bioengineering and Bioprocesses: Needs and Opportunities in Latin America* (eds J.M. Aguilera, R. San Martín and W. Edwardson), Universidad de Santiago, Santiago, pp. 9–27.

136 Priolo, N., Morcelle Del Valle, S., Arribére, M.C., López, L. and Caffini, N. (2000) Isolation and characterization of cysteine protease from the latex of *Araujia hortorum* fruits. *Journal of Protein Chemistry*, **19**, 9–49.

137 Atomi, H. (2005) Recent progress towards the application of hyperthermophiles and their enzymes. *Current Opinion in Chemical Biology*, **9**, 166–173.

138 Bordusa, F. (2002) Proteases in organic synthesis. *Chemical Reviews*, **102**, 4817–4867.

139 Capellas, M., Caminal, G., González, G., Lopez-Santín, J. and Clapés, P. (1997) Enzymatic condensation of cholecystokinin CCK-8 (4–6) and CCK-8 (7–8) peptide fragments in organic media. *Biotechnology and Bioengineering*, **56**, 456–463.

140 Björup, P., Adlercreutz, P. and Clapés, P. (1999) Useful methods in enzymatic synthesis of peptides: a comparative study focusing on kinetically controlled synthesis of Ac-Phe–Leu-NH_2 catalyzed by α-chymotrypsin. *Biocatalysis and Biotransformation*, **17**, 319–345.

141 Voyushina, T.L., Potetinova, J.V., Milgotina, E.I. and Stepanov, V.M. (1999) Synthesis of peptide aldehydes via enzymatic acylation of amino aldehyde derivatives. *Bioorganic & Medicinal Chemistry*, **7**, 2953–2959.

142 Fernandez, M.M., Margot, A.O., Falender, C.A., Blanch, H.W. and Clark, D.S. (1995) Enzymatic synthesis of peptides containing unnatural amino acids. *Enzyme and Microbial Technology*, **17**, 964–971.

143 Krix, G., Eichhorn, U., Jakubke, H.D. and Kula, M.R. (1997) Protease-catalyzed synthesis of new hydrophobic dipeptides containing non-proteinogenic amino acids. *Enzyme and Microbial Technology*, **21**, 252–257.

144 Broos, J., Engbersen, J.F.J., Sakodinskaya, I.K., Verboom, W. and Reinhoudt, D.N. (1995) Activity and enantioselectivity of serine proteases in transesterification reactions in organic media. *Journal of the Chemical Society, Perkin Transactions 1*, 2899–2905.

145 Khmelnitsky, Y.L., Budde, C., Arnold, J.M., Usyatinsky, A., Clark, D.S. and Dordick, J.S. (1997) Synthesis of water-soluble paclitaxel derivatives by enzymatic acylation. *Journal of the American Chemical Society*, **119**, 11554–11555.

146 Liu, C.F. and Tam, J.P. (2001) Subtilisin-catalyzed synthesis of amino acid and peptide esters. Application in a two-step enzymatic ligation strategy. *Organic Letters*, **3**, 4157–4159.

147 Roy, I. and Gupta, M.N. (2003) Non-thermal effects of microwaves on protease-catalyzed esterification and transesterification. *Tetrahedron*, **59**, 5431–5436.

148 Noritomi, H., Nishida, S. and Kato, S. (2007) Protease-catalyzed esterification of amino acid in water-miscible ionic liquid. *Biotechnology Letters*, **29**, 1509–1512.

149 Pogorevc, M. and Faber, K. (2000) Biocatalytic resolution of sterically hindered alcohols, carboxylic acids and esters containing fully substituted chiral centers by hydrolytic enzymes. *Journal of Molecular Catalysis B – Enzymatic*, **10**, 357–376.

150 Roberts, S.M. (2001) Preparative biotransformations. *Journal of the Chemical Society, Perkin Transactions 1*, 1475–1499.

151 Wong, C.H., Schuster, M., Wang, P. and Sears, P. (1993) Enzymic synthesis of *N*- and *O*-linked glycopeptides. *Journal of the American Chemical Society*, **115**, 893–5901.

152 Koeller, K.M. and Wong, C.H. (2000) Synthesis of complex carbohydrates and glycoconjugates: enzyme-based and programmable one-pot strategies. *Chemical Reviews*, **100**, 4465–4494.

153 Witte, K., Sears, P., Martin, R. and Wong, C.H. (1997) Enzymatic glycoprotein synthesis: preparation of ribonuclease glycoforms via enzymatic glycopeptide condensation and glycosylation. *Journal of the American Chemical Society*, **119**, 2114–2118.

154 Walsh, G. (2005) Therapeutic insulins and their large scale manufacture. *Applied Microbiology and Biotechnology*, **67**, 151–159.

155 Kimmerlin, T. and Seebach, D. (2005) 100 years of peptide synthesis: ligation methods for peptide and protein synthesis with applications to β-peptide assemblies. *The Journal of Peptide Research*, **65**, 229–260.

156 Kumar, D. and Bhalla, T. (2005) Microbial proteases in peptide synthesis: approaches and applications. *Applied Microbiology and Biotechnology*, **68**, 726–736.

157 Merrifield, B. (1986) Solid phase synthesis. *Science*, **232**, 341–347.

158 Nilsson, B., Soellner, M. and Raines, R. (2005) Chemical synthesis of proteins. *Annual Review of Biophysics and Biomolecular Structure*, **34**, 91–118.

159 Albericio, F. (2000) Orthogonal protecting groups for N^{α}-amino and C-terminal carboxyl functions in solid-phase synthesis. *Biopolymers (Peptide Science)*, **55**, 123–139.

160 Chan, W.C. and White, P.D. (2000) In *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*, Oxford University Press, Oxford, p. 341.

161 Barlos, K., Chatzi, O., Gatos, D. and Stavropoulos, G. (1991) Studies on anchoring of Fmoc-amino acids and peptide cleavage. *International Journal of Peptide and Protein Research*, **37**, 513–520.

162 Lloyd-Williams, P., Albericio, F. and Giralt, E. (1993) Convergent solid-phase peptide synthesis. *Tetrahedron*, **49**, 11065–11133.

163 Lloyd-Williams, P. and Giralt, E. (2000) Solid-phase convergent approaches to the synthesis of native peptides and proteins, in *Solid-Phase Synthesis: A Practical Guide* (eds S.A. Kates and F. Albericio), Marcel Dekker, New York, pp. 377–418.

164 Chiva, C., Barthe, P., Codina, A., Gairí, M., Molina, F., Granier, C., Pugnière, M., Inui, T., Nishio, H., Nishiuchi, Y., Kimura, T., Sakakibara, S., Albericio, F. and Giralt, E. (2003) Synthesis and NMR structure of P41icf, a potent inhibitor of human cathepsin L. *Journal of the American Chemical Society*, **125**, 1508–1517.

165 Bray, B.L. (2003) Large-scale manufacture of peptide therapeutics by chemical synthesis. *Nature Reviews. Drug Discovery*, **2**, 587–593.

166 Fillon, Y.A., Anderson, J.P. and Chmielewski, J. (2005) Cell penetrating agents based on a polyproline helix scaffold. *Journal of the American Chemical Society*, **127**, 11798–11803.

167 Sanclimens, G., Crespo, L., Giralt, E., Albericio, F. and Royo, M. (2005) Preparation of *de novo* globular proteins based on proline dendrimers. *The Journal of Organic Chemistry*, **70**, 6274–6281.

168 Baca, M., Muir, T.W., Schnölzer, M. and Kent, S.B.H. (1995) Chemical ligation of cysteine-containing peptides: synthesis of a 22 kDa tethered dimer of HIV-1 protease. *Journal of the American Chemical Society*, **117**, 1881–1887.

169 Yan, L.Z. and Dawson, P.E. (2001) Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *Journal of the American Chemical Society*, **123**, 526–533.

170 Lu, Y.A. and Tam, J.P. (2005) Peptide ligation by a reversible and reusable C-terminal thiol handle. *Organic Letters*, **7**, 5003–5006.

171 Nardin, E.H., Calvo-Calle, J.M., Oliveira, G.A., Clavijo, P., Nussenzweig, R., Simon, R., Zeng, W. and Rose, K., (1998) *Plasmodium falciparum* polyoximes: highly immunogenic synthetic vaccines constructed by chemoselective ligation of repeat B-cell epitopes and a universal T-cell epitope of CS protein. *Vaccine*, **16**, 590–600.

172 Tam, J.P., Lu, Y., Liu, C.F. and Shao, J. (1995) Peptide synthesis using unprotected peptides through orthogonal coupling methods. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 2485–12489.

173 Frank, R. (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports: principles and

applications. *Journal of Immunological Methods*, **267**, 13–26.

174 Houghten, R.A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen–antibody interaction at the level of individual amino acids. *Proceedings of the National Academy of Sciences of the United States of America*, **82**, 5131–5135.

175 Ambulos, N.P., Bibbs, L., Bonewald, L.F., Kates, S., Khatri, A., Medzihradszky, K.F. and Weintraub, S.T. (2000) Peptide analysis, in *Solid-Phase Synthesis: A Practical Guide* (eds S.A. Kates and F. Albericio), Marcel Dekker, New York, pp. 751–789.

176 Guzmán, F., Jaramillo, K., Salazar, L.M., Rivera, A. and Patarroyo, M.E. (2003) H-NMR structures of the *Plasmodum falciparum* 1758 erythrocyte binding peptide analogues and protection against malaria. *Life Sciences*, **71**, 2773–2785.

177 Ahn, J.M., Boyl, N.A., Macdonald, M.T. and Jan, K.D. (2002) Peptidomimetics and peptide backbone modifications. *Mini-Reviews in Medicinal Chemistry*, **2**, 463–473.

178 Hamel, E. and Covell, D.G. (2002) Antimitotic peptides and depsipeptides. *Current Medicinal Chemistry – Anti-Cancer Agents*, **2**, 19–53.

179 Sarabia, F., Chammaa, S., Ruiz, A.S., Ortiz, L.M. and Herrera, F.J.L. (2004) Chemistry and biology of cyclic depsipeptides of medicinal and biological interest. *Current Medicinal Chemistry*, **11**, 1309–1332.

180 Lioy, E., Suarez, J., Guzmán, F., Siegrist, S., Pluschke, G. and Patarroyo, M.E. (2001) Synthesis, biological, and immunological properties of cyclic peptides from *Plasmodium falciparum* merozoite surface protein-1. *Angewandte Chemie International Edition*, **40**, 2631–2635.

181 Calvo, J.C., Choconta, K.C., Diaz, D., Orozco, O., Bravo, M.M., Espejo, F., Salazar, L.M., Guzman, F. and Patarroyo, M.E. (2003) An alpha helix conformationally restricted peptide is recognized by cervical carcinoma patients' sera. *Journal of Medicinal Chemistry*, **46**, 5389–5394.

182 Yin, H., Frederick, K.K., Liu, D., Wand, A.J. and DeGrado, W.F. (2006) Arylamide derivatives as peptidomimetic inhibitors of calmodulin. *Organic Letters*, **8**, 223–225.

183 Radzishevsky, I.S., Rotem, S., Bourdetsky, D., Navon-Venezia, S., Carmeli, Y. and Mor, A. (2007) Improved antimicrobial peptides based on acyl-lysine oligomers. *Nature Biotechnology*, **25**, 657–659.

184 Barlos, K. and Gatos, D. (1999) 9-Fluorenylmethyloxycarbonyl/tbutyl-based convergent protein synthesis. *Biopolymers*, **51**, 266–278.

185 Johnson, E.C., Malito, E., Shen, Y., Pentelute, B., Rich, D., Florián, J., Tang, W.J. and Kent, S.B. (2007) Insights from atomic-resolution X-ray structures of chemically synthesized HIV-1 protease in complex with inhibitors. *Journal of Molecular Biology*, **373**, 573–586.

186 Bruckdorfer, T., Marder, O. and Albericio, F. (2004) From production of peptides in milligram amounts for research to multi-tons quantities for drugs of the future. *Current Pharmaceutical Biotechnology*, **5**, 29–43.

187 Patarroyo, J.H. and Guzmán, F. (2004) Peptideos sinteticos como vacinas, in *Biotecnologia e Saude* (eds M.R. De Almeida, A. Borem and G.R. Franco), Folha de Viçosa, Viçosa, pp. 113–139.

188 Guzmán, F., Barberis, S. and Illanes, A. (2007) Peptide synthesis: chemical or enzymatic. *Electronic Journal of Biotechnology*, **10**, 279–314.

189 Andersson, L., Blomberg, L., Flegel, M., Lepsa, L., Nilsson, B. and Verlander, M. (2000) Large-scale synthesis of peptides. *Biopolymers*, **55**, 227–250.

190 Barberis, S. and Illanes, A. (1996) Catálisis enzimática en fase orgánica. *Ingeniera Quimica (Madrid)*, **28**, 165–173.

191 Kasche, V. (1996) Proteases in peptide synthesis, in *Proteolytic Enzymes: A*

Practical Approach (eds R. Beynon and J. Nond), IRL Press, Oxford, pp. 125–143.

192 Filippova, I. and Lysogorskaia, E. (2003) Modified proteinases in peptide synthesis in organic media. *Bioorganicheskaia Khimiia*, **29**, 544–550.

193 Mesiano, A., Beckman, E. and Russell, A. (1999) Supercritical biocatalysis. *Chemical Reviews*, **99**, 623–634.

194 Karmee, S.K., Casiraghi, L. and Greiner, L. (2008) Technical aspects of biocatalysis in non-CO_2-based supercritical fluids. *Biotechnology Journal*, **3**, 104–111.

195 Gill, I. and Vulfson, E.N. (1993) Enzymatic synthesis of short peptides in heterogeneous mixtures of substrates. *Journal of the American Chemical Society*, **115**, 3348–3349.

196 Halling, P.J., Eichhorn, U., Kuhl, P. and Jakubke, H.D. (1995) Thermodynamics of solid-to-solid conversion and application to enzymic peptide synthesis. *Enzyme and Microbial Technology*, **17**, 601–606.

197 Gill, I. and Valivety, R. (2002) Pilot-scale enzymatic synthesis of bioactive oligopeptides in eutectic-based media. *Organic Process Research & Development*, **6**, 684–691.

198 Lombard, C., Saulnier, J. and Wallach, J. (2005) Recent trends in protease-catalyzed peptide synthesis. *Protein and Peptide Letters*, **12**, 621–629.

199 Schellenberger, V., Jakubke, H.D., Zapevalova, N.P. and Mitin, Y.V. (1991) Protease-catalyzed peptide synthesis using inverse substrates: the influence of reaction conditions on the trypsin acyl transfer efficiency. *Biotechnology and Bioengineering*, **38**, 104–108.

200 Kitaguchi, H. (1996) Peptide synthesis, in *Enzymatic Reactions in Organic Media* (eds A.M.P. Koskinen and A.M. Klibanov), Blackie, London, pp. 224–243.

201 Jakubke, H., Kuhl, P. and Konnecke, A. (1985) Basic principles of protease-catalyzed peptide bond formation. *Angewandte Chemie International Edition*, **24**, 85–93.

202 Halling, P.J. (1994) Thermodynamic predictions for biocatalysis in nonconventional media: theory, tests, and recommendations for experimental design and analysis. *Enzyme and Microbial Technology*, **16**, 178–206.

203 Cabral, J.M.S., Aires-Barros, M.R., Pinheiro, H.D. and Prazeres, M.F. (1997) Biotransformation in organic media by enzymes and whole cells. *Journal of Biotechnology*, **59**, 133–143.

204 Illanes, A., Anjarí, S., Arrieta, R. and Aguirre, C. (2002) Optimization of yield in the kinetically controlled synthesis of ampicillin with immobilized penicillin acylase in organic media. *Applied Biochemistry and Biotechnology*, **97**, 165–179.

205 Illanes, A., Cabrera, Z., Wilson, L. and Aguire, C. (2003) Synthesis of cephalexin in ethylene glycol with glyoxyl-agarose immobilised penicillin acylase: temperature and pH optimisation. *Process Biochemistry*, **39**, 111–117.

206 Quiroga, E., Priolo, N., Marchese, J. and Barberis, S. (2005) Behaviour of araujiain, a new cystein phytoprotease, in organic media with low water content. *Electronic Journal of Biotechnology*, **9**, 18–25.

207 Quiroga, E., Priolo, N., Obregón, D., Marchese, J. and Barberis, S. (2008) Peptide synthesis in aqueous–organic media catalyzed by proteases from latex of *Araujia hortorum* (Asclepiadaceae) fruits. *Biochemical Engineering Journal*, **39**, 115–120.

208 Clapés, P., Adlercreutz, P. and Mattiasson, B. (1990) Enzymatic peptide synthesis in organic media: a comparative study of water-miscible and water-immiscible solvent systems. *Journal of Biotechnology*, **15**, 323–338.

209 Castro, R. (2000) Properties of soluble α-chymotrypsin in neat glycerol and water. *Enzyme and Microbial Technology*, **27**, 143–150.

210 Szabó, A., Kotormán, M., Laczkó, I. and Simon, L.M. (2006) Spectroscopic studies of stability of papain in aqueous organic

solvents. *Journal of Molecular Catalysis B – Enzymatic*, **41**, 43–48.

211 Ogino, H., Uchiho, T., Doukyu, N., Yasuda, M., Ishimi, K. and Ishikawa, H. (2007) Effect of exchange of amino acid residues on the surface region of the PST-01 protease on its organic solvent-stability. *Biochemical and Biophysical Research Communications*, **358**, 1028–1033.

212 Levitsky, V.Y., Lozano, P. and Iborra, J.L. (1999) Kinetic analysis of deactivation of immobilized α-chymotripsin by water-miscible organic solvent in kyotorphin synthesis. *Biotechnology and Bioengineering*, **65**, 170–175.

213 Barberis, S., Quiroga, E., Arribére, M.C. and Priolo, N. (2002) Peptide synthesis in aqueous–organic biphasic systems catalyzed by a protease isolated from *Morrenia brachystephana* (*Asclepiadaceae*). *Journal of Molecular Catalysis B – Enzymatic*, **17**, 39–47.

214 Barberis, S., Quiroga, E., Morcelle, S., Priolo, N. and Luco, J.M. (2006) Study of phytoproteases stability in aqueous–organic biphasic systems using linear free energy relationships. *Journal of Molecular Catalysis B – Enzymatic*, **38**, 95–103.

215 Guisán, J.M., Polo, E., Aguado, J., Romero, M.D., Alvaro, G. and Guerra, M.J. (1997) Immobilization-stabilization of thermolysin onto activated agarose gels. *Biocatalysis and Biotransformation*, **15**, 159–173.

216 Illanes, A. and Fajardo, A. (2001) Kinetically controlled synthesis of ampicillin with immobilized penicillin acylase in the presence of organic cosolvents. *Journal of Molecular Catalysis B – Enzymatic*, **11**, 587–595.

217 Illanes, A., Anjarí, S., Altamirano, C. and Aguirre, C. (2004) Optimization of cephalexin synthesis with immobilized penicillin acylase in ethylene glycol medium at low temperatures. *Journal of Molecular Catalysis B – Enzymatic*, **30**, 95–103.

218 Hou, R., Yang, Y., Li, G., Huang, Y.B., Wang, H., Liu, Y.J., Xu, L. and Zhang, X.Z. (2006) Synthesis of a precursor dipeptide of RGDS (Arg–Gly–Asp–Ser) catalysed by the industrial protease alcalase. *Biotechnology and Applied Biochemistry*, **44**, 73–80.

219 Wong, C.H. and Wang, K.T. (1991) New developments in enzymatic peptide synthesis. *Cellular and Molecular Life Sciences*, **47**, 1123–1129.

220 Wubbolts, M.G., Favre-Bulle, O. and Withhold, B. (1996) Biosynthesis of synthons in two-liquid-phase media. *Biotechnology and Bioengineering*, **52**, 301–308.

221 Miyanaga, M., Ohmori, M., Imamura, K., Sakiyama, T. and Nakanishi, K. (2000) Kinetics and equilibrium for thermolysin-catalyzed syntheses of dipeptide precursors in aqueous/organic biphasic systems. *Journal of Bioscience and Bioengineering*, **90**, 43–51.

222 Murakami, Y., Yoshida, T., Hayashi, S. and Hirata, A. (2000) Continuous enzymatic production of peptide precursor in aqueous/organic biphasic medium. *Biotechnology and Bioengineering*, **69**, 57–65.

223 Trusek-Holownia, A. (2003) Synthesis of *Z*-Ala–Phe.OMe, the precursor of bitter dipeptide in the two-phase ethyl acetate–water system catalysed by thermolysin. *Journal of Biotechnology*, **102**, 153–163.

224 Feliú, J., De Mas, C. and López-Santín, J. (1995) Studies on papain action in the synthesis of Gly–Phe in two-liquid-phase media. *Enzyme and Microbial Technology*, **17**, 882–887.

225 Clapés, P., Espelt, L., Navarro, M. and Solans, C. (2001) Highly concentrated water-in-oil emulsions as novel reaction media for protease-catalysed kinetically controlled peptide synthesis. *Journal of the Chemical Society, Perkin Transactions 2*, 1394–1399.

226 Gómez-Puyou, M.T. and Gómez-Puyou, A. (1998) Enzymes in low water systems.

Critical Reviews in Biochemistry and Molecular Biology, **33**, 53–89.

227 Chen, Y.X., Zhang, X.Z., Chen, S.M., You, D.L., Wu, X.X., Yang, X.C. and Guan, W.Z. (1999) Kinetically controlled synthesis catalyzed by proteases in reverse micelles and separation of precursor dipeptides of RGD. *Enzyme and Microbial Technology*, **25**, 310–315.

228 Xing, G.W., Liu, D.J., Ye, Y.H. and Ma, J.M. (1999) Enzymatic synthesis of oligopeptides in mixed reverse micelles. *Tetrahedron Letters*, **40**, 1971–1974.

229 Sergeeva, M.V., Paradkar, V.M. and Dordick, J.S. (1997) Peptide synthesis using proteases dissolved in organic solvents. *Enzyme and Microbial Technology*, **20**, 623–628.

230 Altreuter, D.H., Dordick, J.S. and Clark, D.S. (2003) Solid-phase peptide synthesis by ion-paired α-chymotrypsin in nonaqueous media. *Biotechnology and Bioengineering*, **81**, 809–817.

231 Manabe, K., Mori, Y., Wakabayashi, T., Nagayama, S. and Kobayashi, S. (2000) Organic synthesis inside particles in water: Lewis acid–surfactant-combined catalysts for organic reactions in water using colloidal dispersions as reaction media. *Journal of the American Chemical Society*, **122**, 7202–7207.

232 Zaks, A. and Klibanov, A.M. (1984) Enzymatic catalysis in organic media at 100 °C. *Science*, **224**, 1249–1251.

233 Wescott, C.R. and Klibanov, A.M. (1994) The solvent dependence of enzyme specificity. *Biochimica et Biophysica Acta*, **1206**, 1–9.

234 Ebert, C., Gardossi, L., Linda, P. and Vesnaver, R. (1996) Influence of organic solvents on enzyme chemoselectivity and their role in enzyme–substrate interaction. *Tetrahedron*, **52**, 4867–4876.

235 Kawashiro, K., Sugahara, H., Sugiyama, S. and Hayashi, H. (1997) Effect of organic solvents on enantioselectivity of protease catalysis. *Biotechnology and Bioengineering*, **53**, 26–31.

236 Dickinson, M. and Fletcher, P. (1989) Enzymes in organic solvents. *Enzyme and Microbial Technology*, **11**, 55–56.

237 Schulze, B. and Klibanov, A.M. (1991) Inactivation and stabilization of subtilisins in neat organic solvents. *Biotechnology and Bioengineering*, **38**, 1001–1006.

238 Zhou, Y.Y., Yang, T., Wang, N., Xu, L., Huang, Y.B., Wu, X.X., Yang, X.C. and Zhang, X.Z. (2003) Chemo-enzymatic synthesis of tripeptide RGD in organic solvents. *Enzyme and Microbial Technology*, **33**, 55–61.

239 Basso, A., De Martin, L., Ebert, C., Gardossi, L. and Linda, P. (2000) High isolated yields in thermodynamically controlled peptide synthesis in toluene catalysed by thermolysin adsorbed on Celite R-640. *Chemical Communications*, 467–468.

240 Schmitke, J.L., Stern, L.J. and Klibanov, A.M. (1997) The crystal structure of subtilisin Carlsberg in anhydrous dioxane and its comparison with those in water and acetonitrile. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 4250–4255.

241 Klibanov, A. (1997) Why are enzymes less active in organic solvents than in water? *Trends in Biotechnology*, **15**, 97–101.

242 Zaks, A. and Klibanov, A.M. (1988) Enzymatic catalysis in non-aqueous solvents. *The Journal of Biological Chemistry*, **263**, 3194–3201.

243 Capellas, M., Caminal, G., González, G., Lopez-Santín, J. and Clapés, P. (1996) Enzymic synthesis of a CCK-8 tripeptide fragment in organic media. *Biotechnology and Bioengineering*, **50**, 700–708.

244 Theppakorn, T., Kanawasud, P. and Halling, P. (2004) Activity of immobilized papain dehydrated by *n*-propanol in low-water media. *Biotechnology Letters*, **26**, 133–136.

245 Gill, I., López-Fandiño, R., Jorba, X. and Vulfson, E.N. (1996) Biologically active peptides and enzymatic approaches to

their production. *Enzyme and Microbial Technology*, **18**, 162–183.

246 Barros, R.J., Wehtje, E. and Adlercreutz, P. (1998) Mass transfer studies on immobilized chymotrypsin biocatalysts prepared by deposition for use in organic medium. *Biotechnology and Bioengineering*, **59**, 364–373.

247 Gorman, L.A.S. and Dordick, J.S. (1992) Organic solvents strip water off enzymes. *Biotechnology and Bioengineering*, **39**, 392–397.

248 van Unen, D., Engbersen, F. and Reinhoudt, D. (1998) Large acceleration of α-chymotrypsin-catalyzed dipeptide formation by 18-crown-6 in organic solvents. *Biotechnology and Bioengineering*, **59**, 553–556.

249 van Unen, D., Engbersen, F. and Reinhoudt, D. (2001) Studies on the mechanism of crown-ether-induced activation of enzymes in non-aqueous media. *Biotechnology and Bioengineering*, **75**, 154–158.

250 van Unen, D., Engbersen, F. and Reinhoudt, D. (2002) Why do crown ethers activate enzymes in organic solvents. *Biotechnology and Bioengineering*, **77**, 248–255.

251 De Santis, G. and Jones, J.B. (1999) Chemical modification of enzymes for enhanced functionality. *Current Opinion in Biotechnology*, **10**, 324–330.

252 Hudson, E.P., Eppler, R.K. and Clark, D.S. (2005) Biocatalysis in semi-aqueous and nearly anhydrous conditions. *Current Opinion in Biotechnology*, **16**, 637–643.

253 Sánchez-Porro, C., Mellado, E., Bertoldo, C., Antranikian, G. and Ventosa, A. (2003) Screening and characterization of the protease CP1 produced by the moderately halophilic bacterium *Pseudoalteromonas* sp. strain CP76. *Extremophiles: Life Under Extreme Conditions*, **7**, 221–228.

254 Ferrer, M., Martínez-Abarca, F. and Golyshin, P.N. (2005) Mining genomes and "metagenomes" for novel catalysts. *Current Opinion in Biotechnology*, **16**, 588–593.

255 Yun, J. and Ryu, S. (2005) Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. *Microbial Cell Factories*, **4**, 8.

256 Sears, P.S. and Clark, D.S. (1993) Comparison of soluble and immobilized trypsin kinetics: implications for peptide synthesis. *Biotechnology and Bioengineering*, **42**, 118–124.

257 Afaq, S., Muslim, A. and Iqbal, J. (2001) Immobilization and stabilization of papain on chelating sepharose: a metal chelate regenerable carrier. *Electronic Journal of Biotechnology*, **4**, 1–2.

258 Magnin, D., Dumitriu, S. and Chornet, E. (2003) Immobilization of enzymes into a polyionic hydrogel: ChitoXan. *Journal of Bioactive and Compatible Polymers*, **18**, 355–373.

259 Moeschel, K., Nouaimi, M., Steinbrenner, C. and Bisswanger, H. (2003) Immobilization of thermolysin to polyamide nonwoven materials. *Biotechnology and Bioengineering*, **82**, 190–199.

260 Lei, H., Wang, W., Chen, L.L., Li, X.C., Yi, B. and Deng, L. (2004) The preparation of catalytically active papain immobilized on magnetic composite microspheres. *Enzyme and Microbial Technology*, **35**, 15–21.

261 Rao, S.R., Borkar, P.S., Khobragade, C.N. and Sagar, A.D. (2006) Enzymatic activities of proteases immobilized on tri (4-formyl phenoxy) cyanurate. *Enzyme and Microbial Technology*, **39**, 958–962.

262 Cao, L., van Rantwijk, F. and Sheldon, R. (2000) Cross-linked enzyme aggregates: a simple and effective method for the immobilization of penicillin acylase. *Organic Letters*, **2**, 1361–1364.

263 Cao, L., van Langen, L.M., van Rantwijk, F. and Sheldon, R.A. (2001) Cross-linked aggregates of penicillin acylase: robust biocatalysts for the synthesis of β-lactam antibiotics. *Journal of Molecular Catalysis B – Enzymatic*, **11**, 665–670.

264 Wilson, L., Illanes, A., Abián, O., Pessela, B.C.C., Fernández-Lafuente, R. and

Guisán, J.M. (2004) Co-aggregation of penicillin G acylase and polyionic polymers: an easy methodology to prepare enzyme biocatalysts stable in organic media. *Biomacromolecules*, **5**, 852–857.

265 Davis, B.G. and Boyer, V. (2001) Biocatalysis and enzymes in organic synthesis. *Natural Product Reports*, **18**, 618–640.

266 Abrahamsen, L., Tom, J., Burnier, J., Butcher, K.A., Kossakoff, A. and Wells, J.A. (1991) Engineering subtilisin and its substrates for efficient ligation of peptide bond in aqueous solution. *Biochemistry*, **30**, 4151–4159.

267 Takagi, H., Hirai, K., Maeda, Y., Matsuzawa, H. and Nakamori, S. (2000) Engineering subtilisin E for enhanced stability and activity in polar organic solvents. *Journal of Biochemistry (Tokyo)*, **127**, 617–625.

268 Ogino, H., Uchiho, T., Yokoo, J., Kobayashi, R., Ichise, R. and Ishikawa, H. (2001) Role of intermolecular disulfide bonds of the organic solvent stable PST-01 protease in its organic solvent stability. *Applied and Environmental Microbiology*, **67**, 942–947.

269 Thurst, S. and Koksche, B. (2003) Protease-catalysed peptide synthesis for the site specific incorporation of alpha-fluoroalkyl amino acids into peptides. *The Journal of Organic Chemistry*, **68**, 2290–2296.

270 Pulido, M.A., Koga, Y., Takano, K. and Kanaya, S. (2007) Directed evolution of Tk-subtilisin from a hyperthermophilic archaeon: identification of a single amino acid substitution responsible for low-temperature adaptation. *Protein Engineering Design and Selection*, **20**, 143–153.

271 Varadarajan, N., Gam, J., Olsen, M.J., Georgiou, G. and Iverson, B.L. (2005) Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6855–6860.

272 Miyazawa, T., Tanaka, K., Ensatsu, E., Yanagihara, R. and Yamada, T. (2001) Broadening of the substrate tolerance of α-chymotrypsin by using the carbamoylmethyl ester as an acyl donor in kinetically controlled peptide synthesis. *Journal of the Chemical Society, Perkin Transactions 1*, 87–93.

273 Bordusa, F. (2005) Protease-catalyzed formation of C– N bonds, in *Highlights in Bioorganic Chemistry* (eds C. Schmuck and H. Wennemers), John Wiley & Sons, Inc., New York, pp. 387–403.

274 Fischer, U., Zeitschel, U. and Jakubke, H. (1991) Chymotrypsin-catalyzed peptide synthesis in an acetonitrile–water–system: studies on the efficiency of nucleophiles. *Biomedica Biochimica Acta*, **50**, S131–S135.

275 Thormann, M., Thust, S., Hofmann, H.J. and Bordusa, F. (1999) Protease-catalyzed hydrolysis of substrate mimetics (inverse substrates): a new approach reveals a new mechanism. *Biochemistry*, **38**, 6056–6062.

276 Bordusa, F., Ullmann, D., Elsner, C. and Jakubke, H.D. (1997) Substrate mimetics mediated peptide synthesis: an irreversible ligation strategy that is independent of substrate specificity. *Angewandte Chemie International Edition*, **36**, 2473–2475.

277 Wehofsky, N., Wissmann, J.D., Alisch, M., Bordusa, F. *et al.* (2000) Engineering of substrate mimetics as novel-type substrates for glutamic acid-specific endopeptidases: design, synthesis, and application. *Biochimica et Biophysica Acta*, **1479**, 114–122.

278 Grunberg, R., Domgall, I., Gunther, R., Rall, K., Hofman, H.J. and Bordusa, F. (2000) Peptide bond formation mediated by substrate mimetics. *European Journal of Biochemistry*, **267**, 7024–7030.

279 Cheetham, P.S.J. (1994) Case studies in applied biocatalysis-from ideas to products, in *Applied Biocatalysis* (eds J.M.S. Cabral D.B.L. Boros and J. Tramper), Harwood, Chur, pp. 47–108.

280 Murakami, Y. and Hirata, A. (1997) Continuous enzymatic synthesis of aspartame precursor at low pH using an extractive reaction. *Journal of Fermentation and Bioengineering*, **84**, 264–267.

281 Isono, Y. and Nakajima, M. (1999) Application of highly concentrated substrate to enzymatic synthesis of aspartame precursor. *Food Science and Technology Research*, **5**, 255–258.

282 Isono, Y. (2002) Enzymatic synthesis of aspartame precursor in solvent-free reaction system. *Journal of the Japanese Society for Food Science and Technology*, **49**, 813–817.

283 López-Fandiño, R., Gill, I. and Vulfson, E.N. (1994) Protease-catalyzed synthesis of oligopeptides in heterogenous substrate mixtures. *Biotechnology and Bioengineering*, **43**, 1024–1030.

284 Clapés, P., Camina, G., Feliu, J.A. and López-Santín, J. (2000) Peptide synthesis in non-aqueous media, in *Methods in Non-Aqueous Enzymology* (ed. M.N. Gupta), Birkhäuser, Basel, pp. 110–132.

285 Schwarz, A., Wandrey, C., Steinke, D. and Kula, M.R. (1992) A two-step enzymatic synthesis of dipeptides. *Biotechnology and Bioengineering*, **39**, 132–140.

286 Sareen, R., Bornscheuer, U.T. and Mishra, P. (2004) Synthesis of kyotorphin precursor by an organic solvent-stable protease from *Bacillus licheniformis* RSP-09-37. *Journal of Molecular Catalysis B – Enzymatic*, **32**, 1–5.

287 Kimura, Y., Nakanishi, K. and Matsuno, R. (1990) Enzymatic synthesis of precursor of Leu-enkephalin in water-immiscible organic solvent systems. *Enzyme and Microbial Technology*, **12**, 272–280.

288 Richards, A., Gill, I. and Vulfson, E.N. (1993) Continuous enzymatic production of oligopeptides: synthesis of an enkephalin pentapeptide in a multistage bioreactor. *Enzyme and Microbial Technology*, **15**, 928–935.

289 Zhang, X.Z., Wang, X., Chen, S., Fu, X., Wu, X. and Li, C. (1996) Protease-catalyzed small peptide synthesis in organic media. *Enzyme and Microbial Technology*, **19**, 538–544.

290 Fité, M., Clapés, P., López-Santín, J., Benaiges, M.D. and Caminal, G. (2002) Integrated process for the enzymatic synthesis of the octapeptide PhAcCCK-8. *Biotechnology Progress*, **18**, 1214–1220.

291 Meng, L., Joshi, R. and Eckstein, H. (2006) Application of enzymes for the synthesis of the cholecystokinin pentapeptide (CCK-5): a contribution to green chemistry. *Chimica Oggi – Chemistry Today*, **24**, 50–53.

292 Schwartz, G.J., Netterville, L.A., McHugh, P.R. and Moran, T.H. (1991) Gastric loads potentiate inhibition of food intake produced by a cholecystokinin analogue. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, **261**, R1141–R1146.

293 Rizo, J. and Gierasch, L.M. (1992) Constrained peptides: models of bioactive peptides and protein structures. *Annual Review of Biochemistry*, **61**, 387–418.

294 Kullmann, W. (1982) Enzymatic synthesis of dynorphin (1–8). *The Journal of Organic Chemistry*, **47**, 5300–5303.

295 Jorba, X., Clapés, P., Xaus, N., Clavet, S., Torres, J.L., Valencia, G. and Mata, J. (1992) Optimization and kinetic studies of the enzymatic synthesis of Ac-Phe–Leu-NH_2 in reverse micelles. *Enzyme and Microbial Technology*, **14**, 117–124.

296 Zaks, A. and Dodds, D. (1997) Application of biocatalysis and biotransformations to the synthesis of pharmaceuticals. *Drug Discovery Today*, **2**, 513–531.

297 Liu, P., Tian, G.L., Lee, K.S., Wong, M.S. and Ye, Y.H. (2002) Full enzymatic synthesis of a precursor of bioactive pentapeptide OGP (10–14) in organic solvents. *Tetrahedron Letters*, **43**, 2423–2425.

298 Sinisterra, J. and Alcantara, A. (1993) Synthesis of peptides catalyzed by enzymes: a practical overview. *Journal of Molecular Catalysis*, **84**, 327–364.

299 Ruiz, S., Feliú, J.A., Caminal, G., Alvaro, G., López-Santín, J. *et al*. (1997) Reaction

engineering for consecutive enzymatic reactions in peptide synthesis: application to the synthesis of a pentapeptide. *Biotechnology Progress*, **13**, 783–787.

300 Xiang, H., Xiang, G., Lu, Z., Guo, L. and Eckstein, H. (2004) Total enzymatic synthesis of cholecystokinin CCK-5. *Amino Acids*, **27**, 101–105.

301 Tsuchiyama, S., Doukyu, N., Yasuda, M., Ishimi, K. and Ogino, H. (2007) Peptide synthesis of aspartame precursor using organic-solvent-stable PST-01 protease in monophasic aqueous–organic solvent systems. *Biotechnology Progress*, **23**, 820–823.

302 Illanes, A., Wilson, L., Corrotea, O., Tavernini, L., Zamorano, F. and Aguirre, C. (2007) Synthesis of cephalexin with immobilized penicillin acylase at very high substrate concentrations in fully aqueous medium. *Journal of Molecular Catalysis B – Enzymatic*, **47**, 72–78.

303 Erbeldinger, M., Ni, X. and Halling, P.J. (1998) Effect of water and enzyme concentration on thermolysin-catalyzed solid-to-solid peptide synthesis. *Biotechnology and Bioengineering*, **59**, 68–72.

304 Erbeldinger, M., Ni, X. and Halling, P.J. (1998) Enzymatic synthesis with mainly undissolved substrates at very high concentrations. *Enzyme and Microbial Technology*, **23**, 141–148.

305 Chaiwut, P., Kanasawud, P. and Halling, P.J. (2007) Solid-to-solid peptide synthesis by glycyl endopeptidase. *Enzyme and Microbial Technology*, **40**, 954–960.

306 Hou, R., Zhang, N., Li, G., Huang, Y.B., Wang, H., Xiao, Y.P., Liu, Y.J., Yang, Y., Zhao, L. and Zhang, X.Z. (2005) Synthesis of the tripeptide RGD amide by a combination of chemical and enzymatic methods. *Journal of Molecular Catalysis B – Enzymatic*, **37**, 9–15.

9
Semisynthetic Enzymes

Usama M. Hegazy and Bengt Mannervik

9.1
Usefulness of Semisynthetic Enzymes

Natural enzymes play essential roles in the living cell by accelerating the rate of chemical processes. High catalytic efficiency, substrate specificity, stereoselectivity, and lack of toxicity are properties of enzymes that make them valuable in a multitude of applications, such as the production of fine chemicals [1, 2] and pharmaceuticals [3, 4], food processing [5], bioremediation [6], and mining of metals [7, 8]. Nevertheless, major obstacles in the use of natural enzymes in biotechnology are limitations in solubility, stability, and efficiency under nonphysiological conditions, including the presence of organic solvents or oxidants, extremes of temperature, pressure, or pH. Conventional genetic engineering of enzymes has provided solutions for these obstacles in some cases [9–11]. However, the repertoire of functional groups provided by the genetic code is restricted to the assortment of 20 canonical proteinogenic amino acids [phenylalanine (Phe), leucine (Leu), isoleucine (Ile), methionine (Met), valine (Val), serine (Ser), proline (Pro), threonine (Thr), alanine (Ala), tyrosine (Tyr), histidine (His), glutamine (Gln), lysine (Lys), asparagine (Asn), aspartic acid (Asp), glutamic acid (Glu), cysteine (Cys), tryptophan (Trp), arginine (Arg), and glycine (Gly)], plus selenocysteine (Sec) [12] and pyrrolysine (Pyl) [13]. For many applications it could be advantageous to go beyond the conventional building blocks and incorporate novel chemical groups in available enzymes. In the cell, enzyme functionalities can be expanded by post-translational chemical modifications and incorporation of metal ions or organic cofactors. In protein engineering, semisynthetic enzymes are derived from natural protein scaffolds, which are chemically modified in ways that mimic or go beyond what cellular systems provide. We define a semisynthetic enzyme as an entity that has been procured by insertion of novel chemical groups via genetic engineering or chemical methods in the laboratory. A rigorous demarcation between natural post-translational modification and semisynthesis is difficult to define, but semisynthesis generally implies that the product is not naturally occurring.

The first known attempt to construct active and intact semisynthetic proteins was described by Cowie and Cohen in 1957 [14]. They replaced the amino acid Met in

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

proteins with the analog, selenomethionine (Sem), by growing an *Escherichia coli* strain auxotrophic for Met in a Sem-containing growth medium. In this manner they forced the cells to make an amino acid mis-incorporation by supplementation of the growth medium with the desired noncanonical amino acid in the absence of Met. Thus this method is called "sense codon reassignment by selective pressure incorporation". Selenium-containing proteins are useful in structural analyses, especially by X-ray diffraction and nuclear magnetic resonance spectroscopy.

Before the development of directed mutagenesis in protein engineering, Chapeville *et al*. succeeded in incorporating Ala instead of Cys by chemically modifying the previously cysteinylated transfer RNA (tRNA) [15]. This seminal work demonstrated that tRNAs charged with noncognate amino acids are accepted by the ribosome and that the translation of codons in the messenger RNA (mRNA) sequence is governed by anticodons of the tRNAs rather than by the amino acid attached. Thus, conventional proteinogenic amino acids can be replaced with noncanonical amino acids. This finding formed the basis for later developments of site-specific incorporation of novel amino acid derivatives into proteins by genetic engineering. In 1966, the groups of Bender [16] and Koshland [17] independently succeeded in converting the active site Ser of subtilisin into Cys by straightforward chemical modification. The thiol-subtilisin obtained displayed very low peptidase activity, contrary to expectation. However, the chemically modified enzyme showed enhanced peptide ligase activity [18].

These early accomplishments laid the foundation for development of many techniques currently used in the construction of semisynthetic proteins. Applied to enzymes, semisynthesis may include covalent incorporation of unnatural amino acid residues, or even peptides, as well as other chemical structures into naturally occurring enzymes. The nonproteinogenic parts of semisynthetic enzymes can be incorporated prior to (pretranslational modification) or after mRNA translation (post-translational modification). The pretranslational modification techniques include sense codon reassignment, missense suppression, nonsense suppression, frame-shift suppression (use of four- or five-base codons), and noncanonical base pair techniques, whereas the post-translational modification techniques include chemical peptide synthesis (CPS), chemical ligation, and expressed protein ligation (EPL), as well as chemical modification of protein side-chains. These various methods are used not only to obtain enzymes with novel functions and properties [19, 20], but they can also be employed in mechanistic studies and in mapping the functional structures of enzymes [21–24]. In this chapter, protein semisynthesis will be described with comments on the advantages and limitations of each technique. Also, recent examples of the methods showing the wide-ranging applications of the semisynthetic enzymes will be given.

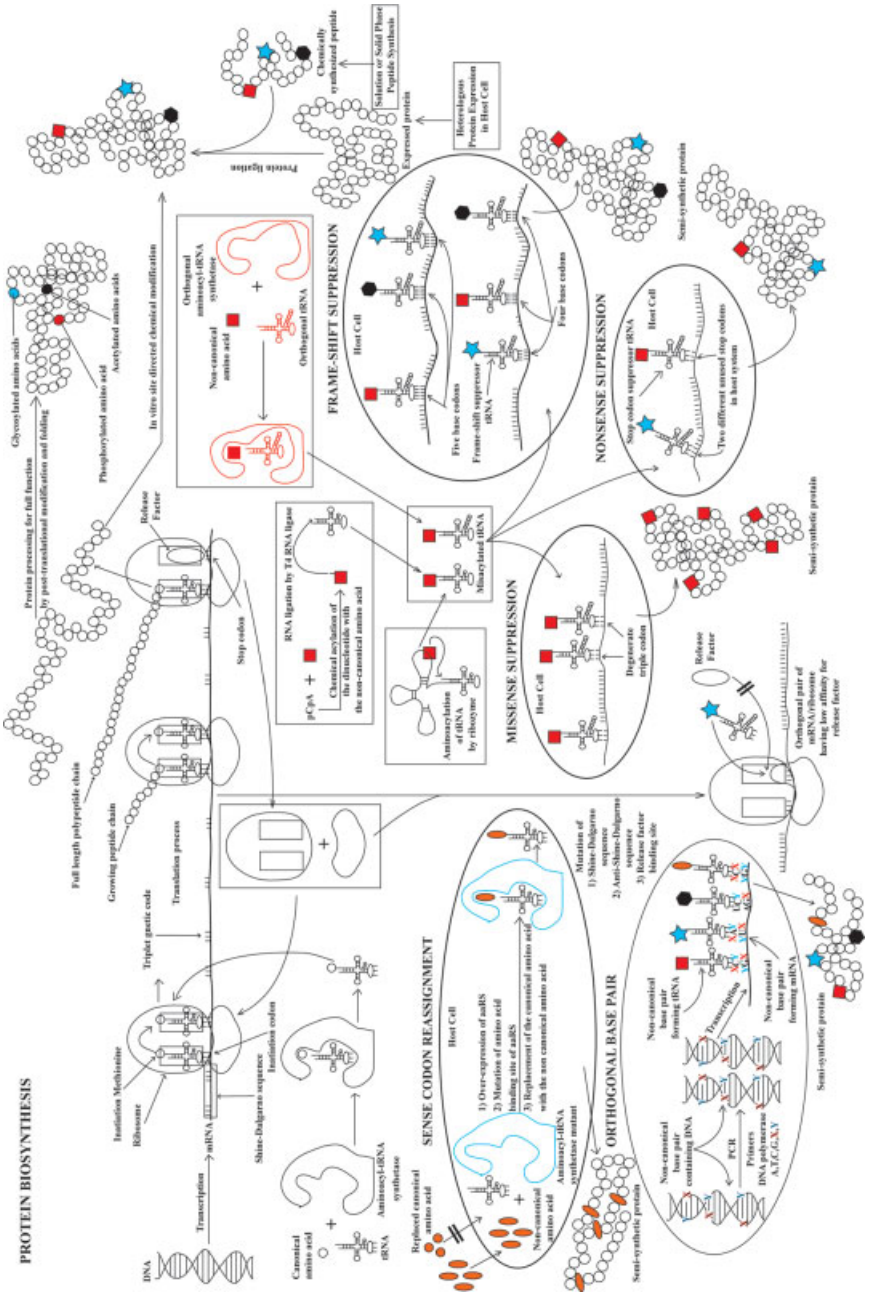
9.2
Natural Protein Biosynthesis

Owing to the close relationship between the biosynthesis of natural protein and the approaches used in the synthesis of semisynthetic protein, we will review the protein biosynthesis process before going through the different versions of semisynthetic

protein synthesis (Figure 9.1). DNA, which stores the structural information of the protein structure, contains four different bases: two purine bases, adenine (A) and guanine (G), and two pyrimidine bases, thymine (T) and cytosine (C). RNA, which translates the information stored in DNA to protein, contains the same bases as DNA except that uracil (U) replaces T. The bases represent the letters of genetic code and form the pairs A–T (or A–U) and G–C governed by the hydrogen-bonding patterns and structural complementarities. The genetic code consists of 64 three-base codons and 61 of these encode the 20 canonical amino acids and are called sense codons. The additional three codons, UAG (amber), UGA (opal), and UAA (ochre), are termination codons, alternatively called "stop" or nonsense codons. All the canonical amino acids, except Met and Trp, are encoded by more than one codon – a property known as the degeneracy of the code. The DNA is transcribed into mRNA, which subsequently is translated into proteins. Protein synthesis requires tRNA, which carries an amino acid to the ribosome, where protein synthesis takes place. The tRNA is charged with an amino acid corresponding to its anticodon (three nucleobases complementary to the codon of the mRNA). The charging of a tRNA with its cognate amino acid is carried out by a specific aminoacyl-tRNA synthetase (aaRS) forming aminoacyl-tRNA (aa-tRNA). Consequently, an aaRS exists for each amino acid [25, 26]. Although anticodons of tRNAs are general for all organisms, aaRS are often species-specific, such that the enzyme from one species does not aminoacylate tRNA from another species [27]. This functional orthogonality is useful in relation to the incorporation of unnatural amino acids into proteins. Translation of mRNA into proteins is carried out by the ribosome in four steps: initiation, elongation, termination, and recycling. In the first step, an initiation complex is formed between the ribosome, mRNA, initiator tRNA (a tRNA charged with Met), initiation factors, and guanosine triphosphate. In the second step, an aa-tRNA is transported by an elongation factor to the mRNA in the ribosome, where the anticodon of the aa-tRNA is matched with the codon of the mRNA. The cognate interaction between anticodon and codon leads to peptide bond formation between the incoming amino acid and Met of the initiator tRNA in the ribosome, thereby generating a dipeptide. The elongation cycle proceeds until a stop codon in the mRNA is encountered, which triggers association of a release factor (RF) to the ribosome. The RF–ribosome interaction leads to hydrolysis of the ester bond between the C-terminal amino acid of the peptide chain and the last aa-tRNA encountered by the ribosome. Consequently, the polypeptide chain is released from the ribosome and folded into a functional structure. In addition, post-translational modifications, such as phosphorylation, acetylation, glycosylation, or covalently linking with prosthetic groups, can take place on amino acid side-chains of protein before or after translocating to its cellular site of function.

9.3
Sense Codon Reassignment

Sense codon reassignment is the oldest technique used to incorporate noncanonical amino acids into proteins. As mentioned in Section 9.1, the complete replacement of


PROTEIN BIOSYNTHESIS
DNA
Transcription
mRNA
Ribosome
Initiation Methionine
Initiation codon
Shine-Dalgarno sequence
Triplet genetic code
Translation process
Growing peptide chain
Full length polypeptide chain
Protein processing for full function by post-translational modification and folding
Glycosylated amino acids
Phosphorylated amino acid
Acetylated amino acids
In vitro site directed chemical modification
Protein Ligation
Chemically synthesized peptide
Solution or Solid Phase Peptide Synthesis
Expressed protein
Heterologous Protein Expression in Host Cell
Semi-synthetic protein
Release Factor
Stop codon
Canonical amino acid
Aminoacyl-tRNA synthetase
tRNA
Orthogonal aminoacyl-tRNA synthetase
Non-canonical amino acid
Orthogonal tRNA
FRAME-SHIFT SUPPRESSION
Host Cell
Five base codons
Frame-shift suppressor tRNA
Four base codons
NONSENSE SUPPRESSION
Stop codon suppressor tRNA
Two different unused stop codons in host system
RNA ligation by T4 RNA ligase
pCpA
Chemical acylation of the dinucleotide with the non-canonical amino acid
Aminoacylation of tRNA by ribozyme
Misacylated tRNA
MISSENSE SUPPRESSION
Degenerate triple codons
SENSE CODON REASSIGNMENT
Replaced canonical amino acid
1) Over-expression of aaRS
2) Mutation of amino acid binding site of aaRS
3) Replacement of the canonical amino acid with the non canonical amino acid
Aminoacyl-tRNA synthetase mutant
Mutation of
1) Shine-Dalgarno sequence
2) Anti-Shine-Dalgarno sequence
3) Release factor binding site
ORTHOGONAL BASE PAIR
Semi-synthetic protein
Non-canonical base pair containing DNA
PCR
Primers
DNA polymerase
A,T,C,G,X,Y
Transcription
Non-canonical base pair forming tRNA
Non-canonical base pair forming mRNA
Orthogonal pair of tRNA/synthetase having less affinity for release factor

Met by Sem in *E. coli* Met auxotroph by Cowie and Cohen [14] was the first success of this technique. Through appropriate engineering of the expression host, a sense codon may be reassigned to a noncanonical amino acid under the conditions used for protein expression (Figure 9.2). The net result is the generation of proteins that contain 19 of the canonical amino acids and one noncanonical amino acid. This technique relies on the use of auxotrophic expression hosts deficient in the biosynthetic pathway needed to produce one or more of the 20 canonical amino acids. By using chemically defined media in conjunction with such strains, it is possible to control the intracellular pools of the amino acids of interest. The amino acid auxotrophic strains can be generated or obtained from public repositories such as the Yale Coli Genetic Stock Center (cgsc.biology.yale.edu). Since leaky expression in the initial growth medium will lead to a protein product that does not contain the desired noncanonical amino acids, it is important to control protein expression with a strongly repressed promoter such as phage-derived promoters T5 and T7. These promoters are found in commercially available protein expression vectors such as the pQE series or the pET series. The suitable amino acid auxotrophic strain is transformed with a plasmid encoding the desired recombinant protein under the control of a strongly inducible promoter. Since most noncanonical amino acids will not support bacterial growth, the expression host is allowed to grow initially to late exponential phase either in a rich medium or in a medium such as M9 or M63 supplemented with all 20 canonical amino acids and an appropriate antibiotic. When the bacterial culture has reached the optical density of 1.0 at wavelength 600 nm (enough mass of cellular material for protein synthesis), the cells are collected by centrifugation and resuspended in sterile isotonic saline (0.9% NaCl). After repeating the washing cycle with 0.9% NaCl to remove all amino acids, the cells are collected and resuspended in a medium supplemented with a desired

Figure 9.1 Overview of protein biosynthesis and methods used for semisynthetic protein synthesis. Protein biosynthesis starts with DNA transcription into mRNA, which is translated into protein (chain of open circles) by ribosomes, aminoacylated tRNAs, and many translation factors, including RF. tRNA is specifically aminoacylated with canonical amino acid (open circle) by aaRSs. The produced protein can be post-translationally modified (different colored circles) and folded into functional protein. Pretranslational methods used for semisynthetic protein synthesis are presented in ovals. Methods used for mischarging tRNA are shown in rectangles. Noncanonical amino acids, which are incorporated into semisynthetic proteins, are shown in colored noncircular shapes. Sense codon reassignment, which is a residue-specific method, replaces one canonical amino acid (orange circle) with a noncanonical amino acid (orange oval) by depletion of the replaced amino acid (double lines), overexpression of RS, or/and mutation of RS (blue structure). Missense, nonsense, and frame-shift suppression as well as orthogonal base pairs all extend the genetic code by use of degenerate codons, one or two unused stop codons, four- or five-base codons, or noncanonical base pairs (shown as red X and blue Y), respectively, in targeted positions for incorporation of desired noncanonical amino acid (s) into proteins. While nonsense suppression cannot incorporate more than two noncanonical amino acids, frame-shift suppression and orthogonal base pairs can incorporate more than two. Orthogonal mRNA/ribosome pairs decrease or eliminate the competition between mischarged stop codon suppressor tRNA and RF (oval), and increase the efficiency of nonsense suppression. Mischarged tRNA can be prepared by chemical synthesis or enzymatic aminoacylation of tRNA with ribozyme or orthogonal RS (red structure). Post-translational methods include *in vitro* SDCM, CPS, and chemical ligation with heterologous expressed protein.

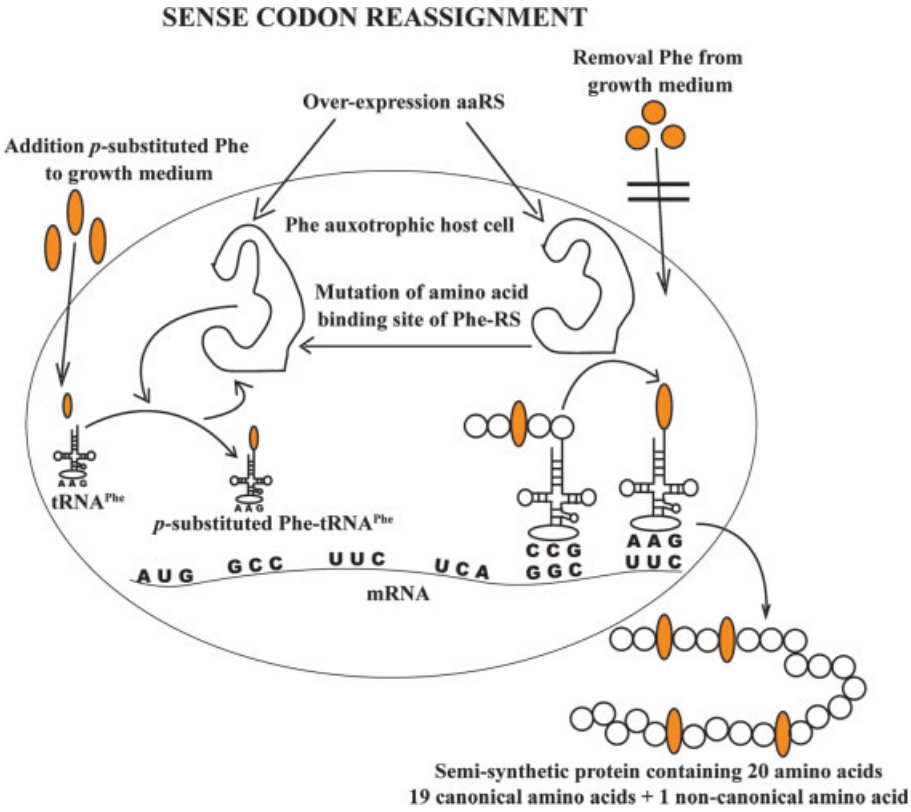


Figure 9.2 Sense codon reassignment is used to replace canonical amino acid (orange circle) by noncanonical amino acid (orange oval) producing semisynthetic protein containing 19 canonical amino acids (open circles) and one noncanonical amino acid. Sense codons (e.g., UUC) may be reassigned to *para*-substituted Phe instead of Phe through growing the expression host, Phe auxotrophic host cells, on chemically defined growth media, containing *para*-substituted Phe and lacking Phe. The efficiency of the technique can be increased by over expression of $tRNA^{Phe}$ synthetases (PheRS) and consequently increase the rate of mischarging of $tRNA^{Phe}$ with *para*-substituted Phe.

noncanonical amino acid and the proper complement of canonical amino acids (lacking the amino acid to be replaced by a noncanonical amino acid). To further deplete the intracellular pool of the original amino acids, the cells are incubated at the expression temperature for 10–30 min. Protein expression is then initiated by addition of an appropriate inducer (isopropyl-β-D-thiogalactopyranoside to a final concentration of 1 mM). Depending on the nature of the recombinant protein, the expression time and temperature may have to be varied. Following protein expression, the cells are collected by centrifugation for lysis and then the expressed protein is isolated by a suitable purification method. The yields of purified proteins expressed in this method vary widely from 10 to 100 mg/l of culture depending on the amino acid analog used as well as the recombinant protein and the purification method. Although this simple method has succeeded in reassigning the sense codons of different amino acids such

as Trp [28, 29], Met [30, 31], Leu [32], Ile [33, 34], Phe [35], Pro [36], and His [37] in recombinant proteins to many amino acid analogs, it has failed to incorporate other noncanonical amino acids in many cases. The fidelity of protein synthesis in the cell is controlled in large part by the aaRS [26], either by activation of the amino acid or by ligation of the activated amino acid to its cognate tRNA. Thus, successful incorporation of noncanonical amino acids into protein depends on the ability of the aaRS to activate the noncanonical amino acid of interest. Kiick *et al.* found that a higher rate of activation of amino acid analogs by aaRS promoted an effective incorporation of the amino acid analog into the recombinant protein [38, 39]. Thus, overexpression of aaRS, which increases the rate of amino acid analog activation, has allowed reassignment of sense codons with difficult-to-incorporate analogs [40, 41]. Some modification of the method should be introduced when aaRS are overexpressed. Thus, it is important to express aaRS either constitutively or with a promoter orthogonal to the promoter used to express the recombinant protein. The nonorthogonal expression of aaRS will result in incorporation of the noncanonical amino acid into aaRS and may render the aaRS inactive. While the sense codon reassignment of many amino acid analogs may become greatly efficient by overexpression of the appropriate wild-type aaRS, many other analogs require further manipulation of the aaRS activity of the cell. The wealth of structural data on the aaRS [14] provides some insight into the origins of the high fidelity characteristic of protein synthesis. The binding pocket of aaRS for the amino acid is highly shape-selective and can discriminate between the similar structures such as Phe and Tyr. Just as DNA and RNA polymerases have a proofreading function by exonuclease activity on mismatched base pairs, aaRS have an editing function by hydrolysis of improperly aminoacylated tRNAs [42–45]. Thus, the ability of a noncanonical amino acid to penetrate the genetic code *in vivo* is dependent upon its ability to pass through both of these checkpoints. The availability of high-resolution crystal structures for most of the aaRS allows engineering and redesign of the amino acid binding pockets and editing sites of these enzymes for the ultimate goal of introducing noncanonical amino acids into proteins. A single-point mutation in the amino acid binding pocket of PheRS can expand the specificity of the binding to include a set of *para*-substituted Phe analogs and benzofuranyl-Ala as good substrates for activation and consequently efficient incorporation into recombinant proteins [46–48]. Also perturbation of the editing site of the aaRS can lead to incorporation of a large set of noncanonical amino acids. The single point mutation of a highly conserved Thr to Ala in the editing site of LeuRS has led to loss of discrimination of Leu from similar amino acids [49]. Furthermore, replacement of the conserved Thr with a larger amino acid such as Tyr impairs the editing function of LeuRS [50] and allows incorporation of several noncanonical amino acids in response to Leu codons [51].

9.4
Missense Suppression

As mentioned in Section 9.2, the genetic code is degenerate and 18 of the 20 proteinogenic amino acids have more than one triplet base codon (sense codon).

Instead of replacement of a canonical amino acid by noncanonical one as in a sense codon reassignment technique (see Section 9.3), using the "surplus" sense codons for incorporation of noncanonical amino acids can substantially expand the genetic code to give an enriched set of building blocks for artificial proteins. The missense codon technique relies on the fact that the Gibbs free energy of melting of Watson–Crick base pairs between codons and anticodons is higher than that of wobble base pair (Figure 9.3). A good example is described by Tirrell and his group [52]. They incorporated a Phe analog, 2-naphthylalanine (Nal), into dihydrofolate reductase (DHFR) as a test protein. Phe is encoded by two codons, UUC and UUU, which are decoded in *E. coli* by a single $tRNA^{Phe}_{GAA}$. The anticodon sequence of this *E. coli* $tRNA^{Phe}_{GAA}$ binds via Watson–Crick base pairing to the UUC codon and via wobble base pairing to UUU. It has been suggested that a heterologous tRNA with an AAA anticodon sequence might compete with *E. coli* $tRNA^{Phe}_{GAA}$ in decoding of the

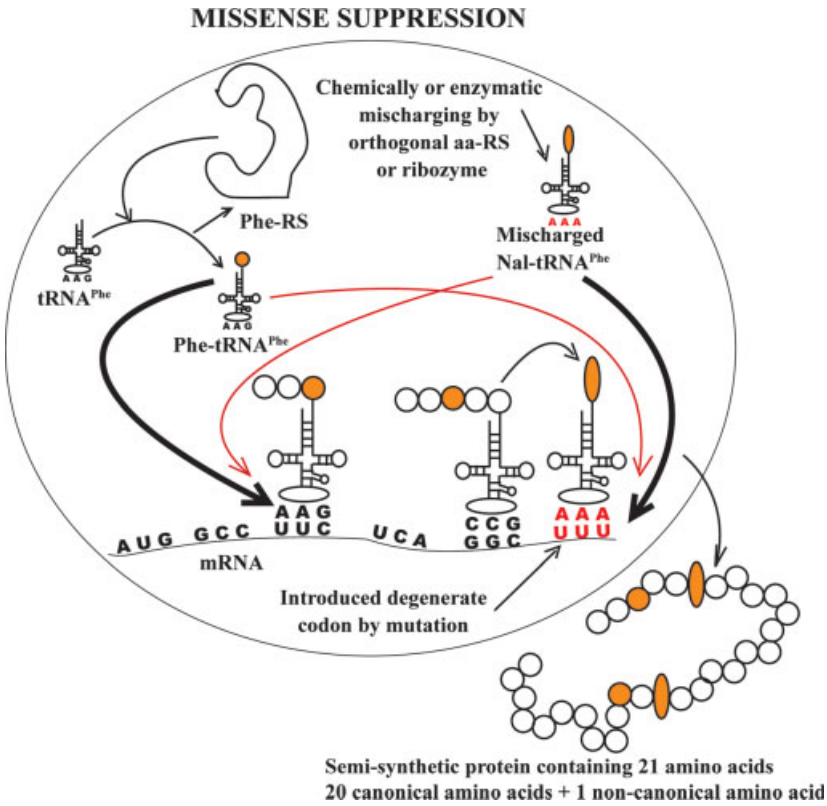


Figure 9.3 Missense suppression is used to incorporate noncanonical amino acid (orange oval) in addition to 20 canonical amino acids (orange and open circles). Introduced degenerate codon such as UUU, coding for Phe, can be used to incorporate noncanonical amino acid at the target position by using mischarged $tRNA^{Phe}$ having an AAA anticodon. There is competition between Phe-$tRNA^{Phe}_{AAG}$ and mischarged $tRNA^{Phe}_{AAA}$ for both UUC and UUU codons (red arrows).

UUU codon. The orthogonal $tRNA^{Phe}_{AAA}$/PheRS pair (see the end of this section) from *Saccharomyces cerevisiae*, $ytRNA^{Phe}_{AAA}$/yPheRS, was introduced into the *E. coli* Phe auxotroph (K10-F6Δ) to specify the incorporation of Nal. The binding site of the yPheRS was engineered (Thr415Gly) to activate Nal. The protocol for this method is quite similar to the protocol for sense codon reassignment (see Section 9.3) except that an *E. coli* Phe auxotroph is transformed with two plasmids. The first plasmid pQE16 carries a gene encoding the protein of interest (DHFR) under the control of the bacteriophage T5 promoter as well as the gene encoding the yPheRS mutant under control of a constitutive *tac* promoter. The second plasmid pREP4 carries a gene encoding $ytRNA^{Phe}_{AAA}$ under control of the constitutive *lpp* promoter. The combination of the yPheRS and this additional ytRNA serves as a 21st aaRS/tRNA pair permitting the incorporation of a 21st amino acid in response to the UUU codon. Induction of these cells in media supplemented with 3 mM Nal leads to highly biased assignment of UUC codons to Phe and of UUU codons to Nal. The assignment of UUC to Phe is essentially quantitative, while UUU sites are assigned to Nal at levels that are typically 80%. The missense incorporation of a noncanonical amino acid can be verified by purification of the recombinant DHFR and digestion with a protease such as trypsin. The mixture of tryptic fragments may then be analyzed by mass spectrometry (MS) to confirm incorporation of Nal. The yield of protein isolable from this experiment is typically an order of magnitude lower than that achievable in sense codon reassignment experiments with yields being on the order of 2–5 mg of DHFR per liter of culture.

9.5
Evolving the Orthogonal aaRS/tRNA Pair

Before going into the nonsense and frame-shift suppression techniques we need to understand the meaning of the orthogonal aaRS/tRNA pair used in missense (Section 9.4), nonsense (Section 9.6), and frame-shift (Section 9.9) suppression (Figure 9.4). The suppression techniques need noncanonical amino acid-carrying suppressor tRNA (misacylated tRNA) for incorporation of noncanonical amino acid into proteins. For *in vitro* translation, suppressor tRNA can be chemically aminoacylated with noncanonical amino acids [53] (see Section 9.6). The low yield of the *in vitro* suppression techniques, which is related mainly to the difficulty of obtaining the high amounts and recycling of misacylated tRNA, is the reason for using the orthogonal aaRS/tRNA pairs in the suppression techniques. The differences in the tRNA identity elements between the tRNA of different species [54, 55], which are found primarily in the acceptor stem and variable arm, in combination with the species-specific aaRSs, form a strong foundation for use of the orthogonal aaRS/tRNA pairs in suppression techniques. The orthogonality of the aaRS/tRNA pair means that the aaRS does not charge any tRNA except its cognate tRNA and does not couple its cognate tRNA with any amino acid except the desired noncanonical amino acid. Furthermore, the orthogonal tRNA, which should function efficiently in a translation system and pair only with a unique codon in mRNA, is not a substrate for

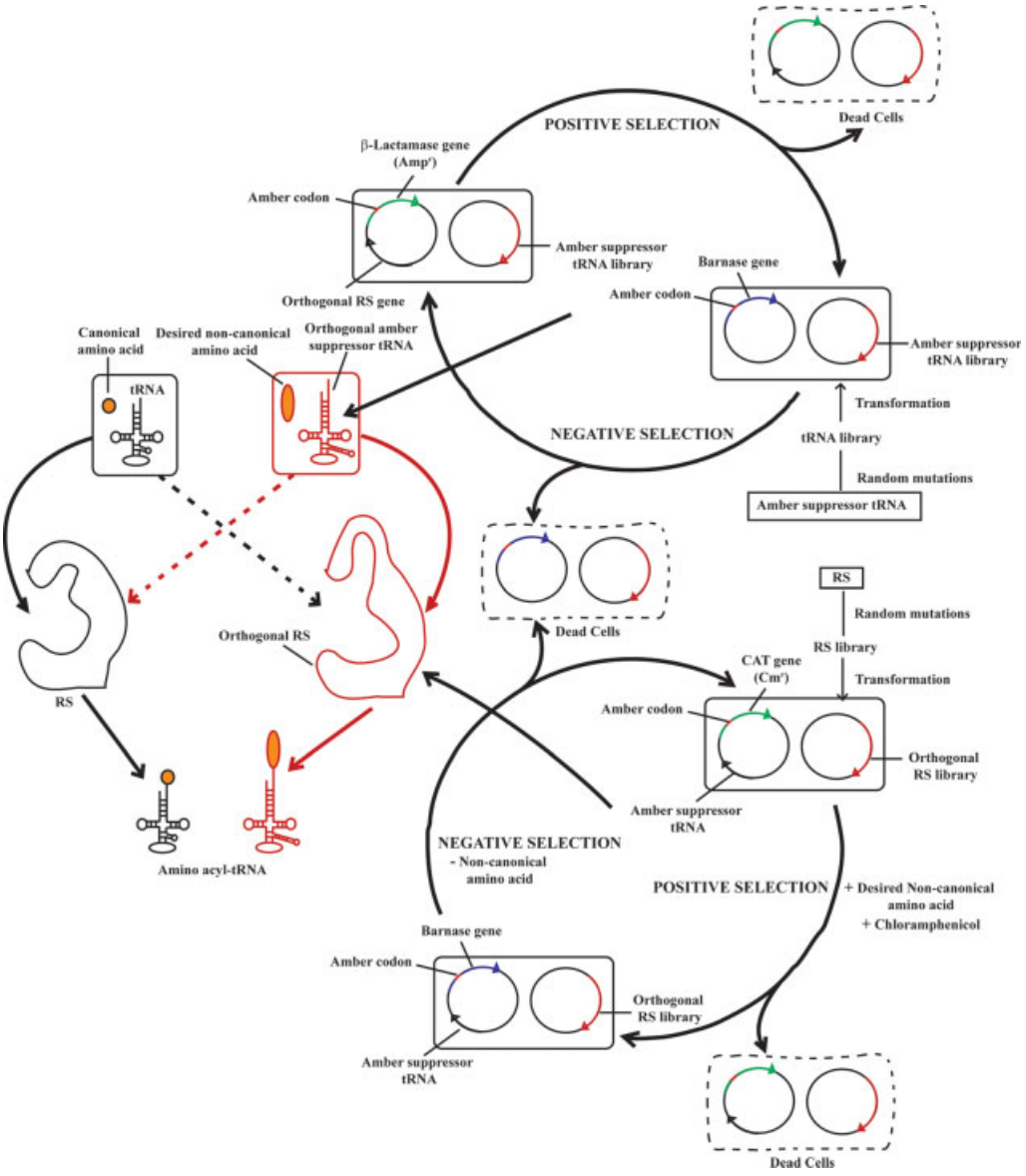


Figure 9.4 Mischarged suppressor tRNA (red structure) with noncanonical amino acid (orange oval) is used for constructing semisynthetic proteins. The orthogonal aaRS (red structure) will not acylate any tRNA (black dashed arrow) except its cognate tRNA and only with the desired noncanonical amino acid (red solid arrow). The orthogonal suppressor tRNA, which base-pairs only with a unique codon in mRNA, is not a substrate for any aaRS (red dashed arrow) except its cognate aaRS. To evolve orthogonal tRNA, an amber suppressor tRNA library (red arrow) can be constructed by random mutations and subjected to negative selection in the absence of the cognate orthogonal aaRS and coexpressed with a mutant barnase gene (blue arrow) in which amber codons (red arc) are introduced at permissive sites. All orthogonal or nonfunctional

any aaRS except its cognate aaRS. The heterologous expression of the orthogonal aaRS/tRNA pair into a host organism can produce a system with 21 non-cross-reacting pairs with 20 proteinogenic amino acids and one noncanonical amino acid. The orthogonality of the aaRS/tRNA pair can be achieved either by expression of an aaRS/tRNA of one species in another distant species or by expressing a mutated host pair that fulfills the orthogonality requirements.

The first attempt to use the orthogonality of aaRS/tRNA pair in suppression was by RajBhandary *et al*. [56]. They coexpressed *E. coli* GlnRS with tRNA suppressor in the mammalian cell line COS-1 to suppress the amber codon by incorporating the amino acid Gln. In 1997, Schultz *et al*. first tried using coexpression of the mutated aaRS, *E. coli* GlnRS, with engineered suppressor tRNA to incorporate a noncanonical amino acid in the expressed protein in *E. coli* [57]. This early attempt was not fully successful in the incorporation of a noncanonical amino acid. The same group subsequently used engineered *Methanococcus jannaschii* $tRNA^{Tyr}_{CUA}$ and its cognate *M. jannaschii* TyrRS in the incorporation of *O*-methyl-Tyr into DHFR expressed in *E. coli*. [58]. Although this orthogonal pair was shown to function efficiently, unfortunately Tyr was also incorporated in the target position to some degree [59]. The specificity and efficiency of the suppression techniques for incorporation of noncanonical amino acids depends on the efficiency of the orthogonal aaRS/tRNA in the host. Thus, evolving new efficient pairs of orthogonal aaRS/tRNA has been in high demand. Schultz *et al*. have improved the strategy of evolving orthogonal aaRS/tRNA pairs by a combination of positive and negative selections for suppressor tRNA [60] and aaRS [58]. To evolve orthogonal tRNA, a library of suppressor tRNAs was constructed by random mutations of tRNA residues that do not interact with the cognate orthogonal aaRS. This tRNA library was subjected to negative selection in the absence of the cognate orthogonal aaRS and coexpressed with a mutant barnase gene in which amber nonsense codons were introduced at permissive sites. When a member of the suppressor tRNA library is aminoacylated by an endogenous *E. coli* aaRS, the amber codons are suppressed and the ribonuclease (RNase), barnase, is produced, resulting in cell death. Thus, only cells harboring orthogonal or

tRNAs from surviving clones (red arrow) are then subjected to positive selection in the presence of the cognate orthogonal aaRS (black arrow) and a β-lactamase gene (green arrow) with an amber codon (red arc) at a permissive site. After many cycles of alternating rounds of negative and positive selections orthogonal tRNA can be obtained. Similarly, a library of aaRSs can be constructed by randomizing residues of aaRSs that interact with the substrate amino acid in the binding pocket. In the presence of a desired noncanonical amino acid and chloramphenicol, the aaRS library (red arrow) is subjected to positive selection by coexpressing with amber suppressor tRNA (black arrow) and CAT gene (green arrow) in which an amber codon (red arc) is introduced at a permissive site. All active aaRS mutants with the noncanonical and endogenous amino acids from surviving clones (red arrow) are then subjected to negative selection in absence of the desired noncanonical amino acid by coexpressing with amber suppressor tRNA (black arrow) and a barnase gene (blue arrow) in which amber codons (red arc) are introduced at permissive sites. After many cycles of alternating rounds of positive and negative selections orthogonal aaRS can be obtained.

nonfunctional tRNAs can survive. All tRNAs from surviving clones are then subjected to positive selection in the presence of the cognate orthogonal aaRS and a β-lactamase gene with an amber codon at a permissive site. tRNAs that function in translation and are good substrates for the cognate orthogonal aaRS are selected on the basis of their ability to suppress the amber codon and produce active β-lactamase. Therefore, only tRNAs that are not substrates for endogenous *E. coli* aaRSs, but can be aminoacylated by the aaRS of interest and function in translation, will survive both selections. Also, libraries of orthogonal aaRS variants were generated by randomizing five or six residues interacting with the amino acid in the binding pocket of aaRS based on an analysis of the X-ray crystal structure. To identify aaRS variants that specifically recognize the noncanonical amino acid and no endogenous host amino acid, the aaRS libraries were subjected to alternating rounds of positive and negative selections. The positive selection is based on resistance to chloramphenicol conferred by suppression of an amber mutation at a permissive site in the chloramphenicol acetyltransferase (CAT) gene; the negative selection uses the barnase gene with amber mutations at permissive sites. When the library of aaRS mutants is passed through the positive selection in the presence of the noncanonical amino acid, those cells with mutant aaRS that can acylate the tRNA with either the noncanonical amino acid or an endogenous amino acid will survive. Plasmids encoding active mutant aaRS are then transformed into the negative selection strain and selections are carried out in the absence of the noncanonical amino acid. Those cells containing mutant aaRS that recognize endogenous amino acids incorporate the latter in response to the amber codons in the barnase gene and die. Repeating rounds of positive and negative selections can lead to the isolation of mutant aaRS that can specifically incorporate the noncanonical amino acid.

9.6 Nonsense Suppression

Suppressor tRNA was not originally used to incorporate noncanonical amino acids into protein, but analysis of its mode of action showed that the incorrect proteinogenic amino acid can be inserted *in vivo* into a specific location of a protein [61, 62]. This observation has laid the foundation for nonsense and frame-shift suppression techniques. The nonsense suppression strategy is based on the fact that only one of three stop (nonsense) codons in the genetic code is necessary for termination of protein synthesis (Figure 9.5). Hence, two of the stop codons can be exploited for the introduction of noncanonical amino acids. One or both of these stop codons can be introduced in the target position of DNA coding for the protein of interest by conventional site-directed mutagenesis. The introduced stop codon in the protein coding sequence can then be decoded as a noncanonical amino acid. This is accomplished by *in vitro* or *in vivo* translation using chemically or enzymatically mischarged suppressor tRNA, which is orthogonal to endogenous aaRS of the translation systems and has anticodon base paired with the introduced stop codon. The amber (UAG) stop codon has been most frequently employed for the

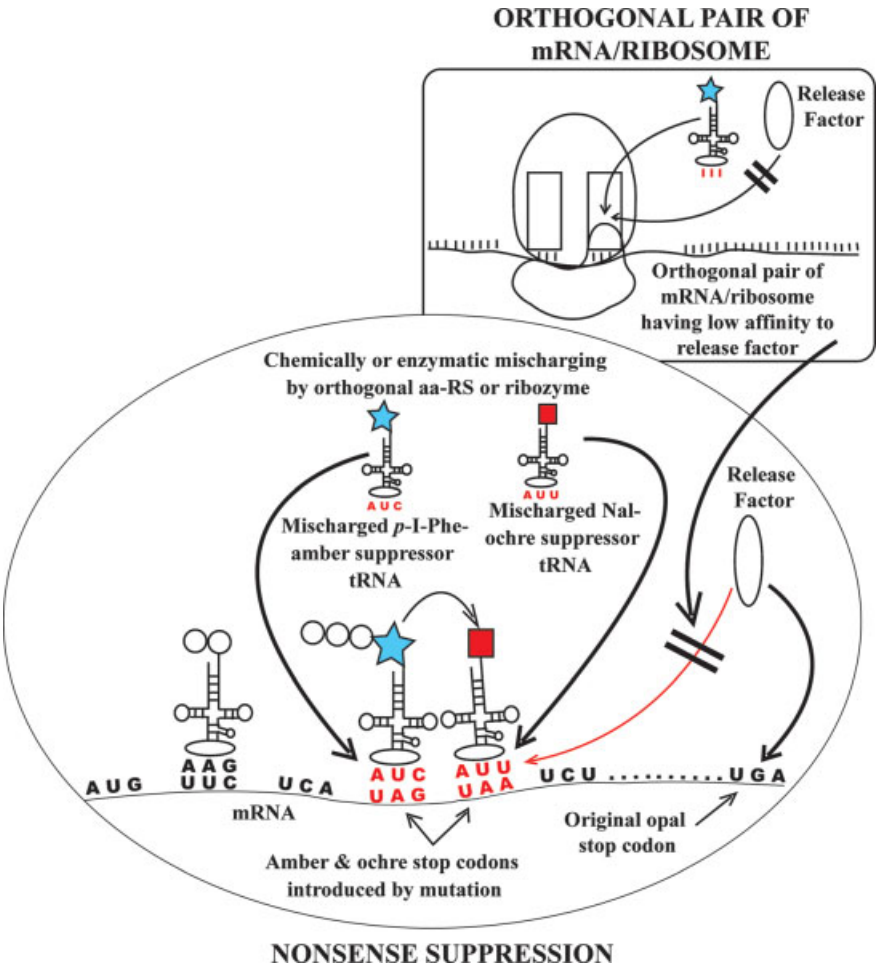


Figure 9.5 Nonsense suppression allows simultaneous incorporation of two noncanonical amino acids (blue star and red square). Two superfluous stop codons (red codons) are introduced at targeted positions and mischarged stop codon suppressor tRNAs base-pair with the introduced stop codons (black arrows) producing full-length semisynthetic protein. The competition of suppressor tRNAs with the RF (red arrow) producing truncated proteins can be minimized or eliminated (double line) by using an orthogonal mRNA/ribosome pair. The orthogonal ribosome has low affinity for the RF due to U531G and U534A mutations in its 16S rRNA.

incorporation of noncanonical amino acids in *E. coli*, because it is the stop codon occurring least frequently and the presence of natural amber suppressors in some *E. coli* strains does not significantly affect cell growth rates [63, 64]. However, the use of the ochre (UAA) stop codon [65] and the potential use of the opal (UGA) stop codon [66] have also been reported. Only one noncanonical amino acid per stop codon can be incorporated into the expressed protein, thus two different noncanonical amino acids can be incorporated at the same time by using two different stop

codons [67]. In 1989, the Chamberlin [68] and Schultz [69, 70] laboratories independently were the first to adopt this strategy for *in vitro* site-specific incorporation of a noncanonical amino acid into protein. They used the flexibility of the synthetic tRNA aminoacylation reaction combined with the specificity of nonsense suppression to incorporate several Phe and Tyr analogs into test proteins *in vitro*. The Phe or Tyr triplet codon was replaced by an amber codon at a specific position of the gene for the test protein. The anticodons of $tRNA^{Phe}$ or $tRNA^{Tyr}$ were modified to CUA for binding to the introduced amber codon. Schultz's group chemically aminoacylated the suppressor $tRNA^{Phe}_{CUA}$ with several Phe analogs (*p*-fluoro-Phe, *p*-nitro-Phe, and homo-Phe), whereas Chamberlin's group chemically modified Tyr – $tRNA^{Tyr}_{CUA}$ with $Na^{125}I$ to generate radiolabeled [^{125}I]Tyr – $tRNA^{Tyr}_{CUA}$. Addition of the modified suppressor tRNA to an *E. coli* or rabbit reticulocyte cell-free translation system containing the targeted gene produced the protein containing the noncanonical amino acid at the position specified by TAG in the coding sequence. The nonsense suppression efficiency obtained was up to 30% and the yield 1–10 μg/ml culture medium. The comparatively low yield was due mainly to both the difficulty of obtaining large amounts of charged suppressor tRNA and the competition between RF and suppressor tRNA [71]. By using a thermally inactivated mutant of RF [71], omitting RF from the reconstituted cell-free translation system [72], or using small interfering RNA-mediated downregulation of RF [73], the yield of noncanonical amino acid containing protein can be increased. The chemical acylation of tRNA was developed [74] and then improved [75] by Hecht *et al.* to expand the utility of misacylated tRNA for noncanonical amino acid incorporation into proteins. They acylated the dinucleotide pCpA with an N-terminally protected amino acid (*N*-pyroglutamylamino acid), which had been activated by carbonyldiimidazole (Figure 9.6). Then the acylated dinucleotide was ligated to the 3′-terminus of a truncated tRNA in the presence of T4 RNA ligase, thereby providing an aa-tRNA charged with a noncanonical amino acid. The *N*-pyroglutamyl protecting group of the aa-tRNA was cleaved by the enzyme pyroglutamate aminopeptidase and the deprotected aa-tRNA became fully functional [53]. This methodology has been further developed by Schultz *et al.* [76]. They have used a variety of photo- and chemo-cleavable protecting groups, such as 4-nitroveratrylcarbamates and *o*-nitrophenylsulfenyl chloride [69, 77, 78], and also replaced the cytidine of the dinucleotide pCpA with a deoxycytidine [76]. Sisido *et al.* have introduced a simplification of this method, whereby the acylation of the dinucleotide is carried out in cationic micelles [79]. An interesting alternative is to employ ribozymes as catalysts for the formation of aa-tRNA. Suga *et al.* have generated a ribozyme that aminoacylates different tRNAs with different noncanonical amino acids by using *in vitro* selection of a random sequence of RNA [80] (see Section 9.7).

The nonsense suppression method has been used to incorporate more than 30 noncanonical amino acids into proteins, representing a large variety of structures and functionalities. Some of these noncanonical amino acids have chemically reactive groups including ketone, azide, acetylene, and thioester groups. Other noncanonical amino acids are photoreactive, such as *p*-benzoyl-Phe and *p*-(3-trifluoromethyl-3 *H*-diazirin-3-yl)-Phe, or contain biophysical probes such as iodo-, fluoro-, or cyano-7-hydroxycoumarin and also dansyl side-chains. In addition, noncanonical amino acids

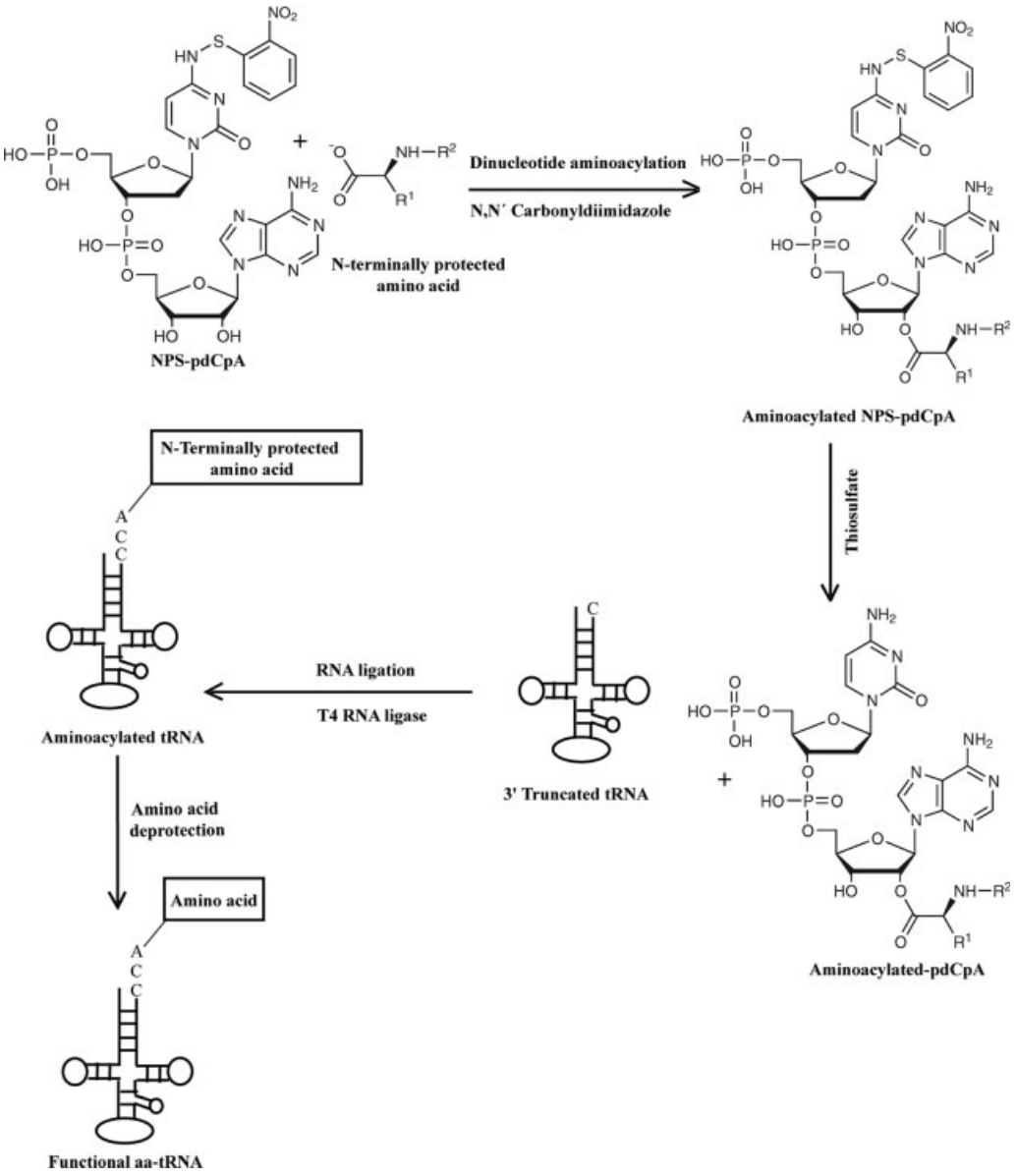
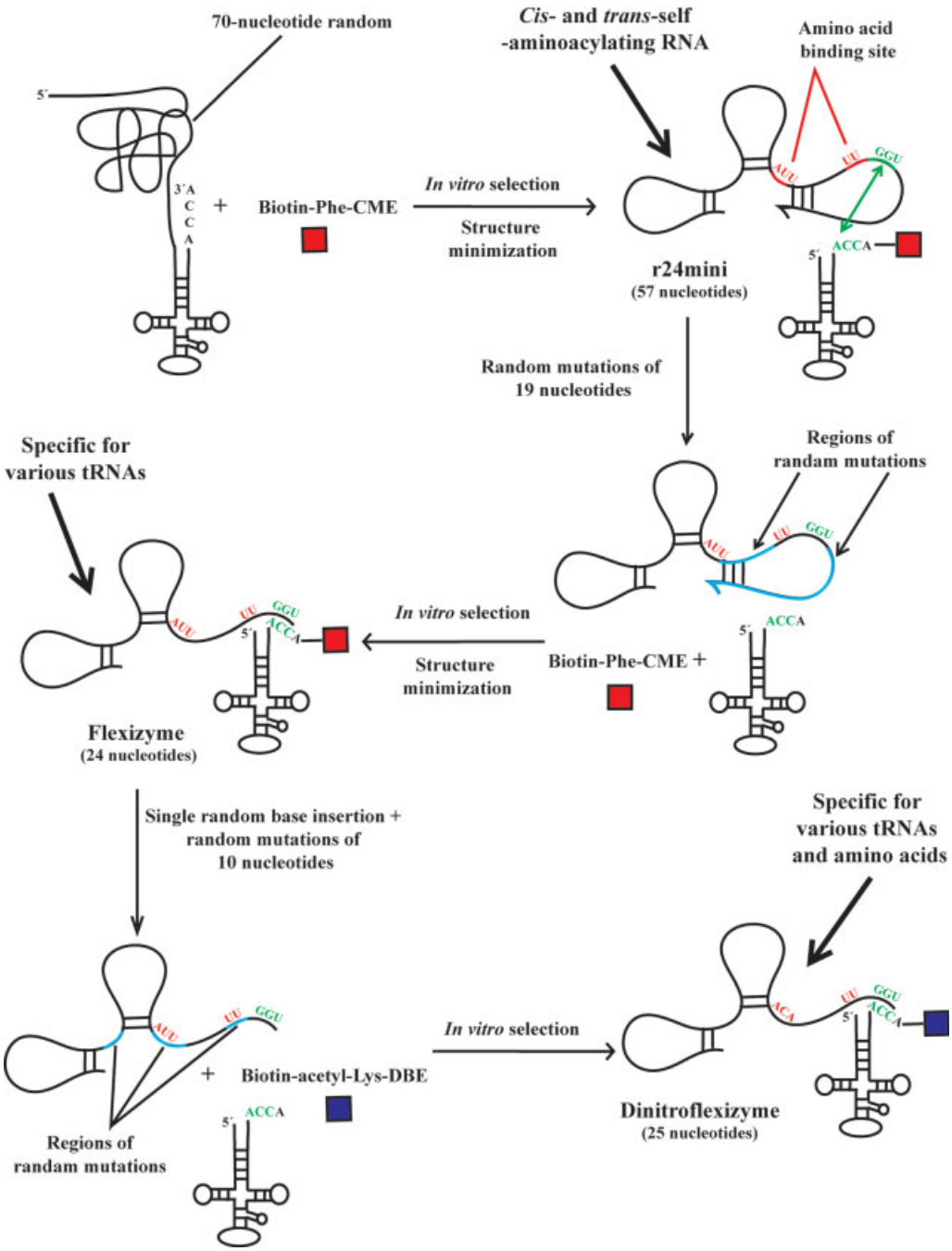


Figure 9.6 Chemical synthesis of mischarged aa-tRNA. *N*-(2-Nitrophenylsulfenyl)-pdCpA (NPS-pdCpA) is aminoacylated with the desired N-terminally blocked amino acid giving aminoacylated NPS-pdCpA. The NPS group is released by a reaction with thiosulfate producing aa-pdCpA, which is ligated to 3′-truncated tRNA by T4 RNA ligase giving mischarged tRNA with the N-terminally blocked amino acid. The functional aa-tRNA is obtained by release of the blocking group from the amino group of the amino acid.


70-nucleotide random
5′
3′A
C
C
A
+
Biotin-Phe-CME
In vitro selection
Structure minimization
Cis- and trans-self -aminoacylating RNA
Amino acid binding site
AUU
UU
GGU
ACCA
5′
r24mini
(57 nucleotides)
Random mutations of 19 nucleotides
Specific for various tRNAs
Regions of randam mutations
In vitro selection
Structure minimization
Biotin-Phe-CME +
Flexizyme
(24 nucleotides)
Single random base insertion + random mutations of 10 nucleotides
Specific for various tRNAs and amino acids
Biotin-acetyl-Lys-DBE
Regions of randam mutations
In vitro selection
ACA
Dinitroflexizyme
(25 nucleotides)

can include redox-active and metal-chelating residues, such as dihydroxy-Phe, 3-amino-Tyr, and bipyridyl-Ala.

The nonsense suppression approach has so far only been used *in vitro*. To apply the method to *in vivo* systems, it could be an attractive alternative to develop and use orthogonal tRNA/aaRS pairs (see Section 9.5). In this approach, the genes encoding an orthogonal aaRS/tRNA pair are engineered in such a way as to make the aaRS/tRNA pair orthogonal to the host organism as well, and so it only recognizes certain noncanonical amino acids and efficiently acylates their cognate tRNA. The orthogonal aaRS/tRNA pair in a separate plasmid is coexpressed in the host cells with the protein of interest. The noncanonical amino acid and expression inducer are added to the growth media when enough cellular material for protein synthesis is present. After incubation of the host cells for 20–24 h, the cells are collected and lysed and the expressed protein is purified by a suitable method and analyzed by MS. Recently, Chin *et al.* evolved orthogonal ribosome/mRNA pairs *in vivo* to overcome the drawback of the competition between the amber suppressor tRNA and RF1 stop codon in an *in vivo* translation system, and increase the suppression efficiency to more than 40% [81] (see Section 9.8).

9.7 Mischarging of tRNA by Ribozyme

In order to circumvent the chemical aminoacylation of suppressor tRNA, Suga *et al.* have designed an RNA library to evolve ribozymes that can aminoacylate suppressor tRNA with noncanonical amino acids (Figure 9.7). The RNA variants were composed of two domains – an amber anticodon-containing *E. coli* $tRNA^{Gln}$ and a varied 70-nucleotide 5′-leader sequence [82, 83]. The selection aimed to find a self-aminoacylating 5′-leader sequence. The pool of RNA variants was subjected to reaction with *N*-biotinylated-Phe-cyanomethyl ester (CME) and the self-aminoacylating RNA was isolated by using streptavidin. The isolated material was treated with RNase P to remove the 5′-leader and get a full sequence of mischarged suppressor tRNA. By aid of chemical and enzymatic mapping of the secondary structure of the ribozyme the

Figure 9.7 Evolution of ribozymes that can aminoacylate suppressor tRNA with noncanonical amino acids. *In vitro* selection and structural minimization of a RNA library to produce the 57-nucleotide r24mini RNA, which has *cis* and *trans* self-aminoacylation activity with aromatic amino acid analogs (red square). Biotin, which is linked to the amino acid analog, is used for selecting active ribozymes. The red bases represent nucleotides at the amino acid-binding site. The green bases represent bases proposed to form base pairs between ribozyme and tRNA, and the green arrow indicates their interaction. *In vitro* selection from an RNA library formed by randomly mutating the region flanked by the two binding sites of amino acid and tRNA and the acceptor stem recognition site (blue lines of r24mini), followed by structural minimization, produced a ribozyme (24-nucleotide Fx) that can aminoacylate different tRNAs. To further evolve a ribozyme able to aminoacylate tRNA with aliphatic amino acid analogs, an RNA library was constructed by single random base insertions and random mutations of 10 bases of the amino acid-binding site of Fx (blue lines of Fx). *In vitro* selection using acetyl-Lys-DBE (blue square) as substrate gave rise to the 25-nucleotide dinitro-Fx.

unnecessary sequence was eliminated, resulting in a short ribozyme called r24mini. The analysis revealed the ribozyme recognition motif (G43–U45), which base pairs with the 3′-end of tRNA (R73–C75, R = A or G), as well as the amino acid binding site (A33–U35 and U40–U41) [84]. Although the r24mini was able to function as a *cis*-acting and *trans*-acting ribozyme, it could not be used as a general catalyst because of its specificity for a particular tRNA and its strict selectivity for aromatic amino acid analogs. Therefore, Suga *et al.* made an additional *in vitro* selection to generate a ribozyme that can aminoacylate various tRNAs [85]. They constructed a pool of r24minis, where 19 nucleotides with random mutations were introduced into the region flanked by two binding sites (one for Phe and one for tRNA) as well as the acceptor stem recognition site. In this pool the variable sequence was linked to the 5′-end of $tRNA^{Asn}$ instead of $tRNA^{Gln}$, for which the tRNAs are quite different in sequence. Depending on the result of the selection, the selected ribozyme was shortened to a 24-nucleotide sequence called flexizyme (Fx), which was more efficient and versatile toward different tRNAs than r24mini. The ribozyme has also been evolved so it is not restricted to aromatic amino acid analogs [86]. By replacing the aminoacyl donor CME group (aliphatic chain) with a 3,5-dinitrobenzyl ester (DBE, aromatic group) the catalyst can freely use nonaromatic amino acid analogs, such as acetyl-Lys-DBE, as substrates. To enhance the binding of the engineered substrate to Fx, a RNA pool was constructed by randomizing the Fx bases involved in amino acid binding with a single random base insertion (a total of 11 nucleotides). A new member of Fx, called dinitro-dFx, was evolved that accepts a wide variety of DBE esters of non-natural α-L-amino acids, α-*N*-acyl L-amino acids, α-D-amino acids, β-amino acids, and α-hydroxy acids. Furthermore, a resin-immobilized Fx, called flexiresin, has been developed for preparative production of aminoacylated suppressor tRNA [87].

9.8 Evolving the Orthogonal Ribosome/mRNA Pair

To avoid the strong competition between mischarged stop codon suppressor aa-$tRNA^{aa}$ and RF, and consequently increase the efficiency of the nonsense suppression technique, Chin *et al.* have evolved new ribosome/mRNA pairs having low affinity for RF, and being orthogonal to endogenous ribosomes and mRNA molecules [81] (Figure 9.5). The orthogonality of this new pair relies on the fact that the ribosomes recognize mRNA through base-pair matching between the Shine–Dalgarno (SD) sequence, located 7–13 nucleotides upstream of the AUG initiation codon in the mRNA, and an anti-SD sequence, near the 3′-end of the 16S ribosomal RNA (rRNA) in prokaryotic ribosomes. Thus the orthogonal ribosome can be targeted to a particular mRNA by changing the SD sequence and concomitantly introducing the appropriate complementary changes in the anti-SD sequence [88, 89]. The evolved ribosome called ribo-X can only translate the cognate mRNA, which is not translated by endogenous natural ribosomes. Based on the structural information that the 530-loop of 16S rRNA in the ribosomal A-site can

interact directly with both tRNA and RF1 [90, 91], an A-site mutant library was designed by randomization of seven nucleotides (529–535) in the 530-loop. An *in vivo* selection system using CAT mRNA with a UAG, together with an appropriate amber suppressor tRNA, yielded the mutant ribosome, ribo-X, carrying U531G and U534A mutations in 16S rRNA. Ribo-X demonstrated increased suppression efficiency for the amber codon as compared with wild-type ribosomes, without loss of fidelity in protein translation.

9.9
Frame-Shift Suppression

All proteinogenic amino acids are encoded by trinucleotide codons, which sets a limit to the number of codons in mRNA translation to 64 specifying 20 amino acids as well as three stop signals. While stop codon suppression makes possible the incorporation of a maximum of two different noncanonical amino acids, the use of codons longer than three nucleotides can expand the genetic code system and give the possibility to incorporate many alternative noncanonical amino acids into protein at the same time. The discovery of frame-shift suppressor tRNA containing four-base anticodon suggested that the ribosomal machinery could accept codon/anticodon pairs that are larger than three nucleotides in length. Building on this discovery, Hohsaka *et al.* have developed a frame-shift strategy for incorporation of noncanonical amino acids into protein by using four- and five-base anticodon suppressor tRNA and an *in vitro* translation system [92–94] (Figure 9.8). In this strategy, an mRNA containing an extended codon consisting of four or five bases is read by a modified aa-tRNA (acylated with a noncanonical amino acid and containing the corresponding extended anticodon) and a full-length protein containing a noncanonical amino acid at the targeted site is obtained. If the extended codon is read as a three-base codon by an endogenous tRNA, the reading frame will be shifted by one base. The frame shift will eventually result in a premature encounter of a stop codon and early termination of protein synthesis, thereby resulting in a truncated protein. In certain species, some naturally occurring codons are rarely used and the amount of their corresponding tRNA is low. This low abundance has been used to design four-base codons to minimize the competition between the four-base anticodon tRNA and endogenous tRNA. Several such four-base codons (AGGU, CGGU, CCCU, CUCU, CUAU, and GGGU) were adequately read by corresponding tRNAs, whereas four-base codons derived from termination codons were not successfully decoded [95, 96]. Among the successful codons, GGGU provided the most efficient recognition, giving 86% yield relative to the wild-type protein. The four-base codon technique has been used to incorporate three different noncanonical amino acids into three different sites of a single protein, thereby showing that four-base codons are not only orthogonal to their host organism, but also to each other [97]. Also, the use of five-base codons has been reported [94]. In this extension, 16 different mRNAs, each containing one of the five-base codons CGGNN (N indicates one of the four bases), were decoded by aa-tRNAs with complementary five-base anticodons to give rise to full-length proteins, each

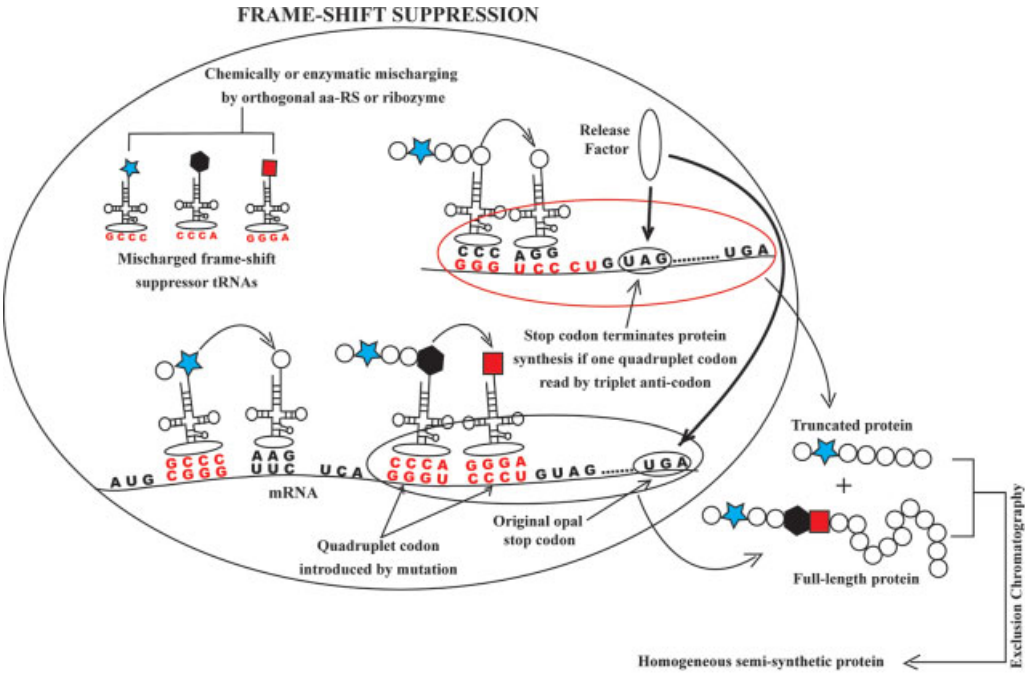


Figure 9.8 Frame-shift suppression technique simultaneously incorporating several noncanonical amino acids (blue star, black hexagonal, and red square) by using quadruplet codons (red codons) at targeted positions. The mischarged frame-shift suppressor tRNAs have quadruplet anticodons and base-pair with the introduced codons to produce full-length semisynthetic proteins (black oval). The competition between triplet anticodon tRNAs with frame-shift suppressor tRNAs (red oval) results in a premature encounter with a stop codon and produces truncated proteins.

containing an unnatural amino acid. Moreover, it was shown that at least two of the five-base codons (CGGUA and CGGUG) were orthogonal to each other.

In addition to the *in vitro* work, Schultz *et al.* have developed four-base codons for translation in *E. coli* [98]. Two specific Ser codons of the β-lactamase gene were independently mutated to four randomized bases and this was followed by a selection strategy. A library of $tRNA^{Ser}$ variants was also constructed. In this library, the anticodon loop of $tRNA^{Ser}$ was randomized. The loop was expanded from the normal seven to either eight or nine bases. Colonies were assessed for growth in the presence of ampicillin, to select for synthesis of β-lactamase. A series of selections led to the identification of 14 different four-base codons, including AGGA, CUAG, UAGU, and CCCU. These have been used in an alternative selection scheme to determine the optimal length for codon/anticodon pairs [99]. It was concluded that *E. coli* can tolerate three-, four-, and five-base codons, with a preference for $(n + 4)$-base anticodon loops (n = number of anticodon bases). Only a few four-base suppressor tRNAs that are orthogonal to endogenous aaRS have been reported [100–102]. To expand the availability of these orthogonal suppressor tRNAs for multiple

incorporation of noncanonical amino acids in a single protein by the four-base codon strategy, Sisido *et al*. evolved three new orthogonal suppressor tRNAs from nonstandard structures, such as some animal mitochondrial tRNAs, and tRNA for Sec and Pyl [103]. The deacylated and chemically acylated nonstandard structure $tRNA_{CCCG}$ mutants have been tested by translation of streptavidin using an *E. coli in vitro* translation system. Although all nonstandard tRNA were orthogonal to the host system, only bovine mitochondrial $tRNA^{Ser}$, *Methanosircina acetivorans* $tRNA^{Pyl}$, and mouse $tRNA^{Sec}$ were efficient in incorporation of three different noncanonical amino acids in the same protein.

9.10
Noncanonical Base Pairs

In protein translation the incorporation of a noncanonical amino acid into the protein may occur in direct competition with canonical amino acids in the approaches of sense codon reassignment (see Section 9.3), the missense and frame-shift suppressions as in three-base anticodon aa-tRNAaa (see Sections 9.4 and 9.9), or with the RF as in nonsense suppression (see Section 9.6), and results in heterogeneity of the product and lowers the efficiency of the technique. The natural genetic code is based on only two base pairs in DNA, A–T and G–C, limiting the genetic alphabet to 64 codons. Introduction of noncanonical base pairs into the genetic system as substrates for transcription and translation offers an alternative approach to genetic code expansion (Figure 9.9). The number of the three-base codons expands to 216 (including the canonical codons) if a single noncanonical base pair is added to the genetic material and is made accessible to the translation apparatus. Such a noncanonical base pair should be orthogonal to the canonical base pairs and have correspondingly high selectivity, efficiency, and stability in the biosynthesis of nucleic acids and proteins. The approaches to designing unnatural base pairs rely on the concepts of nonstandard hydrogen-bonding patterns, shape complementarity, and/or hydrophobic interfaces between the pairing bases. Benner *et al*. have developed noncanonical base pairs, isoguanine (isoG)– isocytosine (isoC) and xanthosine (X)–5-(2,4-diaminopyrimidine) (K), which have different hydrogen-bonding patterns from those of the canonical base pairs. These base pairs can be incorporated into DNA and RNA by appropriate polymerases in a template-directed manner [104, 105]. In the polymerase chain reaction (PCR), use of the base pair X–K prolongs the elongation time of each cycle to 24 h at 37 °C and HIV reverse transcriptase was inactivated at the denaturation step. On the other hand, the tautomer of isoG base pairs with T in PCR amplification by using the Klenow fragment (KF) of *Thermus aquaticus* (Taq) DNA polymerase. To avoid the base-pairing of T with isoG tautomer, Benner's group has used 2-thiothymidine (Ts) instead of T to base-pair with A. Romesberg *et al*. have also developed various hydrophobic base pairs, such as self-pairs of propynylisocarbostyril (PICS–PICS) and 7-azaindole (7AI–7AI), which exhibited high selectivity and efficiency in insertion experiments with the KF [106–108]. Unfortunately, the PICS–PICS pair is poorly extended by KF, and after the PICS incorporation opposite

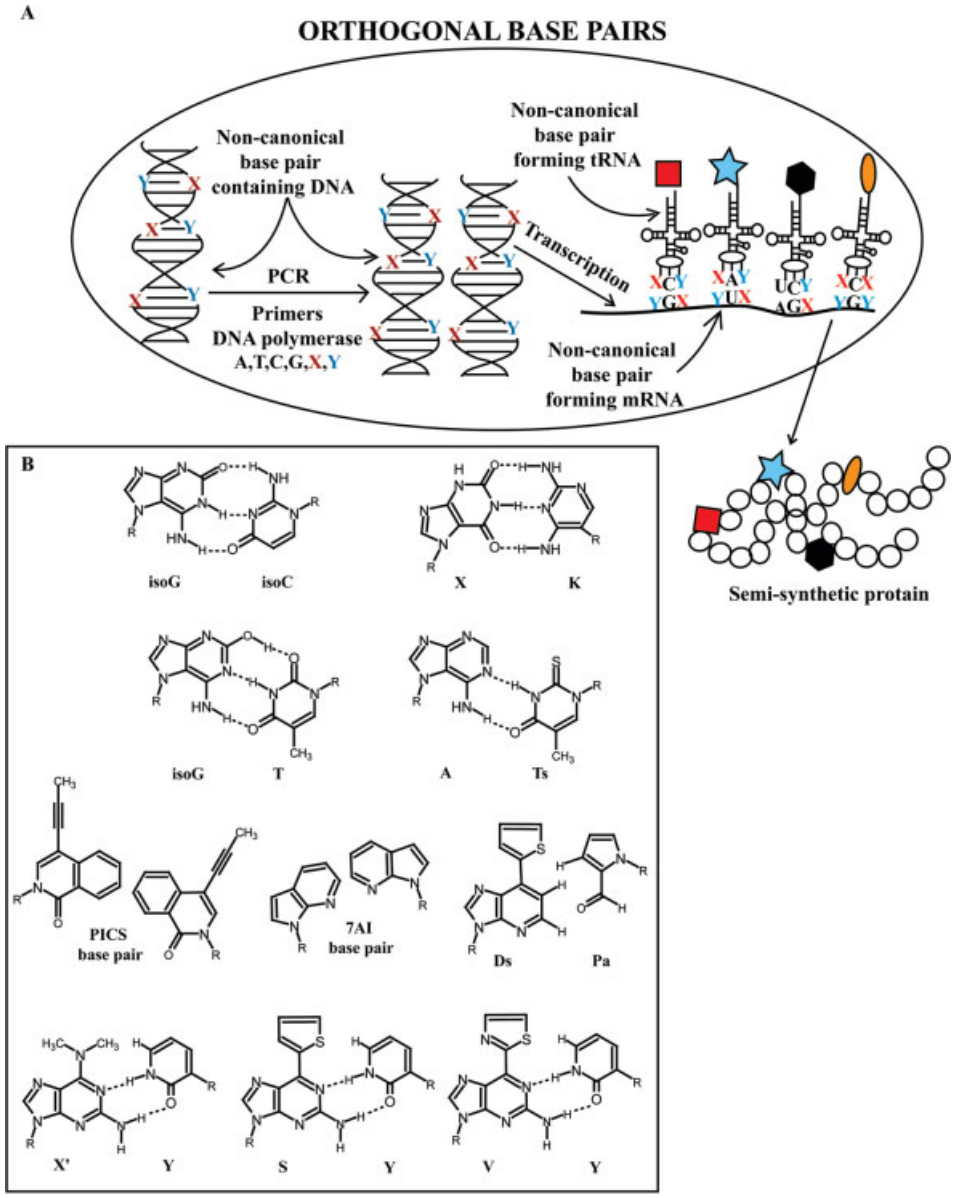


Figure 9.9 Orthogonal base-pair technique simultaneously incorporating several noncanonical amino acids (blue star, black hexagonal, red square, and orange oval). (a) A noncanonical base pair (represented by blue Y and red X), which is orthogonal to A–T and G–C pairs, can be incorporated in DNA, mRNA, and tRNA. (b) The bases of the developed orthogonal base pairs interact through nonstandard hydrogen-bonding patterns (such as isoG–isoC and X–K), shape complementarity (such as PICS–PICS, 7AI–7AI and Ds–Pa), or both (such as X′–Y, S–Y, and V–Y). The Ts base, which is used instead of T to avoid base-pairing with isoG, can base-pair with A.

PICS, further extension is inhibited. To address this problem, Romesberg *et al.* evolved a DNA polymerase, the Stoffel fragment of Taq polymerase, by means of a phage-displayed polymerase library [109]. One of the clones (P2) obtained by four rounds of selection increased the extension at least 30-fold after the PICS–PICS pair, relative to the parental enzyme. In addition, P2 incorporated the substrate of PICS into DNA opposite the template PICS, at least 320-fold more efficiently relative to the parental enzyme. Hirao and Yokoyama and their colleagues have expanded the number of synthetic base pairs that combine the nonstandard hydrogen-bonding patterns and shape complementarity [110–113]. The new bases were 2-amino-6-(*N*,*N*-dimethylamino)purine (X′), pyridin-2-one (Y), 2-amino-6-(2-thiazolyl)purine (V), and 2-amino-6-(2-thienyl)purine (S). Y pairs with X′, S, or V, but not with any of the canonical bases (A, G, C, or U). These bases, which are substrates for template-dependent T7 RNA polymerase, have been incorporated into RNA, demonstrating its utility for the incorporation of noncanonical amino acids into proteins. The large substituent at position 6 of X′, S, and V efficiently excludes pairing with the canonical bases. The amino acid Tyr32 of human c-Ha-Ras protein was chosen as the target site for modification. The mRNA encoding Ras with a YAG codon at position 32 was synthesized by transcription of a special DNA template. The DNA template was generated by inserting (by ligation) a synthetic oligonucleotide containing S at codon 32. A combination of chemical synthesis and RNA ligation was used to convert yeast $tRNA^{Tyr}$ into a new tRNA with a modified CUS anticodon. The $tRNA^{Tyr}_{CUS}$ was aminoacylated with 3-chloro-Tyr by wild-type *S. cerevisiae* TyrRS. The combination of a modified Ras mRNA and Tyr − $tRNA^{Tyr}_{CUS}$ was functional *in vitro* and resulted in production of full-length Ras with 3-chloro-Tyr at position 32. Recently, Hirao and Yokoyama and their colleagues have developed an artificial genetic system with a hydrophobic base pair between 7-(2-thienyl)-imidazo[4,5-*b*]pyridine (Ds) and pyrrole-2-carbaldehyde (Pa), which functions in replication and transcription [114]. The shape of the large purine analog Ds is complementary to that of the small five-member Pa. One critical problem of the Ds–Pa pair is that the substrate of Ds is efficiently incorporated opposite the template Ds and consequently further extension was interrupted after the Ds incorporation opposite the template Ds as found with the PICS–PICS. This problem has been overcome by using a modified substrate, γ-amidotriphosphate of Ds, which remarkably reduced the self-pairing of Ds. Similarly, to reduce the mis-incorporation of ATP opposite Pa, the γ-amidotriphosphate of A was also used in the system. The combination of the conventional triphosphates of Pa, G, C, and T, and the γ-amidotriphosphates of Ds and A enabled PCR amplification of the DNA fragments containing the artificial base pair by a thermophilic DNA polymerase, Vent DNA polymerase, with high fidelity.

9.11
Chemical Ligation

The average length of a protein is approximately 250 amino acids and the polypeptide chain is generally folded into two structural domains of 15 kDa in size [115, 116].

Chemical peptide synthesis (CPS), which will be covered in detail in other chapters, has been used to synthesize proteins with relatively long polypeptide chains, such as Green Fluorescent Protein (238 amino acids) [117] and RNase A (124 amino acids) [118]. However, these proteins are exceptional cases and CPS has limited utility in the synthesis of large proteins. This limitation is related to the solubility of protected peptides, the correct folding of large synthetic proteins, and the sequence homogeneity of the final product. Nevertheless, CPS makes possible the incorporation of nonpeptide linkages, D-amino acids, and many noncanonical amino acids with unrestricted side-chain structures into a peptide chain, which is difficult to accomplish by other techniques. The combination of CPS with a suitable technique for ligation of the product with a biosynthetically produced polypeptide provides a means for the preparation of larger proteins and represents a powerful strategy for producing semisynthetic enzymes. Different chemical ligation approaches allowing the coupling of unprotected peptide fragments have been developed. A key feature is the chemoselective coupling reaction of unprotected peptide fragments in aqueous solution. In 1953, Wieland *et al.* [119] published the first amide bond synthesis in aqueous solution through an intramolecular acyl shift. In 1981, Kemp *et al.* developed the initial peptide ligation methodology "prior thiol capture method" using an intramolecular acyl shift [120, 121] (Figure 9.10). The ligation takes place between an ester group at the C-terminus of one peptide and Cys thiol group at the N-terminus of a second peptide. A subsequent intramolecular acyl transfer from thiol to amino group leads to the formation of a new amide bond between the two peptides. This method relies on a template molecule to bring the acyl and amino groups in close proximity by having a hydroxyl group which esterifies the free acyl group and having a thiol group which forms a mixed disulfide with the Cys thiol group. The extensive study of this template-mediated ligation method led to design of 4-hydroxy-6-mercaptodibenzofuran as the best template [122] and showed that peptide bond making is extremely sensitive to steric interactions [123]. This method has the disadvantage of requiring side-chain protection of groups except those involved in the peptide bond formation. In 1992, Schnölzer and Kent developed the first method for "chemical ligation" for coupling unprotected peptide fragments in aqueous solution [124]. The method is based on the presence of uniquely reactive terminal groups in each peptide fragment to undergo chemoselective ligation. The thiol group of the C-terminal thioacid of one peptide and the alkyl bromide of the N-terminal the second peptide have been used in a nucleophilic substitution reaction to form a nonpeptide linker, a thioester, at the ligation site. This method overcomes all the limitations of the traditional convergent approach, namely poor solubility and difficulty of purifying long chain peptides. By CPS and chemical ligation, the entire HIV-1 protease, which contains a 99-amino-acid polypeptide chain, has been synthesized [124]. Also, a covalently linked dimer of the corresponding protein has been produced [125].

Even though numerous applications of these chemoselective ligation reactions have proven that the non-natural backbones are structurally and functionally well tolerated, the unnatural structure at the ligation site is a major disadvantage of these ligation approaches. Kent *et al.* introduced the first method for "native chemical ligation" (NCL) to couple unprotected peptide fragments in solution through native peptide

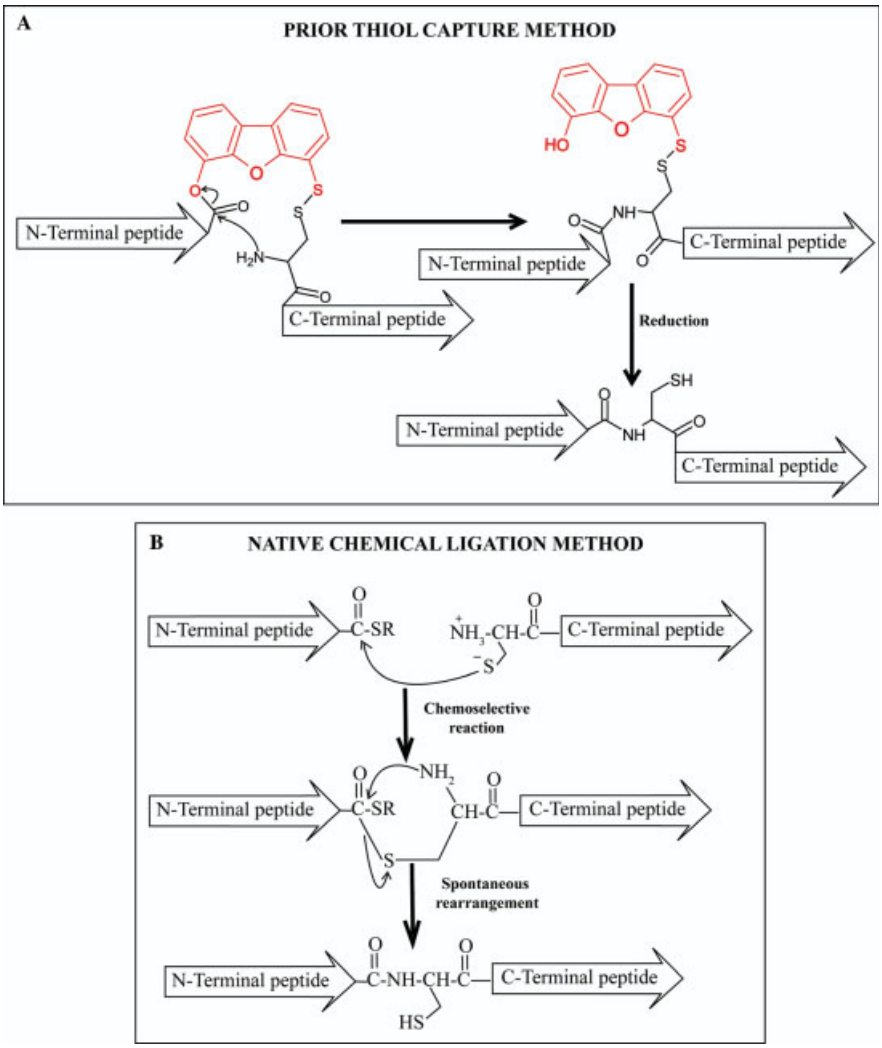


Figure 9.10 Peptide ligation by the prior thiol capture and the NCL methods. (a) In the prior thiol capture method, 4-hydroxy-6-mercaptodibenzofuran (red structure) brings the acyl and amino groups in close proximity. The ligation takes place between an ester group at the C-terminus of one peptide and the amino group of Cys at the N-terminus of a second peptide, forming an amide bond between the two peptides. (b) In the NCL method, trans-thioesterification takes place between the thiol group of the N-terminal Cys of one peptide and a C-terminal thioester moiety of a second peptide, followed by intramolecular acyl shift from thiol to amino group producing the amide bond.

bond formation at a ligation site [126]. NCL is well suited for production of semisynthetic proteins [127]. The trans-thioesterification reaction takes place between the thiol group of N-terminal Cys of one peptide and a C-terminal thioester moiety of a second peptide. The thioester-linked intermediate then undergoes a rapid intramolecular acyl

shift from thiol to amino group producing the amide bond preceding the Cys residue at the ligation site. While the trans-thioesterification is a reversible reaction, the intramolecular acyl shift is irreversible. Thus, there is no side-product formed between C-terminal thioester and any internal Cys residues present in the peptide fragment. The rate of the ligation reaction is dependent on the nature of the thioester leaving group, the C-terminal amino acid, and the pH of the reaction. Benzyl thioesters are used and show coupling rate enhancement with addition of an excess of thiophenol to the reaction [128]. It is believed that the thiophenol replaces the benzyl thiol before the ligation and therefore provides a better leaving group. In addition, the presence of a reducing agent is required to suppress the risk of the oxidation of the N-terminal Cys, which is essential for the ligation reaction. The rate of the coupling is much slower if the C-terminal amino acid in the peptide fragment bearing the C-terminal thioester is Pro or a β-branched amino acid such as Val, Ile, or Thr [129]. While the coupling reaction is rapid at pH 7 and finished after 5 min, the same reaction below pH 6 is only 50% complete after 10 min, suggesting that the thiolate anion of Cys is directly involved in the coupling reaction [126]. While the thioester containing the C-terminal peptide fragment can only be prepared by CPS, the N-terminus Cys peptide fragment can be obtained by either CPS or protein expression. Methods for generating recombinant proteins possessing N-terminal Cys all rely on cleavage of an appropriate precursor protein. In the simplest strategy, a Cys is introduced next to the initiating Met in the protein sequence and the Met is then removed *in vivo* by endogenous methionyl aminopeptidases, thereby generating the N-terminal Cys protein [130]. This approach is often inefficient and gives low yields of the desired protein. Most methods currently used for preparing N-terminal Cys proteins involve the *in vitro* digestion by exogenous proteases such as factor Xa. In this approach, a protease recognition sequence is introduced immediately in front of the cryptic N-terminal Cys in the protein of interest [131]. Treatment of such a recombinant fusion protein with the protease gives the requisite N-terminal Cys protein directly, which can then be used in subsequent ligation reactions. The Cys protease from tobacco etch virus (TEV) can also be used to release N-terminal Cys proteins from suitable precursors [132]. The advantages of using TEV Cys protease in preparation of N-terminal Cys proteins are the high specificity of the protease and the ability to overexpress the protease in *E. coli*. Hundreds of proteins have now been prepared by the NCL method, including medium-sized proteins [133]. However, a limitation of the general use of NCL is the resulting Cys at the ligation site, which can disrupt the protein structure or function. To overcome this restriction, the Cys at the ligation site can be alkylated [134] or desulfurized into Ala by using palladium or Raney nickel with hydrogen [135]. However, these approaches are only suitable for Cys-deficient proteins.

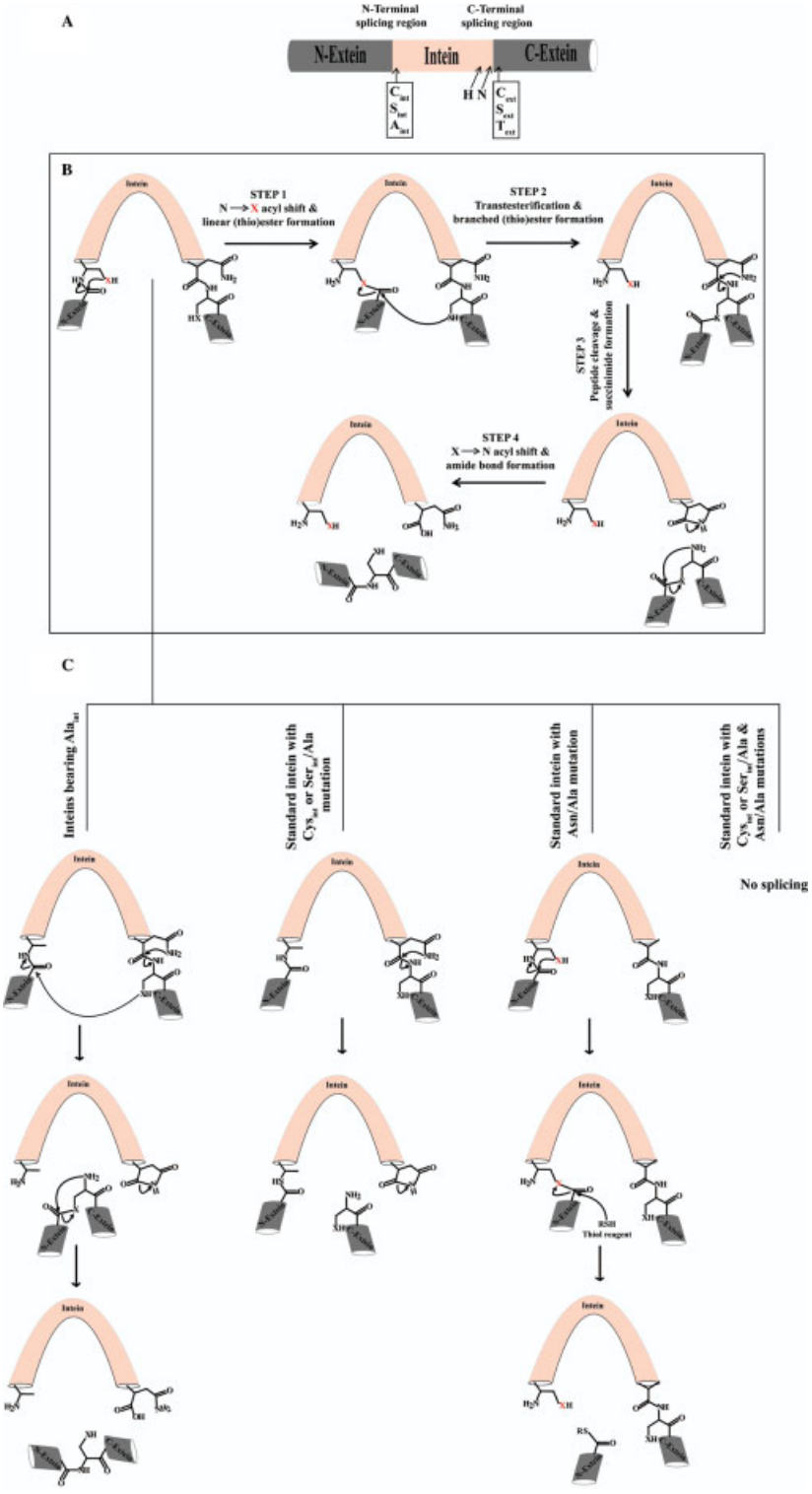
9.12 Inteins

Inteins [136] are internal segments of precursor proteins that catalyze the excision of themselves, in an intramolecular process called protein splicing, with the concurrent

ligation of the two flanking external regions (N- and C-exteins) through a native peptide bond (Figure 9.11). Thus, inteins are analogs of self-splicing RNA introns. Inteins were discovered in 1990 [137, 138] and up to now over 200 inteins have been submitted to the on-line intein database (InBase; http://www.neb.com/neb/inteins.html) [139]. The origin of inteins is not yet clear. Inteins occur in organisms of all three domains of life as well as in viral and phage proteins. There they are predominantly found in enzymes involved in DNA replication and repair. Inteins, which consist of an approximately 135-amino-acid protein-splicing domain split by a nonessential endonuclease domain, can be divided into four classes: the maxi inteins (with integrated endonuclease domain), mini inteins (lacking the endonuclease domain), trans-splicing inteins (where the splicing junctions are not covalently linked) and Ala inteins (Ala as the N-terminal amino acid). Inteins are characterized by several conserved sequence motifs. The N-terminal amino acids of inteins are typically Cys (Cys_{int}), Ser (Ser_{int}), or Ala (Ala_{int}), whereas the C-terminus contains a conserved His/Asn pair. Similarly, the N-terminus of C-extein is Cys (Cys_{ext}), Ser (Ser_{ext}), or Thr (Thr_{ext}).

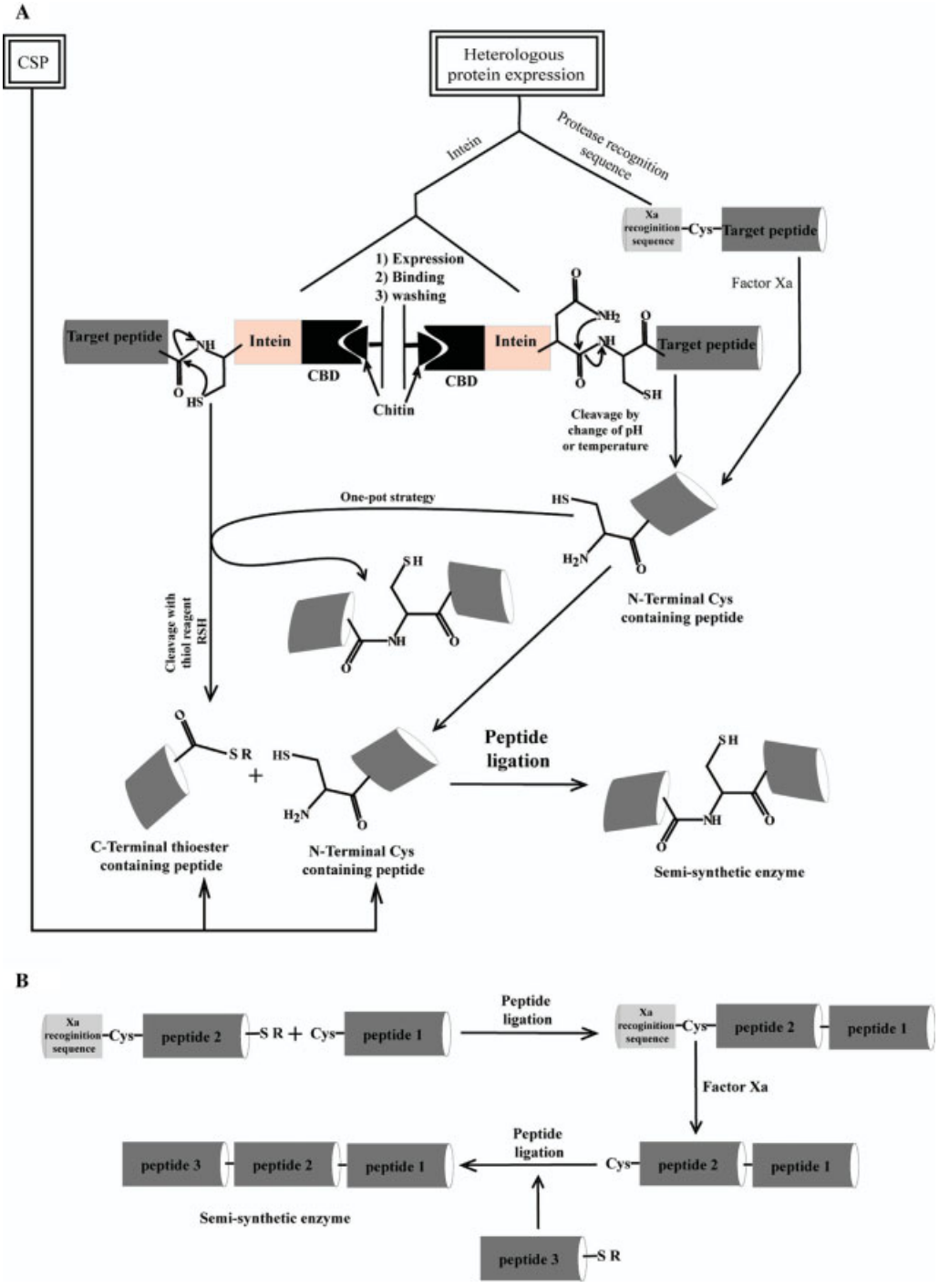
The standard protein-splicing mechanism consists of four steps [140, 141] (Figure 9.11). Initially, a linear thioester or ester intermediate, represented herein as (thio)ester, is formed when Cys_{int} (or Ser_{int}) undergoes an *N*- to *S*- or *N*- to *O*-acyl shift converting the peptide bond at the N-terminal splice junction to a (thio)ester bond (first step). In the second step, Cys_{ext}, Ser_{ext}, or Thr_{ext} cleaves this (thio)ester bond to form a branched (thio)ester intermediate. Cyclization of the intein C-terminal Asn (third step) resolves the branched intermediate. The aminosuccinimide is slowly hydrolyzed to regenerate Asn or Iso-Asn and another acyl shift (fourth step) rapidly forms the amide bond between the exteins. The N-extein is directly ligated to the C-extein during cleavage of the N-terminal splice junction. These four steps are highly coordinated in native systems. Coordination may be accomplished by differences in reaction rates for individual steps, conformational changes induced after completion of prior chemical steps, or a combination of both. Mutation or expression in a foreign context can result in single splice junction cleavage products that are unable to continue on the splicing pathway [141–143]. The C-terminal cleavage occurs when Asn cyclization precedes the branched intermediate formation. The N-terminal cleavage occurs when the (thio)ester in the linear or branched (thio)ester intermediates is hydrolyzed or attacked by an exogenous nucleophile, such as a thiol reagent. Replacement of Cys_{int} or Ser_{int} and the Asn at the C-terminus, ends with complete loss of splicing activity of the intein [142, 143] (Figure 9.11). In cases of inteins bearing Ala_{int}, the first step is circumvented [144] and the Cys_{ext}, Ser_{ext}, or Thr_{ext} attacks the carbonyl carbon of the peptide's N-terminal splicing junction (Figure 9.11). This rearrangement seems to be thermodynamically highly unfavorable, but the molecular architecture of the intein forces the scissile peptide bond into a twisted conformation of higher energy and thereby pushes the equilibrium to the (thio)ester side [140].

The elucidation of the splicing mechanism and the identification of the key amino acid residues involved in the scission and ligation of the peptide bonds [140, 141] facilitated the molecular engineering of artificial inteins as tools for different


A
N-Terminal splicing region
C-Terminal splicing region
N-Extein
Intein
C-Extein
B
STEP 1
N → X acyl shift & linear (thio)ester formation
STEP 2
Transtesterification & branched (thio)ester formation
STEP 3
Peptide cleavage & succinimide formation
STEP 4
X → N acyl shift & amide bond formation
C
Inteins bearing Ala
Standard intein with Cys or Ser/Ala mutation
Standard intein with Asn/Ala mutation
Standard intein with Cys or Ser/Ala & Asn/Ala mutations
No splicing
Thiol reagent

applications in protein chemistry [140, 145]. These applications include protein purification, trans-splicing, protein semisynthesis, synthesis of cyclic proteins, and the detection of protein–protein interactions. Semisynthesis of protein via NCL requires the specific generation of C-terminal thioester-tagged proteins allowing ligation with a second peptide or protein containing an N-terminal Cys or Ser residue. The efficient synthesis of C-terminal thioesters of bacterially expressed proteins was found through studies of the N-terminal cleavage mechanism of inteins. Generally, the cleavage of the peptide bonds at either the N-terminus or the C-terminus of the intein can occur independently. Replacement of the C-terminal Asn by Ala blocked the splicing process in the GB-D and VMA inteins [139, 146]. However, the lack of succinimide formation did not affect the preceding *N*- to *S*- or *N*- to *O*-acyl rearrangement at the N-terminal splicing junction. Incubation of this modified inteins with thiols, such as dithiothreitol, releases the corresponding free C-terminal thioester-tagged extein from the N-terminal splicing junction through trans-thioesterification [146]. This thiol-inducible cleavage activity of an engineered intein initiated the extensive use of other intein mutants in biotechnology to obtain mono-thioester-labeled proteins and N-terminal Cys proteins [147]. The IMPACT (Intein-Mediated Purification with an Affinity Chitin-binding Tag) system [146] is commercially available and allows single column isolation of protein thioesters by utilizing the thiol-induced self-cleavage activity of various inteins (Figure 9.12). In this system, the gene of the target protein is cloned into an expression vector followed by genes of a modified intein and a chitin-binding domain (CBD). In this case the expressed protein is composed of the fusion of the protein of interest to the N-terminus of the modified intein and the fusion of the CBD to the C-terminus of the intein. The presence of the CBD enables the affinity purification of the expressed three segmental fusion proteins. All other cell proteins can be washed away from the absorbed fusion protein, and after initiation of the cleavage with an excess of thiol and overnight incubation, the protein of interest can be eluted as a

Figure 9.11 ` Protein-splicing structures and mechanisms. (a) A protein-splicing precursor is composed of an intein (red cylinder) flanked by N- and C-terminal exteins (gray cylinders). The conserved amino acid residues of protein-splicing elements are present in the N- and C-terminal splicing regions. The N-terminal amino acids of inteins are Cys (Cys_{int}), Ser (Ser_{int}), or Ala (Ala_{int}) and the C-terminus contains a conserved His/Asn pair. Also, the N-terminus of the C-extein is Cys (Cys_{ext}), Ser (Ser_{ext}), or Thr (Thr_{ext}). (b) Standard mechanism of protein splicing consisting of four steps. In the first step the linear (thio)ester is formed between the carbonyl of the C-terminal amino acid of N-extein and Cys_{int} or Ser_{int} of intein by an *N*- to *S*- or *N*- to *O*- (*S* or *O* shown as red X) acyl shift. In the second step the branched (thio)ester is formed by transferring of the N-extein from Cys_{int} or Ser_{int} to Cys_{ext}, Ser_{ext}, or Thr_{ext}. In the next step the linear (thio)ester of N- and C-exteins is released from intein by cyclization of the intein C-terminal Asn and formation of succinimide. The (thio)ester bond of the released exteins is converted to an amide bond by *S*- or *O*- (shown as black X) to *N*-acyl shift in the fourth step. (c) Different mechanisms and final products of nonstandard and mutants of inteins. The protein splicing of inteins bearing Ala_{int} starts from the second step of the standard mechanism (b). N-terminal Cys proteins can be produced by Cys_{int} or Ser_{int}/Ala mutants. C-terminal thioester proteins can be produced by an Asn/Ala mutant in the presence of a suitable thiol reagent. Mutations of Cys_{int} or Ser_{int} and Asn to Ala result in no splicing.


A
CSP
Heterologous protein expression
Intein
Protease recognition sequence
Xa recoginition sequence
Cys
Target peptide
1) Expression
2) Binding
3) washing
Factor Xa
Target peptide
Intein
CBD
Chitin
CBD
Intein
Target peptide
Cleavage by change of pH or temperature
One-pot strategy
N-Terminal Cys containing peptide
Cleavage with thiol reagent RSH
Peptide ligation
C-Terminal thioester containing peptide
N-Terminal Cys containing peptide
Semi-synthetic enzyme
B
Xa recoginition sequence
Cys
peptide 2
S R
Cys
peptide 1
Peptide ligation
Xa recoginition sequence
Cys
peptide 2
peptide 1
Factor Xa
peptide 3
peptide 2
peptide 1
Peptide ligation
Cys
peptide 2
peptide 1
Semi-synthetic enzyme
peptide 3
S R

C-terminal thioester from the chitin resin. Several inteins are available that differ with respect to the thiols used at 4 °C [139]. The modified mini inteins are commonly used with a mutation of Asn to Ala from the genes of *Mycobacterium xenopi* DNA gyrase A (*Mxe* GyrA), *S. cerevisiae* vacuolar ATPase (*Sce* VMA1), *Methanobacterium thermoautotrophicum* (*Mth* RIR1), and *Synechocystis* sp. PCC6803 (*Ssp* DnaB). The cleavage takes place only at the N-terminus of the intein because of the absence of the Asn cyclization. These inteins can be cleaved through induction with various thiols with great efficiency [148–150]. This is an important chemical aspect for ongoing protein ligation together with the thioester stability. For the thiolysis of the intein fusion proteins, a broad range of thiols have been investigated [148–150]. The choice of a certain thiol depends on the accessibility of the catalytic pocket of the intein/extein splicing domain and the properties of the target protein of interest. In general, the thiols should be small, nucleophilic molecules that can enter the catalytic pocket to attack the thioester bond connecting the extein and the intein. For further application of protein thioesters in EPL two things have to be considered to be dependent on the synthesis strategy. On one hand, the protein thioester should be stable to hydrolysis in order to be isolated. On the other hand, the thioester should also be reactive enough in EPL. Simple alkyl thioesters are quite stable to hydrolysis, but not very reactive. Mixtures of alkylthiols and thiophenol [148, 149] or 2-mercaptoethanesulfonic acid (MESNA) [150] improve the reactivity. If there is no need for thioester isolation, MESNA or thiophenol could be used directly for the induction of the cleavage and the subsequent reaction. Instead of thiols, other nucleophiles like hydroxylamine [151] can be used to induce protein splicing and the isolation of the target protein. Additionally, recombinant inteins such as *Mth* RIR1, *Ssp* DnaB, and *Mxe* GyrA inteins have all been mutated such that cleavage at the C-terminal splice junction (i.e., between the intein and the C-extein) is induced in a pH- and temperature-dependent fashion [139] (Figure 9.12). Release of these inteins allows the production of an N-terminal Cys-containing protein without the need to resort to exogenous protease treatment, which can often result in cleavage at nonspecific sites. This can be applied to protein purification or for the synthesis of the N-terminal Cys-containing proteins [152]. The EPL requires a peptide or protein that contains an amino terminal Cys besides

Figure 9.12 Expressed protein ligation. (a) The routes toward semisynthetic enzymes. N-terminal Cys protein (dark gray cylinder) can be prepared through three routes: CPS, heterologous expression of protein fused with a protease recognition sequence (light gray cylinder) in front of Cys, or with an intein–CBD (red/black cylinders, IMPACT). The C-terminal thioester protein can be prepared through two routes: CPS and heterologous expression protein fusion with intein–CBD. In the case of IMPACT, protein expression and purification through a chitin-affinity gel is followed by peptide ligation between the fusion protein of the C-terminal thioester peptide and the released N-terminal Cys peptide (one-pot strategy) or between the released N-terminal Cys and the C-terminal thioester peptides. (b) Sequential EPL strategy for preparing semisynthetic protein from more than two peptides. The nonterminal peptides contain a C-terminal thioester group and a cryptic N-terminal Cys masked by a factor Xa-cleavable pro-sequence. After ligation of C-terminal thioester with N-terminal Cys of the C-terminal peptide, the protecting group of the cryptic N-terminal Cys is removed to allow ligation with the N-terminal peptide.

the α-thioester moiety for protein semisynthesis. To synthesize proteins possessing Cys, the cDNAs of interest protein can be cloned into various commercially available vectors as mentioned above (IMPACT). The expressed protein is composed of the fusion of the protein of interest to the C-terminus of the modified intein and the fusion of the CBD to the N-terminus of the intein. Thus, after expression, the intein–CBD fusion protein can be purified on a chitin column and cleaved by changing pH or temperature. This will lead to the free Cys proteins. One drawback in the intein-based synthesis of Cys proteins is the possible spontaneous cleavage, which results in a loss of the purification tag [153].

9.13 EPL

One of the largest obstacles of NCL in biotechnology is producing the N-terminal thioester peptide by CPS, which gives low yields of product in comparison with the high yield of product of heterologous protein expression (HPE). The other major obstacle, which limits the use of NCL in synthesis of semisynthetic proteins, arises if the noncanonical part of interest should be incorporated at the C-terminal end of the protein or in the middle part of the primary structure of a high molecular mass protein. For example, at least five synthetic peptide segments (50 amino acids in length) are needed to produce a 250-amino-acid protein having a C-terminus noncanonical amino acid by NCL. Also, five to six synthetic peptide segments and one recombinant polypeptide are needed to construct a 500-amino-acid protein having the noncanonical amino acid in the middle of the protein sequence. EPL, which can be considered an extension of the NCL method, overcomes these limitations by combining the diversity and flexibility of CPS and the high yield of HPE. The EPL and NCL methods differ in the ability to produce either an N-terminal thioester peptide, a C-terminal Cys peptide, or both, by using inteins (see Section 9.12). In EPL the short noncanonical part containing peptide segment (N- or C-terminal segment) can be synthesized and chemically ligated to the unlimited size of recombinant protein, producing high yields of a pure product. An excellent example of the success of EPL as a synthetic method for semisynthetic protein is the preparation of analogs of the *E. coli* RNA polymerase β subunit [154], which contains 1407 amino acids and is beyond the range of total synthesis even using sophisticated sequential NCL [129, 155].

EPL, like NCL, produces a Cys at the ligation site. This means that a Cys must be located every 50 residues or less in the primary structure. Although many hundreds of proteins meet this requirement, producing Cys at the ligation site is the major restriction of EPL. In cases where there is not a native Cys at a suitable position, the required Cys can be introduced into a suitable position in such a way that it does not affect the structure and function of the protein [156, 157]. The effect of a Cys mutation on protein structure and function can be evaluated prior to the semisynthesis [158]. In general, the mutation is designed to be as conservative as possible, such as mutation of Ala or Ser to Cys [159] or introducing Cys into a linker or loop region [156].

Furthermore, mutating highly conserved residues should be avoided. Another factor in choosing where to introduce a Cys for ligation is the identity of the preceding amino acid; this residue will be at the C terminus of the N-terminal thioester fragment (see Section 9.11). One should try to avoid using Thr–Cys, Ile–Cys, and Val–Cys ligation junctions. EPL reactions can be performed via two strategies – either the thiolysis and NCL reactions are carried out in one-pot or the recombinant protein N-terminal thioester is isolated first (Figure 9.12). The former approach limits, to some extent, the type of additives that can be present in the reaction mixture, since the intein must remain folded during the ligation reaction. Nonetheless, these one-pot EPL reactions have been successfully performed in the presence of detergents [160], guanidinium chloride up to 2 M [161], urea up to 4 M [162], mixed aqueous/organic solvent systems [161], and a host of different thiols [147, 149, 150, 156]. If the protein N-terminal thioester is first isolated, the subsequent NCL step can be performed with a much broader range of additives, including agents that fully denature the proteins [163]. The advantage of additives such as denaturants or detergents that increase solubility is that they allow high concentrations (millimolar) of the reactant polypeptides to be used, thereby improving the ligation yields.

Most applications of EPL to date have involved just two polypeptides and thus a single ligation reaction. This requires that the region of interest in the protein be relatively close (within 50 residues) to the native N- or C-terminus. To address this limitation, a sequential ligation strategy was developed to allow multiple polypeptides to be linked together in series [164]. The ability to ligate just three peptide fragments together in a spatially controlled manner opens up the entire primary structure of a protein to chemical modification (Figure 9.12). In the sequential EPL method, the synthetic peptide insert contains an N-terminal thioester group and a cryptic C-terminal Cys masked by a factor Xa-cleavable pro-sequence. The reversible Cys protection is necessary to prevent the peptide from reacting with itself in either an intra- or intermolecular fashion. Removal of the protecting group sequence following the first ligation reaction allows the second ligation to be performed in a controlled and directed fashion. Other Cys protection strategies have been used in sequential ligation reactions [129, 155, 165, 166]. However, the advantage of the proteolytic approach is that the protecting group sequence can be genetically encoded. This advantage has allowed recombinant protein inserts to be used in sequential EPL reactions [167]. EPL has been applied to many different protein classes, including kinases [167], phosphatases [168], and polymerases [154], and introduced a variety of noncanonical residues into these proteins.

9.14
Post-Translational Chemical Modification

In the cell, post-translational chemical modification of amino acid side-chains or incorporation of prosthetic groups or metal ions into protein structures are used to expand the functional diversity, which is otherwise limited to a few functional groups (hydroxyl, phenolic, thiol, amine, imidazole, guanidinium, and carboxylic acid

groups) of the canonical amino acids. The post-translational chemical modification of amino acid side-chains can be considered one of the oldest techniques used in the preparation of semisynthetic enzymes. Chemical modification of protein originated with the interest in quantitative determination of proteins and their constituent amino acids. Later, chemical modifications were used to identify the particular amino acid residues required for the protein functionality such as catalysis or ligand binding. For example, active-site residues of different proteases were identified by inhibition the proteolytic activity by residue-specific modifying reagents. During the last decade increasing attention to the post-translational chemical modification of protein has been promoted by practical interests. Examples include possible pharmacological, medical, environmental, and biotechnological applications, such as conversion of protein toxins into toxoids [169], or production of enzymes with enhanced properties [170, 171] or with useful new functions [172, 173]. In 1966, and previous to the invention of the powerful methods of site-directed mutagenesis, the conversion of an active-site Ser residue of the Ser protease, subtilisin, to Cys by Polgar and Bender [16] and Neet and Koshland [17] was a pivotal step for redirecting the aim of chemical modification from identification of the functional residues of a protein to functionalization of the protein structure. The post-translational chemical modification technique relies on the differences in the chemical reactivity of the canonical amino acid side-chains toward different modifying reagents as well as possible changes in chemical reactivity of a given amino acid side-chain due to different environments.

The essentially unlimited structural and functional diversity of modifying reagents, the post-translation chemical modification technique can in principle generate a large number of noncanonical amino acids in a protein in an easy, fast, inexpensive, and quantitative manner avoiding many limitations of the previous methods. However, there are many considerations regarding the targeted amino acid side-chain and reagent that should be taken into consideration for the design of a successful modification. The protein of interest should have a unique amino acid side-chain in the target position or the target amino acid residue has an environment in which the target residue becomes enhanced in comparison with other similar residues. Also, the target amino acid side-chain needs to be accessible to the modifying reagent. In addition, the chemical reaction between the modifying reagent and the target amino acid side-chain should be fast, highly selective for the target side-chain, not produce side-chain byproducts, and occur under conditions in which the protein structure and reagent are stable.

Many useful modifying reagents have been developed over the years to meet the varied and specific needs of protein chemistry. These reagents, which are provided by many chemical and molecular probe suppliers, are side-chain selective reagents, and react under certain specified conditions with single or a limited number of side-chain groups, such as thiol, amine, hydroxyl, phenolic, and carboxylic acid residues. The carboxyl group of Glu or Asp can be specifically converted into amides, acyl hydrazides, or hydroxamic acids by coupling with amines, hydrazines, and hydroxylamines, respectively, in aqueous solution by using water-soluble carbodiimides as condensing agents [174]. To reduce intra- and interprotein coupling to Lys residues,

which is a common side-reaction, carbodiimide-mediated coupling should be performed in a concentrated protein solution at a low pH, using a large excess of the nucleophile. The carboxyl group can also be esterified by using diazoalkanes without the need of a catalyst [174, 175].

The hydroxyl group of Ser or Thr has extremely low chemical reactivity in aqueous media. Few reagents are specific for hydroxyl groups in aqueous solution, especially in the presence of more reactive nucleophiles such as thiols and amines. It is therefore difficult to selectively modify Ser and Thr in proteins except when they exhibit unusual reactivity, such as by residing at an enzyme active site. An activated phenolic group of Tyr can react with 4-chloro-7-nitrobenz-2-oxa-1,3-diazole [176]. Tyr in some proteins can be selectively modified by initial nitration of the *ortho* position of its phenol using tetranitromethane, nitric oxide or nitrosamines. Subsequent reduction of the *o*-nitro-Tyr with sodium dithionite will form an *o*-amino-Tyr, which between pH 5 and 7 can react with most amine-reactive reagents, except succinimidyl esters [174, 177, 178].

The amine-reactive reagents are mostly alkylating or acylating reagents that form secondary amines, carboxamides, sulfonamides, or thioureas upon reaction with amines [174, 179–181]. The kinetics of the reaction depends on the reactivity and concentration of both the reagent and the amine [174]. The reaction medium should not contain buffers with free amines, such as Tris and glycine, or high concentrations of nucleophilic thiols. The most significant factors affecting reactivity are the chemical structure and the basicity of the amine. The majority of proteins contain Lys residues and most have a free α-amino group at the N-terminus. Lys is moderately basic and reactive toward most acylating reagents. However, below pH 8 the concentration of the free base form of aliphatic amines is very low. Thus, the kinetics of acylation reactions of amines by isothiocyanates, succinimidyl esters, and other reagents are optimal between pH 8.5 and 9.5 [174]. In contrast, the α-amino group at the N-terminus usually has a pK_a of around 7, such that it can sometimes be selectively modified by reaction at near neutral pH. In aqueous solution, acylating reagents are unreactive with the amide group of peptide bonds and the side-chain amides of Gln and Asn, the guanidinium group of Arg, and the imidazolium group of His. Isocyanates, which are very susceptible to deterioration during storage, can be obtained from acyl azides and react with amines to form ureas. Isothiocyanates, which are moderately reactive, but quite stable in water and most solvents, form thioureas upon reaction with amines [181]. Succinimidyl esters are excellent reagents for amine modification, because the amide bonds they form are as stable as peptide bonds, and show good reactivity with aliphatic amines and very low reactivity with Ser, Thr, Tyr, and His [182–184]. Succinimidyl esters can react with Cys in organic solvents to form thioesters that may transfer the acyl moiety to a nearby amine. Sulfosuccinimidyl esters have higher water solubility than simple succinimidyl esters and eliminate the need for organic solvents in the conjugation reaction [185]. However, they are also more polar than succinimidyl esters, which makes them less likely to react with buried amines in proteins. Like succinimidyl esters, carbonyl azides are active esters that can react with amines to yield amides. Sulfonyl chlorides are highly reactive and quite unstable in water, especially at the higher pH required

for reaction with aliphatic amines [186]. Protein modification by this reagent is best done at low temperature. Once conjugated, however, the sulfonamides that are formed are extremely stable. Aldehydes condense with primary amines to form imines, Schiff bases, which can be reduced by relatively strong reducing agents such as sodium cyanoborohydride or borohydride to give stable secondary amines [187, 188]. Instead of using strong reducing agents, which also break accessible disulfide bonds, McFarland and Francis have used mild transition metal hydrogenation catalyzed by iridium hydride to produce the secondary amines following condensation of Lys with aldehydes [189]. Although this mild hydrogenation takes days to complete, the reduction reaction takes place under physiological conditions and does not cleave disulfide bonds.

The common thiol-reactive reagents are primarily alkylating reagents, including halogenoacetamides, maleimides, alkyl halides, benzylic halides, and bromomethylketones [174]. Reaction of any of these functional groups with thiols usually proceeds rapidly at or below room temperature in the physiological pH range (pH 6.5–8.0) to yield chemically stable thioethers. Alkyl thiosulfates and mercurials are suitable reagents for modifying thiol groups in a reversible fashion. Thiosulfates are similar to disulfides in that they react with thiols to form mixed disulfides, but an excess of the reagent is required to obtain a stoichiometrical conversion of the thiol [190]. *p*-Chloromercuribenzoate and other mercurial reagents rapidly react with thiols to form mercaptide adducts, and the reaction does not need an excess of the reagent [191, 192]. Disulfide and mercury-thiol bonds can be cleaved by thiol-containing agents. Iodoacetamides readily react with thiols to form thioethers [193]. However, when Cys residues are blocked or absent, iodoacetamides can sometimes modify Met, His, and Tyr residues [174]. Although iodoacetamides can react with the free base form of amines, most aliphatic amines, except the amino group at a protein's N-terminus, are protonated and thus relatively unreactive below pH 8. Iodoacetamides are intrinsically light-sensitive and especially unstable in solution; reactions should therefore be carried out in the dark. Maleimides are excellent reagents for thiol-selective modification [194]. The thiol is added across the double bond of the maleimide to yield a thioether. Maleimides apparently do not react with Met, His or Tyr, but a reaction of maleimides with amines can occur at a higher pH than the normal reaction of maleimides with thiols [174]. Above pH 9 thiol-maleimide adducts can hydrolyze to a mixture of isomeric maleamic acid adducts, which can ring-open by nucleophilic reaction with an adjacent amine to yield cross-linked products.

In spite of the fact that most of the modifying reagents are side-chain selective under certain conditions, they are generally not site-directed reagents due to the multiplicity of the side-chains in the target protein. Heterogeneity of the product, which results not only from the multiplicity of the target residues but also from incomplete chemical modification of the target residue(s), is a difficulty commonly encountered. A heterogeneous product can be purified by separating the desired modified protein from undesired side products. In some cases a distinguishing property resulting from the modification can be used for the separation. For example, targeting the ligand-interacting residue with chemical modification can change the

binding affinity of the protein for its ligand and targeting the subunit interface residues of the multimer protein with cross-linking reagents will increase the molecular mass of the protein [195, 196].

Alternatively, a site-directed chemical modification (SDCM) strategy – a combination of site-directed genetic mutagenesis and chemical modification – can direct the chemical modification to a desired site producing a homogeneous product [171–173]. In general, this strategy relies on using the specificity of site-directed mutagenesis in preparing a protein containing a chemically reactive and unique amino acid at a desired site followed by chemical modification of the incorporated amino acid with a side-chain-selective reagent. Several proteins are Cys deficient or have very few Cys residues, which can be removed by mutation, if they do not play a critical role in the protein function or structure. Thus, the SDCM strategy commonly introduces a Cys residue at the target site in a thiol-free background. The protein is then chemically modified by a thiol-reactive reagent to produce the desired side-chain-modified homogenous product. Similarly, SDCM can target certain positions by selective modification of low-abundancy Trp residues. Using transition metal-catalyzed modification, Antos and Francis have successfully modified Trp in a specific manner by using a rhodium carbene [197].

Post-translational chemical modification techniques have been used extensively for increasing enzyme stability, solubility in organic solvent, changing substrate specificity, immobilizing on solid support, or introducing new enzyme activity. Furthermore, this technique has been used not only in incorporation of noncanonical amino acids, but also in incorporating large prosthetic groups to proteins [173, 198].

9.15
Examples of Semisynthetic Enzymes

There are hundreds of examples of noteworthy semisynthetic enzymes and proteins produced by the techniques presented in this chapter. This section highlights a few cases showing the highly efficient use of the previously mentioned methods in producing proteins as new functional enzymes.

Production of enantiopure compounds is a major objective of fine chemical and pharmaceutical industries. While many chemical syntheses produce racemic mixtures of enantiomers, the products of corresponding enzymatic syntheses are often highly enantiopure. Enzymes may act on only one enantiomer of a chiral substrate, or produce only one enantiomer or diastereomer from a prochiral substrate. This specificity is related to the asymmetric structure of the enzyme molecule, including the three-dimensional folding of the polypeptide backbone and the direction of the amino acid side-chains in the folded protein molecule. Naturally occurring L-enzymes are composed of L-amino acids ("L-enzymes") and may show stereospecificity in the production of a desired enantiomeric product. This raises the question about the substrate stereospecificity of the corresponding "D-enzyme" composed of corresponding D-amino acids. Would it be possible to synthesize a D-enzyme with similar catalytic activity as the L-enzyme but with reversed enantiomeric selectivity?

Kent *et al*. answered this question by using CPS, which has the advantage over other methods in its ability to incorporate D-amino acids into protein, to synthesize an all D-amino acid form of HIV-1 protease [199]. They structurally and functionally characterized the chemically synthesized D-HIV-1 protease of 99 amino acids. The ion-spray MS and circular dichroism spectra showed that the D-form of HIV-1 protease had the same molecular mass and secondary structure as the L-form. The catalytic activity of D-HIV-1 protease toward the fluorogenic D-hexapeptide substrate was the same as that of L-HIV-1 protease toward the L-hexapeptide analog. Similarly, affinity of D-HIV-1 protease for D-form of pseudopeptide inhibitor, MVT101, was the same as that of L-HIV-1 protease for the L-form of MVT101. On the other hand, D-HIV-1 protease had no activity or affinity toward the L-form of the substrate and inhibitor, respectively. In contrast, the achiral inhibitor Evans Blue inhibited both enantiomers of HIV-1 protease, suggesting that a D-enzyme working on an achiral substrate should be fully functional *in vivo*. D-Enzymes have important therapeutic potential, since they have long half-lives *in vivo*, due to their resistance to naturally occurring L-proteases. The synthesis of D-variants of industrially important enzymes for producing desired enantiopure compounds can also be envisaged.

Chemical and conformational stability of enzymes are essential to meet requirements for biotechnical applications. Replacement of the α-peptidic backbone with a highly stable non-α-peptidic structure may increase protein stability. β-Peptides have greater secondary structural stability than do conventional peptides [200]. Raines *et al*. incorporated a di-β-peptide segment into a 124-amino-acid RNase A by using EPL [201]. Incorporation of this segment can be done neither by pretranslation modification techniques nor by CPS. They replaced the dipeptide Asn–Pro segment in a β-turn comprising Gly–Asn–Pro–Tyr of the native RNase A with a di-β-peptide (*R*)-nipecotic acid (Nip)–(*S*)-Nip. RNase A fragment 1–94 was expressed as a fusion protein with the *Mxe* intein and CBD, and was then ligated to a chemically synthesized peptide 95–124 containing the di-β-peptide. RNase A fragment 1–94 had no detectable ribonucleolytic activity. In contrast, the semisynthetic variant containing the dipeptide (*R*)-Nip–(*S*)-Nip had identical activity to that of the wild-type enzyme. Incorporation of the dipeptide (*R*)-Nip–(*S*)-Nip increased the conformational stability by 1.2 °C. On the other hand, replacement of Pro in the β-turn with other canonical α-amino acids decreased the thermal stability by 10 °C. In contrast, incorporation of the isomeric (*R*)-Nip–(*R*)-Nip in the β-turn unfolded the protein structure. The β-turn is a favored motif for a polypeptide chain reverse turn, as required for its folding into a compact structure [202]. Reverse turns are critical to protein conformational stability and many protein–protein interactions. In addition, turns are exposed structures and often sites for degradation by proteolytic enzymes. Modification of β-turns can be useful to stabilize other enzyme structures.

Broadening or redirecting the specificity of enzymes are among the most important goals of enzyme engineering. These goals can be realized by diverse strategies. One involves changing the conformation of the ligand binding site to fit the desired substrate. Another is based on linking a highly specific binder to the catalytic

structure to promote selective binding and catalytic efficiency. Schultz *et al*. used the latter strategy for designing semisynthetic DNA-cleaving enzymes [203]. They modulated the sequence specificity of the staphylococcal nuclease through linking an oligonucleotide of defined sequence to the enzyme using the SDCM method. A unique Cys was introduced near the active site of the nuclease, which produced a low sequence specificity nuclease. An oligonucleotide with a spacer was then linked to the introduced Cys via a disulfide bond. The resulting conjugate displayed a significant change in sequence specificity and cleaved a single-stranded DNA segment 64 nucleotides in length almost exclusively at one site. However, the nuclease activity of this semisynthetic variant showed high nonspecific background activity. Two additional mutations and shortening of the spacer length between the oligonucleotide and the enzyme were required to produce a highly specific and efficient variant with a K_M of 120 nM and k_{cat} of 1.2 min^{-1} at a temperature equal to the melting temperature of the oligonucleotide–substrate complex [204]. Also, a semisynthetic staphylococcal nuclease was developed to cleave RNA [205] and double-stranded DNA [206]. To accomplish cleavage of the doubled-stranded DNA, a polypyrimidine-containing oligonucleotide was linked to staphylococcal nuclease via a disulfide bond with a uniquely introduced Cys. The oligonucleotide recognition element was designed to hybridize to double-stranded DNA via triple-helix formation through Hoogsteen hydrogen bonding. Upon activation with Ca^{2+}, the resulting semisynthetic enzyme cleaved a 70-base-pair double-stranded DNA fragment adjacent to the 5′-end of the homopurine target with 75% efficiency at room temperature within 1 h. By the same strategy of linking a specificity-determining element to a catalytically active structure novel enzymes were constructed. Sigman *et al*. engineered a novel nuclease by linking a DNA-binding protein, *E. coli* Trp repressor with the chemical nuclease 1,10-phenanthroline (OP)-Cu coordinate complex [207]. OP-Cu achieves the nucleic acid backbone cleavage by oxidative attack of a Cu-oxo species on the C_1–H of the deoxyribose within the minor groove of DNA producing the 3′- and 5′-phosphomonoester, free base, and a 5-methylene furanone [208]. Minor groove-directed chemistry of OP-Cu is particularly advantageous because the contacts between the specific target sequence and DNA-binding proteins are primarily in the major groove. Thus, attachment of the chemical cleavage functionality does not disrupt crucial binding determinants. Sigman *et al*. used SDCM to construct this semisynthetic nuclease by introducing a unique Cys at the target site prior to alkylation with 5-(iodoacetamido)-OP. The resulting conjugate bound tightly to the operator sequence regulated by the Trp repressor with a K_D of approximately 7×10^{-9} M and with cleavage efficiencies of 80–90% for single- and double-strand DNA [209].

A major limitation of using Cys-containing enzymes in biotechnology is the chemical instability of these enzymes toward electrophiles and oxidizing conditions. Replacement of structurally critical Cys with Ala or Ser without redesign of the Cys environment can disturb the enzyme structure and subsequently have an inactivation effect on the enzyme. SDCM has an important advantage over other techniques in that it provides the ability to construct vast chemical libraries that are useful for protein redesign. Recently, Mannervik *et al*. have applied SDCM in re-engineering

glutathione transferase (GST) producing a highly stable and efficient Cys-free variant [195].

GSTs are multifunctional dimeric proteins catalyzing many reactions including the nucleophilic attack of glutathione (GSH) on an electrophilic carbon, nitrogen, sulfur, or oxygen atom of a second substrate such as alkyl and aryl halides, activated alkenes, quinones, nitrate esters, thiocyanates, hydroperoxides, and epoxides [210]. Furthermore, certain GSTs catalyze the isomerization reaction of the double bond through GSH. Pi class human GST P1-1 has four Cys residues, one of them, Cys48, is responsible for the sensitivity of the enzyme to chemical and oxidative inactivation [211]. Conventional mutagenesis of Cys48 to a homologous analog Ser or to Ala decreased the catalytic efficiency of the enzyme [212]. This decrease in the catalytic efficiency of GST P1-1 results from the disturbance of the structure of the catalytic α2-loop, the GSH-binding loop consisting of residues 36–52 [213]. The α2-loop, which is controlled by fitting of the side-chain of Tyr50 in a hydrophobic pocket of the neighboring subunit of GST P1-1, modulates GSH affinity and controls the rate of catalysis [214]. Mutation of Tyr50 to several nonaromatic amino acids inactivated the enzyme, indicating an essential role of the Tyr50 side-chain for catalytic competence [215]. In order to investigate the structural requirements of the side-chain of residue 50, Tyr50, was replaced by Cys to allow chemical modifications by diverse reagents. All Cys residues, including the chemical instability factor Cys48, were first mutated to produce a Cys-free mutant [195]. The mutation of Cys48 into Ala, the Cys-free mutant displayed a 2-fold increase of the K_M value for GSH and a 24- to 41-fold decrease of K_i for the active-site ligands *S*-methyl- and *S*-hexyl-GSH, respectively, in comparison with the that contained Cys in position 48 only [213]. Tyr50 was then mutated to Cys producing a Y50C mutant to allow redesign of the α2-loop through chemical modification of Cys50 with a library of different structures of aromatic and aliphatic thiol-reactive compounds [195]. Y50C showed a 3000-fold decrease of the catalytic efficiency, in comparison with Cys-free mutant, due to disturbance of the α2-loop structure. Applying combinatorial chemistry, Y50C was then treated with three different libraries each containing 10 thiol-reactive compounds different in structure but similar in chemical reactivity. Several reagents clearly activated the enzyme by alkylation of the sulfur of Cys50. Deconvolution of the most active compound library, by testing the components of the library individually, identified *n*-butyl bromide as the optimal modifier producing a *S*-butyl-Cys50 mutant that is stable toward oxidizing conditions and electrophiles like the Cys-free mutant as well as being as catalytically efficient as wild-type GST P1-1. Interestingly, several aromatic compounds providing structures similar to the Tyr side-chain failed to restore the catalytic efficiency of the wild-type to any noteworthy degree, while some aliphatic compounds activated the enzyme to different extents depending on their structure. This study also allowed discrimination between effects on the two kinetic constants, K_M and k_{cat}, by comparing the kinetics of different semisynthetic GST P1-1 [213]. This successful molecular surgery involving replacement of Tyr by *S*-butyl-Cys can possibly be applied to other proteins for substitution of undesired amino acid side-chains with retention of functional properties.

9.16 Conclusions

This chapter has focused on different available techniques that can be used for construction of semisynthetic proteins with altered properties. The history of semisynthetic enzymes began with chemical modification of active-site residues producing inactivated enzymes. The techniques subsequently developed for residue-specific chemical modification produce enzymes with modified properties, but are mostly undirected and therefore give heterogeneous products. Nowadays, several techniques, largely based on recombinant DNA methods, can site-specifically incorporate a synthetic part to a protein and produce homogeneous and rationally designed semisynthetic enzymes. All of these methods, as indicated in this chapter, have advantages and limitations. Selecting a suitable approach for constructing the desired semisynthetic enzyme depends on different factors, including the properties of the targeted protein, the desired structure of the product, and the scale of synthesis. No method serves all purposes and it is ultimately the creativity of the scientist that sets limits to what can be achieved in the field of semisynthetic proteins.

References

1 Panke, S., Held, M. and Wubbolts, M. (2004) Trends and innovations in industrial biocatalysis for the production of fine chemicals. *Current Opinion in Biotechnology*, **15**, 272–279.

2 Matsumura, S., Soeda, Y. and Toshima, K. (2006) Perspectives for synthesis and production of polyurethanes and related polymers by enzymes directed toward green and sustainable chemistry. *Applied Microbiology and Biotechnology*, **70**, 12–20.

3 Margolin, A.L. (1993) Enzymes in the synthesis of chiral drugs. *Enzyme and Microbial Technology*, **15**, 266–280.

4 Panke, S. and Wubbolts, M. (2005) Advances in biocatalytic synthesis of pharmaceutical intermediates. *Current Opinion in Chemical Biology*, **9**, 188–194.

5 James, J. and Simpson, B.K. (1996) Application of enzymes in food processing. *Critical Reviews in Food Science and Nutrition*, **36**, 437–463.

6 Alcalde, M., Ferrer, M., Plou, F.J. and Ballesteros, A. (2006) Environmental biocatalysis: from remediation with enzymes to novel green processes. *Trends in Biotechnology*, **24**, 281–287.

7 Rawlings, D.E., Dew, D. and du Plessis, C. (2003) Biomineralization of metal-containing ores and concentrates. *Trends in Biotechnology*, **21**, 38–44.

8 Valenzuela, L., Chi, A., Beard, S., Orell, A., Guiliani, N., Shabanowitz, J., Hunt, D.F. and Jerez, C.A. (2006) Genomics, metagenomics and proteomics in biomining microorganisms. *Biotechnology Advances*, **24**, 197–211.

9 Arnold, F.H. (1993) Protein engineering for unusual environments. *Current Opinion in Biotechnology*, **4**, 450–455.

10 Chirumamilla, R.R., Muralidhar, R., Marchant, R. and Nigam, P. (2001) Improving the quality of industrially important enzymes by directed evolution. *Molecular and Cellular Biochemistry*, **224**, 159–168.

11 Hibbert, E.G. and Dalby, P.A. (2005) Directed evolution strategies for improved enzymatic performance. *Microbial Cell Factories*, **4**, 29–35.

12 Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B. and Zinoni, F. (1991) Selenocysteine: the 21st amino acid. *Molecular Microbiology*, **5**, 515–520.

13 Srinivasan, G., James, C.M. and Krzycki, J.A. (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science*, **296**, 1459–1462.

14 Cowie, D.B. and Cohen, G.N. (1957) Biosynthesis by *Escherichia coli* of active altered proteins containing selenium instead of sulfur. *Biochimica et Biophysica Acta*, **26**, 252–261.

15 Chapeville, F., Lipmann, F., Von Ehrenstein, G., Weisblum, B., Ray, W.J. Jr and Benzer, S. (1962) On the role of soluble ribonucleic acid in coding for amino acids. *Proceedings of the National Academy of Sciences of the United States of America*, **48**, 1086–1092.

16 Polgar, L. and Bender, M.L. (1966) A new enzyme containing a synthetically formed active site. *Journal of the American Chemical Society*, **88**, 3153–3154.

17 Neet, K.E. and Koshland, D.E. (1966) The conversion of serine at the active site of subtilisin to cysteine: a "chemical mutation". *Proceedings of the National Academy of Sciences of the United States of America*, **56**, 1606–1611.

18 Braisted, A.C., Judice, J.K. and Wells, J.A. (1997) Synthesis of proteins by subtiligase. *Methods in Enzymology*, **289**, 298–313.

19 Luo, G.M., Ren, X.J., Liu, J.Q., Mu, Y. and Shen, J.C. (2003) Towards more efficient glutathione peroxidase mimics: substrate recognition and catalytic group assembly. *Current Medicinal Chemistry*, **10**, 1151–1183.

20 Votchitseva, Yu.A., Efremenko, E.N. and Varfolomeyev, S.D. (2006) Insertion of an unnatural amino acid into the protein structure: preparation and properties of 3-fluorotyrosine-containing organophosphate hydrolase. *Russian Chemical Bulletin (International Edition)*, **55**, 369–374.

21 Schwarzer, D., Zhang, Z., Zheng, W. and Cole, P.A. (2006) Negative regulation of a protein tyrosine phosphatase by tyrosine phosphorylation. *Journal of the American Chemical Society*, **128**, 4192–4193.

22 Parsons, J.F., Xiao, G., Gilliland, G.L. and Armstrong, R.N. (1998) Enzymes harboring unnatural amino acids: mechanistic and structural analysis of the enhanced catalytic activity of a glutathione transferase containing 5-fluorotryptophan. *Biochemistry*, **37**, 6286–6294.

23 Seyedsayamdost, M.R., Yee, C.S., Reece, S.Y., Nocera, D.G. and Stubbe, J. (2006) pH rate profiles of F*n*Y356-R2s ($n = 2, 3, 4$) in *Escherichia coli* ribonucleotide reductase: evidence that Y356 is a redox-active amino acid along the radical propagation pathway. *Journal of the American Chemical Society*, **128**, 1562–1568.

24 Thompson, P.R., Wang, D.X., Wang, L., Fulco, M., Pediconi, N., Zhang, D., An, W., Ge, Q., Roeder, R.G., Wong, J., Levrero, M., Sartorelli, V., Cotter, R.J. and Cole, P.A. (2004) Regulation of the p300 HAT domain via a novel activation loop. *Nature Structural & Molecular Biology*, **11**, 308–315.

25 Ibba, M. and Söll, D. (2001) The renaissance of aminoacyl-tRNA synthesis. *EMBO Reports*, **2**, 382–387.

26 Ibba, M. and Söll, D. (2000) Aminoacyl-tRNA synthesis. *Annual Review of Biochemistry*, **69**, 617–650.

27 Wakasugi, K., Quinn, C.L., Tao, N. and Schimmel, P. (1998) Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant. *The EMBO Journal*, **17**, 297–305.

28 Budisa, N., Alefelder, S., Bae, J.H., Golbik, R., Minks, C., Huber, R. and Moroder, L. (2001) Proteins with beta-(thienopyrrolyl) alanines as alternative chromophores and pharmaceutically active amino acids. *Protein Science* **10**, 1281–1292.

29 Bae, J.H., Alefelder, S., Kaiser, J.T., Friedrich, R., Moroder, L., Huber, R. and Budisa, N. (2001) Incorporation of beta-selenolo[3,2-*b*]pyrrolyl-alanine into proteins for phase determination in

protein X-ray crystallography. *Journal of Molecular Biology*, **309**, 925–936.

30 van Hest, J.C.M. and Tirrell, D.A. (1998) Efficient introduction of alkene functionality into proteins *in vivo*. *FEBS Letters*, **428**, 68–70.

31 Kiick, K.L., Saxon, E., Tirrell, D.A. and Bertozzi, C.R. (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 19–24.

32 Tang, Y., Ghirlanda, G., Petka, W.A., Nakajima, T., DeGrado, W.F. and Tirrell, D.A. (2001) Fluorinated coiled-coil proteins prepared *in vivo* display enhanced thermal and chemical stability. *Angewandte Chemie International Edition*, **40**, 1494–1496.

33 Wang, P., Tang, Y. and Tirrell, D.A. (2003) Incorporation of trifluoroisoleucine into proteins *in vivo*. *Journal of the American Chemical Society*, **125**, 6900–6906.

34 Mock, M.L., Michon, T., van Hest, J.C. and Tirrell, D.A. (2006) Stereoselective incorporation of an unsaturated isoleucine analogue into a protein expressed in *E. coli*. *ChemBioChem*, **7**, 83–87.

35 Yoshikawa, E., Fournier, M.J., Mason, T.L. and Tirrell, D.A. (1994) Genetically engineered fluoropolymers: synthesis of repetitive polypeptides containing *p*-fluorophenylalanine residues. *Macromolecules*, **27**, 5471–5475.

36 Renner, C., Alefelder, S., Bae, J.H., Budisa, N., Huber, R. and Moroder, L. (2001) Fluoroprolines as tools for protein design and engineering. *Angewandte Chemie International Edition*, **40**, 923–925.

37 Beiboer, S.H., van den Berg, B., Dekker, N., Cox, R.C. and Verheij, H.M. (1996) Incorporation of an unnatural amino acid in the active site of porcine pancreatic phospholipase A_2. Substitution of histidine by 1,2,4-triazole-3-alanine yields an enzyme with high activity at acidic pH. *Protein Engineering*, **9**, 345–352.

38 Kiick, K.L. and Tirrell, D.A. (2000) Protein engineering by in vivo incorporation of non-natural amino acids: Control of incorporation of methionine analogues by methionyl-tRNA synthetase. *Tetrahedron*, **56**, 9487–9493.

39 Kiick, K.L., van Hest, J.C.M. and Tirrell, D.A. (2000) Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host. *Angewandte Chemie International Edition*, **39**, 2148–2152.

40 Kiick, K.L., Weberskirch, R. and Tirrell, D.A. (2001) Identification of an expanded set of translationally active methionine analogues in *Escherichia coli*. *FEBS Letters*, **502**, 25–30.

41 Tang, Y. and Tirrell, D.A. (2001) Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host. *Journal of the American Chemical Society*, **123**, 11089–11090.

42 Fersht, A.R. and Kaethner, M.M. (1976) Enzyme hypersensitivity. Rejection of threonine by the valyl-tRNA synthetase by misacylation and hydrolytic editing. *Biochemistry*, **15**, 3342–3346.

43 Englisch, S., Englisch, U., Vonderhaar, F. and Cramer, F. (1986) The proofreading of hydroxy analogues of leucine and isoleucine by leucyl-tRNA synthetases from *E. coli* and yeast. *Nucleic Acids Research*, **14**, 7529–7539.

44 Schmidt, E. and Schimmel, P. (1994) Mutational isolation of a sieve for editing in a transfer RNA synthetase. *Science*, **264**, 265–267.

45 Roy, H., Ling, J., Irnov, M. and Ibba, M. (2004) Post-transfer editing *in vitro* and *in vivo* by the beta-subunit of phenylalanyl-tRNA synthetase. *The EMBO Journal*, **23**, 4639–4648.

46 Ibba, M. and Hennecke, H. (1995) Relaxing the substrate specificity of an aminoacyl-tRNA synthetase allows *in vitro* and *in vivo* synthesis of proteins containing unnatural amino acids. *FEBS Letters*, **364**, 272–275.

47 Kirshenbaum, K., Carrico, I.S. and Tirrell, D.A. (2002) Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues. *ChemBioChem*, **3**, 235–237.

48 Bentin, T., Hamzavi, R., Salomonsson, J., Roy, H., Ibba, M. and Nielsen, P.E. (2004) Photoreactive bicyclic amino acids as substrates for mutant Escherichia coli phenylalanyl-tRNA synthetases. *The Journal of Biological Chemistry*, **279**, 19839–19845.

49 Mursinna, R.S., Lincecum, T.L. and Martinis, S.A. (2001) A conserved threonine within *Escherichia coli* leucyl-tRNA synthetase prevents hydrolytic editing of Leu-tRNALeu. *Biochemistry*, **40**, 5376–5381.

50 Mursinna, R.S. and Martinis, S.A. (2002) Rational design to block amino acid editing of a tRNA synthetase. *Journal of the American Chemical Society*, **124**, 7286–7287.

51 Tang, Y. and Tirrell, D.A. (2002) Attenuation of the editing activity of the *Escherichia coli* leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins *in vivo*. *Biochemistry*, **41**, 10635–10645.

52 Kwon, I., Kirshenbaum, K. and Tirrell, D.A. (2003) Breaking the degeneracy of the genetic code. *Journal of the American Chemical Society*, **125**, 7512–7513.

53 Roesser, J.R., Xu, C., Payne, R.C., Surratt, C.K. and Hecht, S.M. (1989) Preparation of misacylated aminoacyl-tRNAPhe's useful as probes of the ribosomal acceptor site. *Biochemistry*, **28**, 5185–5195.

54 Doctor, B.P. and Mudd, J.A. (1963) Species specificity of amino acid acceptor ribonucleic acid and aminoacyl soluble ribonucleic acid synthetases. *The Journal of Biological Chemistry*, **238**, 3677–3681.

55 Kwok, Y. and Wong, J.T. (1980) Evolutionary relationship between *Halobacterium cutirubrum* and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. *Canadian Journal of Biochemistry*, **58**, 213–218.

56 Drabkin, H.J., Park, H.J. and RajBhandary, U.L. (1996) Amber suppression in mammalian cells dependent upon expression of an *Escherichia coli* aminoacyl-tRNA synthetase gene. *Molecular and Cellular Biology*, **16**, 907–913.

57 Liu, D.R., Magliery, T.J., Pastrnak, M. and Schultz, P.G. (1997) Engineering a tRNA and aminoacyl tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 10092–10097.

58 Wang, L., Brock, A., Herberich, B. and Schultz, P.G. (2001) Expanding the genetic code of *Escherichia coli*. *Science*, **292**, 498–500.

59 Wang, L., Magliery, T.J., Liu, D.R. and Schultz, P.G. (2000) A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the *in vivo* incorporation of unnatural amino acids into proteins. *Journal of the American Chemical Society*, **122**, 5010–5011.

60 Wang, L. and Schultz, P.G. (2001) A general approach for the generation of orthogonal tRNAs. *Chemistry & Biology*, **8**, 883–890.

61 Goodman, H.M., Abelson, J., Landy, A., Brenner, S. and Smith, J.D. (1968) Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA. *Nature*, **217**, 1019–1024.

62 Andoh, T. and Ozeki, H. (1968) Suppressor gene Su3+ of *E. coli*, a structural gene for tyrosine TRNA. *Proceedings of the National Academy of Sciences of the United States of America*, **3**, 792–799.

63 Benzer, S. and Champe, S.P. (1962) A change from nonsense to sense in the genetic code. *Proceedings of the National Academy of Sciences of the United States of America*, **48**, 1114–1121.

64 Garen, A. and Siddiqi, O. (1962) Suppression of mutations in the alkaline

phosphatase structural cistron of *E. coli*. *Proceedings of the National Academy of Sciences of the United States of America*, **48**, 1121–1127.

65 Köhrer, C., Xie, L., Kellerer, S., Varshney, U. and RajBhandary, U.L. (2001) Import of amber and ochre suppressor tRNAs into mammalian cells: a general approach to site-specific insertion of amino acid analogues into proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 14310–14315.

66 Anderson, J.C. and Schultz, P.G. (2003) Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry*, **42**, 9598–9608.

67 Köhrer, C., Yoo, J.H., Bennett, M., Schaack, J. and RajBhandary, U.L. (2003) A possible approach to site-specific insertion of two different unnatural amino acids into proteins in mammalian cells via nonsense suppression. *Chemistry & Biology*, **10**, 1095–1102.

68 Bain, J.D., Diala, E.S., Glabe, C.G., Dix, T.A. and Chamberlin, A.R. (1989) Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *Journal of the American Chemical Society*, **111**, 8013–8014.

69 Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C. and Schultz, P.G. (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. *Science*, **244**, 182–188.

70 Anthony-Cahill, S.J., Griffith, M.C., Noren, C.J., Suich, D.J. and Schultz, P.G. (1989) Site-specific mutagenesis with unnatural amino acids. *Trends in Biochemical Sciences*, **14**, 400–403.

71 Short, G.F. III, Golovine, S.Y. and Hecht, S.M. (1999) Effects of release factor 1 on *in vitro* protein translation and the elaboration of proteins containing unnatural amino acids. *Biochemistry*, **38**, 8808–8819.

72 Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nature Biotechnology*, **19**, 751–755.

73 Carnes, J., Jacobson, M., Leinwand, L. and Yarus, M. (2003) Stop codon suppression via inhibition of eRF1 expression. *RNA*, **9**, 648–653.

74 Hecht, S.M., Alford, B.L., Kuroda, Y. and Kitano, S. (1978) "Chemical aminoacylation" of tRNA's. *The Journal of Biological Chemistry*, **253**, 4517–4520.

75 Heckler, T.G., Chang, L.H., Zama, Y., Naka, T., Chorghade, M.S. and Hecht, S.M. (1984) T4 RNA ligase mediated preparation of novel "chemically misacylated" tRNAPheS. *Biochemistry*, **23**, 1468–1473.

76 Robertson, S.A., Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C. and Schultz, P.G. (1989) The use of 5′-phospho-2 deoxyribocytidylylriboadenosine as a facile route to chemical aminoacylation of tRNA. *Nucleic Acids Research*, **17**, 9649–9660.

77 Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C.J. and Schultz, P.G. (1991) Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. *Methods in Enzymology*, **202**, 301–336.

78 Cload, S.T., Liu, D.R., Froland, W.A. and Schultz, P.G. (1996) Development of improved tRNAs for *in vitro* biosynthesis of proteins containing unnatural amino acids. *Chemistry & Biology*, **3**, 1033–1038.

79 Nenomiya, K., Kurita, T., Hohsaka, T. and Sisido, M. (2003) Facile aminoacylation of pdCpA dinucleotide with a non-natural amino acid in cationic micelle. *Chemical Communications*, **17**, 2242–2243.

80 Bessho, Y., Hodgson, D.R. and Suga, H. (2002) A tRNA aminoacylation system for non-natural amino acids based on a programmable ribozyme. *Nature Biotechnology*, **20**, 723–728.

81 Wang, K., Neumann, H., Peak-Chew, S.Y. and Chin, J.W. (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature Biotechnology*, **25**, 770–777.

82 Saito, H., Kourouklis, D. and Suga, H. (2001) An *in vitro* evolved precursor tRNA with aminoacylation activity. *The EMBO Journal*, **20**, 1797–1806.

83 Saito, H. and Suga, H. (2001) A ribozyme exclusively aminoacylates the 3′-hydroxyl group of the tRNA terminal adenosine. *Journal of the American Chemical Society*, **123**, 7178–7179.

84 Saito, H., Watanabe, K. and Suga, H. (2001) Concurrent molecular recognition of the amino acid and tRNA by a ribozyme. *RNA*, **7**, 1867–1878.

85 Murakami, H., Saito, H. and Suga, H. (2003) A versatile tRNA aminoacylation catalyst based on RNA. *Chemistry & Biology*, **10**, 655–662.

86 Murakami, H., Ohta, A., Ashigai, H. and Suga, H. (2006) A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods*, **3**, 357–359.

87 Murakami, H., Kourouklis, D. and Suga, H. (2003) Using a solid-phase ribozyme aminoacylation system to reprogram the genetic code. *Chemistry & Biology*, **10**, 1077–1084.

88 Hui, A. and De Boer, H.A. (1987) Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 4762–4766.

89 Jacob, W.F., Santer, M. and Dahlberg, A.E. (1987) A single base change in the Shine–Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 4757–4761.

90 Carter, A.P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Wimberly, B.T. and Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, **407**, 340–348.

91 Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883–896.

92 Hohsaka, T., Ashizuka, Y., Murakami, H. and Sisido, M. (1996) Incorporation of non-natural amino acids into streptavidin through *in vitro* frame-shift suppression. *Journal of the American Chemical Society*, **118**, 9778–9779.

93 Hohsaka, T., Kajihara, D., Ashizuka, Y., Murakami, H. and Sisido, M. (1999) Efficient incorporation of non-natural amino acids with large aromatic groups into streptavidin in in-vitro protein synthesizing systems. *Journal of the American Chemical Society*, **121**, 34–40.

94 Hohsaka, T., Ashizuka, Y., Murakami, H. and Sisido, M. (2001) Five-base codons for incorporation of non-natural amino acids into proteins. *Nucleic Acids Research*, **29**, 3646–3651.

95 Hohsaka, T., Ashizuka, Y., Taira, H., Murakami, H. and Sisido, M. (2001) Incorporation of non-natural amino acids into proteins by using various four-base codons in an *Escherichia coli in vitro* translation system. *Biochemistry*, **40**, 11060–11064.

96 Sisido, M. and Hohsaka, T. (2001) Introduction of specialty functions by the position-specific incorporation of non-natural amino acids into proteins through four-base codon/anticodon pairs. *Applied Microbiology and Biotechnology*, **57**, 274–281.

97 Muranaka, N., Hohsaka, T. and Sisido, M. (2006) Four-base codon mediated mRNA display to construct peptide libraries that contain multiple non-natural amino acids. *Nucleic Acids Research*, **34**, e7.

98 Magliery, T.J., Anderson, J.C. and Schultz, P.G. (2001) Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in *Escherichia coli*. *Journal of Molecular Biology*, **307**, 755–769.

99 Anderson, J.C., Magliery, T.J. and Schultz, P.G. (2002) Exploring the limits of codon

and anticodon size. *Chemistry & Biology*, **9**, 237–244.

100 Anderson, J.C. and Schultz, P.G. (2003) Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for fourbase, amber, and opal suppression. *Biochemistry*, **42**, 9598–9608.

101 Anderson, J.C., Wu, N., Santoro, S.W., Lakshman, V., King, D.S. and Schultz, P.G. (2004) An expanded genetic code with a functional quadruplet codon. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 7566–7571.

102 Murakami, H., Kourouklis, D. and Suga, H. (2003) Using a solid phase ribozyme aminoacylation system to reprogram the genetic code. *Chemistry & Biology*, **10**, 1077–1084.

103 Ohtsuki, T., Manabe, T. and Sisido, M. (2005) Multiple incorporation of non-natural amino acids into a single protein using tRNAs with non-standard structures. *FEBS Letters*, **579**, 6769–6774.

104 Switzer, C., Moroney, S.E. and Benner, S.A. (1989) Enzymatic incorporation of a new base pair into DNA and RNA. *Journal of the American Chemical Society*, **111**, 8322–8323.

105 Piccirilli, J.A., Krauch, T., Moroney, S.E. and Benner, S.A. (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature*, **343**, 33–37.

106 McMinn, D.L., Ogawa, A.K., Wu, Y., Liu, J., Schultz, P.G. and Romesberg, F.E. (1999) Efforts toward expansion of the genetic alphabet: DNA polymerase recognition of a highly stable, self-pairing hydrophobic base. *Journal of the American Chemical Society*, **121**, 11585–11586.

107 Ogawa, A.K., Wu, Y., McMinn, D.L., Liu, J., Schultz, P.G. and Romesberg, F.E. (2000) Efforts toward the expansion of the genetic alphabet: information storage and replication with unnatural hydrophobic base pairs. *Journal of the American Chemical Society*, **122**, 3274–3287.

108 Wu, Y., Ogawa, A.K., Berger, M., McMinn, D.L., Schultz, P.G. and Romesberg, F.E. (2000) Efforts toward expansion of the genetic alphabet: optimization of interbase hydrophobic interactions. *Journal of the American Chemical Society*, **122**, 7621–7632.

109 Leconte, A.M., Chen, L. and Romesberg, F.E. (2005) Polymerase evolution: efforts toward expansion of the genetic code. *Journal of the American Chemical Society*, **127**, 12470–12471.

110 Ohtsuki, T., Kimoto, M., Ishikawa, M., Mitsui, T., Hirao, I. and Yokoyama, S. (2001) Unnatural base pairs for specific transcription. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4922–4925.

111 Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., Nakayama, H., Takio, K., Yabuki, T., Kigawa, T., Kodama, K., Yokogawa, T., Nishikawa, K. and Yokoyama, S. (2002) An unnatural base pair for incorporating amino acid analogs into proteins. *Nature Biotechnology*, **20**, 177–182.

112 Mitsui, T., Kimoto, M., Harada, Y., Yokoyama, S. and Hirao, I. (2005) An efficient unnatural base pair for a base-pair-expanded transcription system. *Journal of the American Chemical Society*, **127**, 8652–8658.

113 Hirao, I. (2006) Placing extra components into RNA by specific transcription using unnatural base pair systems. *BioTechniques*, **40**, 711–717.

114 Hirao, I., Kimoto, M., Mitsui, T., Fujiwara, T., Kawai, R., Sato,1 A., Harada, Y. and Yokoyama, S. (2006) An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nature Methods*, **3**, 729–735.

115 Berman, A.L., Kolker, E. and Trifonov, E.N. (1994) Underlying order in protein sequence organization. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 4044–4047.

116 Gerstein, M. (1998) How representative are the known structures of the proteins in a complete genome? A comprehensive

structural census. *Folding & Design*, **3**, 497–512.

117 Nishiuchi, Y., Inui, T., Nishio, H., Bódi, J., Kimura, T., Tsuji, F.I. and Sakakibara, S. (1998) Chemical synthesis of the precursor molecule of the *Aequorea* green fluorescent protein, subsequent folding, and development of fluorescence. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 13549–13554.

118 Gutte, B. and Merrifield, R.B. (1917) The synthesis of ribonuclease A. *The Journal of Biological Chemistry*, **246**, 1922–1941.

119 Wieland, T., Bokelmann, E., Bauer, L., Lang, H.U. and Lau, H. (1953) Ueber peptidsynthesen. 8. Mitteilung: bildung von S-haltigen peptiden durch intramolekulare wanderung von aminoacylresten. *Justus Liebigs Annalen der Chemie*, **583**, 129–149.

120 Kemp, D.S., Leung, S.L. and Kerkman, D.J. (1981) Models that demonstrate peptide bond formation by prior thiol capture. I. Capture by disulfide formation. *Tetrahedron Letters*, **22**, 181–184.

121 Kemp, D.S. and Kerkman, D.J. (1981) Models that demonstrate peptide bond formation by prior thiol capture. II. Capture by organomercury derivatives. *Tetrahedron Letters*, **22**, 185–186.

122 Kemp, D.S., Galakatos, N.G., Bowen, B. and Tan, K. (1986) Peptide synthesis by prior thiol capture. 2. Design of templates for intramolecular *O, N*-acyl transfer. 4,6-Disubstituted dibenzofurans as optimal spacing elements. *The Journal of Organic Chemistry*, **51**, 1829–1838.

123 Kemp, D.S., Galakatos, N.G., Dranginis, S., Ashton, C., Fotouhi, N. and Curran, T.P. (1986) Peptide synthesis by prior thiol capture. 4. Amide bond formation: the effect of side-chain substituent on the rates of intramolecular *O, N*-acyl transfer. *The Journal of Organic Chemistry*, **51**, 3320–3324.

124 Schnölzer, M. and Kent, S.B.H. (1992) Constructing proteins by dovetailing unprotected synthetic peptides: backbone engineered HIV protease. *Science*, **256**, 221–225.

125 Baca, M., Muir, T.W., Schnolzer, M. and Kent, S.B.H. (1995) Chemical ligation of cysteine-containing peptides: synthesis of a 22 kDa tethered dimer of HIV-1 protease. *Journal of the American Chemical Society*, **117**, 1881–1887.

126 Dawson, P.E., Muir, T.W., Clark-Lewis, I. and Kent, S.B.H. (1994) Synthesis of proteins by native chemical ligation. *Science*, **266**, 776–779.

127 Vogel, E.M. and Imperiali, B. (2007) Semisynthesis of unnatural amino acid mutants of paxillin: protein probes for cell migration studies. *Protein Science* **16**, 550–556.

128 Dawson, P.E., Churchill, M.J., Ghadiri, M.R. and Kent, S.B.H. (1997) Modulation of reactivity in native chemical ligation through the use of thiol additives. *Journal of the American Chemical Society*, **119**, 4325–4329.

129 Hackeng, T.M., Griffin, J.H. and Dawson, P.E. (1999) Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 10068–10073.

130 Hirel, P.H., Schmitter, J.M., Dessen, P., Fayat, G. and Blanquet, S. (1989) Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 8247–8251.

131 Erlanson, D.A., Chytil, M. and Verdine, G.L. (1996) The leucine zipper domain controls the orientation of AP-1 in the NFAT–AP-1–DNA complex. *Chemistry & Biology*, **3**, 981–991.

132 Tolbert, T.J. and Wong, C.H. (2002) New methods for proteomic research: preparation of proteins with N-terminal cysteines for labeling and conjugation. *Angewandte Chemie International Edition*, **41**, 2171–2174.

133 Becker, C.F., Hunter, C.L., Seidel, R., Kent, S.B., Goody, R.S. and Engelhard, M. (2003) Total chemical synthesis of a functional interacting protein pair: the protooncogene H-Ras and the Ras binding domain of its effector c-Raf1. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5075–5080.

134 Kochendoerfer, G.G., Chen, S.Y., Mao, F., Cressman, S., Traviglia, S., Shao, H., Hunter, C.L., Low, D.W., Cagle, E.N., Carnevali, M., Gueriguian, V., Keogh, P.J., Porter, H., Stratton, S.M., Wiedeke, M.C., Wilken, J., Tang, J., Levy, J.J., Miranda, L.P., Crnogorac, M.M., Kalbag, S., Botti, P., Schindler-Horvat, J., Savatski, L., Adamson, J.W., Kung, A., Kent, S.B. and Bradburne, J.A. (2003) Design and chemical synthesis of a homogeneous polymer modified erythropoiesis protein. *Science*, **299**, 884–887.

135 Yan, L.Z. and Dawson, P.E. (2001) Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *Journal of the American Chemical Society*, **123**, 526–533.

136 Perler, F.B., Davis, E.O., Dean, G.E., Gimble, F.S., Jack, W.E., Neff, N., Noren, C.J., Thorner, J. and Belfort, M. (1994) Protein splicing elements: inteins and exteins – a definition of terms and recommended nomenclature. *Nucleic Acids Research*, **22**, 1125–1127.

137 Hirata, R., Ohsumk, Y., Nakano, A., Kawasaki, H., Suzuki, K. and Anraku, Y. (1990) Molecular structure of a gene, VMA1, encoding the catalytic subunit of H^+-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry*, **265**, 6726–6733.

138 Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M. and Stevens, T.H. (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H^+-adenosine triphosphatase. *Science*, **250**, 651–657.

139 Perler, F.B. (2002) InBase: the Intein Database. *Nucleic Acids Research*, **30**, 383–384.

140 Perler, F.B. (2005) Protein splicing mechanisms and applications. *IUBMB Life*, **57**, 469–476.

141 Anraku, Y., Mizutani, R. and Satow, Y. (2005) Protein splicing: its discovery and structural insight into novel chemical mechanisms. *IUBMB Life*, **57**, 563–574.

142 Xu, M.Q. and Perler, F.B. (1996) The mechanism of protein splicing and its modulation by mutation. *The EMBO Journal*, **15**, 5146–5153.

143 Kawasaki, M., Nogami, S., Satow, Y., Ohya, Y. and Anraku, Y. (1997) Identification of three core regions essential for protein splicing of the yeast Vma1 protozyme: a random mutagenesis study of the entire VMA1-derived endonuclease sequence. *The Journal of Biological Chemistry*, **272**, 15668–15674.

144 Southworth, M.W., Benner, J. and Perler, F.B. (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO Journal*, **19**, 5019–5026.

145 David, R., Richter, M.P.O. and Beck-Sickinger, A.G. (2004) Expressed protein ligation: method and applications. *European Journal of Biochemistry*, **271**, 663–677.

146 Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A., Pelletier, J.J., Paulus, H. and Xu, M.-Q. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, **192**, 271–281.

147 Muir, T.W. (2003) Semisynthesis of proteins by expressed protein ligation. *Annual Review of Biochemistry*, **72**, 249–289.

148 Dawson, P.E., Churchill, M., Ghadiri, M.R. and Kent, S.B.H. (1997) Modulation of reactivity in native chemical ligation through the use of thiol additives. *Journal*

of the American Chemical Society, **119**, 4325–4329.

149 Muir, T.W., Sondhi, D. and Cole, P.A. (1998) Expressed protein ligation: a general method for protein engineering. *Proceedings of the National Academy of Sciences of the United States of America*, **1998**, **95**, 6705–6710.

150 Evans, T.C. Jr, Benner, J. and Xu, M.-Q. (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Science* **7**, 2256–2264.

151 Evans, T.C. Jr and Xu, M.-Q. (2002) Mechanistic and kinetic considerations of protein splicing. *Chemical Reviews*, **102**, 4869–4883.

152 Mathys, S., Evans, T.C., Chute, I.C., Wu, H., Chong, S., Benner, J., Liu, X.Q. and Xu, M.Q. (1999) Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene*, **231**, 1–13.

153 Southworth, M.W., Amaya, K., Evans, T.C. Jr, Xu, M.-Q. and Perler, F.B. (1999) Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein. *Biotechniques*, **1999**, **27**, 110–120.

154 Mukhopadhyay, J., Kapanidis, A.N., Mekler, V., Kortkhonjia, E., Ebright, Y.W. and Ebright, R.H. (2001) Translocation of sigma70 with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell*, **106**, 453–463.

155 Muir, T.W., Dawson, P.E. and Kent, S.B.H. (1997) Protein synthesis by chemical ligation of unprotected peptides in aqueous solution. *Methods in Enzymology*, **289**, 266–298.

156 Severinov, K. and Muir, T.W. (1998) Expressed protein ligation, a novel method for studying protein–protein interactions in transcription. *The Journal of Biological Chemistry*, **273**, 16205–16209.

157 Alexandrov, K., Heinemann, I., Durek, T., Sidorovitch, V., Goody, R.S. and Waldmann, H. (2002) Intein-mediated synthesis of geranylgeranylated Rab7 protein *in vitro*. *Journal of the American Chemical Society*, **124**, 5648–5649.

158 Valiyaveetil, F.I., MacKinnon, R. and Muir, T.W. (2002) Semisynthesis and folding of the potassium channel KcsA. *Journal of the American Chemical Society*, **124**, 9113–9120.

159 Xu, R., Ayers, B., Cowburn, D. and Muir, T.W. (1999) Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 388–393.

160 Camarero, J.A., Shekhman, A., Campbell, E., Chlenov, M., Gruber, T.M., Bryant, D.A., Darst, S.A., Cowburn, D. and Muir, T.W. (2002) Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 8536–8541.

161 Ayers, B., Blaschke, U.K., Camarero, J.A., Cotton, G.J., Holford, M. and Muir, T.W. (2000) Introduction of unnatural amino acids into proteins using expressed protein ligation. *Biopolymers*, **51**, 343–354.

162 Sydor, J.R., Mariano, M., Sideris, S. and Nock, S. (2002) Establishment of intein-mediated protein ligation under denaturing conditions: C-terminal labeling of a single-chain antibody for biochip screening. *Bioconjugate Chemistry*, **13**, 707–712.

163 Dawson, P.E. and Kent, S.B.H. (2000) Synthesis of native proteins by chemical ligation. *Annual Review of Biochemistry*, **69**, 923–960.

164 Cotton, G.J., Ayers, B., Xu, R. and Muir, T.W. (1999) Insertion of a synthetic peptide into a recombinant protein framework: a protein biosensor. *Journal of the American Chemical Society*, **121**, 1100–1101.

165 Canne, L.E., Botti, P., Simon, R.J., Chen, Y., Dennis, E.A. and Kent, S.B.H. (1999) Chemical protein synthesis by solid phase

ligation of unprotected peptide segments. *Journal of the American Chemical Society*, **121**, 8720–8727.

166 Brik, A., Keinan, E. and Dawson, P.E. (2000) Protein synthesis by solid-phase chemical ligation using a safety catch linker. *The Journal of Organic Chemistry*, **65**, 3829–3835.

167 Hofmann, R.M., Cotton, G.J., Chang, E.J., Vidal, E., Veach, D., Bornmann, W. and Muir, T.W. (2001) Fluorescent monitoring of kinase activity in real time: development of a robust fluorescence-based assay for Abl tyrosine kinase activity. *Bioorganic & Medicinal Chemistry Letters*, **11**, 3091–3094.

168 Lu, W., Gong, D., Bar-Sagi, D. and Cole, P.A. (2001) Site-specific incorporation of a phosphotyrosine mimetic reveals a role for tyrosine phosphorylation of SHP-2 in cell signaling. *Molecular Cell*, **8**, 759–769.

169 Robinson, J.P., Picklesimer, J.B. and Puett, D. (1975) Tetanus toxin. The effect of chemical modifications on toxicity, immunogenicity, and conformation. *The Journal of Biological Chemistry*, **250**, 7435–7442.

170 Vandertol-Vanier, H.A., Vazquez-Duhalt, R., Tinoco, R. and Pickard, M.A. (2002) Enhanced activity by poly(ethylene glycol) modification of *Coriolopsis gallica* laccase. *Journal of Industrial Microbiology & Biotechnology*, **29**, 214–220.

171 DeSantis, G. and Jones, J.B. (1999) Chemical modification of enzymes for enhanced functionality. *Current Opinion in Biotechnology*, **10**, 324–330.

172 Tann, C.-M., Qi, D. and Distefano, M.D. (2001) Enzyme design by chemical modification of protein scaffolds. *Current Opinion in Chemical Biology*, **5**, 696–704.

173 Qi, D., Tann, C.-M., Haring, D. and Distefano, M.D. (2001) Generation of new enzymes via covalent modification of existing proteins. *Chemical Reviews*, **101**, 3081–3111.

174 Lundblad, R.L. (2005) *Chemical Reagents for Protein Modification*, 3rd edn, CRC Press, Boca Raton, FL.

175 Hoare, D.G. and Koshland, D.E. Jr (1966) A procedure for the selective modification of carboxyl groups in proteins. *Journal of the American Chemical Society*, **88**, 2057–2058.

176 Sigg, R.H., Luisi, P.L. and Aboderin, A.A. (1977) Alkaline structural transition of 4-nitrobenz-2-oxa-1,3-diazolyl-lysozyme. Kinetic and spectroscopic investigations. *The Journal of Biological Chemistry*, **252**, 2507–2514.

177 Herriott, R.M. (1947) Reactions of native proteins with chemical reagents. *Advances in Protein Chemistry*, **3**, 169–173.

178 Riordan, J.F., Sokolovsky, M. and Vallee, B.L. (1966) Tetranitromethane. A reagent for the nitration of tyrosine and tyrosyl residues in proteins. *Journal of the American Chemical Society*, **88**, 4104–4105.

179 Fujita, Y. and Noda, Y. (1992) Effect of alkylation with different sized substituents on thermal stability of lysozyme. *International Journal of Peptide and Protein Research*, **40**, 103–109.

180 Cervenansky, C., Engstrom, A. and Karlsson, E. (1994) Study of structure-activity relationship of fasciculin by acetylation of amino groups. *Biochimica et Biophysica Acta*, **1199**, 1–5.

181 Chang, J.Y. (1989) Binding of heparin to human antithrombin III activates selective chemical modification at lysine 236. *The Journal of Biological Chemistry*, **264**, 3111–3115.

182 Yem, A.W., Zurcher-Neely, H.A., Richard, K.A., Staite, N.D., Heinrikson, R.L. and Deibel, M.R. Jr (1989) Biotinylation of reactive amino groups in native recombinant human interleukin-1. *The Journal of Biological Chemistry*, **264**, 17691–17697.

183 Wilchek, M. and Bayer, E. (1990) Biotin-containing reagents. *Methods in Enzymology*, **184**, 123–138.

184 Bayer, E. and Wilchek, M. (1990) Protein biotinylation. *Methods in Enzymology*, **184**, 138–160.

185 Salvucci, M.E. (1993) Covalent modification of a highly reactive and

essential lysine residue of ribulose-l,5-bisphosphate carboxylase/oxygenase activase. *Plant Physiology*, **103**, 501–508.

186 Zanetti, G. (1976) A lysyl residue at the NADP binding site of ferredoxin-NADP reductase. *Biochimica et Biophysica Acta*, **445**, 14–24.

187 Wilson, G. (1978) Effect of reductive lactosamination on the hepatic uptake of bovine pancreatic ribonuclease A dimer. *The Journal of Biological Chemistry*, **253**, 2070–2072.

188 Bunn, H.F. and Higgins, P.J. (1981) Reaction of monosaccharides with proteins: possible evolutionary significance. *Science*, **213**, 222–224.

189 McFarland, J.M. and Francis, M.B. (2005) Reductive alkylation of proteins using iridium catalyzed transfer hydrogenation. *Journal of the American Chemical Society*, **127**, 13490–13491.

190 Stauffer, D.A. and Karin, A. (1994) Electrostatic potential of the acetylcholine binding sites in the nicotinic receptor probed by reactions of binding-site cysteines with charged methanethiosulfonates. *Biochemistry*, **33**, 6840–6849.

191 Boyer, P.D. (1954) Spectrophotometric study of the reaction of protein sulfhydryl groups with organic mercurials. *Journal of the American Chemical Society*, **76**, 4331–4337.

192 Bai, Y. and Hayashi, R. (1979) Properties of the single sulfhydryl group of carboxypeptidase Y. *The Journal of Biological Chemistry*, **254**, 8473–8479.

193 Hempel, J.D. and Pietruszko, R. (1981) Selective chemical modification of human liver aldehyde and dehydrogenases E_1 and E_2 by iodoacetamide. *The Journal of Biological Chemistry*, **256**, 10889–10896.

194 Bednar, R.A. (1990) Reactivity and pH dependence of thiol conjugation to *N*-ethylmaleimide: detection of a conformational change in chalcone isomerase. *Biochemistry*, **29**, 3684–3690.

195 Hegazy, U.M., Hellman, U. and Mannervik, B. (2006) Replacement surgery with unnatural amino acids in the lock-and-key joint of glutathione transferase subunits. *Chemistry & Biology*, **13**, 929–936.

196 Tyagi, R. and Gupta, M.N. (1998) Chemical modification and chemical cross-linking for protein/enzyme stabilization. *Biochemistry*, **63**, 334–344.

197 Antos, J.M. and Francis, M.B. (2004) Selective tryptophan modification with rhodium carbenoids in aqueous solution. *Journal of the American Chemical Society*, **126**, 10256–10257.

198 Hudson, E.P., Eppler, R.K. and Clark, D.S. (2005) Biocatalysis in semi-aqueous and nearly anhydrous conditions. *Current Opinion in Biotechnology*, **16**, 637–643.

199 Milton, R.C., Milton, S.C.F. and Kent, S.B.H. (1992) Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show demonstration of reciprocal chiral substrate specificity [Corrected]. *Science*, **256**, 1445–1448.

200 Cheng, R.P., Gellman, S.H. and DeGrado, W.F. (2001) β-Peptides: from structure to function. *Chemical Reviews*, **101**, 3219–3232.

201 Arnold, U., Hinderaker, M.P., Nilsson, B.L., Huck, B.R., Gellman, S.H. and Raines, R.T. (2002) Protein prosthesis: a semisynthetic enzyme with a β-peptide reverse turn. *Journal of the American Chemical Society*, **124**, 8522–8523.

202 Serrano, L. (2000) The relationship between sequence and structure in elementary folding units. *Advances in Protein Chemistry*, **53**, 48–85.

203 Corey, D.R. and Schultz, P.G. (1987) Generation of a hybrid sequence-specific single-stranded deoxyribonuclease. *Science*, **238**, 1401–1403.

204 Corey, D.R., Pei, D. and Schultz, P.G. (1989) Generation of a catalytic sequence-specific hybrid DNase. *Biochemistry*, **28**, 8277–8286.

205 Zuckermann, R.N., Corey, D.R. and Schultz, P.G. (1988) Site-selective cleavage of RNA by a hybrid enzyme.

Journal of the American Chemical Society, **110**, 1614–1615.

206 Pei, D., Corey, D.R. and Schultz, P.G. (1990) Site-specific cleavage of duplex DNA by a semisynthetic nuclease via triple-helix formation. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 9858–9862.

207 Sutton, C.L., Mazumder, A., Chen, C.H. and Sigman, D.S. (1993) Transforming the *Escherichia coli* Trp repressor into a site-specific nuclease. *Biochemistry*, **32**, 4225–4230.

208 Goyne, T.E. and Sigman, D.S. (1987) Nuclease activity of 1,10-phenanthroline-copper ion. Chemistry of deoxyribose oxidation. *Journal of the American Chemical Society*, **109**, 2846–2848.

209 Pan, C.Q., Landgraf, R. and Sigman, D.S. (1994) DNA-binding proteins as site-specific nucleases. *Molecular Microbiology*, **12**, 335–342.

210 Josephy, P.D. and Mannervik, B. (2006) *Molecular Toxicology*, Oxford University Press, New York.

211 Lo Bello, M., Petruzzelli, R., De Stefano, E., Tenedini, C., Barra, D. and Federici, G. (1990) Identification of a highly reactive sulphydryl group in human placental glutathione transferase by a site-directed fluorescent reagent. *FEBS Letters*, **263**, 389–391.

212 Ricci, G., Lo Bello, M., Caccurri, A.M., Pastore, A., Nuccetelli, M., Parker, M.W. and Federici, G. (1995) Site-directed mutagenesis of human glutathione transferase P1-1. Mutation of Cys-47 induces a positive cooperatively in glutathione transferase P1-1. *The Journal of Biological Chemistry*, **270**, 1243–1248.

213 Hegazy, U.M., Tars, K., Hellman, U. and Mannervik, B. (2008) Modulating catalytic activity by unnatural amino acid residues in a GSH-binding loop of GST P1-1. *Journal of Molecular Biology*, **376**, 811–826.

214 Hegazy, U.M., Mannervik, B. and Stenberg, G. (2004) Functional role of the lock and key motif at the subunit interface of glutathione transferase p1-1. *The Journal of Biological Chemistry*, **279**, 9586–9596.

215 Stenberg, G., Abdalla, A.M. and Mannervik, B. (2000) Tyrosine 50 at the subunit interface of dimeric human glutathione transferase P1-1 is a structural key residue for modulating protein stability and catalytic function. *Biochemical and Biophysical Research Communications*, **271**, 59–63.

10
Catalysis by Peptide-Based Enzyme Models

Giovanna Ghirlanda, Leonard J. Prins, and Paolo Scrimin

10.1
Introduction

The preparation of artificial proteins capable of catalyzing (non-)biological reactions with an efficiency and specificity comparable to enzymes is one of the most intellectually challenging tasks in the field of peptide chemistry. Therefore, the development of small peptide-based catalysts that mimic enzymes continues to be a very popular theme in research. Chemists have adopted different strategies to reach this ambitious target. Generally, one can differentiate between "natural" systems and (almost) entirely synthetic systems, both of which have yielded interesting results. Natural systems are based on the naturally available set of amino acids, whereas in synthetic systems this set has been expanded with a wide variety of man-made building blocks. The use of artificial building blocks allows for an increase in variety of the catalysts and catalytic reactions that can be targeted, and allows also for a (partial) compensation of the loss of protein framework in stabilizing secondary structures.

De novo protein design presents a distinct approach to obtaining functional proteins [1–4]. In *de novo* design, investigators utilize information obtained from structural studies of natural proteins as well as data derived from computational modeling to guide the design of novel miniaturized proteins. These models are experimentally evaluated and the data obtained are utilized to improve the design; the cycle is repeated iteratively until a satisfactory model is achieved [5, 6]. The purpose of this procedure is (i) to gain accurate knowledge of the system by reproducing its spectroscopic and chemical properties, and (ii) once the key principles are elucidated, to design novel proteins with activities beyond those of natural proteins.

In this chapter we present, from our own perspective, examples of peptide-based catalysts (both natural and synthetic) focusing on two types of reaction: the hydrolysis of carboxylate and phosphate esters, and redox reactions mimicking heme proteins.

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

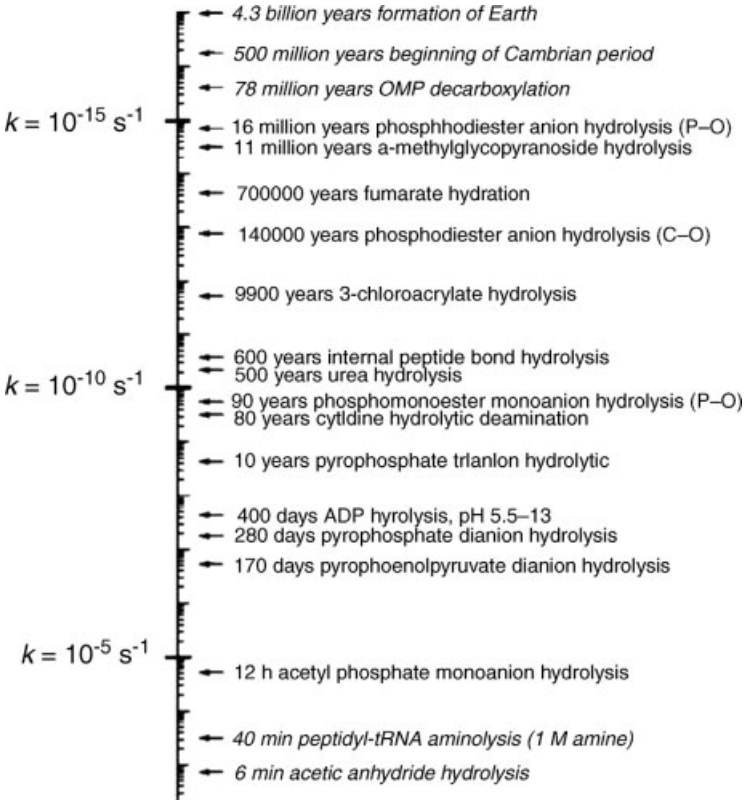


Figure 10.1 Rate constants and half-lives for selected hydrolysis and hydration reactions. (Reprinted with modification from [7].)

10.2
Peptide Models of Hydrolytic Enzymes

One of the most abundant classes of enzymes is that constituted by hydrolytic proteins. They are particularly proficient catalysts with very high k_{cat}/K_M and a rate-determining step that is close to the diffusion control limit. The reason is that the reactions they catalyze are, typically, very slow. Figure 10.1 shows some of these rates with the half-lives of the spontaneous reaction [7]. It is thus not surprising that the modeling of the activity of such enzymes has attracted considerable interest among scientists. Among these models, many are based on oligopeptide sequences and we will briefly review those that are, in our opinion, the most relevant in this regard.

10.2.1
Ester Hydrolysis and Acylation

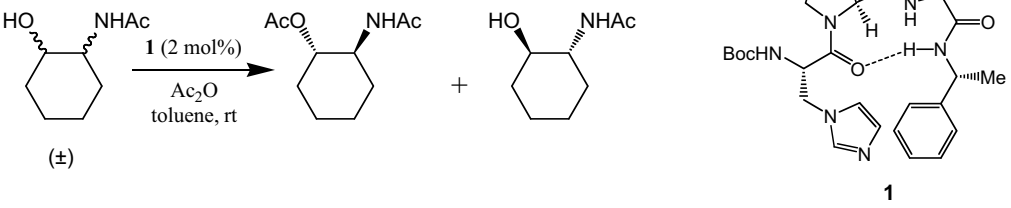
The development of artificial peptide-based catalysts for ester hydrolysis and acylation has attracted enormous attention. These reactions are very amenable to catalysis

and allow even relatively small catalytic activities to be determined accurately, very often by spectrophotometric means. Consequently, the catalysis of ester hydrolysis and acylation is very attractive for proof-of-concept studies. In the following subsections we will highlight peptide-based catalysts in which the activity results from the proper spatial organization of catalytic units, just like in enzymes. Strategies that will be discussed are the use of small oligopeptides with a persistent secondary structure, the spontaneous self-assembly of simple peptidic fragments into large aggregates, and the incorporation of peptide fragments in multivalent structures such as dendrimers and nanoparticles.

10.2.1.1 **Catalytically Active Peptide Foldamers**

Inspired by enzyme models, Miller *et al.* started by embedding nucleophiles in conformationally restricted peptidic structures in order to discover catalysts for asymmetric acyl-transfer reactions [8]. Histidine residues served as nucleophiles in a series of β-turn-type peptides for the kinetic resolution of *trans*-(±)-*N*-(2-hydroxycyclohexyl)acetamide. As an example, tripeptide **1** afforded a k_{rel} of 17 in favor of the (*S*,*S*) enantiomer (Scheme 10.1). The enantioselectivity results from the secondary structure, which allows the formation of preferential interactions of one enantiomer with the peptide backbone. Control peptides lacking the secondary structure gave no selectivity at all. Importantly, a strong correlation was observed between the rigidity of the β-hairpin structure and the enantioselectivity of the catalyst [9].

This hypothesis was confirmed by studies by Toniolo *et al.* of related catalysts in which the β-turn inducing α-amino isobutyric acid (Aib) residue was replaced by other C^{α}-tetrasubstituted amino acids [10]. Conformational analysis by ^{1}H-nuclear magnetic resonance (NMR) and Fourier transform infrared spectroscopy confirmed that the conformational constraint plays a key role in the enantioselectivity of the catalyst. It was concluded that the changes in the selectivity factors are not due to weaker interactions, but rather to the existence of a larger number of possible interaction modes that are not all selective for one of the two enantiomers. As suggested by conformational studies, the existence of such multiple interaction modes is clearly due to the absence of hydrogen bonds strong enough to stabilize the β-hairpin conformation (Figure 10.2), thus allowing the *N*-methylimidazole to reside above (Figure 10.2, left) or below (Figure 10.2, right) the catalyst.



Scheme 10.1 Kinetic resolution of *trans*-(±)1,2-acetamidocyclohexanol using stereoselective acylation.

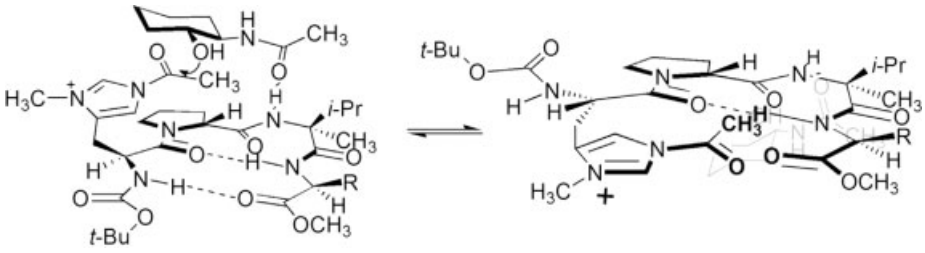
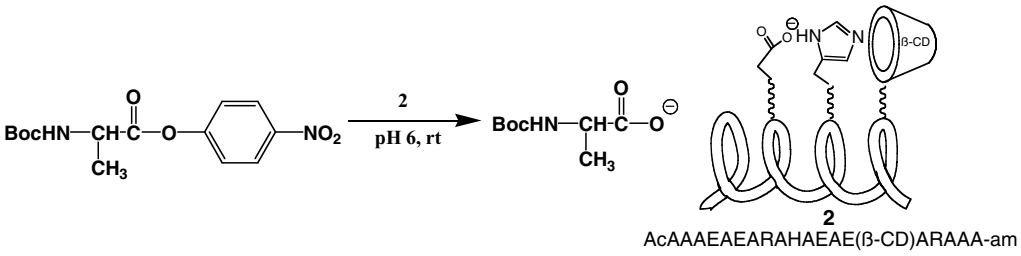


Figure 10.2 Conformational freedom allows the approach of the substrate from different faces of the histidine *N*-methylimidazole peptide catalyst.

Ueno *et al.* reported on a series of cyclodextrin (CD)–peptide hybrids in which three functional groups (β-CD, imidazole, and carboxylic acid) were present on the same face of an α-helical peptide [11]. In this design, the β-CD would bind substrates and place the ester group in close proximity to the imidazole and carboxylic acid functions, which would cooperatively cleave the ester moiety. The catalytic activity of peptide **2** was tested on both the D- and L-enantiomers of the *p*-nitrophenylester of Boc-alanine (Scheme 10.2). Kinetic parameters were determined by measuring the initial rate of hydrolysis as a function of substrate concentration. It was observed that hydrolysis of the D-enantiomer was favored over the L-enantiomer both in terms of k_{cat} (3.7×10^{-3} versus $1.7 \times 10^{-3}\ s^{-1}$) and K_M (3.33 versus 10.5 M). Interestingly, the kinetic parameters were significantly affected by the order in which the three functional groups were positioned on the α-helix. For instance, reversing the triad led to a 7-fold drop in the k_{cat}/K_M value for the hydrolysis of Boc-D-AlaONp. Furthermore, a reduced catalytic activity was observed in the absence of the carboxylic acid, leading to the postulation that the mechanism indeed involves a catalytic triad in which the carboxylic acid activates the imidazole unit via hydrogen bonding.

The group of Baltzer, exploiting the conformational preference of designed 42-amino-acid sequences for a helix–loop–helix (HLH) conformation [12], has systematically modified specific residues in key positions of the oligopeptide in order to catalyze the hydrolysis and transesterification reactions of *p*-nitrophenyl esters (Figure 10.3) [13–15]. Imidazole-functionalized peptides obtained by introducing several histidines in the sequence were able to provide substrate recognition



Scheme 10.2 Hydrolysis of the *p*-nitrophenylester of Boc-alanine by CD–peptide hybrid **2**.

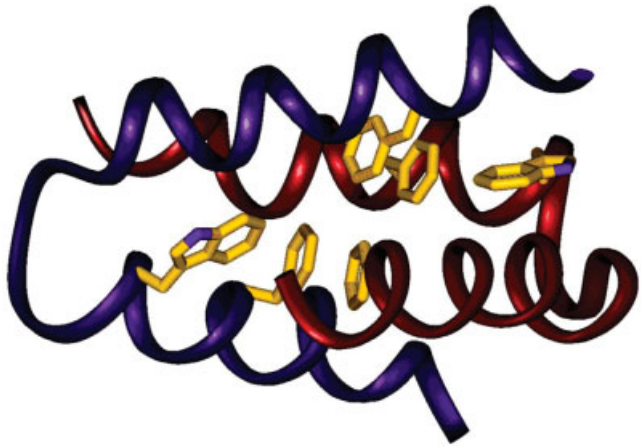


Figure 10.3 The HLH-forming 42mer peptide motif used by Baltzer for the design of catalysts for the cleavage of carboxylate esters (the dimeric form is shown here).

and accelerations exceeding three orders of magnitude compared to *N*-methylimidazole. For example, the sequence depicted in Figure 10.3 hydrolyzes 2,4-dinitrophenyl acetate (DNPA) with a second-order rate constant of $0.18\,M^{-1}\,s^{-1}$ at pH 3.1 compared with $9.9 \times 10^{-5}\,M^{-1}\,s^{-1}$ for *N*-methylimidazole. Interestingly, it has been demonstrated that the reaction mechanism takes advantage of the cooperativity of two adjacent histidines – one acting as the nucleophile and the other one as a general base, as shown in Figure 10.4. In the case of flanking His–Lys sequences the unprotonated form of the histidine attacks the ester in the rate-determining step of the process followed by subsequent transacylation of the lysine. If several lysine residues are present, at low pH, only those that flank the histidine are acylated. This leads to site-selective incorporation of an acyl residue in a natural sequence [16, 17]. Direct alkylation of lysines occurs in sequences devoid of histidines. In this case the nucleophilicity of the lysines could be controlled by site-selective pK_a depression [18]. As stated by the authors, the elucidation of the principles that control reactivity in such elementary processes is the first step toward the construction of biomimetic catalysts that accelerate reactions for which naturally occurring enzymes do not exist.

Hecht *et al*. have developed a combinatorial approach towards the *de novo* design of artificial enzymes [19]. Since most proteic structures obtained from a random combination of amino acids will not be structures, focused libraries were prepared based on binary patterns of polar (P) and apolar (A) residues in which the patterns encode secondary structure elements [20]. For instance, the binary pattern APAPAPA favors amphiphilic β-strands. Libraries of proteins are generated by expressing the corresponding synthetic genes in *Escherichia coli*. This way a 102-residue protein was obtained that folds into a four-helix bundle. Histidine units were present in the protein and their catalytic activity in the hydrolysis of *p*-nitophenyl acetate was evaluated. In comparison to 4-methylimidazole, a 100-fold rate acceleration was

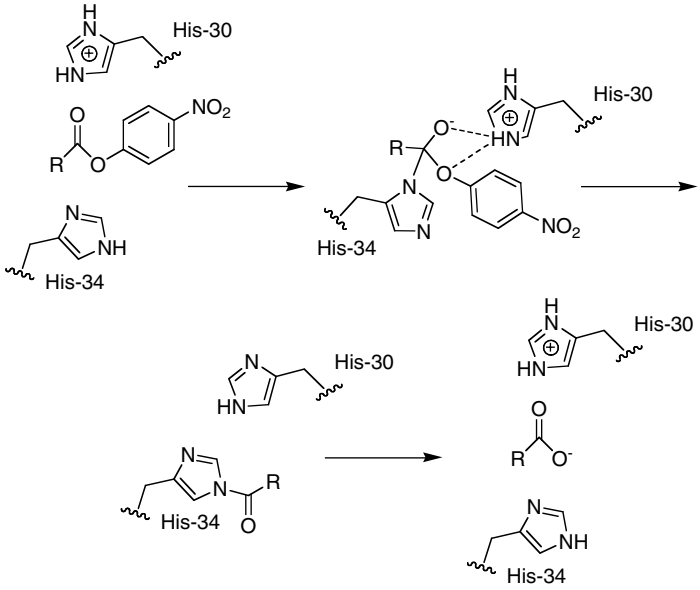


Figure 10.4 Cooperative mechanism involving two histidines in the cleavage of a carboxylate ester reported by Baltzer.

observed (k_2) and the bell-shaped curve of rate constant against pH indicated the involvement of two units in the catalytic cleavage.

Mayo *et al.* have reported on the development of enzyme-like proteins, called "protozymes," using a computational design procedure [21]. Their procedure takes as a starting point the catalytically inert 108-residue *E. coli* thioredoxin protein as scaffold. By computational means, an active-site scan was performed to identify favorable sites for the introduction of histidine units. In separate calculations, the high-energy state of the generated structures involved in histidine-mediated hydrolysis of *p*-nitrophenylacetate was calculated. Two sequences that emerged from the computational studies were selected for experimental analysis. In the best case a 25-fold increase in activity was observed (second-order rate constant k_{cat}/K_M) with respect to 4-methylimidazole. By itself, this is not an impressive number, but these studies clearly reveal the potential of using computational methods for the design of artificial catalysts.

10.2.1.2 Self-Organizing Catalytic Peptide Units

The spontaneous self-assembly of oligopeptides into a well-defined large structure is one of the most straightforward strategies to bring catalytic units in close proximity. For this purpose micellar and vesicular systems have been studied extensively since 1970. Surfactants functionalized with di- or tripeptides were used to rapidly generate catalytic sites on the surface of the aggregates. Frequently, these peptide sequences incorporated one histidine residue for the cleavage of esters (in cooperation with a second residue located on an adjacent surfactant) combined with additional amino acids for the creation of a chiral microenvironment. Key contributions in this area

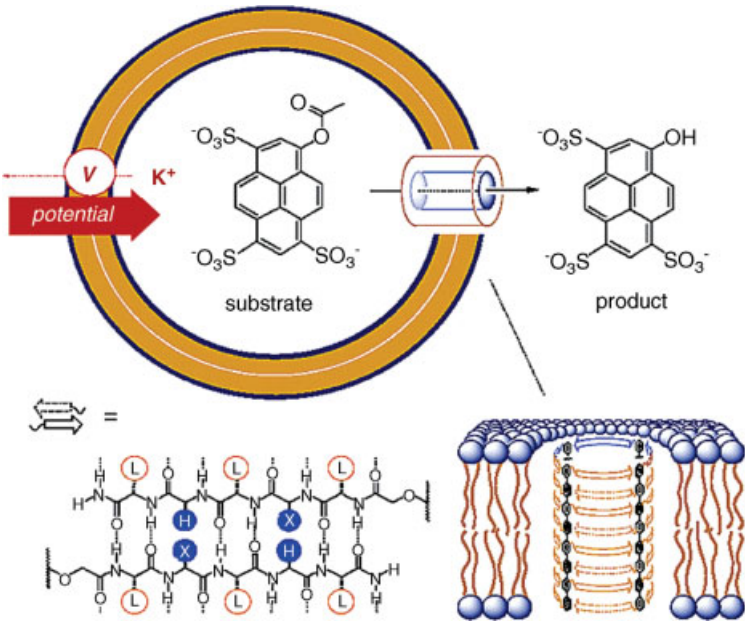


Figure 10.5 Synthetic catalytic pores. α-Amino acid residues pointing to the exterior of the barrel are depicted in black on white, whereas residues pointing to the interior are depicted in white on blue (L = leucine; H = histidine; X = arginine). (From [28].)

came from the groups of Moss [22], Ueoka [23], and Ohkubo [24], which have been extensively reviewed elsewhere recently [25, 26].

A very elegant example is provided by Matile *et al.*, who exploited the self-assembly of peptidic structures into synthetic pores that span lipid bilayers (Figure 10.5) [27]. Small peptide units are laterally attached to rigid *p*-octiphenyl backbones that end up forming the side-walls of the pores. The hydrophobicity of the octiphenyl backbone and some of the amino acid residues induces incorporation of the staves into a lipid bilayer, after which β-sheet formation between peptide fragments induces the spontaneous formation of the pore. The diameter of the so-called rigid-rod β-barrel pores could be varied from 2 to 25 Å depending on the peptide sequence. Synthetic catalytic pores are defined as pores constructed from abiotic scaffolds that catalyze substrate conversion *during* substrate translocation across the same pore [28]. It was shown that a rigid-rod β-barrel pore with internal Arg–His dyads catalyzed the esterolysis of 8-acetoxypyrene-1,3,6-trisulfonate present inside large unilamellar vesicles composed of egg-yolk phosphatidylcholine. Product formation after the addition of the synthetic pore to the vesicles was monitored continuously by following the increase in fluorescence emission of the product. It was observed that the initial velocity of product formation depended in a nonlinear manner on substrate concentration, which allowed determination of the Michaelis–Menten parameters k_{cat} and K_M.

Stupp *et al.* have extensively reported on the self-assembly of peptide amphiphiles into nanofibers driven by a hydrophobic collapse of the alkyl substituents in

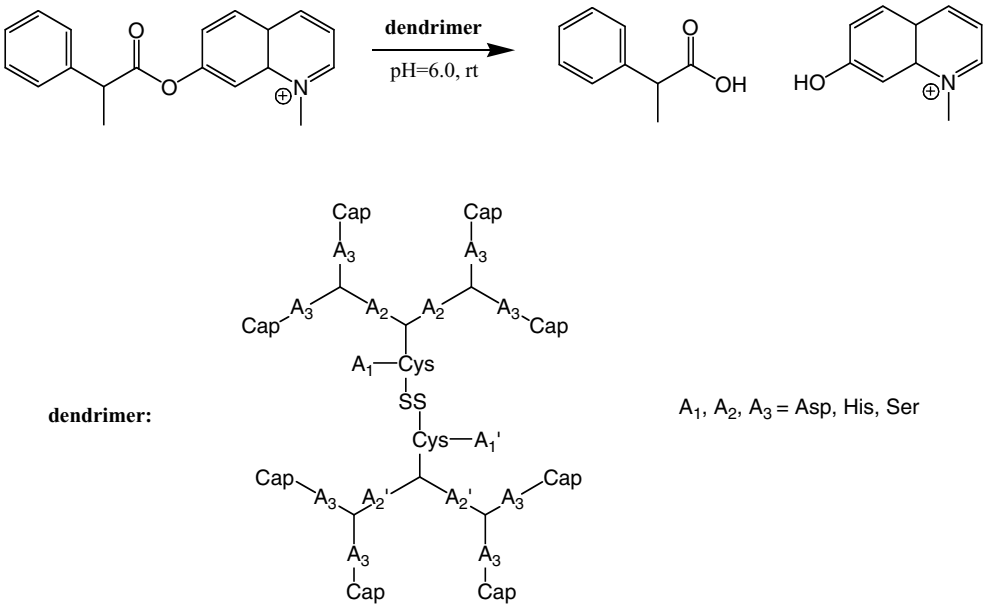
combination with β-sheet formation of the peptidic segment [29]. Recently, they showed that insertion of histidine residues in the peptidic segment resulted in the formation of nanofibers with a high imidazole density [30]. This resulted in a modest 8-fold rate acceleration in the hydrolysis of DNPA with respect to free imidazole. More interestingly, a higher catalytic activity was observed for peptide amphiphiles that formed nanofibers over those that aggregated into spherical aggregates in which the components are less ordered.

10.2.1.3 **Multivalent Catalysts**

Although the self-assembly of small components into large aggregates using non-covalent interactions is an attractive approach, it has the drawback that the obtained structures are not robust. This may pose restrictions on their application in catalysis, especially regarding the type of reaction medium used and the lowest concentration of catalyst accessible without causing disaggregation of the assembly. Here two alternative types of multivalent structures will be discussed: dendrimers and self-assembled monolayers (SAMs) on gold nanoparticles. Dendrimers are monodispersed, hyperbranched polymers that expose a high surface density of functional groups that depends on the generation. Whereas dendrimer synthesis requires high-yielding covalent synthesis, SAMs spontaneously form on gold nanoparticles via strong thiol–Au interactions. Thus, SAMs combine the advantage of a facile synthesis of the monomeric units and the self-assembly of a high-density surface of functional groups, with a robustness far superior to those of micelles, for instance. Both systems have been extensively explored for the development of artificial esterases.

The first catalytic peptide dendrimers were reported by Reymond *et al.* for the hydrolytic cleavage of esters [31–33]. These dendrimers were prepared by disulfide dimerization of second-generation dendritic peptides containing all possible combinations of the catalytic triad of the serine proteases (aspartic acid, histidine, and serine) resulting in 21 dimeric dendrimers (Scheme 10.3). Screening with a fluorogenic substrate showed that peptide dendrimers with triad amino acids having eight histidine groups at the surface were catalytically active and displayed enzyme-like Michaelis–Menten kinetics with substrate binding ($K_M \sim 0.1$ mM) and rate acceleration ($k_{cat}/k_{uncat} \sim 10^3$). From a systematic alanine scan it was concluded that catalysis most likely originated from the cooperative action of two imidazoles, whereas the serine residues did not appear to play a significant role [32].

Therefore, in a subsequent study a series of dendrimers of different generation were prepared in which histidine units were repeated in each generation (Figure 10.6) [34]. The catalytic activity was studied using a pyrene trisulfonate ester as a substrate. A systematic study of the dendritic effect in peptide dendrimer catalysis revealed that the catalytic rate constant k_{cat} and substrate-binding constant $1/K_M$ both increased with increasing generation number. The dendrimers showed rate accelerations up to $k_{cat}/k_{uncat} = 20\,000$ and $K_M \sim 0.1$ mM. The reactivity of histidine side-chains within the dendrimer is increased up to 4500-fold when compared to 4-methylimidazole. A bell-shaped pH rate profile around pH 5.5 in the dendrimer-catalyzed reactions suggests that catalytic activity results from cooperative action of two histidines (adding both a nucleophilic and a general base component to catalysis).



Scheme 10.3 Peptide-catalyzed ester hydrolysis.

Grafting small peptide fragments on gold nanoparticle surfaces is a straightforward method to induce cooperativity between the chemical functions as illustrated by the following examples by Scrimin *et al.* [35]. Thiol **3** (Figure 10.7) contains a His–Phe-OH terminal sequence and was used to passivate a gold nanoparticle in conjunction with a tris-ethyleneglycol methyl ether (TEG)-containing thiol, which makes the system water-soluble [36]. As a reference catalyst, Ac-His–Phe-OH was used, which contains the same functional groups, but is unable to aggregate. Positive proof of cooperativity was found by self-assembling the functionalities on gold nanoparticles. The complementary role of a carboxylate and an imidazolium ion was demonstrated by studying the hydrolysis at low pH.

Figure 10.7 reports the activity against pH of these functional nanoparticles and the monomeric catalyst in the hydrolysis of 2,4-dinitrophenylbutanoate (DNPB). At all pH values the gold nanoparticle catalyst out-competes the monomeric catalyst but, very interestingly, the two curves show strikingly different profiles. The monomeric catalyst behaves as a system in which a catalytically relevant nucleophile is generated with pK_a 6.6, which is consistent with the basicity of the imidazole. On the contrary, the nanoparticle shows a more complex profile: a first nucleophilic species is generated with pK_a 4.2, then the curve flattens up to pH 7 where a second nucleophile is generated with pK_a 8.1. These pK_a values can be assigned to the carboxylic acid and the imidazolium, respectively. The reason for the higher value of the pK_a of the imidazolium in the nanoparticle is due to the anionic nature of the nanoparticle that disfavors the deprotonation of the imidazolium cation. What is particularly remarkable is the high activity, at acidic pH, of the nanoparticle-based catalyst, showing over 300-fold rate acceleration with respect to the acetylated dipeptide.

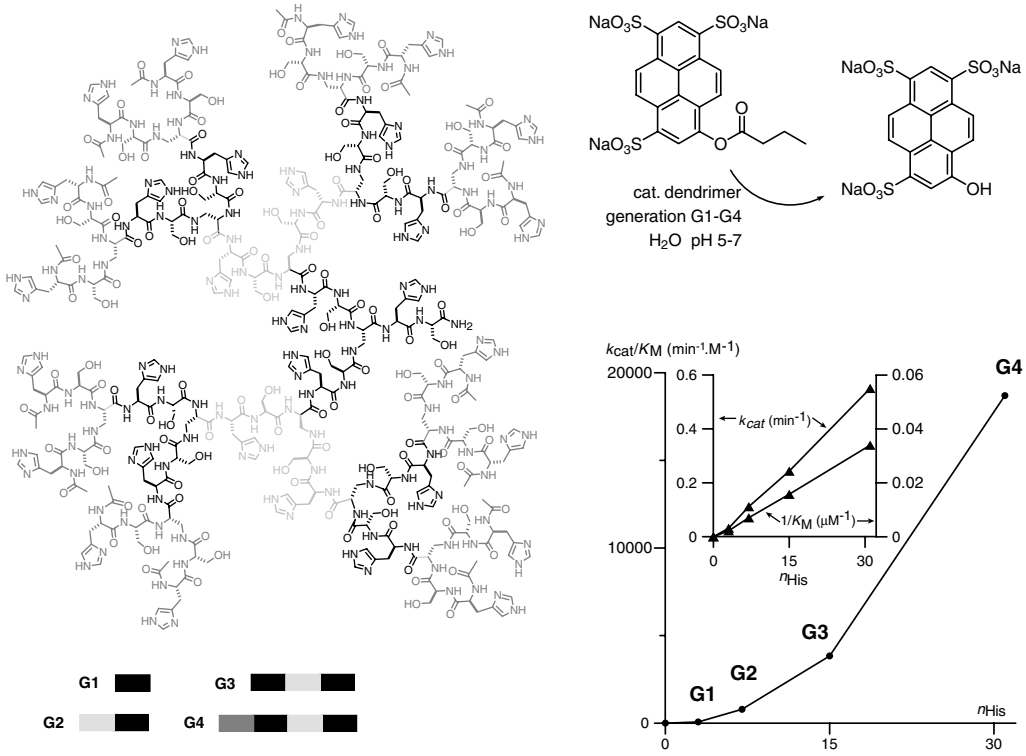


Figure 10.6 A strong positive effect in peptide dendrimer-catalyzed ester hydrolysis. Higher-generation dendrimers show a stronger substrate binding *and* stronger catalysis, resulting in a large enhancement of the specificity constant k_{cat}/K_M (graph at lower right). (From [33].)

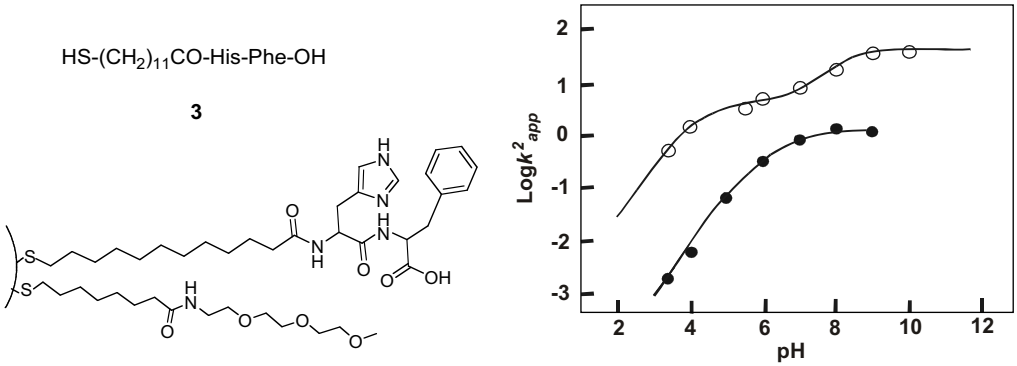


Figure 10.7 Dipeptide functionalized nanoparticle and dependence of the rate of cleavage of DNPB on pH for the nanoparticle-based catalysts (empty circles) and the acetylated monomeric catalyst (filled circles).

Mechanistically, this has been interpreted by involving a carboxylate anion in the cleavage that acts as a general base deprotonating a water molecule and a protonated imidazole acting as a general acid. The absence of this mechanism in the monomeric system clearly indicates that this behavior results from the confinement of the dipeptide on the monolayer covering the nanoparticle. Any influence of the TEG units appears to be highly unlikely.

Moving on towards real "nanozymes" (i.e., nanoparticle-based models of enzymes), a dodecapeptide was grafted on gold nanoparticles in a collaborative study with the group of Baltzer (see Figure 10.8) [37]. A combination of a histidine, two arginines, and a lysine residue was expected to enable nucleophilic, general acid, and/or general base catalysis, but also stabilization of the negatively charged transition state that arises along the pathway of ester hydrolysis. Gold nanoparticles

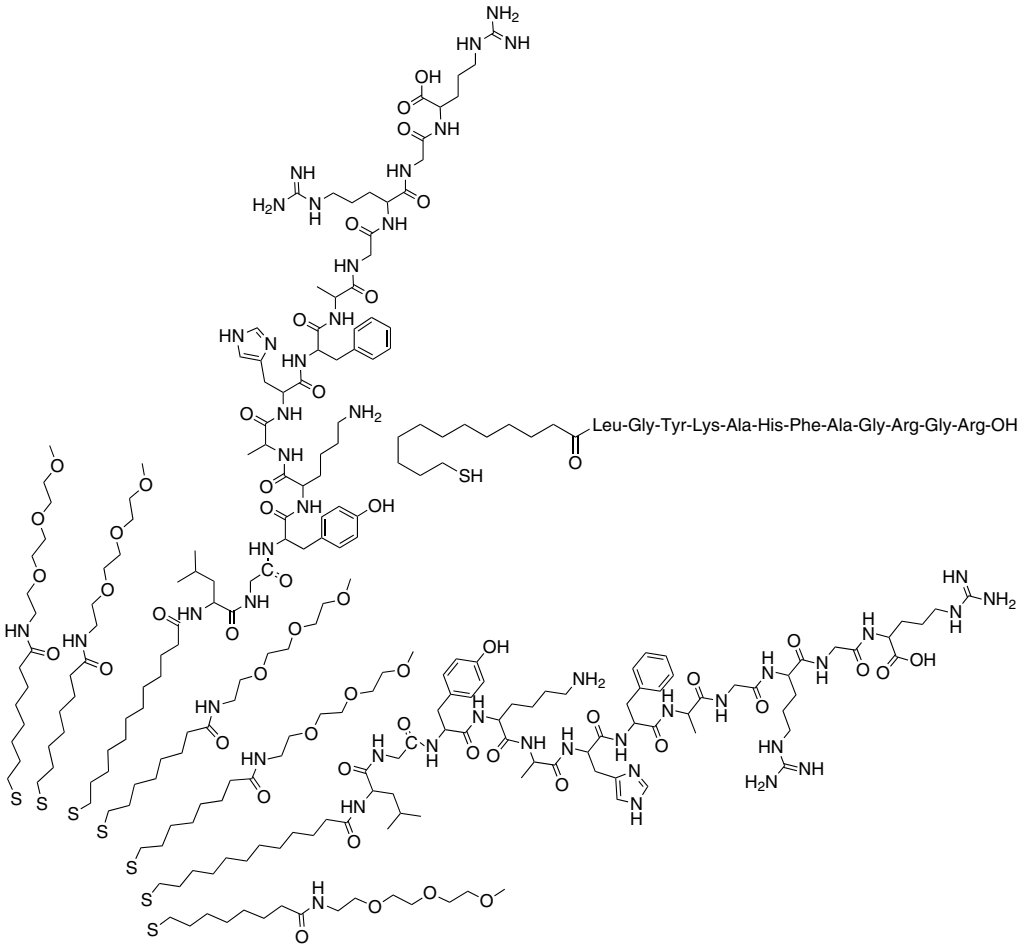


Figure 10.8 Dodecapeptide-functionalized nanoparticle and thiolated dodecapeptide used for the grafting on the gold cluster.

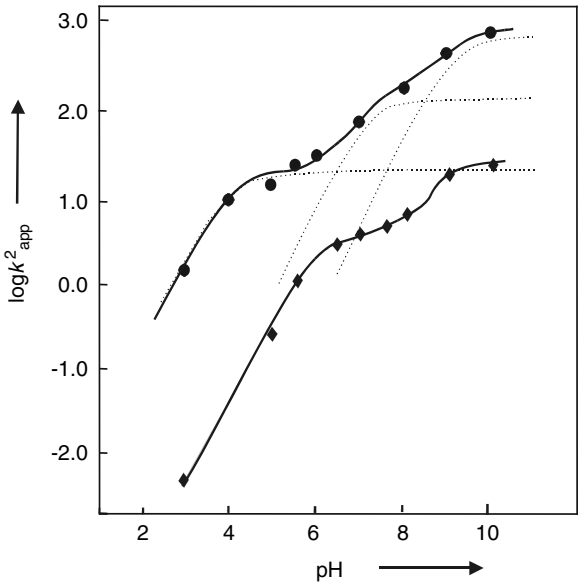


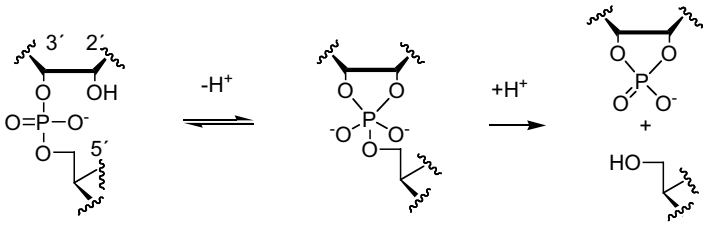
Figure 10.9 Dependence of the second-order rate constant on pH for the cleavage of DNPB for dodecapeptide-functionalized nanoparticles (filled circles) and the acetylated monomeric peptide (see Figure 10.7 for its structure).

were functionalized in a similar way as before and the catalytic activity of the resulting nanosystem in the hydrolysis of DNPB was studied as a function of pH. The results are reported in Figure 10.9 together with those related to the activity of the monomeric *S*-acetylated peptide. At low pH values the nanoparticle-based catalyst behaves very similarly to the previous dipeptide-based system, although the dodecapeptide-nanoparticle has an additional 10-fold gain in activity. This amounts to a 3000-fold rate acceleration over that exerted by the simple dipeptide. The larger catalytic efficiency is ascribed to a stronger acidity of the protonated imidazole group, which in the dodecapeptide has a pK_a value 0.9 units lower than in the dipeptide. This is due to the fact that now the nanoparticle no longer has a net negative charge as this is compensated by the presence of the cationic guanidinium groups.

At higher pH values, the activity of the peptide-nanoparticle increases significantly with respect to the dipeptide-nanoparticle reaching an additional 40-fold rate acceleration. This can be ascribed to the presence of an additional nucleophile (the phenoxide of tyrosine) with an apparent pK_a of 9.9.

10.2.2
Cleavage of the Phosphate Bond

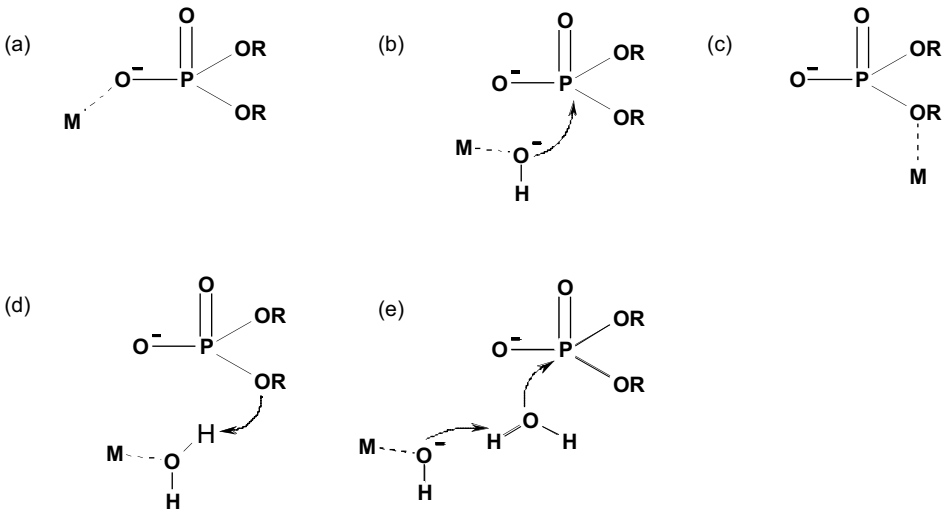
The phosphate bond of phosphoric acid diesters is particularly resistant to hydrolytic cleavage under physiological conditions (i.e., at pH close to 7) when these substrates are present as anions [38]. Apart from the intrinsic reactivity of the functional group



Scheme 10.4 Intramolecular cleavage of a phosphate bond in RNA.

the problem is compounded, under these conditions, by the electrostatic repulsion of any anionic nucleophile (e.g., OH^-), a situation that does not apply to the peptide bond. This sluggish reactivity is highlighted by the half-lives reported for the hydrolysis of the most common phosphate bonds present in the biological world – those of DNA and RNA [7]. In the first case the half-life is around 10^{10} years, while for the second one it is around 10^2 years. The reason why a phosphate of RNA is more reactive than that of DNA is that the bond cleavage is due to an intramolecular attack of the proximal OH of the ribose (Scheme 10.4). This situation is obviously not possible with DNA [39].

Many enzymes devoted to the hydrolysis of RNA or DNA present in their catalytic site a metal ion as cofactor which is essential in the case of hydrolytic DNA cleavage. The role played by the metal ions may be summarized as follows (Scheme 10.5) [38]. (i) The coordination of the $P(=O)-O^-$ group to the metal ion dissipates the negative charge, thus removing the electrostatic repulsion for the attack by OH^- (or any other anionic nucleophile), and provides Lewis acid-like catalysis. (ii) Coordination of a



Scheme 10.5 Possible roles played by metal ions (M) in the catalysis of the hydrolysis of a phosphate diester. For (*a*)–(*c*) see text; (*d*) and (*e*) represent examples of general acid or general base catalysis.

water molecule decreases its pK_a, thus providing an increased amount of (metal-bound) OH^- at physiological pH. (iii) Coordination of the leaving group facilitates its departure also by decreasing its pK_a. The seminal work by Chin *et al.* [38] has provided compelling evidence that, by summing up all these contributions, up to a 10^{18}-fold rate acceleration can be achieved, comparable to that observed with phosphatases. For this purpose at least two metal ions must be present in the catalytic site.

An important aspect of any synthetic catalyst is its ability to place the functional groups relevant to the catalytic process in the correct place. Thus, in the case of two metal ions, their relative distance is of critical importance as well as that of any nucleophile present in the catalytic site. Diffractometric studies indicate that, for dinuclear phosphatases, for instance, the distance between the two metal centers is in the range 3.5–5 Å [40]. This aspect of the design of a catalyst imposes on anyone embarking on such a demanding challenge, the selection of molecular architectures whose conformation is not flexible, but nevertheless can be easily controlled.

10.2.2.1 DNA and DNA Models as Substrates

One of the most attractive approaches for the design of an efficient catalyst is constituted by the modification of an existing protein [41]. Early examples have been reported by Schultz and Pei [42, 43]. By using a combination of both chemical and genetic approaches they were able to convert a relatively nonspecific phosphodiesterase, staphylococcal nuclease, into a molecule capable of sequence-specifically hydrolyzing RNA, single-stranded DNA, and duplex DNA.

Slightly more complex systems have been described by conjugating two different protein fragments: one able to bind to DNA, the other able to hydrolyze it [44–46]. Chimeric restriction enzymes were obtained by conjugating a zinc finger DNA-binding domain and the nonspecific DNA-cleavage domain (Fn) from the natural restriction enzyme *Fok*I. The zinc fingers are rather flexible in DNA recognition, thus by tuning their selectivity it was possible to cleave DNA in a sequence-specific manner. This "proof of principle" opens the way to generate "artificial" nucleases (zinc finger nucleases, see Figure 10.10) that may cut DNA near a predetermined site. In fact, Carroll *et al.* has been able to achieve site-specific DNA double-strand breaks in genomic DNA, which are then substrates for cellular repair mechanisms [47, 48].

Owing to the high activity of metal ion complexes in catalyzing the hydrolytic cleavage of phosphates (see Section 10.1), Franklin and her group [50, 51] have designed chimeric 33- and 34-residue peptides that comprise helix 2 and 3 of engrailed homeodomain [the helix-turn-helix (HTH) region], and the 12-residue calcium-binding loop of an EF-hand from calmodulin. This protein binds quite strongly to Ca^{2+} and this ion can be exchanged with a Ln(III), leading to equally strong metal complexes. Ln(III) ions are among the most powerful Lewis acid catalysts for phosphate cleavage [52]. Franklin's reasoning was that Ln(III)-binding peptides able to act as catalysts for the cleavage of bis-*p*-nitrophenylphosphate (BPNPP, a model for a DNA phosphate bond) would work with DNA as well. This turned out, in fact, to be the case and she has shown that HTH/EF-hand chimeras that bind lanthanides (specifically Eu^{3+}) have a metal-dependent solution structure, and cleave BPNPP and DNA in a sequence selective manner (see Figure 10.11) [53].

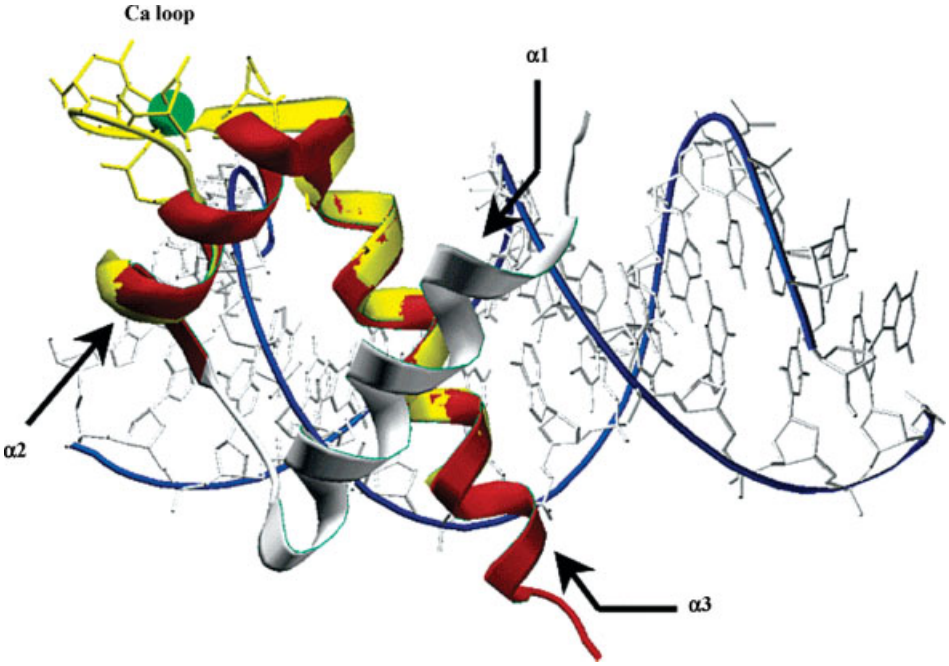


Figure 10.10 A three-finger zinc finger linked to the Fn domain through a flexible peptide linker. (From [49].)

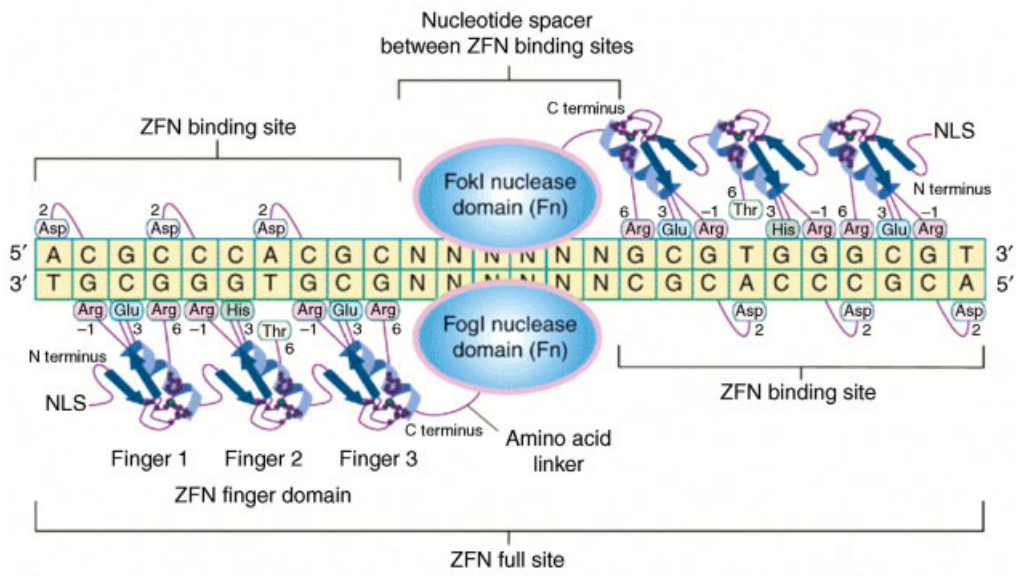


Figure 10.11 Computer-generated model of the interaction of chimeric peptides with DNA reported by Franklin. (From [53].)

All the above systems were based on fragments of native proteins or on peptide sequences based on a known natural one. Even more challenging is the design of totally synthetic sequences that are active as catalysts. Although different laboratories have shown that very short di- or tripeptides form metal complexes that are hydrolytically active in the cleavage of model substrates [54–56], we will focus our attention on more complex sequences where the activity is not just the result of the mere formation of a metal complex. Barton and her group [57] reported the activity of zinc-binding peptides tethered to rhodium intercalators. In these systems the intercalator provides DNA binding affinity while the metal-binding peptide contributes the reactivity.

The peptides described by Barton, although showing good activity, suffer limited conformational control. In the case of the helical sequence this amounted to no more than 30%. Certainly the possibility to control the conformation of the sequence in a more rigorous way could provide a way to better control the arrangement of the active groups in the putative catalytic site. With this in mind we have designed [58] a very simple sequence comprising two copies of a synthetic metal-binding amino acid, (*S*)-2-amino-3-[1-(1,4,7-triazacyclononane)]propanoic acid (ATANP) [59], and five copies of Aib. Aib is the prototype of the class of C^{α}-tetrasubstituted amino acids that are known to impart helical conformation even to very short sequences [60] because of the conformational constraint imposed to the main chain by the geminal substituents at the α-carbon. In the case of the heptapeptide of Figure 10.12 the conformation is that of a 3_{10}-helix. This conformation differs from the α-helix in the longer pitch (6.2 instead of 5.5 Å) and also the fact that each helical turn requires just three amino acids instead of 3.6. In this way the lateral arms of two amino acids placed in relative position i and $i + 3$ in the sequence face each other. This is exactly what happens to the aza-crowns of the two ATANP so that when they bind two metal ions (specifically, Zn^{2+}) these are placed in the right position for cooperating in the cleavage of a phosphate bond in a similar way to what happens in hydrolytic enzymes. This was the case for the dizinc complex of this peptide that turned out to be a very good catalyst for the hydrolysis of plasmid DNA with clear evidence of cooperativity between the two metal centers. The reactivity profile as a function of pH showed a maximum close to pH 7.3, supporting a mechanism in which one metal ion coordinates a phosphate anion while the other coordinates a water molecule. The conjugate base of the latter is the actual nucleophile. Thus, the increase of activity up to pH 7.3 accounts for the deprotonation of this water molecule, while the decrease of activity above this pH is related to the difficulty a phosphate has to compete with an hydroxide for coordination to the metal center. This is a common feature of many catalysts working by similar mechanisms.

A totally different approach to the discovery of a metal ion-based peptide catalyst was reported by Berkessel and Herault [61] by exploiting the screening of a combinatorial library rather than via rational design. The basic steps for the selection of the active systems are: (i) split-and-pool synthesis of the ligand library containing 625 solid-phase-bound undecapeptides, (ii) complexation of the ligand library with Lewis acidic transition metals (Cu^{2+}, Zn^{2+}, Fe^{3+}, Co^{3+}, Eu^{3+}, Ce^{4+}, and Zr^{4+}), and (iii) screening of the library with chromogenic test substrates. This latter aspect

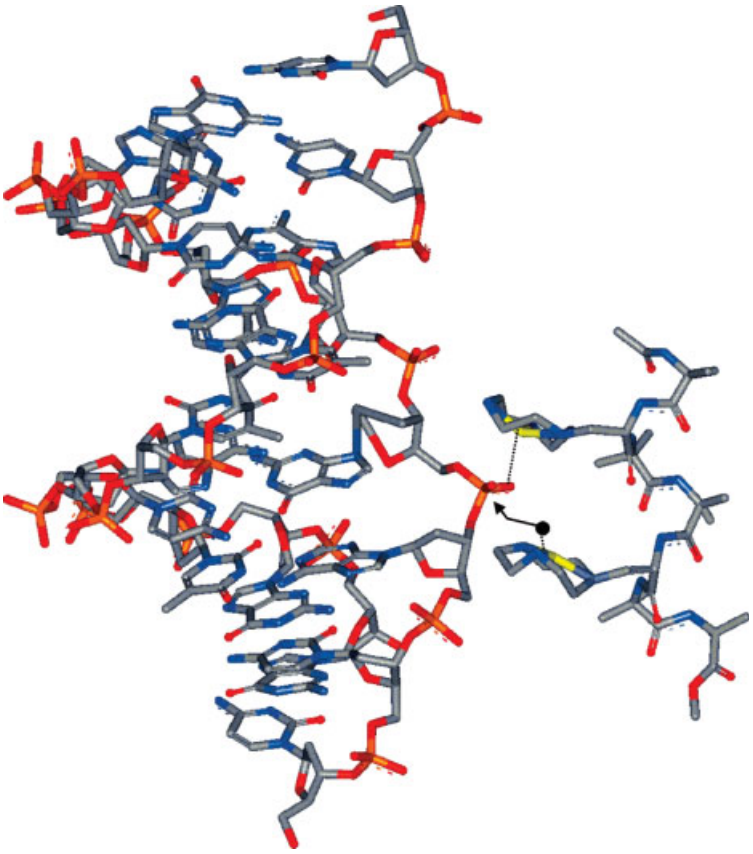


Figure 10.12 Aib-rich octapeptide functionalized with triazacyclonane-bearing amino acid ATANP is active in the cleavage of plasmid DNA.

was performed on beads – a challenging task that was accomplished by using a substrate that yields a product easily oxidized by air exposure with formation of a colored, insoluble dye that sticks to the pellet holding the active catalyst. By this procedure Zr^{4+} was found to be the best metal ion, while the following were pinpointed as the most active ligands (the amino acids marked in bold were subject to combinatorial variation):

H_2N-**Ser**-Gly-Gly-**His**-Gly-Gly-**Arg**-Gly-Gly-**His**-Phe-COOH
H_2N-**Ser**-Gly-Gly-**Ser**-Gly-Gly-**Ser**-Gly-Gly-**His**-Phe-COOH
H_2N-**Ser**-Gly-Gly-**Arg**-Gly-Gly-**His**-Gly-Gly-**His**-Phe-COOH

The best catalysts were eventually independently synthesized and their activity confirmed by homogeneous solution cleavage experiments.

Combinatorial screening of catalysts for the (metal-free) hydrolysis of phosphate monoesters was also reported by Schmuck and Dudaczek [62]. The design of the

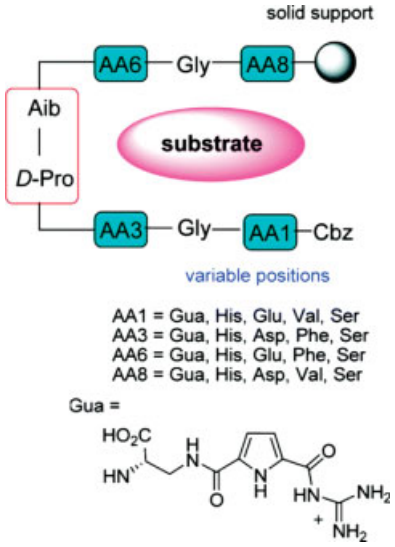


Figure 10.13 Combinatorial peptide screening reported by Schmuck and Dudaczek. (From [62].)

octapeptide library is shown in Figure 10.13. A D-Pro–Aib-turn element was incorporated at the central positions 4 and 5 of the octapeptide to hopefully help induce a more folded conformation in solution. Positions 1 and 3 as well as 6 and 8 were chosen as the variable positions with a glycine in between to reduce the steric bulk within the oligopeptide. For each of the four variable positions five different amino acids were used in the split-and-pool protocol giving rise to a total library size of 625 members. For these five amino acids different combinations of six proteinogenic amino acids (His, Ser, Glu, Asp, Phe, and Val) as well as an artificial arginine analog Gua [63] were used. This choice of amino acids was based on the idea that the polar amino acids (His, Glu, Asp, and Ser) could function as catalytically active residues or help in structuring the oligopeptide (e.g., via ion-pair formation between the arginine analog Gua and Glu or Asp). The two nonpolar amino acids, phenylalanine and valine, could provide a less polar microenvironment further favoring substrate binding or catalysis. The most active peptide found, after direct screening on resin has been:

$$\text{Cbz} - \text{His} - \text{Gly} - \text{Gua} - \text{D} - \text{Pro} - \text{Aib} - \text{Gua} - \text{Gly} - \text{Val} - \text{NHR}\ (\text{R} = \text{TentaGel resin})$$

In general, the most active sequences in this screen all contained the artificial arginine analog Gua in combination with either serine or histidine.

10.2.2.2 RNA and RNA Models as Substrates

RNA is as appealing (if not more) than DNA as a cleavage target [64]. It is substantially less prone to oxidative cleavage [65] as a consequence of the higher stability of the glycoside bond in ribonucleotides compared to that in deoxyribonucleotides. On the basis of the arguments presented in the introduction it is much less stable to hydrolytic

cleavage. For this reason the hydrolysis of the phosphate bond in this system can be successfully catalyzed not only by metal ions, but also by ammonium ions.

In this regard, Brack and Barbier [66] presented experimental evidence for the accelerated hydrolysis of polyribonucleotides by polycationic polypeptides containing basic and hydrophobic amino acids. Previously, strong ribonuclease (RNase) activity was reported for a 70-residue synthetic polypeptide analog of RNase S-protein [67] and for a 34-residue polypeptide [68]. Brack and Barbier have shown that the activity is not only due to the presence of lysine or arginine side-chains, but also to the spatial geometry of the backbone. The cleavage proceeds by the classical base-induced hydrolysis mechanism of RNAs. Thus, once the complex between the peptide and the oligoribonucleotide is formed through ionic interactions, the cleavage occurs via a nucleophilic attack of the free base of the side-chain, affording a cyclic phosphate. Subsequently the 2′–3′ cyclic phosphate is hydrolyzed to give both 2′- and 3′-phosphates.

One of the ways to secure interaction with the target substrate is to use intercalation. This was the strategy used by Stein *et al.* [69] who reported on the activity of a series of peptide–acridine conjugates. The catalytic tripeptides had been designed on the basis of three features of the proposed catalytic mechanism of RNase A: 2′-proton abstraction by His12, proton donation to the leaving 5′-oxygen by His119, and stabilization of the pentacoordinated phosphorus transition state by Lys41. A lysine side-chain served as the connector between the catalytic unit and the intercalating acridine. The three tripeptides studied (HGH, HPH and GGH) all had similar hydrolytic activity with a mechanism most likely involving only the abstraction of the 2′-proton and stabilization of the transition state. Furthermore, activity was observed with single-stranded RNA and ribosomal RNA as substrates but not with double-stranded RNA and transfer RNA.

As we have already seen zinc finger peptides are well-explored polypeptide motifs that have found many applications in synthetic systems mostly for their ability to bind metal ions and to interact with oligonucleotides. In this context the report by Lima and Crooke [70] of the hydrolytic cleavage by a zinc finger peptide *devoid* of any metal ion constitutes a remarkable example. The system they have studied, a 30-mino-acid sequence, is based on a catalytic mechanism very similar to the one discussed above and does not rely on metal ion catalysis. The peptide sequence corresponds to a single zinc finger of the human male-associated ZFY protein – a transcription factor belonging to the Cys_2His_2 family. The RNase-active structure was determined to be a homodimer of the peptide resulting from the formation of a single intermolecular disulfide between two of the four cysteines. In this case too, activity was observed only with single-stranded RNA with preference for the dinucleotide sequence 5′-pyr-A-3′. The exceptionally high activity of the system is manifested by the fact that initial rates of cleavage, V_o, were comparable to rates observed for human RNase and the k_{cat} was comparable to the rates observed for the group II intron ribozymes.

By using a peptide sequence derived from that reported in Section 10.2.2 we have reported, in collaboration with the group of Baltzer, a 42-mer peptide analogous to the one reported in Figure 10.4, with the difference that these new sequences incorpo-

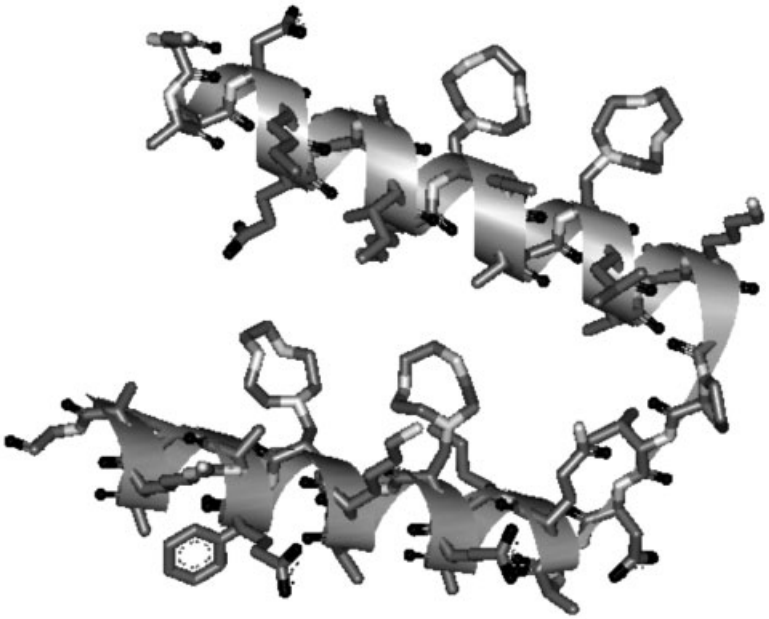


Figure 10.14 Hairpin-folded 42mer peptide incorporating four ATANP amino acids active in the cleavage of HPNP.

rated up to four copies of ATANP (Figure 10.14) [71]. Also, these new peptides form HLH motifs and bind Zn(II) ions with the triazacyclononane subunits present in the lateral arms of ATANP. It was observed that metal complexation causes a decrease in the helical content of the peptide. However, even upon a part unfolding of the structure, an acceleration of the cleavage of 2-hydroxypropyl-*p*-nitrophenyl phosphate (HPNP – an RNA model substrate) was observed. The mechanism calls for the cooperation of at least two metal ions due to their placement in close proximity due to the folded structure of the sequence.

A similar sequence was used by Baltzer's group to catalyze the cleavage of uridine 3′-2,2,2-trichloroethylphosphate [72]. The active site on the surface of the folded catalysts is in this case composed of two histidine and four arginine residues, with the capacity to provide general acid, general base, and/or nucleophilic catalysis as well as transition state stabilization. In this case too, metal ions are not involved in the catalytic process.

We have also developed [73] a rather complex RNase mimic system consisting of a polypeptide comprising again the ATANP ligand amino acid and C^{α}-tetrasubstituted amino acids which is subject to allosteric control [74]. This last point is a feature common to many proteins and constitutes an important element of regulation of enzymatic activity. Thus we have synthesized the heptapeptide H-Iva-Api-Iva-ATANP-Iva-Api-Iva-$NHCH_3$ (P1), where Iva = (*S*)-isovaline and Api = 4-amino-4-carboxypiperidine, and verified that its conformation in aqueous solution is essentially that of a 3_{10}-helix. By connecting three copies of P1 to a functionalized

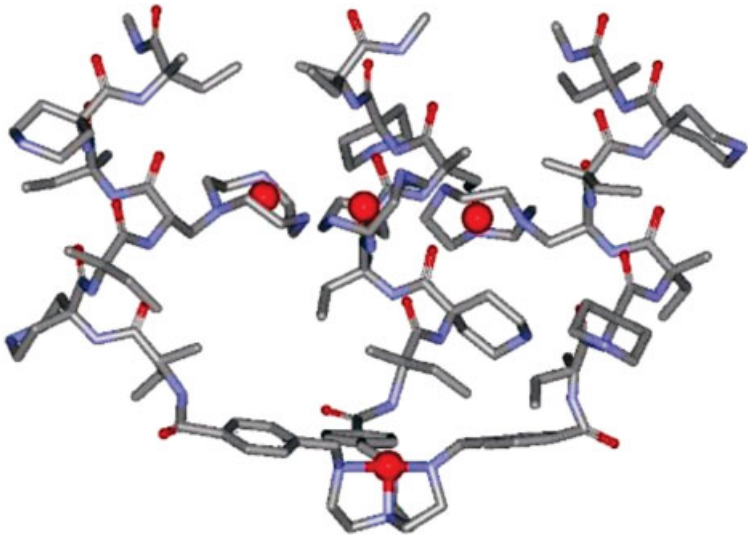


Figure 10.15 Tripodal catalysts incorporating three helical unnatural sequences comprising ATANP amino acids and active in the cleavage of HPNP.

tris(2-aminoethyl)amine (Tren) platform we obtained a new peptide template, $T(P1)_3$, that is able to bind up to four metal ions [Cu(II) or Zn(II)]: one in the Tren subsite and three in the triazacyclononane subunits. The binding of the metals to the Tren platform induces a change from an open to a closed conformation in which the three short, helical peptides are aligned in a parallel manner with the azacyclonane units pointing inward within the pseudocavity they define (Figure 10.15). $T(P1)_3$ shows a peculiar behavior in the transphosphorylation of phosphate esters; the tetrazinc complex is a catalyst of the cleavage of HPNP, whereas the free ligand is a catalyst of the cleavage of an oligomeric RNA sequence with selectivity for pyrimidine bases. In the case of HPNP, Zn(II) acts as a positive allosteric effector by enhancing the catalytic efficiency of the system. In the case of the polyanionic RNA substrate, Zn(II) switches off the activity, thus behaving as a negative allosteric regulator. We suggested that the opposite behavior of the catalyst induced by Zn(II) is associated with the change of conformation of the Tren platform, and consequently of the relative spatial disposition of the three linked peptides, that occurs after binding of the metal ion. In this system the secondary conformation of the three peptides defines the position of the functional groups and orients the peptides in the template while the Tren platform controls the tertiary conformation, thus providing the system with many of the basic elements present in natural enzymes, including a well-defined catalytic site. The mechanism in this system is clearly different for the two substrates investigated: in the case of HPNP the metal ions cooperate providing substrate binding and activation of the nucleophile (the alcoholic function of HPNP), while in the case of the oligomeric RNA sequence it is general

acid–base catalysis that operates, although some involvement of a metal ion cannot be fully ruled out. Conformational tuning was also confirmed recently by using simplified systems incorporating in a similar platform only ATANP-like amino acids [75].

Accelerated cleavage of HPNP was also observed by using the same dinuclear Zn^{2+} complex of the peptide described in the Section 10.2.2.1 [58, 76]. Although in this case too, cooperativity between the two metal centers was demonstrated, its effect was modest because the substrate has three possible coordination modes to the catalyst, only one of them allowing interaction with both metal centers.

Poly-aza-crowns have been shown to be able to act as catalysts of the cleavage of RNA with activities unrelated to metal ion coordination as shown in the previous example. A notable example is that described by Michaelis and Kalesse [77] who studied 6mer and 9mer peptide–cyclen conjugates and tested their activity against a 31mer RNA model of transactivation responsive (TAR) RNA of HIV-1. The latter is recognized by the HIV-1 regulatory protein Tat via interaction of the arginine residues with the bulge region of TAR RNA (lysine-rich peptides are known to bind to DNA, see for instance [78]). It is known that the binding of the Tat protein to the TAR RNA is a crucial event in the transcription of viral DNA. The sequences of the two peptides were such to maximize the interaction with the bulge of the RNA substrate studied and, in fact, the cleavage occurs almost exclusively at those positions. The cleavage experiment came as a complete surprise for the investigators who were actually looking for metal ion catalysis and eventually discovered that the most important hydrolysis mechanism was not based on metal ions. The facts that Eu^{3+} and Zn^{2+} both inhibit the reaction with the conjugate, that no hydrolysis was observed in cleavage experiments with peptides lacking the cyclen moiety, and that hydrolysis decreases at higher pH values indicate that the presence of an ammonium species is a requisite for the hydrolysis reaction. Thus the catalytic species is a bis cation on the cyclen moiety. This was subsequently confirmed with other catalysts comprising similar cyclen units [79].

The catalytic activity exerted by polyamines in the cleavage of RNA was exploited also by van Boom *et al*. [80]. They have prepared a peptide nucleic acid (PNA)–diethylenetriamine (DETA) conjugate and shown that it cleaves RNA at micromolar concentrations under physiological conditions and very specifically. The PNA 10mer sequence was such that it bound to a 25mer RNA substrate characterized by a central sequence complementary to that of the putative catalyst. Additional features of the substrate were the presence of two potential neighboring C–A scission sites. Degradation studies were performed at 40 °C and pH 7, and the analysis of the hydrolysis products revealed the presence of two major 5′-end-labeled RNA fragments. Both fragments resulted from hydrolysis at the 3′-side of C17 and C19, respectively. Complete cleavage of the substrate occurred in 24 h and almost 30% of it was already cleaved after 4 h. The polyamine (polyammonium at the experimental pH) is required for catalysis as the same PNA sequence devoid of the DETA subunit is completely inactive. This work nicely complements that reported above by Michaelis and Kalesse, where RNA recognition was performed by a small, natural peptide. Here is a short PNA that binds to the target substrate.

A recurrent feature of many catalysts described above is the importance of the control of the conformation. To tackle this problem, Kawai *et al*. [81]. have synthesized a cyclic decapeptide modeled after the ionophoric cyclic peptide gramicidin S. The natural system possesses a stable antiparallel β-sheet conformation with two type II′ turns at the D-Phe–Pro sequences. Notably, the amino groups of the two Orn residues are located at one side of the β-sheet and are suitable for the introduction of different functional groups. In fact, the two amines could be easily functionalized by attaching to each of them two methylpyridine units to make two facing metal ion-binding units. The activity of the dinuclear Zn^{2+} complex of this derivative turned out to be extremely high with rate accelerations for the cleavage of HPNP very close to four orders of magnitude with respect to the uncatalyzed reaction.

As the last example a very complex system resulting from the incorporation of ATANP into thiolated units used to passivate the surface of gold nanoparticles is shown (Figure 10.16) [82]. These nanoclusters were used, as Zn(II) complexes, as

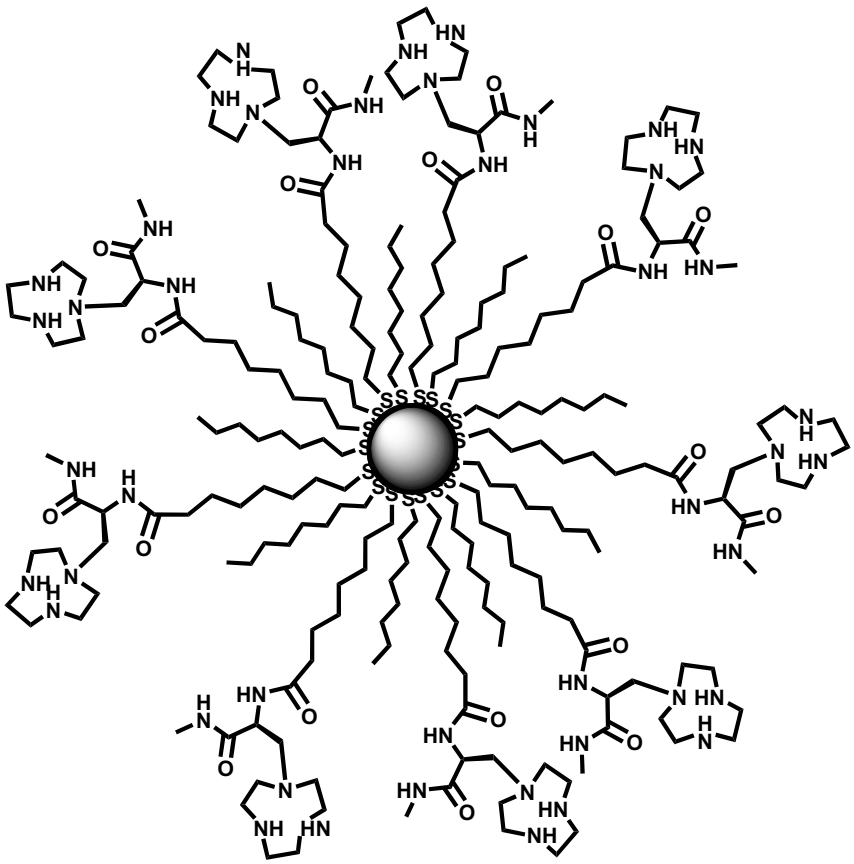


Figure 10.16 Nanoparticle-based RNase.

catalysts for the cleavage of phosphate esters in the mimicry of a RNase. A thorough analysis of the system was carried out by using HPNP as the substrate. The reactivity profile obtained by progressively adding Zn(II) ions to a solution of these nanoparticles up to the saturation of the metal ion binding subunits reveals: (i) the most active system is the one fully loaded with Zn(II) ions and (ii) the sigmoidal profile of the curve supports a cooperativity model between the metal centers as the catalytic efficiency becomes much higher after the first 30% of Zn(II) ions is added. The real catalytic nature of the process was assessed by carrying out experiments with excess substrate. No formation of an intermediate was detected and well-behaved first-order kinetics were observed up to the complete cleavage of all substrate present. By varying the initial substrate concentration a kinetic profile towards saturation was observed. These kinetics allowed the determination of the apparent Michaelis–Menten parameters K_M and k_{cat}, 0.93 mM and $4.2 \times 10^{-3}\,s^{-1}$, respectively. The formal second-order rate constant for HPNP cleavage (k_{cat}/K_M) by the Zn(II)-saturated nanocluster has been determined to be $4.4\,s^{-1}\,M^{-1}$, which is more than 600 times higher than that measured under identical conditions for the mononuclear catalyst corresponding to the "active unit" on the surface of the gold monolayer-protected clusters. The kinetic profile indicated that the role of the metal ions was that of stabilizing the complexed substrate towards the transition state where a further negative charge develops and in facilitating deprotonation of the nucleophilic species.

With such an outstanding catalyst in hand we turned to more appealing substrates like RNA dinucleotides (3′,5′-NpN), namely ApA, CpC, and UpU. Their uncatalyzed cleavage is extremely slow with rate constants (pH 7) ranging from $9.8 \times 10^{-9}\,s^{-1}$ (UpU) to $1.7 \times 10^{-9}\,s^{-1}$ (ApA) (i.e., about two orders of magnitude less reactive than HPNP). As in the case of HPNP, the process is an intramolecular transesterification, in this case by the hydroxyl group at the 2′-position of the ribose. At pH 7.5 (5 mM HEPES buffer) and 40 °C, Zn(II)-gold monolayer-protected clusters cleaved ApA, CpC, and UpU with second-order rate constants of 3.0×10^{-4}, 3.6×10^{-4}, and $1.2 \times 10^{-2}\,s^{-1}\,M^{-1}$, respectively. Thus, the nanoparticle-based catalyst was also fairly active in the cleavage of RNA dinucleotides. Interestingly, it showed a remarkable selectivity in the cleavage of UpU, likely due to a preferential binding of this substrate on the functional surface of the nanoparticles.

10.3 Peptide Models of Heme Proteins

A promising approach to the design of functional proteins exploits the intrinsic reactivity of metal centers, as observed in natural metalloproteins and also illustrated in the previous sections [1, 3, 83–86]. The protein matrix serves the dual purpose of tuning the reactivity of the metal by controlling the number and types of ligands, and controlling substrate access to the active site, providing specificity when required. In particular, transition metals form the active site in the majority of redox-active enzymes.

From an historic perspective, early efforts have focused in particular on iron porphyrin binding model proteins [87–89]. These systems are attractive because of the strong coordination of the porphyrin ring to the metal center, which simplifies the design of the first ligand sphere to two axial positions. More recently, investigators have moved their attention to mononuclear metal proteins, dinuclear centers, multicofactor systems, and complex metalloclusters. While most of these model systems are water soluble, recent developments have focused on designed membrane proteins.

The initial focus of model metalloproteins has been on the design of folded protein structures, the control of redox potential, and, more recently, the introduction of function. Examples of functional proteins span from electron transfer to binding of oxygen to peroxidase activity to more complex oxidations of phenols. The following subsections provide a summary of examples of artificial redox catalysts of which the activity results from the organization of the metal centers.

10.3.1 Heme Proteins

In nature, proteins exploit the intrinsic reactivity of a limited number of metal centers to serve an array of biological functions. Iron porphyrins constitute a particularly versatile type of metal-based cofactor [87]: the heme is at the core of processes ranging from reversible binding of oxygen and nitric oxide (hemoglobin, myoglobin, and guanylyl cyclase) to electron transport (cytochromes *b* and *c*, and cytochrome *c* oxidase) and catalysis of several biotransformations, such as the disruption of hydrogen peroxide (catalase) and the oxidation of alkenes (cytochrome P450). Depending on their function, these proteins can be water-soluble, or associated with membranes (cytochromes *f* and P450) or localized across a membrane (cytochrome bc_1 complex). While the core of each protein is mainly hydrophobic, the outer surface ranges in polarity from hydrophilic to amphiphilic to lipophilic.

10.3.1.1 Early Heme-Peptide Models: Porphyrin as Template

Studies on natural proteins have shown that the protein environment modulates the specific function in a hierarchical manner [87]: the primary coordination sphere (i.e., the presence of one or both axial ligands, and their type and donor properties) controls the intrinsic reactivity of the metal center. Factors such as the local dielectric constant, hydrophobicity, hydrogen bonding, and steric interactions between the primary sphere and its immediate surroundings (i.e., the second coordination sphere) contribute to the next level of control. Long-range electrostatic interactions provide the fine-tuning of the reactivity and can have a surprisingly large effect on the redox potential [90].

Rational design is ideally suited to design model heme-proteins: typically, a minimalist model is built, and more elaborate design features can be added and evaluated sequentially. Thus, these systems are simpler than their natural counterparts, yet can recapitulate all the essential features of metalloproteins [3, 87].

The first model system presented in the literature, helichrome [88, 89], utilized the porphyrin core as both the active site and the template to assemble the protein matrix

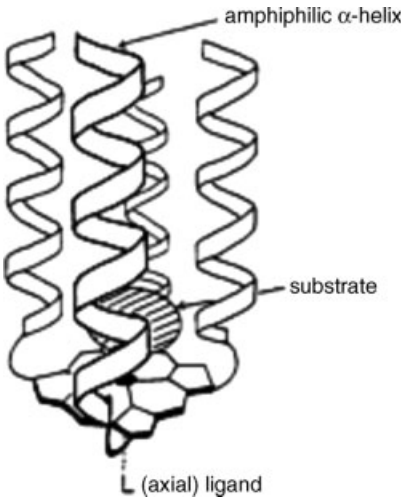


Figure 10.17 Model of helichrome. (From [88].)

(Figure 10.17). Four amphiphilic peptide chains were covalently linked to coproporphyrin 1, forming a four-helix bundle capped at one side by the porphyrin ring. The bundle defined a cavity that could accommodate a substrate in close proximity to the porphyrin ring. While the peptide alone is unfolded in solution, templating the peptide chains on the porphyrin ring induced the formation of partial helical structure. Helichrome catalyzed the oxidation of aniline to aminophenol with $k_{cat} = 0.02\,min^{-1}$ and $K_M = 5.0\,mM$.

The concept of covalently templating peptide chains onto a porphyrin ring was later applied by Akerfeldt and DeGrado to design an artificial membrane channel [91]. The group utilized tetraphenylporphyrin to covalently organize four hydrophobic helical peptides, whose sequence was derived from membrane channels. As the parent peptide, tetraphilin functioned as a proton-selective channel. The attachment to the template stabilized the conducting state of the channel, resulting in a two orders of magnitude improvement in lifetimes.

10.3.1.2 **Bishistidine-Coordinated Models**

Several groups have presented model systems in which the porphyrin is coordinated to amino acid side-chains, generally histidines. The protein is usually designed to assume an amphiphilic α-helical conformation, with the porphyrin localized in the hydrophobic core [92–97]. Histidine residues are placed in the hydrophobic core to coordinate the iron through their ε nitrogen, similarly to what is observed in natural proteins such as cytochromes b_5, b/b_6, and c_{554}, and cytochrome *c* oxidase.

10.3.1.2.1 **Water-Soluble Models: Heme Sandwich** A first class of derivatives encompasses relatively simple models in which one heme is sandwiched between two

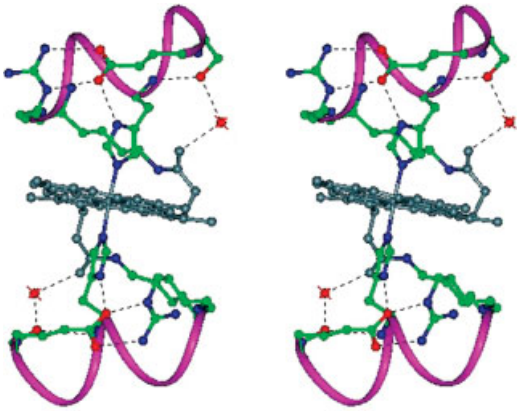


Figure 10.18 Crystal structure of Co(III)-mimochrome IV. (From [101].)

helical peptides, which provide a minimalist hydrophobic core [85, 94, 96, 98–105]. Both covalent and noncovalent strategies have been used to link the porphyrin to the protein matrix. The structures of two Co(III)-mimochromes, in which the porphyrin is covalently attached to both helices, have been determined (Figure 10.18) [101, 102].

Preorganized peptide systems composed of two helices linked together at one or both termini by disulfide bonds demonstrated the role of noncovalently bound cofactors, in this case Co(III)- or Fe(III)-coproporphyrin I, in inducing structure and stability in unfolded peptides [94]. In turn, preorganizing the peptide scaffold in tweezers or cyclic peptides resulted in dissociation constants in the low micromolar range, with an improvement of two to three orders of magnitude over the linear peptides. Cyclic systems in which the peptide sequence is further stabilized by internal salt bridges display native-like folding similar to small natural proteins, with free energy of folding, ΔG, of about 3 kcal/mol. The native-like structure of the complex allowed the determination of the solution structure of the Co(III)-coproporphyrin I complex [Protein Data Bank (PDB) ID: 1pbz] [105].

Although these minimalist models are not functional, they provide a synthetically accessible ground on which to test the effect of simple substitutions in the amino acid sequence on the peptide stability and on the redox potential of the cofactor. Most models exhibit very negative redox potentials, typical of solvent-exposed porphyrins. However, the redox potential of the cofactor and its binding affinity are correlated with the stability of the peptide in the free and bound state.

This observation has been exploited by Benson *et al.*, who demonstrated a significant stabilizing effect in the helical structure of sandwiched hemoprotein models due to edge to face tryptophan–heme interactions [103]. Similarly, two groups [92, 106] have observed a linear correlation between the free energy of

formation of peptide–heme conjugates and the relative hydrophobicity of the mutants within the context of a water-soluble dimeric four helix bundle and of noncovalently bound sandwiched peptide complexes. The relationship with the redox potential is more complex: any analysis of the effect of peptide hydrophobicity must in fact factor in the difference in ligand binding in the ferric versus the ferrous complexes [83, 92, 107]. In general, ferric porphyrins bind histidines more tightly than their ferrous counterparts, thus an increase in binding affinity might in fact result in a stabilization of the oxidized state and a shift for the redox potential towards more negative values. On the other hand, a more hydrophobic environment may also stabilize the neutral ferrous complexes. The two conflicting trends were elegantly dissected by Benson and Gibney [104] who calculated the former effect to account for a −90 mV shift, and the opposing hydrophobic stabilization in about 35 mV: the experimentally measured redox potential shift is −56 in the Ala → Phe mutant. Clearly, any systematic study on the effect of mutations on the redox potential must also consider the affinity for both the ferric and the ferrous porphyrin. In the two studies on bishistidine-ligated complexes reviewed here [83, 92, 107], the inherent preference for the oxidized state prevails over the cost of burying a positive charge; this latter effect lessens the impact on the redox potential to varying degrees. However, other factors might be involved, such as a change in the coordination sphere and/or a structural reorganization of the protein, either by rotation of the helices or by side-chain movements. The determination of the structure of reduced and oxidized complexes could address this issue.

10.3.1.2.2 **Water-Soluble Models: Four-Helix Bundles** A second class of heme protein "maquettes" involves a four-helix bundle [95, 97, 108] (Figure 10.19), in analogy to the topology of many natural heme binding proteins such as cytochrome b_{562} and the central core of cytochrome bc_1. These model systems can bind between one and four heme moieties, depending on the design. Many early systems were highly dynamic, showing some of the properties of a molten globule. A protein that is well structured in the apo state was obtained by systematically changing the hydrophobic side-chains of an early design (PDB ID: 1m3w) [109]. No preformed heme binding pocket was observed, leading to the conclusion that the protein must undergo some rearrangement to accommodate the cofactor. These scaffolds were also used to investigate the redox potential modulation resulting from the presence of charged amino acids in the heme vicinity as a function of pH [110]. However, the redox potential values reported for these models are consistent with a partially water exposed porphyrin ring.

More recently proteins derived from a binary patterned library of four helix bundles [111, 112] and not explicitly designed to bind heme have shown several promising characteristics: binding of carbon monoxide, redox potentials in the −110 to −170 mV range, and nascent peroxidase activity in solution and on electrodes [111, 113]. These proteins were designed by specifying the sequence locations of polar and nonpolar amino acids, but not the identities of these side-chains. This is possible by the use of appropriate degenerate codons in the synthetic genes corresponding to the protein sequence: NAN, used when a polar residue is desired,

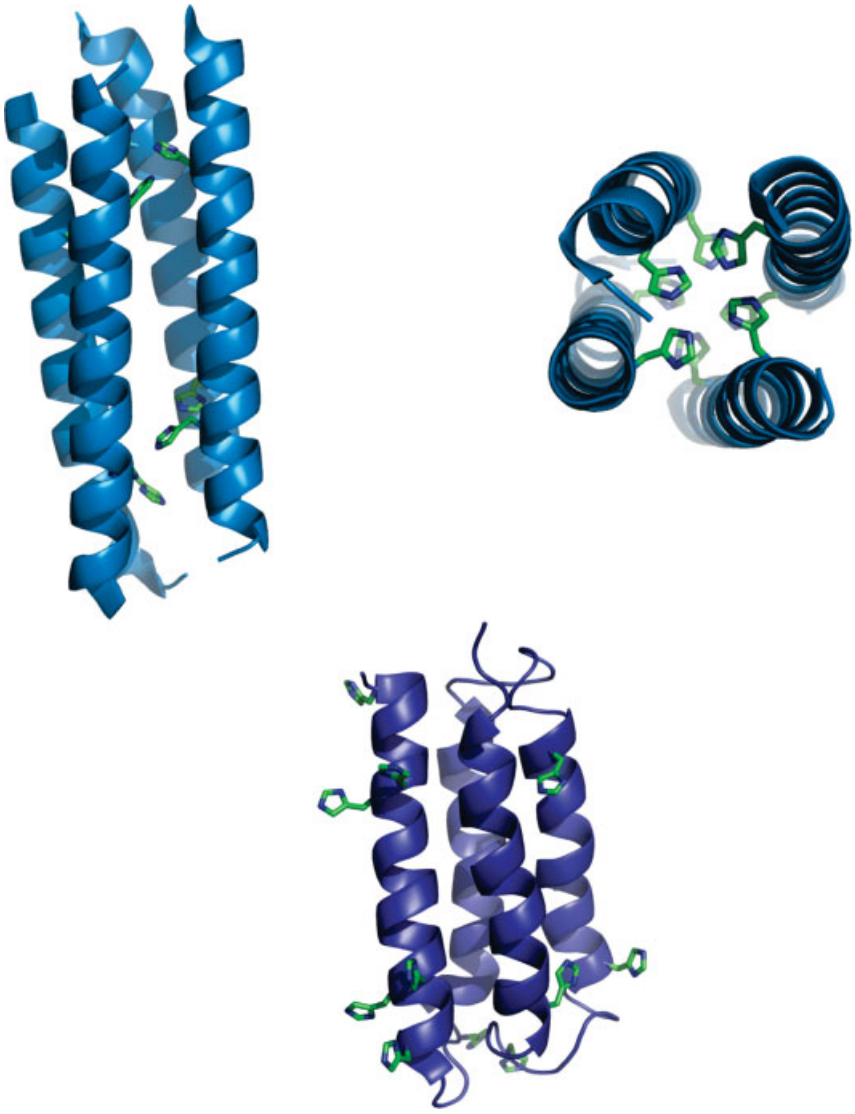
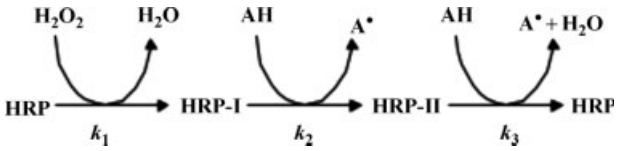


Figure 10.19 (Top) Structure of a "maquette", PDB ID: 1m3w (side and top view), showing the overall architecture of the bundle, held together by disulfide bridges, and the positions of the designed histidines in the core. (Bottom) NMR structure of a single-chain four helix bundle derived from a combinatorial library (PDB ID: 1p68). The coordinating histidines are located on the protein surface.



Scheme 10.6 General mechanism for peroxidase activity. In HRP, the kinetics are analyzed using a "ping-pong" mechanism (see text).

codes for Lys, His, Glu, Gln, Asp, or Asn, while NTN codes for apolar amino acids, specifically Met, Leu, Ile, Val or Phe. Several members of this family catalyzed the peroxide-dependent oxidation of substrates such as 3,3,5,5-tetramethylbenzidine and 2,2-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (ABTS). In particular, Protein 86, the best performer in the series, is about 10 times faster than microperoxidase, a peptide derived from the enzymatic digestion of cytochrome c, which is covalently linked to the porphyrin cofactor. When analyzed according to a ping-pong kinetic model [114] (Scheme 10.6), Protein 86 shows a k_1 of 500 $M^{-1} s^{-1}$, which is significantly lower than horseradish peroxidase (HRP). This rate describes the binding of hydrogen peroxide to form the activated Compound I in peroxidases. Compound I undergoes a one-electron reduction to Compound II, utilizing an organic substrate, in this case ABTS, as reducing agent. However, Protein 86 has a k_3 of 11 000 $M^{-1} s^{-1}$, about 50 times faster than HRP. This parameter describes the rate at which the activated Compound II oxidizes a second molecule of substrate to regenerate the resting state of the cofactor and reflects an active site readily accessible by the substrate.

A second-generation library in which the helices were elongated by approximately one heptad, corresponding to a full turn of the helix, resulted in increased stability. The solution structure of the protein in the free form reveals a stable four-helix bundle in which hydrophobic residues are positioned in the core and polar residues, which includes histidines, decorate the outside (Figure 10.19, PDB ID: 1p68). As in the maquette case, binding of heme presumably requires a rearrangement of the structure. Members of this protein family in the apo form showed catalytic activity in the hydrolysis of activated esters [19] (see Section 10.2.1.1).

10.3.1.2.3 **Membrane-Soluble Heme-Binding Systems** In contrast to water-soluble model systems, very little work on membrane soluble systems has been published after the initial report of tetraphylin by Arkerfeldt and DeGrado [91], possibly due to the experimental difficulties associated with membrane systems. One attractive feature of membrane proteins is that transmembrane helices are inherently more stable than their water-soluble counterparts of similar length, because the α-helical structure can effectively sequester the polar peptide bonds from the nonpolar environment via intrahelical backbone hydrogen bonds [115, 116]. Thus, it should be possible to design peptide scaffolds that are preorganized in the apo form. However, the *de novo* design of membrane-soluble sequences is still in its infancy, due to the paucity of membrane protein structures and of computational design approaches that can handle membrane proteins.

Ways to circumvent this obstacle include the incremental addition of a transmembrane domain to a preorganized water-soluble one, which would then drive the association of the transmembrane part, or utilizing a natural membrane protein as scaffold.

Discher *et al.* have designed a series of amphiphilic maquettes formed by the fusion of an hydrophilic maquette to the transmembrane portion of the M2 influenza ion channel, which is a tetramer and thus matches the topology of the water-soluble portion [117–120]. The heme-binding sites are located at various positions, with some variants carrying a heme-binding site in the lipophilic portion. These systems can be inserted vectorially into membranes and could potentially shuttle electrons from one side to the other of the membrane in a controlled manner.

Ghirlanda *et al.* designed a functional membrane protein, ME1, by engineering a bishistidine-binding site into glycophorin A [121]. Glycophorin A is a small membrane protein formed by the dimerization of two identical transmembrane helices, which cross at an angle of -40° at the N-terminus and define a small cavity at the C-terminus (Figure 10.20).

The protein has been used extensively as a model system to study helix–helix interactions in membrane proteins; through these studies, the GXXXG motif has been identified as an absolute requirement for dimerization, while other mutations are well tolerated.

ME1 binds Fe(III)-protoporphyrin IX with submicromolar affinity, has a redox potential of -128 mV, and displays nascent peroxidase activity. Two single-point mutants, G25F and A22T, were later designed to test whether specific factors such as aromatic-porphyrin interactions and hydrogen bonding to the coordinating histidine

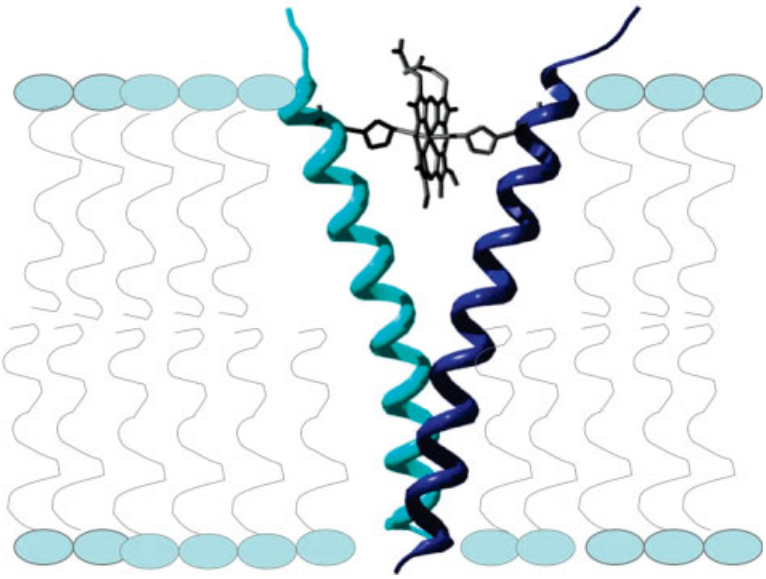


Figure 10.20 Ribbon model of ME1 in complex with protoporphyrin IX; the side-chains of His 26, which provide the axial ligand for the iron center, are shown in stick form. The sequence of glycoporphyrin A is compared with ME1 and the five amino acids changed are highlighted in red.

($N\delta$) can modulate the properties of the heme complex in the context of a membrane protein. G25F and A22T bind Fe(III)-protoporphyrin IX with apparent dissociation constants, $k_{d,app}$, of 1.05×10^{-7} and 6.9×10^{-7} M, and exhibit a redox midpoint potentials of -285 and -196 mV versus standard hydrogen electrode, respectively. The peroxidase activity observed in the series increases in the order G25F $<$ A22T $\leq$ ME1, which is the reverse of the trend observed for the binding affinity to the cofactor. The most dramatic effects on the properties of the cofactor can be ascribed to the stabilization of the Fe(III) form observed in G25F, which results in a more negative redox potential and in reduced peroxidase activity, while the A22T mutation has a lesser effect. The electron paramagnetic resonance (EPR) spectrum of the ferric form of A22T reveals specific signatures typical of a rigid orientation of the imidazole rings providing the axial ligands. In contrast, the EPR spectra of ME1 and G25F suggest free rotation of the imidazole rings with respect to each other. This evidence provides strong support for the designed hydrogen bond of Thr22 to the coordinating histidine. Thus, relatively complex design features aimed at controlling the properties of bound cofactors can be easily introduced in a minimalist membrane hemeprotein model. Future work will focus on designing analogs with altered first sphere ligands (e.g., monohistidine pentacoordinate analogs, His/Met analogs) and exploring the reactivity of the resulting porphyrin complexes.

10.3.2
Diiron Model Proteins: the Due-Ferri Family

A second major family of functional metal centers in proteins involves dinuclear metalloproteins, in particular diiron and dimanganese sites. In nature, these centers catalyze a vast array of reactions, such as hydroxylation, desaturation, and epoxidation, on varied substrates; moreover, the two iron centers are able to transport and activate oxygen. Although the reactivity of these enzymes is quite diverse, there are common structural features both at the protein and at the metal site levels. DeGrado and Lombardi have conducted a careful retrostructural analysis of this class of proteins [122–124].

Generally, the topology of the core protein scaffold is that of an antiparallel four-helix bundle, which contains the active site; the two metal centers are coordinated by two histidines and four glutamic acids, two of which bridge the metals and two of which each chelate a single metal. The site is highly symmetric: two of the ligands, a histidine and a glutamic acid, are housed on a single helix forming a conserved Glu–Xxx–Xxx–His motif, while the chelating glutamic acid is housed in a neighboring helix which is oriented antiparallel to the first one. This hairpin is repeated with pseudo-C_2 symmetry to form the second half of the site. DeGrado and Lombardi used this information to guide the design of a family of minimal models of diiron proteins – the Due-Ferri (DF) series [125–132]. The DF family is characterized by a four-helix bundle, which can be formed by individual helices, or by the dimerization of a helical hairpin, or by a single chain of amino acids. Each of these constructs presents advantages and disadvantages; for example, a tetrameric construct is well suited for combinatorial studies of mutants, while a single-chain construct is best for solution

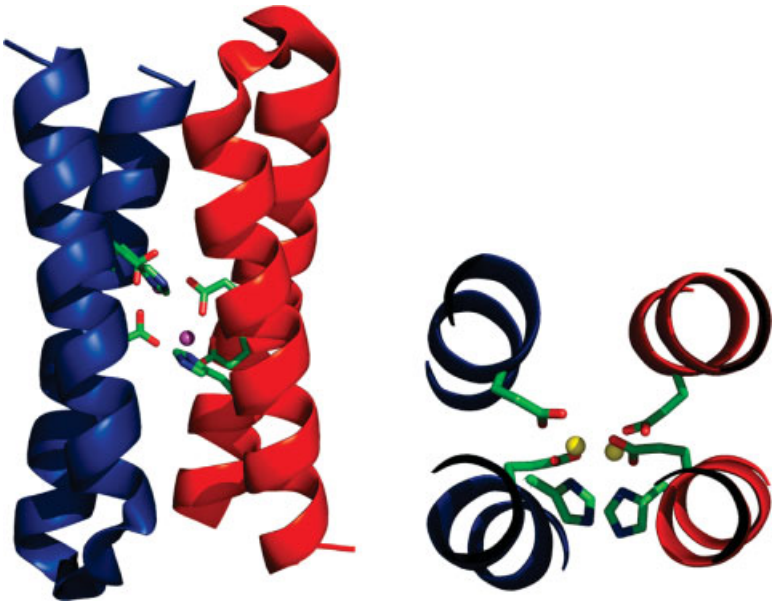


Figure 10.21 Crystal structure of L13A-DF_1.

state characterization. Several members of this family have been structurally characterized by X-ray crystallography or by NMR, in the free and metal-bound state (Co, Zn, Mn, and Fe). As an example, Figure 10.21 shows the overall architecture of the bundle in a dimeric hairpin construct and a detailed view of the binding site in a dimanganese binding construct (PDB ID: 1jmb) [129].

Once the basic features of the structure were verified, the investigators moved their attention to the design of more elaborate features and to optimizing the sequence for structural studies. Finally, the investigators tackled the explicit design of functional variants, specifically by introducing second shell hydrogen-bonding networks to modulate the active site and by sculpting cavities adjacent to the active site that could be easily accessed by substrates. Each mutant was characterized by rigorous spectroscopic studies and when possible by structural analysis, informing the design of successive generations. These efforts resulted in variants able to reversibly bind oxygen and substrates, and most notably in variants that can catalyze the oxidation of 4-aminophenol to the corresponding quinone monoimine in the presence of atmospheric oxygen (Figure 10.22) [133]. The series was designed by modeling a 4-aminophenol coordinated to the two iron atoms of the active site and by sculpting the core of the bundle to accommodate the substrate. The most successful mutant, G_4-DF_{tet}, catalyzes the two-electron oxidation of the substrate with k_{cat}/K_M of $1500\,M^{-1}\,min^{-1}$ for about 100 turnovers and displays saturation kinetics. The reaction is quite specific – no signs of oxygen radical generation were observed. Moreover, subtle changes in the steric hindrance of the hydrophobic core resulted in diminished activity. The catalytic parameters displayed by G_4-DF_{tet} compare very well

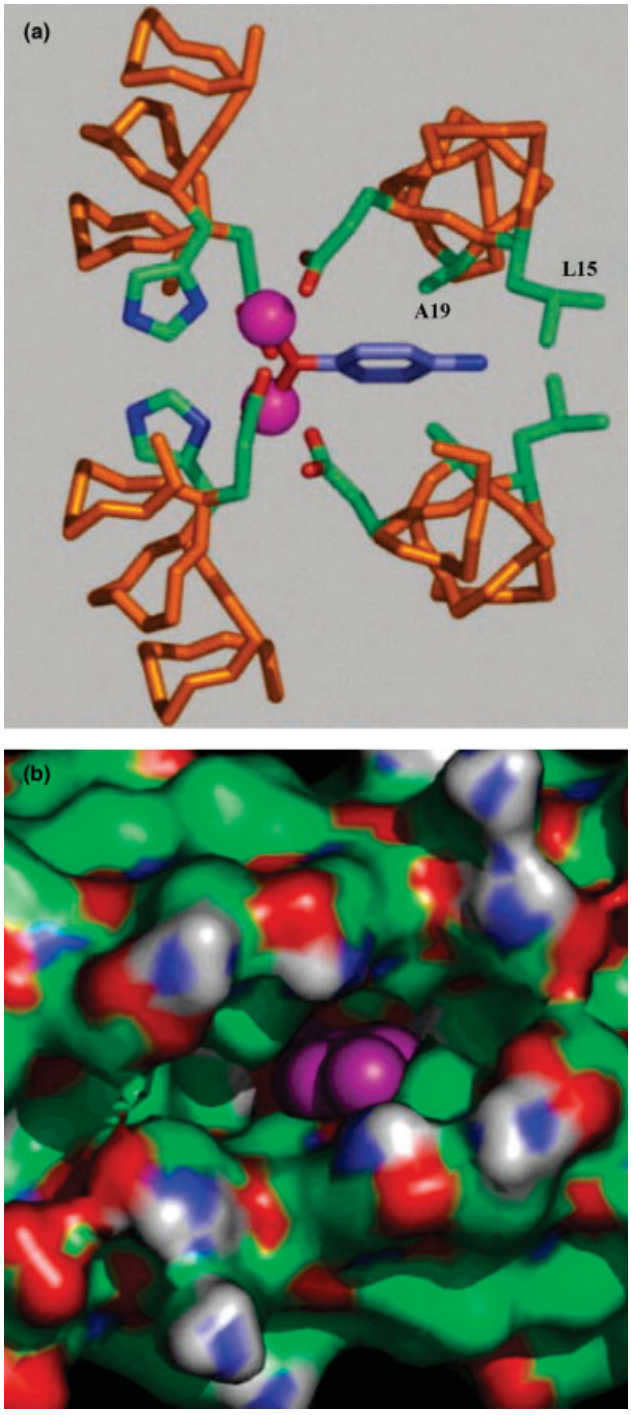


Figure 10.22 (a) Computer model of 4-aminophenol bound to the active site of $DF_{tet}A_2B_2$. (b) Solvent-accessible surface associated with G4–DF_{tet}. The aminophenol ring into the pocket is represented in purple. (Reprinted from [123].)

with catalytic antibodies. To date, G_4-DF_{tet} is the most efficient peptide-based catalyst of a redox reaction.

10.4 Conclusions

Although far from being exhaustive, the selected examples discussed in this chapter illustrate the significant progress made in synthesizing short oligopeptide sequences that display enzyme-like catalysis. Although nature remains the champion to be beaten, artificial model systems have shown impressive efficiencies in mimicking their biological counterparts. In addition, they have revealed valuable mechanistic insights in these processes. In this respect, the fact that chemists are not restricted to the natural set of amino acids is of great help and appears of fundamental importance for designing peptide structures that function also outside the boundaries of the natural system. We would have accomplished our goal if this chapter fosters new interest and research in this field.

References

1 Benson, D.E., Wisz, M.S. and Hellinga, H.W. (2000) *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 6292.

2 Bolon, D.N., Voigt, C.A. and Mayo, S.L. (2002) *Current Opinion in Chemical Biology*, **6**, 125.

3 DeGrado, W.F., Summa, C.M., Pavone, V., Nastri, F. and Lombardi, A. (1999) *Annual Review of Biochemistry*, **68**, 779.

4 Baltzer, L. and Nilsson, J. (2001) *Current Opinion in Biotechnology*, **12**, 355.

5 Bryson, J.W., Betz, S.F., Lu, H.S., Suich, D.J., Zhou, H.X., O'Neil, K.T. and DeGrado, W.F. (1995) *Science*, **270**, 935.

6 Dahiyat, B.I. and Mayo, S.L. (1997) *Science*, **278**, 82.

7 Wolfenden, R. (2006) *Chemical Reviews*, **106**, 3379.

8 Miller, S.J., Copeland, G.T., Papaioannou, N., Horstmann, T.E. and Ruel, E.M. (1998) *Journal of the American Chemical Society*, **120**, 1629.

9 Copeland, G.T., Jarvo, E.R. and Miller, S.J. (1998) *The Journal of Organic Chemistry*, **63**, 6784.

10 Licini, G., Bonchio, M., Broxterman, Q.B., Kaptein, B., Moretto, A., Toniolo, C. and Scrimin, P. (2006) *Biopolymers*, **84**, 97.

11 Tsutsumi, H., Hamasaki, K., Mihara, H. and Ueno, A. (2000) *Journal of the Chemical Society, Perkin Transactions 2*, 1813.

12 Olofsson, S., Johansson, G. and Baltzer, L. (1995) *Journal of the Chemical Society, Perkin Transactions 2*, 2047.

13 Nilsson, J. and Baltzer, L. (2000) *Chemistry – A European Journal*, **6**, 2214.

14 Baltzer, L., Broo, K.S., Nilsson, H. and Nilsson, J. (1999) *Bioorganic and Medicinal Chemistry*, **7**, 83.

15 Broo, K.S., Nilsson, H., Nilsson, J. and Baltzer, L. (1998) *Journal of the American Chemical Society*, **120**, 10287.

16 Anderson, L.K., Dolphin, G.T. and Baltzer, L. (2002) *ChemBioChem*, **3**, 741.

17 Andersson, L.K., Dolphin, G.T., Kihlberg, J. and Baltzer, L. (2000) *Journal of the Chemical Society, Perkin Transactions 2*, 459.

18 Andersson, L.K., Caspersson, M. and Baltzer, L. (2002) *Chemistry – A European Journal*, **8**, 3687.

19 Wei, Y. and Hecht, M.H. (2004) *Protein Engineering Design and Selection*, **17**, 67.

20 Kamtekar, S., Schiffer, J.M., Xiong, H., Babik, J.M. and Hecht, M.H. (1993) *Science*, **262**, 1680.

21 Bolon, D.N. and Mayo, S.L. (2001) *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 14274.

22 Moss, R.A. and Sunshine, W.L. (1974) *The Journal of Organic Chemistry*, **39**, 1083.

23 Ueoka, R., Matsumoto, Y., Moss, R.A., Swarup, S., Sugii, A., Harada, K., Kikuchi, J.-I. and Murakami, Y. (1988) *Journal of the American Chemical Society*, **110**, 1588.

24 Ohkubo, K., Matsumoto, N. and Ohta, H. (1982) *Journal of the Chemical Society, Chemical Communications*, **13**, 738.

25 Davie, E.A.C., Mennen, S.M., Xu, Y. and Miller, S.J. (2007) *Chemical Reviews*, **107**, 5759.

26 Dwars, T., Pätzold, E. and Öhme, G. (2005) *Angewandte Chemie International Edition*, **44**, 7174.

27 Sakai, N. and Matile, S. (2003) *Chemical Communications*, 2514.

28 Sakai, N., Sordé, N. and Matile, S. (2003) *Journal of the American Chemical Society*, **125**, 7776.

29 Silva, G.A., Czeisler, C., Niece, K.L., Beniash, E., Harrington, D.A., Kessler, J.A. and Stupp, S.I. (2004) *Science*, **303**, 1352.

30 Guler, M.O. and Stupp, S.I. (2007) *Journal of the American Chemical Society*, **129**, 12082.

31 Esposito, A., Delort, E., Lagnoux, D., Djojo, F. and Reymond, J.-L. (2003) *Angewandte Chemie International Edition*, **42**, 1381.

32 Douat-Casassus, C., Darbre, T. and Reymond, J.-L. (2004) *Journal of the American Chemical Society*, **126**, 7817.

33 Kofoed, J. and Reymond, J.-L. (2005) *Current Opinion in Chemical Biology*, **9**, 656.

34 Delort, E., Darbre, T. and Reymond, J.-L. (2004) *Journal of the American Chemical Society*, **126**, 15642.

35 Guarise, C., Manea, F., Zaupa, G., Pasquato, L., Prins, L.J. and Scrimin, P. (2008) *Journal of Peptide Science*, **14**, 174.

36 Pengo, P., Polizzi, S., Pasquato, L. and Scrimin, P. (2005) *Journal of the American Chemical Society*, **127**, 1616.

37 Pengo, P., Baltzer, L., Pasquato, L. and Scrimin, P. (2007) *Angewandte Chemie International Edition*, **46**, 400.

38 Williams, N.H., Takasaki, B., Wall, M. and Chin, J. (1999) *Accounts of Chemical Research*, **32**, 485.

39 Mancin, F., Scrimin, P., Tecilla, P. and Tonellato, U. (2005) *Chemical Communications*, 2540.

40 (a) Wilcox, D.E. (1996) *Chemical Reviews*, **96**, 2435; (b) Sträter, N., Lipscomb, W.N., Klabunde, T. and Krebs, B. (1996) *Angewandte Chemie (International Edition in English)*, **35**, 2025.

41 Qi, D., Tann, C.-M., Haring, D., Distefano, M.D., (1996) *Chemical Reviews*, **101**, (2001) 3081.

42 Pei, D. and Schultz, P.G. (1991) *Journal of the American Chemical Society*, **113**, 9398.

43 (a) Zuckerman, R.N. and Schultz, P.G. (1989) *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 1766 and references therein; (b) Pei, D., Corey, D.R. and Schultz, P.G. (1990) *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 9858.

44 (a) Kim, Y.-G., Cha, J. and Chandrasegaran, S. (1996) *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 1156; (b) Chandrasegaran, S. and Smith, J. (1999) *Biological Chemistry*, **380**, 841.

45 Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S., Carroll, D. (2000) *Nucleic Acids Research*, **28**, 3361.

46 (a) Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.-G. and Chandrasegaran, S. (2001) *Molecular and Cellular Biology*, **21**, 289; (b) Bibikova, M., Golic, M., Golic, K.G. and Carroll, D. (2002) *Genetics*, **161**, 1169.

47 Morton, J., Davis, M.W., Jorgensen, E.M., Carroll, D. (2001) *Proceedings of the*

National Academy of Sciences of the United States of America, **103**, (2006) 16370.

48 Carroll, D., Morton, J.J., Beumer, K.J. and Segal, D.J. (2006) *Nature Protocols*, **1**, 1329.

49 Porteus, M.H. and Carroll, D. (2005) *Nature Biotechnology*, **23**, 967.

50 Welch, J.T., Sirish, M., Lindstrom, K.M. and Franklin, S.J. (2001) *Inorganic Chemistry*, **40**, 1982.

51 Welch, J.T., Kearney, W.R. and Franklin, S.J. (2003) *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 3725.

52 (a) Franklin, S.J. (2001) *Current Opinion in Chemical Biology*, **5**, 201; (b) Komiyama, M., Takeda, N. and Shigekawa, H. (1999) *Chemical Communications*, 1443.

53 Kovacic, R.T., Welch, J.T., Franklin, S.J., (2001) *Journal of the American Chemical Society*, **125**, (2003) 6656.

54 (a) Li, Y., Zhao, Y., Hatfield, S., Wan, R., Zhu, Q., Li, X., McMills, M., Ma, Y., Li, J., Brown, K.L., He, C., Liu, F. and Chen, X. (2000) *Bioorganic and Medicinal Chemistry*, **8**, 2675; (b) Li, Y., Hatfield, S., Li, J., McMills, M., Zhao, Y. and Chen, X. (2002) *Bioorganic and Medicinal Chemistry*, **10**, 667.

55 Madhavaiah, C., Verma, S., (2000) *Bioconjugate Chemistry*, **12**, (2001) 855.

56 Ren, R., Yang, P., Zheng, W. and Hua, Z. (2000) *Inorganic Chemistry*, **39**, 5454.

57 (a) Fitzsimons, M.P. and Barton, J.K. (1997) *Journal of the American Chemical Society*, **119**, 3379; (b) Copeland, K.D., Fitzimons, M.P., Houser, R.P. and Barton, J.K. (2002) *Biochemistry*, **41**, 343.

58 Rossi, P., Felluga, F., Tecilla, P., Formaggio, F., Toniolo, C., Scrimin, P., (1997) *Journal of the American Chemical Society*, **121**, (1999) 6948.

59 Rossi, P., Felluga, F. and Scrimin, P. (1998) *Tetrahedron Letters*, **39**, 7159.

60 (a) Toniolo, C., Crisma, M., Formaggio, F., Valle, G., Cavicchioni, G., Précigoux, G., Aubry, A. and Kamphuis, J. (1993) *Biopolymers*, **33**, 1061; (b) Crisma, M., Bisson, W., Formaggio, F., Broxterman, Q.B. and Toniolo, C. (2002) *Biopolymers*, **64**, 236; (c) Bellanda, M., Mammi, S., Geremia, S., Demitri, N., Randaccio, L., Broxterman, Q.B., Kaptein, B., Pengo, P., Pasquato, L. and Scrimin, P. (2007) *Chemistry – A European Journal*, **13**, 407.

61 Berkessel, A., Hèrault, D.A., (1993) *Angewandte Chemie International Edition*, **38**, (1999) 102.

62 Schmuck, C. and Dudaczek, J. (2007) *Organic Letters*, **9**, 5389.

63 Schmuck, C. and Geiger, L. (2005) *Chemical Communications*, 772.

64 Silnikov, V.N. and Vlassov, V.V. (2001) *Russian Chemical Reviews*, **70**, 491.

65 Thorp, H.H. (2000) *Chemistry & Biology*, **7**, R33.

66 (a) Barbier, B. and Brack, A. (1988) *Journal of the American Chemical Society*, **110**, 6880; (b) Barbier, B. and Brack, A. (1992) *Journal of the American Chemical Society*, **114**, 3511.

67 Gutte, B., (1988) *The Journal of Biological Chemistry*, **250**, (1975) 889.

68 Gutte, B., Daumigen, M. and Wittschieber, E. (1979) *Nature*, **281**, 650.

69 Tung, C.H., Wei, Z., Leibowitz, M.J. and Stein, S. (1992) *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 7114.

70 Lima, W.F. and Crooke, S.T. (1999) *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 10010.

71 Rossi, P., Tecilla, P., Baltzer, L. and Scrimin, P. (2004) *Chemistry – A European Journal*, **10**, 4163.

72 Razkin, J., Nilsson, H. and Baltzer, L. (2007) *Journal of the American Chemical Society*, **129**, 14752.

73 Scarso, A., Scheffer, U., Göbel, M., Broxterman, Q.B., Kaptein, B., Formaggio, F., Toniolo, C. and Scrimin, P. (2002) *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 5144.

74 Shinkai, S., Ikeda, M., Sugasaki, A. and Takeuchi, M. (2001) *Accounts of Chemical Research*, **34**, 494.

75 Scarso, A., Zaupa, G., Bodar Houillon, F., Prins, L.J. and Scrimin, P. (2007) *The Journal of Organic Chemistry*, **72**, 376.

76 Rossi, P., Felluga, F., Tecilla, P., Formaggio, F., Crisma, M., Toniolo, C. and Scrimin, P. (2000) *Biopolymers*, **55**, 496.

77 Michaelis, K. and Kalesse, M. (1999) *Angewandte Chemie International Edition*, **38**, 2243.

78 Bergstrom, R.C., Mayfield, L.D. and Corey, D.R., (2001) *Chemistry & Biology*, **8**, 199.

79 Michaelis, K. and Kalesse, M. (2001) *ChemBioChem*, **2**, 79.

80 Verheijen, J.C., Deiman, B.A.L.M., Yeheskiely, E.Y., van der Marel, G.A. and Van Boom, J.H. (2000) *Angewandte Chemie International Edition*, **39**, 369.

81 Yamada, K., Takahashi, Y., Yamamura, H., Araki, S., Kawai, M. and Saito, K. (2000) *Chemical Communications*, 1315.

82 Manea, F., Bodar Houillon, F., Pasquato, L. and Scrimin, P. (2004) *Angewandte Chemie International Edition*, **43**, 6165.

83 Kennedy, M.L. and Gibney, B.R. (2001) *Current Opinion in Structural Biology*, **11**, 485.

84 Lu, Y., Berry, S.M. and Pfister, T.D. (2001) *Chemical Reviews*, **101**, 3047.

85 Nastri, F., Lombardi, A., D'Andrea, L.D., Sanseverino, M., Maglio, O. and Pavone, V. (1998) *Biopolymers*, **47**, 5.

86 Benson, D.E., Wisz, M.S., Liu, W. and Hellinga, H.W. (1998) *Biochemistry*, **37**, 7070.

87 Lombardi, A., Nastri, F. and Pavone, V. (2001) *Chemical Reviews*, **101**, 3165.

88 Sasaki, T. and Kaiser, E.T. (1989) *Journal of the American Chemical Society*, **111**, 380.

89 Sasaki, T. and Kaiser, E.T. (1990) *Biopolymers*, **29**, 79.

90 Rodgers, K.K. and Sligar, S.G. (1991) *Journal of Molecular Biology*, **221**, 1453.

91 Akerfeldt, K.S., Kim, R.M., Camac, D., Groves, J.T., Lear, J.D. and Degrado, W.F. (1992) *Journal of the American Chemical Society*, **114**, 9656.

92 Huffman, D.L. and Suslick, K.S. (2000) *Inorganic Chemistry*, **39**, 5418.

93 Shifman, J.M., Gibney, B.R., Sharp, R.E. and Dutton, P.L. (2000) *Biochemistry*, **39**, 14813.

94 Rosenblatt, M.M., Huffman, D.L., Wang, X., Remmer, H.A. and Suslick, K.S. (2002) *Journal of the American Chemical Society*, **124**, 12394.

95 Gibney, B.R., Rabanal, F., Reddy, K.S. and Dutton, P.L. (1998) *Biochemistry*, **37**, 4635.

96 Benson, D.R., Hart, B.R., Zhou, X. and Doughty, M.B. (1995) *Journal of the American Chemical Society*, **117**, 8502.

97 Rojas, N.R.L., Kamtekar, S., Simons, C.T., McLean, J.E., Vogel, K.M., Spiro, T.G., Farid, R.S. and Hecht, M.H. (1997) *Protein Science*, **6**, 2512.

98 Nastri, F., Lombardi, A., Morelli, G., Pedone, C., Pavone, V., Chottard, G., Battioni, P. and Mansuy, D. (1998) *Journal of Biological Inorganic Chemistry*, **3**, 671.

99 Nastri, F., Lombardi, A., Morelli, G., Maglio, O., D'Auria, G., Pedone, C. and Pavone, V. (1997) *Chemistry – A European Journal*, **3**, 340.

100 Lombardi, A., Nastri, F., Marasco, D., Maglio, O., De Sanctis, G., Sinibaldi, F., Santucci, R., Coletta, M. and Pavone, V. (2003) *Chemistry – A European Journal*, **9**, 5643.

101 Costanzo, L., Geremia, S., Randaccio, L., Nastri, F., Maglio, O., Lombardi, A. and Pavone, V. (2004) *Journal of Biological Inorganic Chemistry*, **9**, 1017.

102 D'Auria, G., Maglio, O., Nastri, F., Lombardi, A., Mazzeo, M., Morelli, G., Paolillo, L., Pedone, C. and Pavone, V. (1997) *Chemistry – A European Journal*, **3**, 350.

103 Liu, D., Williamson, D.A., Kennedy, M.L., Williams, T.D., Morton, M.M. and Benson, D.R. (1999) *Journal of the American Chemical Society*, **121**, 11798.

104 Kennedy, M.L., Silchenko, S., Houndonougbo, N., Gibney, B.R., Dutton, P.L., Rodgers, K.R. and Benson, D.R. (2001) *Journal of the American Chemical Society*, **123**, 4635.

105 Rosenblatt, M.M., Wang, J. and Suslick, K.S. (2003) *Proceedings of the National*

Academy of Sciences of the United States of America, **100**, 13140.

106 Sakamoto, S., Obataya, I., Ueno, A. and Mihara, H. (1999) *Journal of the Chemical Society, Perkin Transactions 2*, 2059.

107 Reedy, C.J., Kennedy, M.L. and Gibney, B.R. (2003) *Chemical Communications*, 570.

108 Gibney, B.R., Huang, S.S., Skalicky, J.J., Fuentes, E.J., Wand, A.J. and Dutton, P.L. (2001) *Biochemistry*, **40**, 10550.

109 Huang, S.S., Gibney, B.R., Stayrook, S.E., Dutton, P.L. and Lewis, M. (2003) *Journal of Molecular Biology*, **326**, 1219.

110 Shifman, J.M., Moser, C.C., Kalsbeck, W.A., Bocian, D.F. and Dutton, P.L. (1998) *Biochemistry*, **37**, 16815.

111 Moffet, D.A., Certain, L.K., Smith, A.J., Kessel, A.J., Beckwith, K.A. and Hecht, M.H. (2000) *Journal of the American Chemical Society*, **122**, 7612.

112 Moffet, D.A. and Hecht, M.H. (2001) *Chemical Reviews*, **101**, 3191.

113 Das, A. and Hecht, M.H. (2007) *Journal of Inorganic Biochemistry*, **101**, 1820.

114 Dunford, H.B. (1991) In *Peroxidases in Chemistry and Biology*, vol. II (eds J. Everse, K.E. Everse and M.B. Grisham), CRC Press, Boca Raton, FL, pp. 1–24.

115 DeGrado, W.F., Gratkowski, H. and Lear, J.D. (2003) *Protein Science*, **12**, 647.

116 Ubarretxena-Belandia, I. and Engelman, D.M. (2001) *Current Opinion in Structural Biology*, **11**, 370.

117 Discher, B.M., Koder, R.L., Moser, C.C. and Dutton, P.L. (2003) *Current Opinion in Chemical Biology*, **7**, 741.

118 Discher, B.M., Noy, D., Strzalka, J., Ye, S., Moser, C.C., Lear, J.D., Blasie, J.K. and Dutton, P.L. (2005) *Biochemistry*, **44**, 12329.

119 Noy, D., Discher, B.M., Rubtsov, I.V., Hochstrasser, R.M. and Dutton, P.L. (2005) *Biochemistry*, **44**, 12344.

120 Ye, S., Discher, B.M., Strzalka, J., Xu, T., Wu, S.P., Noy, D., Kuzmenko, I., Gog, T., Therien, M.J., Dutton, P.L. and Blasie, J.K. (2005) *Nano Letters*, **5**, 1658.

121 Cordova, J.M., Noack, P.L., Hilcove, S.A., Lear, J.D. and Ghirlanda, G. (2007) *Journal of the American Chemical Society*, **129**, 512.

122 Calhoun, J.R., Nastri, F., Maglio, O., Pavone, V., Lombardi, A. and DeGrado, W.F. (2005) *Biopolymers*, **80**, 264.

123 Maglio, O., Nastri, F., de Rosales, R.T.M., Faiella, M., Pavone, V., DeGrado, W.F. and Lombardi, A. (2007) *Comptes Rendus Chimie*, **10**, 703.

124 Lombardi, A., Summa, C.M., Geremia, S., Randaccio, L., Pavone, V. and DeGrado, W.F. (2000) *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 6298.

125 Geremia, S., Di Costanzo, L., DeGrado, W.F., Lombardi, A., Pavone, V. and Randaccio, L. (2001) *Journal of Inorganic Biochemistry*, **86**, 232.

126 Maglio, O., Nastri, F., Calhoun, J.R., Lahr, S., Wade, H., Pavone, V., DeGrado, W.F. and Lombardi, A. (2005) *Journal of Biological Inorganic Chemistry*, **10**, 539.

127 Maglio, O., Nastri, F., Pavone, V., Lombardi, A. and DeGrado, W.F. (2003) *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 3772.

128 Wade, H., Di Constanzo, L., Pavone, V., Lombardi, A. and DeGrado, W.F. (2002) *Biochemistry*, **41**, 8981.

129 Di Costanzo, L., Wade, H., Geremia, S., Randaccio, L., Pavone, V., DeGrado, W.F. and Lombardi, A. (2001) *Journal of the American Chemical Society*, **123**, 12749.

130 Lahr, S.J., Engel, D.E., Stayrook, S.E., Maglio, O., North, B., Geremia, S., Lombardi, A. and DeGrado, W.F. (2005) *Journal of Molecular Biology*, **346**, 1441.

131 Pasternak, A., Kaplan, J., Lear, J.D. and DeGrado, W.F. (2001) *Protein Science*, **10**, 958.

132 Marsh, E.N.G. and Degrado, W.F. (2002) *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 5150.

133 Kaplan, J. and Degrado, W.F. (2004) *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 11566.

11
Substrate Recognition

Keith Brocklehurst, Sheraz Gul, and Richard W. Pickersgill

11.1
Recognition, Specificity, Catalysis, Inhibition, and Linguistics

Specific molecular recognition processes by which molecules recognize and discriminate between closely related partners are involved in essentially all aspects of biological function, ranging in complexity from enzyme–inhibitor and enzyme–substrate systems to phenomena such as gene expression, the immune system, and synaptic transmission, and these are important considerations in drug design [1]. This chapter is concerned with general concepts exemplified by a selection of relatively well-defined enzyme–substrate and, by extension, enzyme–inhibitor systems. The original concept of recognition by complementarity in the enzyme–substrate adsorptive (Michaelis) complex was suggested by Emil Fischer in 1894 using his well-known lock-and-key analogy [2]. It is now generally accepted that this has now been superseded by complementarity in the reaction transition state, based on the ideas reported by J.B.S. Haldane [3] in 1930 and elaborated by Linus Pauling [4] in 1946. The catalytic potential of enzymes derives from stabilization in the transition state relative to any stabilization in the Michaelis complex. The latter is anticatalytic in that the free energy expended in stabilizing the adsorptive complex offsets the transition-state stabilization energy (e.g., [5]).

Valuable chapters and books on enzyme catalysis and inhibition are available (e.g., [5–11]) and detailed mechanistic studies on many individual enzymes are described in Sinnot [12]. Macromolecular catalysis includes also catalysis by catalytic antibodies [13] and by synthetic polymers [14], including molecularly imprinted polymers [15]. The contributions of selections of various types of chemical catalysis such as acid–base, nucleophilic, and electrophilic catalysis [5, 6, 8] are almost always insufficient to account for the high catalytic efficiency of enzymes. A particularly important additional contribution is the entropic advantage that derives from the fact that enzyme catalysis occurs within an enzyme–substrate complex. The complex is essentially a single molecular species and the catalytic act involves one or more unimolecular steps during which there is no loss of translational or rotational entropy

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

in the transition state. This advantage is compensated for by the loss of these entropies during formation of the Michaelis complex i.e. by the binding energy (the favorable interaction between enzyme and substrate). An important aim of the study of enzyme catalysis is to understand how that part of the binding energy associated with the nonreacting parts of enzyme and substrate is able to be expressed in the transition state rather than in the Michaelis complex (e.g., [5, 8, 16]).

The establishment of contributions by binding to catalytic site function requires consideration of a variety of types of enzyme–substrate system. Enzymes such as tyrosyl-tRNA synthetase lie at one end of the mechanistic spectrum. This enzyme catalyzes nucleophilic attack of one substrate on another by using an array of differential ground-state and transition-state binding interactions without the intervention of catalytic acids, bases, or nucleophiles provided by the enzyme [17]. In potentially more complex systems, a series of chemical events involving enzyme and substrate occur in addition to the existence of opportunities for differential binding interactions. Many proteinases, which both have active centers with large binding areas made up of multiple subsites and employ various chemical mechanisms, might be expected to be in this category, and there is considerable evidence for functional coupling of binding and catalytic sites despite their spatial separation, such as in papain [18–20], trypsin [21], and subtilisin [22, 23].

The concept of the transition state [24–26] is of central importance in understanding and describing chemical mechanisms. Tight binding of enzyme and substrate in the transition state is not merely one of the possible contributors to catalysis, but is the cardinal phenomenon regardless of the detail of the chemical and conformational phenomena that contribute to the complete description of mechanism [27, 28]. The fundamental property of a catalyst is its ability to discriminate in favor of its participation in binding in the transition state rather than in an initial ground-state complex. This description of "transition-state stabilization" as against the often used "the enzyme stabilizes the transition state" serves to emphasize the importance of the mutual stabilization of substrate and enzyme by perturbation in the transition state [8]. Where stable intermediate states exist, optimal activity usually requires that these states are not excessively stabilized [29, 30]; however, for example, see accounts of the stabilization of the aminoacyl-adenylate intermediate tightly bound to the active center of the aminoacyl-tRNA synthetase (aaRS) [5, 17]). Schowen [28] refers to a formulation of catalysis that omits or de-emphasizes the transition state, but is correct in principle and, therefore, is transformable to transition-state language, as a "canonical" formulation. By contrast, the other extreme, which uses transition-state language exclusively, is referred to as a "fundamentalist" formulation. As an example of the semantic nature of the two types of description, he translates the important "canonical" account by Bender *et al.* [31] of the dissection of the "anatomy" of the catalysis by chymotrypsin into "fundamentalist" terms. One area in which consideration of transition-state structure is necessarily explicitly of central importance is the design of transition-state inhibitors [10], which is based on the early ideas of Haldane [3] and Pauling [4] that were developed by Jencks [32], Wolfenden (e.g., [10]) and Lienhard [33]. A particularly interesting application of such inhibitors is their use in distinguishing between alternative mechanisms of inhibition when the structures

of the complexes with their target enzymes are available (see [10] for examples). Some approaches to the investigation of transition-state structures and their properties, including quantum mechanical and isotope studies, are described in detail in Gandour and Schowen [11]. The two fundamental properties of enzymes are substrate recognition and rate enhancement. It seems probable that a complete description of enzyme catalytic mechanism will lie between what Knowles [34] referred to as the two linguistic extremes whereby recognition and catalysis are described either completely separately or, as in the classical treatments by Polanyi [35], Haldane [3] and Pauling [4], as a single phenomenon. This is nicely illustrated by the realization that specificity (discrimination between potential substrates) is a function of both substrate binding and catalytic activity, being quantified by the value of k_{cat}/K_M (e.g., [5]). It is widely assumed that enzyme and substrate bind initially in their ground-state forms, and then together approach the transition state, although as Hammett [36] pointed out, the mechanism by which this occurs may not be capable of experimental investigation, at least by reaction kinetics. For catalyses with relatively low values of k_{cat}/K_M, however, the possibility remains that initial enzyme–substrate combinations might result from minor, more activated forms, of substrate and/or enzyme within the restriction of a diffusion-limited encounter [37].

It is important not to confuse specificity as defined above with maximization of rate of catalysis. The former is defined by a rate constant, k_{cat}/K_M, the apparent second-order rate constant for product formation by reaction of the uncomplexed enzyme with the uncomplexed substrate, whereas rate requires also one or more concentration terms. Thus, although complementarity in the transition state maximizes k_{cat}/K_M, this is not a sufficient condition for maximization of rate. For a given substrate concentration, the additional condition is a high value of K_M ($K_M > [S]$), which maximizes the concentration of enzyme in its uncomplexed form [5, 38]. This contradicts the traditional view, possibly still held by some, that effective catalysis by enzymes is associated with tight binding in the Michaelis complex leading to a low K_M value. The failure to distinguish rates from rate constants can lead to inappropriate conclusions being drawn more generally from kinetic analysis. The question arises as to what phenomena lead to high K_M values. The case for destabilization of the Michaelis complex to facilitate formation of the transition state is made by Jencks [39, 40]. Whereas strain, induced fit, and nonproductive binding cannot affect specificity as quantified by the rate constant k_{cat}/K_M [5], strain and "entropic strain" (i.e., loss of entropy consequent on binding) is essential to achieve the large rates typical of enzyme catalysis [39, 40]. An exception to the requirement for a high K_M value, discussed by Fersht [5], occurs with enzymes whose role is not to provide high rates of catalysis, but rather to regulate metabolic pathways. Such enzymes are often controlled by the K_M values for particular substrates and a low K_M value can be advantageous for the first enzyme in the pathway to control the rate of entry of substrate to the pathway.

An approach complementary to the direct study of catalysis by the use of substrates and analogous binding inhibitors involves the use of time-dependent substrate-derived inhibitors. One of the most sensitive and direct ways of investigating transition-state geometry is the study of chemical reactivity. This has been applied

particularly effectively to the investigation of dynamic aspects of molecular recognition (i.e., the dependence of catalytic site chemistry on specific binding interactions) by using two-protonic-state inhibitors as reactivity probes [41] together with pH-dependent [42] and temperature-dependent [43] kinetics. The reactivity of a catalytic site nucleophile that plays a central role in the catalytic act is a prime target for such investigations, and studies on the cysteine proteinase papain and its natural variants [20] using specifically designed and synthesized thiol-specific probes demonstrated such dependence by large changes in pH–*k* profiles [44] as discussed in Section 11.3. Even if the inhibitors are not substrate analogs, experiments can be designed to use specific noncovalent interactions to reveal key aspects of the catalytic mechanism (e.g., [45]).

The sections that follow exemplify techniques and approaches that have been used in the study of molecular recognition by a selection of enzymes and the advances produced in understanding its roles in specificity, catalysis, and associated catalytic-site chemistry. References to theoretical and practical aspects of the commonly used methods for studying the kinetics of catalysis that can provide a quantitative assessment of the effects of molecular recognition motifs can be found in Gul *et al.* [46]. The choices of the enzyme types discussed in Sections 11.2–11.7 relate to the facets of recognition revealed particularly clearly in each case. Section 11.2 deals with serine proteinases (foundation of transition-state binding and roles of non-reacting parts of substrates), Section 11.3 with cysteine proteinases (early recognition of specific binding interactions in enzyme catalysis, the general P/S designation of subsite recognition, kinetic specificity in catalysis, and characterization of the influence of binding interactions on catalytic site chemistry by the use of time-dependent inhibitors), Section 11.4 with glycoside hydrolases (contributions of X-ray crystallography, stereochemical recognition, substrate distortion, and *exo* versus *endo* specificity for polymeric substrates), Section 11.5 with protein kinases (recognition by remote binding sites with physiological protein substrates and role of specific binding partners such as cyclins in provision of versatility in substrate selection), Section 11.6 with aaRSs (accuracy of substrate selection, proofreading, and contributions of protein engineering and pre-steady-state kinetics to the study of recognition and mechanism and to testing fundamental theories of enzyme catalysis), and Section 11.7 with lipases (interfacial recognition). Valuable detailed accounts of both substrate recognition and catalytic mechanism of most enzyme classes are presented in volumes I–III of Sinnot [12].

11.2
Serine Proteinases

The serine proteinases were among the first enzymes to attract detailed investigation [47] and have been the subject of particularly extensive study. Comprehensive reviews of the structural basis of their substrate specificities and catalytic mechanisms were published by Perona and Craik in 1995 [48] and by Wharton in 1998 [49]. The continuing substantial interest in these enzymes is due to the wealth of

structural and mechanistic information in the literature, which provides the basis for the ever more detailed studies required at atomic resolution to understand dynamic aspects of mechanism (e.g., [50]) and the growing awareness of their roles in diverse physiological processes (e.g., [51–53]). The former includes the development of techniques such as time-resolved X-ray crystallography [54–56] and ultrarapid laser spectroscopy [57]. The two main structural classes of serine proteinases [58], chymotrypsin-like (mainly β-structure, turns and loops with little α-helix) and subtilisin-like (predominantly α-helix), have been joined by wheat serine carboxypeptidase II [59], which is homodimeric with the α/β-fold typical of some other hydrolytic enzymes [60], but shares most of the features of the catalytic sites of the other two classes. The major features of the mechanism of the hydrolysis of amides and esters common to the enzymes of the chymotrypsin and subtilisin families were deduced initially mainly by kinetic study of the effects of variation in substrate structure (e.g., references in [49]). The insights gained from these studies were confirmed and extended subsequently by structural evidence from X-ray (e.g., [61–63] and other references cited in [50]) and neutron [64] crystallography, nuclear magnetic resonance (NMR; e.g., [65–67]) and vibrational [57, 68, 69] spectroscopies, and protein engineering (mutational analysis; see references cited in [48]). Following Michaelis complex formation, the acyl group of the substrate is transferred to the hydroxyl group of the catalytic-site serine residue assisted by general base catalysis provided by the imidazole group of the catalytic-site histidine residue. Deacylation, the hydrolysis of the acylenzyme intermediate, proceeds by an analogous mechanism in which the second substrate, water, replaces the serine hydroxyl group in the nucleophilic attack at carbonyl carbon. Two uncertainties, the possible existence of tetrahedral intermediates in the acylation and deacylation steps and the possible role of Asp102 in a triad relay mechanism (Ser195, His57, and Asp102) as against a diad mechanism without Asp102, are discussed in detail [48, 49]. The current view of the former uncertainty favors nonaccumulating tetrahedral intermediates [70, 71]; that of the latter is more uncertain. Thus, variation in the geometries of the relay systems of chymotrypsin, subtilisin, trypsin and serine carboxypeptidase suggests that the aspartic acid can occupy a variety of positions relative to the Ser–His diad in different enzymes and that these perhaps should be regarded as two diads (Ser–His and His–Asp) that can operate concertedly [59, 72]. This might account for the results from proton inventory measurements suggesting that for the reaction of chymotrypsin with relatively nonspecific substrates containing few recognition sites, only one proton is "in flight" in the transition state, whereas with more specific substrates containing extended acyl groups, two protons are "in flight" [73]. A related possibility, disputed by Warshel *et al*. [74], is that a low-barrier hydrogen bond between His57 and Asp102 in the putative catalytic triad becomes stronger in the transition state, and might render His57 more effective as a base and Ser195 more nucleophilic by moving the pK_a values of their side-chains closer together [65, 75, 76]. This low-barrier hydrogen bond controversy is discussed by Wharton [49], who emphasizes the relevance in enzyme catalysis more generally for its rate enhancement role for which there is theoretical support and structural evidence [77].

The additional substantial insights into both substrate recognition and catalytic mechanism gained by the many protein engineering studies that have been carried out on serine proteinases are comprehensively documented [48]. An important advance to result from mutation of charged residues, which has implications for enzymes more generally, is the evidence that long-range electrostatic interactions can contribute to stabilization of transition states (e.g., [78–81]). A complementary approach to structural variation in enzyme structure is the use of natural variants. A striking example of this approach, discussed in Section 11.3, revealed a previously unrecognized key aspect of cysteine proteinase mechanism [82] (see also [44]). An example of the use of natural serine proteinase variants relates to the evaluation of the "oxyanion hole," the hydrogen-bonding systems that bind the anionic tetrahedral transition state/intermediate. In chymotrypsin both hydrogen bond donors are contributed by backbone peptide NH groups and thus are unable to be changed by site-directed mutagenesis. By contrast, in subtilisin, one hydrogen bond donor is the amide side-chain of Asn155 and mutation of this residue to those of various other amino acids results in large decreases in k_{cat}/K_M (e.g., [83]). Earlier crystallographic studies [84, 85] had suggested a weak hydrogen bond involving Asn155 in the Michaelis complex that becomes stronger in the transition state for acylation. The existence of the oxyanion hole in chymotrypsin (Gly193 and Ser195) proposed by Henderson [86] received support from crystal structures of combinations with transition state analogs [87] and from ^{13}C-NMR studies by the Malthouse group (e.g., [88–91]) on the enzymes derivatized by substrate-derived inhibitors. It is of particular interest that the oxyanion pK_a values of chymotrypsin and subtilisin in the tetrahedral adducts depend on binding interactions in the recognition sites of these enzymes [90].

Early work on substrate recognition by serine proteinases such as chymotrypsin and elastase demonstrated the existence of the now well-known enzyme subsites ($S_nS'_n$) within a surface cleft that bind complementary amino acid side-chains ($P_nP'_n$) on either side of the scissile bond of a polypeptide substrate and provided evidence for enzyme-transition-state complementarity and the utilization of binding energy. Some of the data from the 1970s compiled in Fersht [5] demonstrate that binding energy is used to increase k_{cat}/K_M as the size of the leaving group in peptide substrates for chymotrypsin increases and as the length of the peptide chain of substrates for elastase increases. It is noteworthy that the additional binding energy is used to increase k_{cat} rather than to lower K_M. References to additional compilations of the pioneering work on chymotrypsin, elastase, and α-lytic protease from the 1970s are also given in Fersht [5]. An account of the structural and kinetic basis of the diversity of substrate specificity in the chymotrypsin and subtilisin families is given in Perona and Craik [48], and is exemplified below.

The catalytic sites and substrate-binding clefts of the members of the subtilisin family are contained within a single domain [72, 92]. Crystal structures of the enzymes [92–97] and of various enzyme–inhibitor combinations identify the subsite structures of the active centers, Crystal structures of subtilisin–inhibitor complexes include those involving chloromethyl ketones [84, 98], the protein eglin C [72, 93], chymotrypsin inhibitor and subtilisin inhibitor [72, 93, 99, 100], and various peptide

inhibitors [101]. These structures demonstrate that, in general, subtilisins exhibit broad substrate specificity profiles, often with a preference for large hydrophobic groups at the P_1 position [102], with k_{cat}/K_M increasing with increasing hydrophobicity [22, 103]. Mutation of Gly166, which lies at the base of the S_1 pocket, shows that increasing the side-chain volume at this position results in decrease in k_{cat}/K_M as a result of the decrease in the available space in the S_1 pocket [104]. The other specificity determinant for substrates containing residues distant from the scissile bond has been identified for subtilisin BPN′ as the S_4 subsite [105]. The preference for S_4 decreases in the order Phe > Leu, Ile, Val > Ala for cleavage of peptidyl amide substrates [106].

For the chymotrypsin family crystal structures are available for bovine chymotrypsin [107], porcine pancreatic elastase [108], bovine, rat, and *Streptomyces griseus* trypsins [109–111], rat tonin [112], kallikrein [113], rat mast cell protease I1 [114], human neutrophil elastase [115], thrombin [116], factor Xa [117], and complement factor D [118]. Additionally, structures are available for four microbial enzymes: *Streptomyces griseus* proteases A, B and E [119–121], and the *Lysobacter enzymogenes* α-lytic protease [122]. For enzymes of this family, specificity is conferred mainly by interactions in the S_1 subsite. In enzymes with trypsin-like specificity for substrates containing arginine or lysine residues at P_1 there is a highly conserved aspartic acid residue at the bottom of the S_1 subsite [123]. In chymotrypsin and elastase class enzymes this residue is serine or a residue with a relatively small side-chain which results in specificity towards substrates with aromatic or small hydrophobic side-chains, respectively. Other important S_1 subsite residues at positions 189, 190, 216, and 226 extend into the base of the pocket and provide additional modulation of the specificity profile. Residues at positions 216 and 226 are usually glycine in both trypsin- and chymotrypsin-like enzymes; larger side-chains at these positions would partially or fully block access of large substrate side-chains to the base of the pocket. Elastases possess larger, usually nonpolar, residues at these positions, which provide specificity for small hydrophobic residues at P_1.

Evaluation of recognition characteristics by kinetic analysis (e.g., [49]) makes use of both well-known conventional chromogenic and fluorogenic substrates [124, 125], and more sophisticated approaches, exemplified below. For proteinases without an identified protein substrate, libraries can be used to identify potential substrates; for example, a fluorescent peptide-based library has been used to identify substrates for endopeptidases [126]. Fluorescence resonance transfer methodology has been used [127] to allow the use of physiological substrates (i.e., relatively large proteins) that contain all of the potential molecular recognition sites often lacking in smaller peptide substrates. Cleavage of the protein substrate results in the generation of a product with an epitope that is recognized by a europium-labeled specific antibody. Other approaches use mass spectrometric or high-performance liquid chromatography analysis [128], or calorimetry when catalytic reactions have relatively large heat changes [129]. Although investigation of substrate recognition by proteinases using peptide substrates has been successful in defining their primary subsites, these other approaches can detect other enzyme–substrate interactions that may be of unidentified physiological significance. An example of a serine proteinase whose

physiological function has still not been fully elucidated is proprotein convertase subtilisin kexin 9 (PCSK9). This protein is implicated in the control of concentrations of low-density lipoprotein (LDL) in serum [130]. It is highly expressed in hepatocytes as a zymogen where it undergoes processing which results in cleavage of its pro-domain, but without loss of its high-affinity binding capability [131]. The mechanism by which PCSK9 produces a decrease in blood LDL concentrations is not fully understood, but there is evidence that it acts as a chaperone and directs the intracellular trafficking of the LDL receptor [132].

11.3 Cysteine Proteinases

The enzymes contained in the latex of the tropical fruit tree, *Carica papaya*, occupy a central place in the development of studies on cysteine proteinases, and on some aspects of enzymology in general and on substrate recognition in particular. The suggestion by Wurtz in 1880 that in fibrin digestion, papaïne (as it was then known) acts by becoming bound to fibrin, preceded the proposal by Emil Fischer in 1894 than an enzyme acts on a substrate by first associating with it in a specific manner. A comprehensive review of the older literature on cysteine proteinases [133] includes an account of the historical development of studies on these enzymes. According to Vines [134], the first record of the digestive action of papaya latex is contained in Griffith Hughes's *Natural History of Barbados* published in 1750. The importance of the latex as a source of enzymes appears to have been first recognized by G.C. Roy, who published his report in 1873 in the *Calcutta Medical Journal*, and the results of the first well-controlled scientific investigation of the partially purified cysteine proteinases were published by Wurtz and Bouchet in 1879 [135].

The cysteine proteinases are now known to constitute a large superfamily of six families of thiol-dependent endopeptidases [136, 137], which have been isolated from animals, plants, bacteria, and viruses. Most of the well-characterized and intensively studied cysteine proteinases are members of the papain family [136]. Isolation procedures include those for papain itself [138], chymopapains A and B [139], chymopapain M (glycyl endopeptidase, papaya proteinase IV, and proteinase Gly-C) [140], and caricain (*Papaya proteinase* Ω) [141], all from the latex of *C. papaya*, and other variants that have proved useful in studies on kinetic specificity and mechanism, such as actinidin from *Actinidia chinenisis* [142], ficin from *Ficus glabrata* [143], cathepsins B, H and L [144], and the hexameric enzyme, bleomycin hydrolase [145]. Production of the fully active enzymes by covalent chromatography by thiol-disulfide interchange, introduced for papain [146, 147], is reviewed in [148]. A particular value of papain itself and some other members of the family relates to the possibility of using a combination of specially designed thiol-specific reactivity probes [41], the variation in kinetic characteristics within the family [20, 44], and computer modeling, to investigate active center chemistry and its dependence on electrostatic effects and key binding interactions involved in dynamic aspects of substrate recognition.

The minimal mechanism relates to a three-step acyl-enzyme model, but differs from that of the serine proteinases (Section 11.2) in using a catalytic Cys–His–Asn triad rather than Ser–His–Asp. There is considerable interest also in some of the members of other families such as the human rhinovirus-14 3C proteinase [149] and the human adenovirus proteinase [150], both of which contain Cys–His–Glu catalytic triads, and interleukin-1β-converting enzyme [151], which contains Ser236 instead of a carboxylate analogous to Asp158 of papain and other differences in the electrostatic fields to which the catalytic sites are exposed [81].

Papain (reviewed in [20, 44, 133, 152]) has long been considered the cysteine proteinase archetype. There is considerable variation among members of the papain family, however, in substrate specificity and catalytic-site reactivity [81, 153–158], which reveals that this view is correct only for low-resolution aspects of the mechanism, such as the roles of the cysteine and histidine components of the catalytic-site ion pair. The results of detailed mechanistic investigations on papain and some of its natural variants (discussed in [44, 45]) have revealed that papain exhibits characteristics at one end of a spectrum of chemical behavior with actinidin at the other, and with caricain and ficin occupying intermediate positions. Key results include the identification of kinetically influential pK_a values, both those associated with the catalytic site (Cys25)-S^-/(His159)-Im^+H ion pair (papain numbering) and others more remote [44], the coupling of molecular recognition interactions with catalytic site chemistry, and postacylation protein dynamics.

Studies using reactivity probes, particularly disulfides containing the 2-mercaptopyridine leaving group as two-protonic-state electrophiles [41, 81], including substrate-derived probes [20, 44], a 4-pyrimidyl probe that remains unprotonated over a wide range of low pH values [44] and 4-chloro-7-nitrobenzofurazan [159], pH-dependent kinetics [42], computer modeling including normal mode analysis [45] and electrostatic potential calculations [81], and natural variants of the papain family, have changed the perception of the catalytic mechanisms of these enzymes and of the influence of binding interactions and electrostatic effects (see [20, 44] for detailed discussion).

A long-standing, generally assumed, concept of central importance had been that catalytic competence reflected in k_{cat}/K_M develops synchronously with and consequent on formation of the Cys/His ion pair state by protonic dissociation across pK_a 4. The more recent studies, however, established that:

(i) The Cys/His ion pair, although essential, is not sufficient for catalytic competence in that it exists at low pH (<4) as an inactive "intimate" ion pair and another pH-dependent event associated with pK_a 4 is required to endow it with catalytic ability [81, 140, 160]; this cannot be assigned to Asp158 because its pK_a value is now known to be 2.8 in papain and 2.0 in caricain [44, 159], and reinterpretation of kinetic studies on catalysis by papain mutants is required [161].

(ii) Perturbation of the mutual solvation of the ion pair components, required to release the nucleophilic character in Cys25 as the initial catalytic event, appears to involve decrease in the extent of hydrophobic shielding by movement of the active center cleft around the Trp177 region [45].

(iii) At least for actinidin and ficin, there are additional kinetically significant conformational changes, detected by pre-steady-state kinetic studies using *N*-Ac-Phe-Gly-methylthionoester, required to allow the release of methanol (P_1) from the $ES'{\cdot}P_1$ complex and its replacement by water for deacylation [162, 163].

(iv) Transition-state geometries are determined by a combination of specific binding interactions related to the molecular recognition process discussed below and electrostatic effects.

A review of the older literature on cysteine proteinases, published in 1987 [133], includes an account of their primary and secondary specificity characteristics [164] (updated in [20]). The specificity characteristics of papain were studied extensively by Schechter and Berger [165–167] using diastereoisomeric peptides of alanine from Ala2 to Ala6. They concluded that the papain active center comprises seven subsites, each capable of accommodating a single amino acid residue of substrate. This classic study gave rise to the now widely used "P" and "P′" notation for oligomeric substrates and the related "S" and "S′" notation for the assumed complementary enzyme subsites, although the possibility of a continuum of flexibility needs to be kept in mind [164]. The P_1 and P'_1 residues contribute the carbonyl and amino groups respectively of the scissile $P_1P'_1$ bond. In addition to the approach used by Schechter and Berger, detailed indications of possibilities for interaction with substrate-derived covalent inhibitors such as chloromethylketone derivatives of oligopeptides have been provided by X-ray crystallography [168, 169]. These possibilities include those in binding areas relatively remote from catalytic sites. The crystallographic data for papain [168] permit the construction of models of the acyl-enzyme and tetrahedral intermediates – conclusions supported by the conformations of thionoester substrates bound to papain deduced by resonance Raman spectroscopy combined with crystallography [170, 171].

By contrast with chymotrypsin, in which substrate recognition depends on the hydrophobic S_1 subsite, the major recognition site in papain (and many other cysteine proteinases) is the hydrophobic pocket at S_2 comprising, in papain, mainly Trp67, Pro68, Trp69, Phe207, Val133, and Val157. Although P_2/S_2 contacts confer a strong preference for substrates with bulky nonpolar side-chains at this position, this can be modulated by requirements at other subsites leading to a wide range of cleavage points in different substrates. The interdependence of some of these papain–substrate and analogous papain–inhibitor interactions is demonstrated by their incremental specificity energies [158, 172, 173]. The claim that papain exerts absolute stereospecificity for L-amino acid residues at the S_2 subsite [152, 167, 174] was demonstrated subsequently to be incorrect [158, 173, 175, 176]. Thus, this subsite can bind both L- and D-phenylalanine residues, at least in *N*-Ac-Phe derivatives, in accord with the early literature on papain catalyzed synthesis of anilides and phenylhydrazides (see references cited in [175]). The incorrect conclusion that strict stereospecificity is exerted by the S_2 subsite of papain derives from the model of the enzyme–substrate complex assumed by Berger and Schechter [167] and an incorrect assumption about binding modes made by Lowe and Yuthavong [174] before the crystallographic data were available (see [175] for a detailed analysis). As discussed in

Section 11.1, differences in specificity may be detected by comparison of values of k_{cat}/K_M. Whereas papain and actinidin have similar values of the specificity constant for substrates with aliphatic *N*-acyl groups, the values for substrates with aromatic substituents approximately in the P_2 position are 10–110 times smaller for actinidin than for papain [177, 178]. This difference in specificity derives from a key structural difference between the S_2 subsites. In papain, Ser205 is located at the bottom of the hydrophobic pocket [179], whereas in actinidin the analogous residue is Met211 whose side-chain extends across the pocket making it shorter [180].

Most members of the papain family exert primary specificity for hydrophobic residues at P_2 and in some cases such as cathepsin B [181] at P_3, with relatively little recognition of P_1 side-chains [133], although there are exceptions [20]. One such exception is chymopapain M, the unusual enzyme eluted by ion exchange as the last component of the chymopapain band, which exerts high specificity for glycine at P_1. The structural origin of this S_1 subsite specificity and the lack of inhibition of chymopapain M by cystatins were revealed by the homology-modeled structure [140, 182] and confirmed by the crystal structure refined to 2.1 Å [183]. Chymopapain M has glutamic acid at position 23 and arginine at position 65, which form a barrier across the S_1 subsite, as against glycine at each of these positions in papain. Other noteworthy exceptions to the otherwise common lack of substantial P_1–S_1 specificity are cathepsins B, H, and L [184–186], clostripain [187], and α-gingivain [188].

The leaving group of the substrate is bound in the S'_1 subsite, which in papain exhibits a preference for hydrophobic side-chains, particularly those of isoleucine and tryptophan [189]. This probably derives from contact with Trp177, which as our recent reactivity probe studies [45] suggest, may have an important role in perturbation of the hydrophobic shielding of the Cys25/His159 ion pair required to generate nucleophilic character in its Cys25 component. A different specificity might be predicted for bromelain [190] in which Trp177 is replaced by lysine. More recent studies on S'_1 specificity involve the use of intramolecularly quenched fluorogenic substrates [191] and of the crystal structures of recombinant human stefin B complexed with carboxymethylated papain [192]. The first and second β-hairpin loops of the inhibitor bind to the S'_1 subsite and the conserved Glu-Val-Val-Ala-Gly region of the first loop is a prime target for protein engineering studies. A possible role for Asp158 (*cf.* [45]) might be to prevent carboxypeptidase activity by papain by repulsion of the terminal carboxylate group of the substrate [18]. This would be in accord with the dipeptidylcarboxypeptidase activity of cathepsin B [193], in which Asp158 is replaced by glycine.

Substrate recognition and catalytic mechanism are often inextricably linked (see Section 11.1), and studies on papain using variation in structure of both substrates and thiol-specific two-protonic-state reactivity probes [41] have provided clear illustrations of how this linkage can be manifested. Lowe [152] discussed the kinetic specificity of papain and demonstrated that it is exerted essentially in the acylation step of the catalytic act [174]. Thus in the catalyzed hydrolysis of esters and anilides of *N*-Ac-Gly, *N*-benzoyl-Gly and *N*-Ac-Phe-Gly, the specificity for the phenylalanine residue is exerted in the acylation step and essentially not at all in deacylation. In

acylation, relevant ratios of k_{cat}/K_M are greater than 1×10^7, whereas analogous ratios of deacylation rate constants are less than 10. Papain is able to bind the P_1–P_2 amide bond of a substrate by hydrogen bonding of its NH group with the backbone carbonyl oxygen of Asp158 and of the carbonyl oxygen of the amide with the NH group of Gly66. In this way the substrate should be able to bridge the two walls of the active center cleft and possibly transmit effects of P_2/S_2 hydrophobic contacts to the catalytic site.

This sort of binding site-catalytic communication, together with conformational changes initiated by electrostatic effects [45, 81], results in the modulation of the "intimate" (Cys25)-S^-/(His159)-Im^+H ion pair to allow each of its partners to optimize its role as nucleophile and hydrogen bond (and proton) donor, respectively. Clear evidence for this phenomenon [141, 155–157, 173–197] is provided by the nature of the pH–*k* profiles for reactions of papain with substrate-derived 2-pyridyl disulfides used as thiol-specific, two-protonic-state reactivity probes [41]. These probes contain in one part of the molecule various potential binding sites and, in the other, the 2-mercaptopyridine leaving group to serve as a detector of the relative disposition of the thiolate and imidazolium components of the Cys25/His159 ion pair. Whereas the pH–*k* profile for the reaction of papain with the "featureless" probe, *n*-propyl 2-pyridyl disulfide, has a rate constant minimum at pH 6.5–7.0 (with $k \sim 1000$ l/mol/s) and a rate constant maximum at pH 4 (with $k \sim 1.5 \times 10^4$ l/mol/s), the corresponding reaction with the substrate-derived probe containing both a P_1–P_2 amide bond and a Phe residue at P_2 as a potential occupant of the S_2 subsite, *N*-Ac-[Phe]-NH-$[CH_2]_2$-S-S-2-Py, has a rate constant maximum at pH 6.5, as occurs with k_{cat}/K_M for catalysis (with $k \sim 5 \times 10^6$ l/mol/s). By using structural variants of the complete substrate derived probe (e.g., containing the P_1–P_2 amide bond, but not P_2–Phe) the corresponding P_1–P_2 ester, and the P_1–P_2 ester analog of the complete substrate-derived probe, it was established that the change in transition-state geometry that provides association of the imidazolium component of the ion pair with the leaving group rely on the $(P_1)-NH\cdots O{=}C<$ Asp158 hydrogen bond and the hydrophobic P_2/S_2 contacts. The rate constant maxima at pH 3–4 observed with the *n*-propyl and P_1–P_2 ester probes derive from the coexistence of significant concentrations of ion-pair and protonated probe, R-S-S-2-Py^+H. By contrast, the rate constant maximum at pH 6–7, observed with the highest reactivity for the complete substrate-derived probe, results from assistance to reaction of the unprotonated probe by hydrogen bond donation from the imidazolium component of the conformationally perturbed ion pair to the 2-mercaptopyridine leaving group. The lack of analogous rate constant maxima in reactions with the corresponding substrate-derived 4-pyridyl disulfides such as *N*-Ac-[Phe]-NH-$[CH_2]_2$-S-S-4-Py [198] in which the pyridyl N atom is remote from the ion pair Im^+H group provides evidence for this mechanism.

Non-signaling binding interactions can be distinguished from signaling binding interactions by detailed analysis of the family of pH–*k* profiles [157]. A central role in the binding-site catalytic-site communication in papain is provided by the hydrogen bond between the carbonyl component of the P_1–P_2 amide bond and Gly66. This needs to be supported by a neighboring interaction to provide the imidazolium-

assisted interaction and either the NH component of the same amide bond or a hydrophobic P_2 residue can serve this purpose. In the latter circumstance, the hydrogen bonding of the amide NH to Asp158 does not further assist the change in transition-state geometry, even though it improves the binding.

Temperature dependence studies suggest that key binding features of the reactivity probes produce displacement of parts of the ordered water network of the papain active center such that the transition state of the modification reaction is less ordered than the reactant state. Thus, for the reaction of papain with the "featureless" probe, CH_3-S-S-2-Py, $\Delta S^{\ddagger}$ is large and negative (-78 ± 3 J/mol/K), whereas when the probe is provided with the opportunity for trans-cleft hydrogen bonding (*N*-Ac-NH-$[CH_2]_2$-S-S-2-Py), the value of $\Delta S^{\ddagger}$ decreases (-59 ± 2 J/mol/K), and when, additionally, a potential occupant for the S_2 subsite is present, as in *N*-Ac-[Phe]-NH-$[CH_2]_2$-S-S-2-Py, $\Delta S^{\ddagger}$ becomes large and positive ($+122 \pm 6$ J/mol/K) [43].

11.4 Glycoside Hydrolases

Glycoside hydrolases (glycosidases) are a large, well-characterized group of glycosyl transfer enzymes. Glycosyl transfer more generally is reviewed extensively with emphasis on details of mechanism by Davies *et al.* [199]. Enzyme-catalyzed glycosyl transfer exists from and to a variety of nucleophiles. Enzyme nucleophiles are usually carboxylates, which react to produce relatively unstable intermediates [200]. The general reaction is nucleophilic substitution at the saturated anomeric carbon atom. It involves either retention of configuration, (r), by double displacement involving an enzyme or substrate nucleophile or inversion of configuration, (i), by a single chemical step, hydrolysis (i.e., O $\rightarrow$ O transfer) in the case of glycosidases. Important stereochemical factors involve the ring size of the sugar (five-membered, furanose, or six-membered, pyranose) and the configuration of the leaving group (axial, a, or equatorial, e) in its preferred conformation; see [199, 201, 202] for the designation of pyranosyl-transferring enzymes as e $\rightarrow$ e, a $\rightarrow$ a, e $\rightarrow$ a, and a $\rightarrow$ e, and for that of furanosyl-transferring enzymes as f(r) and f(i). Davies *et al.* [199] suggest that this type of designation is preferable to the commonly used α- and β-glycosidses because these are not always equivalent to "leaving group equatorial" and "leaving group axial," respectively. The standard mechanisms common to many glycosidases, despite the existence of many different protein folds, exemplified in [199], involve two carboxy groups contributed by the enzyme (aspartic acid or glutamic acid) [203]. In the double-displacement mechanism, one provides general acid–base catalysis and the other nucleophilic catalysis, whereas in the single-displacement mechanism, one provides general acid catalysis and the other general base catalysis. In the latter mechanism, the carboxy groups are separated typically by 9–10 Å, which allows the bound water molecule to attack on the side of the sugar ring opposite to the glycosidic bond [204, 205]. By contrast, in the double-displacement mechanism, which provides overall retention of anomeric configuration and involves formation of a covalent glycosyl-enzyme intermediate that is subsequently hydrolyzed via an oxocarbenium

cation ($-^{+}O{=}C$) transition state, the catalytic carboxy groups are typically separated by about 5.5 Å. The carboxy group that acts as the general acid in the formation of the intermediate acts as a base in its subsequent hydrolysis, with possible perturbation of its pK_a value [206]. The geometries of the oxocarbenium transition states consistent with the stereoelectronic requirement discussed by Davies *et al.* [199] are the envelope conformation of a five-membered ring, and the half-chair and classical boat conformations of a six-membered ring. Evidence from conformational analysis of aldonolactone analogs of pyranosyl oxocarbenium cations suggests that the two favored conformations should be of comparable energy [207].

Studies on glycosidases have provided a number of important contributions to understanding substrate recognition, described below, notably the development of enzyme crystallography, the dominance of stereochemistry in enzyme–carbohydrate recognition, discrimination between *endo* and *exo* catalytic action on polymeric substrates, and evidence of substrate distortion.

Glycoside hydrolases occupy a special place in the history of enzymology in that the first crystal structure of an enzyme to be determined was that of hen egg-white lysozyme in 1965 [208] and this proved to be of central importance in the elucidation of a novel catalytic mechanism. In view of the considerable structural diversity among naturally occurring carbohydrates, it is not surprising that a similar extent of diversity has been found to exist also in the range of enzymes that catalyze their hydrolysis [203]. Currently 112 sequence-based families are known to exist [209] and crystal structures of representatives of 48 of these (tabulated in [199]; http://www.cazy.org) have been reported. Many glycosidases have a multidomain architecture with a catalytic domain linked to other domains, some of which may provide binding sites for extended polysaccharide substrates. Most of the crystal structures are of isolated catalytic domains, although there are examples of multidomain structures where the domains are linked by relatively rigid sequences that facilitate crystallization [210, 211]. The classification of glycosidases based on amino acid similarities of catalytic domains ([209], reviewed in [212]) reveals that many of the families contain enzymes with many different substrate specificities – a phenomenon that provides valuable opportunities in the study of recognition. The increase in the number of crystal structures of glycosidases allows classification in terms of structure, which is more highly conserved than sequence, and facilitates homology modeling. Currently, structural classification makes use of the clan concept (similar three-dimensional structures, conserved catalytic groups and stereochemistry, and similar substrate configuration) [209, 212], by far the largest clan, GH-A, being characterized by a classic (α/β) eight barrel. The catalytic acid–base and nucleophilic groups are on β-strands 4 and 7, and act on equatorially linked sugars by the double-displacement mechanism [213] to provide overall retention of configuration [209, 212, 214, 215].

Examples of variation from the standard double-displacement glycosidase mechanism revealed by X-ray crystallography are those of myrosinase from white mustard (*Sinapin alba*) and LacZ β-galactosidase from *Escherichia coli*. In myosinase, the nucleophile is shown to be contributed by Glu426, but the residue expected to provide acid–base catalysis is Gln187 [216]. Since the leaving group of glucosinolate substrates such as sinigrin is a mercaptan, general acid assistance to

its departure is not required and hydrolysis of the intermediate might be assisted by hydrogen bonding of the water molecule to the carbonyl oxygen of Gln187 if the glycosyl ring became distorted. The crystal structure of the high M_r tetrameric β-galactosidase [210] is consistent with a double-displacement mechanism with the nucleophilic carboxylate being provided by Glu537, but with a magnesium cation coordinated to Glu416, Glu461, and His418 acting as an electrophilic catalyst instead of an aspartic acid or glutamic acid as a general acid. This mechanism, however, is controversial (see [199, 202] for detailed discussion of a range of key experimental evidence). One interesting point is that Glu416 is fully conserved in this enzyme family and most members, including another enzyme for which a crystal structure is available [217], are not dependent on metal ions for catalytic activity.

A potentially most important variation is seen in the sialidases, whose mechanism has always been a puzzle; they act with retention, but do not posses an aspartate or glutamate in the correct position to act as nucleophile in a double-displacement mechanism. Interest in sialidase, also known as neuraminidase, is intense as it is the influenza enzyme that cleaves the terminal sialic acid from the other sugars releasing the newly budded viral particles (e.g., in the avian influenza virus H5N1). The most successful antivirals. including Tamiflu and Relenza. are neuraminidase inhibitors. A partial solution to this puzzle comes from elegant crystallography and chemistry showing that the mechanism of trans-sialidase from the pathogen *Trypanosomi cruzi* involves the formation and breakdown of a tyrosyl-enzyme intermediate [218, 219]. There is also evidence that the binding of sialic acid triggers a change in the position of two tyrosine residues; the active-site tyrosine moves into position to act as nucleophile and a more distal tyrosine moves to enhance the affinity of enzyme/donor complex for the sugar acceptor [218].

A recent mechanistic surprise was the discovery of a novel NAD^+-dependent glycosidase mechanism of the GH4 family glycosidases that harness NAD^+ as a transient redox catalyst [220]. Substrate distortion to the reducing side of the cleaved bond is a key feature of the mechanism of glycoside hydrolases that catalyze cleavage of the (-linked sugars by retention (i.e., e → e) which includes lysozyme and the GH-A clan enzymes. The reason for the distortion is simple – it brings the leaving group from an equatorial to an axial orientation approximately normal to, and above, the plane of the sugar ring, allowing the nucleophile to attack from below in-line with the bond to be cleaved. This distortion of the substrate from its ground state chair ($^4C_1$) conformation to envelope (^4E), skew-boat ($^1S_3$), or boat ($^{1,4}B$) conformer has been revealed by crystal structures of Michaelis complexes. It is consistent with the dictates of the antiperiplanar lone-pair hypothesis in which the leaving group is encouraged to leave by the lone pair of the ring oxygen being down and the leaving group up. Where does the energy to distort the sugar residue binding to the reducing side of the cleaved bond come from? This is contributed by favorable binding interactions more remote from the active center. Nucleophilic attack generates the undistorted glycosyl-enzyme intermediate that is both preceded and succeeded on the reaction trajectory by oxocarbenium-ion-like transition states [221]. Distortion in these enzymes is a consequence of the chemistry of the S_N2 reaction. The precise

distortion that occurs is a result of subtle changes in the recognition site of the enzyme and the chemistry of the sugar [222].

11.5 Protein Kinases

The mechanism of phosphoryl transfer is discussed in Fersht [5] and reviewed in Hengge [223]. This section deals with substrate recognition by protein kinases. The eukaryotic kinases contain a catalytic region consisting of approximately 250 amino acid residues comprising an N-terminal domain of β-sheets and a C-terminal domain of α-helices [224–235]. The phosphoryl donor, ATP, binds in a cleft between the two domains such that the hydrophobic pocket is able to accommodate the adenosine moiety with the phosphate backbone oriented towards the enzyme surface. The protein substrate binds along the cleft and conserved residues within the catalytic region catalyze the transfer of the terminal γ-phosphoryl group of ATP to the hydroxy oxygen of a serine, threonine, or tyrosine residue of the substrate. There are two classes of protein kinases: the serine/threonine- and tyrosine-selective (although not absolutely specific) kinases; the tyrosine kinases possess a deeper catalytic cleft needed to accommodate the larger side-chain. Some protein kinases exhibit high selectivity in site phosphorylation, whereas others can produce multiple phosphorylations in a variety of substrates. Mapping of the interactions of protein kinases with their substrates [236] has been carried out using kinetics of catalysis and X-ray crystallography. Free amino acids are poor substrates, the minimal requirement being for short peptides with four residues on either side of the phosphorylation site [237, 238], both sets of residues usually contributing substantially to substrate recognition. Protein kinases recognize consensus sequences [239–243]; examples include protein kinase A (R-R-X-S/T-hydrophobic residue) and cyclin-dependent kinase (CDK; S/T-P-X-K/R) [242], extracellular-regulated kinase-2 (P-X-S/T-P) [244, 245], glycogen synthase kinase-3 (GSK3; S-X-X-X-pS) [244, 246], epidermal growth factor receptor (E-E-E-Y-F) and insulin receptor tyrosine kinase (Y-M-M-M) [244, 247], protein kinase B (R-X-R-X-X-S/T) [248], and protein kinase D (L/I-X-R-X-X-S/T) [243]. Despite the existence of these consensus sequences, physiological protein substrates may experience substantially different specificities contributed by interactions remote from the active center of the kinase [249]. As a result, kinase activity measured using peptides based on limited sequences of their physiological substrates may overlook these aspects of substrate recognition. Docking sites that vary in composition and physical separation from the kinase active center have been identified in a number of kinases, including phosphorylase kinase [235], GSK3 [250], phosphoinositide-dependent kinase-1 [251], and CDK [252, 253].

Substrate recognition by kinases is known to be influenced also by specific protein binding partners that provide versatility in the targeting of different specific substrates. For example, cell cycle regulatory CDKs can form complexes with several different cyclin proteins; the cyclins allosterically activate the CDK and, in addition, contain a docking domain that facilitates the recruitment of appropriate substrates.

The docking domain is a small hydrophobic patch located on the cyclin partner about 40 Å away from the active center of the CDK [233, 254, 255] and interacts with an RXL motif of specific substrates [253, 256]. Progression through the cycle is driven by the sequential activities of different CDKs that require association with their respective cyclins and phosphorylation for activation; for example, CDK2 in association with cyclin A drives cells through the S phase, phosphorylating a wide range of substrates that include the tumor suppressor proteins and transcription factors. Despite this change in specificity of the kinase during each stage of the cycle, they maintain a consensus recognition sequence containing the RXL or KXL motif. Structural studies with the p27 inhibitor [233] and a peptide substrate from p107 [257] have shown that the RXL recognition site is located on the cyclin A molecule and involves recognition by residues located mostly on the helix H1 of the cyclin. These provide an extensive, exposed, non-polar site, located about 40 Å from the catalytic site that is conserved in cyclins A, B, D, and E. Studies on the p107 substrate of CDK2/cyclin A have shown that one of the roles of the RXL motif is to convert an otherwise poor substrate into a good substrate [258]. The function of the RXL motif can be explained either by the fact that it effectively localizes the kinase/cyclin complex in proximity to its substrate, and therefore serves to increase the local substrate concentration [253], or that the catalytic site and RXL motif are inextricably linked [259]. In addition, RXL-containing peptides effectively inhibit CDK activity, presumably by blocking access of the natural substrate to the cyclin recruitment site [260] and this has been exploited in the designed of allosteric inhibitors of the kinase. The sequences around the RXL motif of a number of key substrates reveal variation in these sequences and in the position of the RXL motif with respect to the phosphorylation site. Most RXL motifs are on the C-terminal side of this site, with a minimum separation observed to date of about 20 residues. A combination of X-ray crystallography and isothermal calorimetry has been used to investigate recognition features and the factors that govern affinity using the phospho-CDK/cyclin A system with a peptide from E2F as substrate. The results show that the cyclin A recruitment site is dominated by recognition of the leucine residue of RXL, which docks into a nonpolar pocket on the surface of the cyclin [261]. The recruitment site is able to recognize diverse but conformationally constrained target sequences. These observations have implications for the further identification of physiological substrates of the CDK/cyclin A system and the design of specific inhibitors [262].

Given recent advances in recombinant protein production technologies and the commercial availability of tools such as phospho-specific antibodies for the phosphorylation sites in full-length protein substrates, there is a growing number of studies on kinase activity towards such substrates, and it is noteworthy that many differences between the results obtained using peptide and protein substrates are being revealed [263]. Myelin basic protein, which is a relatively small highly basic protein, has often been used as a substrate for kinases. Assays are usually carried out using [γ-^{32}P]ATP and the rate of incorporation of its terminal radiolabeled phosphoryl group into substrate is used to determine site specificity [264]. This substrate, although still commonly used for protein kinase assays [265–267], has been replaced by assays that use peptides based on the protein kinase physiological target (i.e., the

physiological substrate itself or truncated forms). Such studies have already provided valuable advances in understanding substrate recognition in the protein kinase family of enzymes [228, 236, 263].

11.6 aaRSs

The aaRSs, members of a heterogeneous family of enzymes that catalyze the covalent attachment of amino acids to the 3′-ends of their cognate tRNAs, via a tightly bound aminoacyladenylate [268], are extensively reviewed [269] and discussed succinctly in Fersht [5]. The family is composed of two classes, distinguished by their three-dimensional structures including different specific motifs and functionally by the initial site of aminoacylation on the ribose ring. The particularly high specificity that is essential in protein biosynthesis means that these enzymes have proved key to understanding discrimination with high precision between very similar substrates. Situations discussed in Fersht [5] include: (i) the expected steric exclusion of a larger substrate by an enzyme with a binding site for a smaller substrate, (ii) the discrimination against a substrate that is smaller than or isosteric with the specific substrate due to the absence of a feature that contributes to the binding energy, and (iii) the discrimination against an isosteric substrate that contains a substituent chemically inconsistent with the nature of the binding site. An example of (i) is the fact that the rate of activation of isoleucine by valyl-tRNA synthetase is 2×10^5 times lower than that for its reaction with the substrate that is smaller by one methylene group, valine [270]. Examples of (ii) are the weaker binding of valine to isoleucyl-tRNA synthetase by a factor of 150 [271] and the discrimination against glycine by alanyl-tRNA synthetase by a factor of 250 [272]. An example of (iii) is the weaker binding of threonine by valyl-tRNA synthetase by a factor of 100–200 due to the replacement of one of the methyl groups of valine by the isosteric hydroxyl group as a reluctant occupant for the hydrophobic pocket of the enzyme [273].

tRNA aminoacylation involves two stages: activation of the amino acid substrate by ATP to form the enzyme–aminoacyladenylate complex via a pentacoordinate transition state, followed by transfer of the aminoacyl group (directly with class II enzymes and indirectly with class I enzymes) to the 3′-hydroxy group of the terminal ribose of the tRNA. Specificity is achieved at each stage of the reaction [274, 275]. In stage 1, differential binding affinities discriminate to some extent between the activation of cognate and noncognate amino acids [276, 277]. The accuracy of amino acid recognition, however, is seriously limited by the insufficiency of steric exclusion as a mechanism of specificity. The necessary accuracy is achieved only by the existence of the phenomenon of editing – proofreading [5]. This was identified initially in connection with studies on the discrimination in favor of isoleucine over valine in protein biosynthesis, which is much more accurate than predicted from the discrimination by isoleucyl-tRNA synthetase [278, 279]. In general terms, editing increases the accuracy of the specific aminoacylation of tRNA by the hydrolytic destruction of reaction intermediates in the reactions of noncognate amino acids [5, 280, 281]. The

energetic cost of the editing process derives from the decreased rate of tRNA aminoacylation as a whole [282, 283]. Fersht [5] discusses the cost and benefit of editing in both DNA replication and protein biosynthesis, the possibility of a double-check editing mechanism [284, 285], and the "double-sieve" analogy [270] in connection with the amino acid selection process. Editing involves the destruction of the noncognate amino acid adenylate before the transfer, in a slow step, of its aminoacyl moiety to tRNA, with most of the small fraction that is transferred dealt with by the esterase action of the enzyme. In the double-sieve analogy, amino acids more sterically demanding than the cognate substrate are rejected by the "coarse sieve" and the smaller or isosteric amino acids are activated but more slowly as a result of weaker binding. The "fine sieve" then isolates selectively the products of the reaction of the cognate substrate either because of their size or because of the lack of specific binding interactions that permit an isostere to bind in the hydrolytic site of the enzyme. The molecular mechanism of the double sieve is illustrated by the structural studies on the isoleucyl-tRNA synthetase from *Thermus thermophilus* [286, 287].

Recognition of the tRNA substrate is extensively discussed [269]. By contrast with the specific recognition of the cognate amino acid substrates, which is sensitive to small differences in their side-chains, the binding of tRNA is determined mainly by the relatively nonselective electrostatic interactions involving the essentially common anionic sugar phosphate backbone of the tRNA and cationic side-chains of the aaRS over a large surface area (e.g., [288, 289]). This suggests that binding affinity would not be significantly affected by changes in a small number of nucleotides in the tRNA molecule. As a consequence, discrimination between cognate and noncognate tRNA substrates is determined by differential kinetic characteristics rather than by the differential binding mechanisms that apply to the amino acid substrates (e.g., [290, 291]). Thus, recognition of the tRNA substrate involves those specific interactions between the enzyme and its cognate tRNA substrate that differentially stabilize the transition state for the transfer stage of the overall reaction [292]. An important aspect of the coupling of tRNA recognition with catalysis by aaRSs involves interactions between different domains of the enzyme. Interdomain communication involves interactions both between tRNA recognition domains, and between recognition domains and the catalytic site [276, 292–295]. Communication occurs in both class I and class II enzymes although the structural mechanism differs among members of the family (compare, for example, the mechanisms of the class I enzyme, methionyl-tRNA synthetase [296], and the class II enzyme, seryl-tRNA synthetase [276]).

The large number of experimental investigations that have been carried out on tyrosyl-tRNA synthetase (reviewed in [269]) makes this enzyme the obvious choice from which to gain a detailed understanding of all aspects of the catalytic mechanism of the amino acid activation stage of tRNA aminoacylation by the class I synthetases. This enzyme, like lysozyme (see Section 11.4), has a special place in the development of enzymology. These two enzymes provide an interesting contrast [5]. The latter revealed the basis of a previously unknown catalytic mechanism when its crystal structure, the first enzyme structure to be reported (in 1965), was determined [208]. By contrast, the crystal structure of tyrosyl-tRNA synthetase, published in 1982 [297],

provided no insights into its catalytic mechanism. Insights were revealed only after the initial pioneering investigation by protein engineering was published, also in 1982 [298]. Subsequently, these insights were extended and refined by information from a combination of crystal structures of complexes relevant to various stages of the mechanism [297, 299, 300], investigation of the reactions of intermediate states by pre-steady-state kinetics [298, 301–306], and identification and characterization of the mechanistic roles of the side-chains of key amino acid residues by kinetic analysis of a range of site-specific mutants [307–311]. The combination of steady-state and pre-steady-state kinetics, using both wild-type enzyme and various mutants, permits the construction of the difference free energy diagrams [306, 307] that facilitate identification of the mechanistic roles of the side-chains selected for mutation (see [5]). The contributions that studies on tyrosyl-tRNA synthetase have made to provision of experimental evidence relating to the understanding of the fundamentals of enzyme catalysis (see [5]) include demonstration of enzyme-transition state and enzyme-intermediate complementarity [17, 305], detection of an induced-fit process [312] and the concept of compromise between enzyme-transition state complementarity and access of substrate to the active center [311].

It is noteworthy that the enzyme–substrate interactions involved in the binding of tyrosine by tyrosyl-tRNA synthetase observed by crystallography correspond well with those deduced by site-directed mutagenesis and pre-steady-state kinetics [309]. High specificity is achieved by avoiding nonsolvated charged side-chains and unfavorable steric clashes in the active center region of the enzyme [313]. An example of high specificity exerted by this enzyme is its effective discrimination against phenylalanine, the value of k_{cat}/K_M for the activation of tyrosine being 1×10^5 times greater than for the activation of phenylalanine [314]. Unusually, discrimination is due only to the different binding affinities for the two amino acid substrates because tyrosyl-tRNA synthetase lacks an editing mechanism. Another unusual specificity feature is that the binding pocket for the tyrosine side-chain is larger than would appear to be necessary [300] and this enzyme is unique among aaRSs in being able to activate both stereoisomers of tyrosine [299, 300].

11.7 Lipases

The particular interest in lipases from the viewpoint of the present chapter is that their substrates, lipid esters, are water-insoluble and catalysis of their hydrolysis involves interfacial recognition. A comprehensive review of lipases and other esterases for which crystal structures are available is that by Quinn and Feaster [315]. Structures are available for enzymes of three families: the α/β hydrolases, the pancreatic lipases, and phospholipase A_2. Although enzymes of the pancreatic lipase family are usually considered to be members of the α/β hydrolase fold family, Quinn and Feaster [315] suggest that they are sufficiently different in sequence, structure, and function to be considered as a separate family. The active centers of the lipases of the α/β hydrolase family [60] are insulated from bulk solvent ([315] and e.g., [316])

and catalyze the hydrolysis of water-insoluble substrates such as triacylglycerols and cholesteryl esters contained in supramolecular assemblies such as lipid micelles or lipoproteins. Catalysis involves recognition and binding at the lipid/water interface either prior to or simultaneously with formation of the enzyme substrate complex. The steady-state rate equation is analogous to the Michaelis–Menten equation, but with more complex kinetic parameters [317].

Details of interfacial recognition mechanisms by lipases of this family are revealed by crystal structures of a number of the enzymes and complexes with phosphonate and sulfonate transition-state analogs [318, 319]. In some cases such as the *Candida rugosa* enzyme, the catalytic Ser/His/Glu triad is buried under surface loops and helices and is not accessible to the bulk solution until interaction with the amphipathic lipid/aqueous interface triggers a hinge-like conformational change. By contrast, the active centers of the mammalian cholesterol esterases are not markedly hydrophobic and are accessible to the bulk solution. These enzymes exert specificity for negatively charged lipids (e.g., [320]) and homology modeling [321] suggests interfacial recognition involving long-range electrostatic effects.

Crystal structures of the pancreatic lipases [322–324] reveal differences from the other members of the α/β hydrolase family that arise from minimal sequence similarities which result, for example, in differences in the catalytic triads and the oxyanion holes. The hinge-like conformational change involving a surface Ω-disulfide loop required to create access to the active center is common to both families of lipases, but is constructed from different regions of the sequence (*cf.* [322] with [318, 319]. An important feature of the pancreatic lipases is that active center access is controlled by interaction of the Ω-disulfide loop with a cofactor protein, colipase [323, 324].

Phospholipases not only have a digestive function but also generate signaling molecules or their immediate precursors (e.g., release of arachidonate by phospholipase A_2). Phospholipase A_2 catalyzes the stereospecific hydrolysis of the secondary ester moiety of phospholipids by a mechanism fundamentally different from the acyl-enzyme mechanism of the other lipases. Thus, the imidazole component of the Asp/His diad provides base catalysis of the direct attack of water at the carbonyl carbon of the substrate and the oxyanion hole contains a Ca^{2+} cation. The investigation of specificity and mechanism of phospholipase A_2, particularly by crystallography [325, 326] and protein engineering (e.g., [327, 328]), and the development of kinetic methods for studies at lipid/water interfaces (surface dilution kinetics [329] and "scooting" kinetics [330] and references therein) are extensively reviewed [315]. Use of analog substrates with modified phospholipid head-groups and protein engineering experiments suggest that the interaction of the head-group with a tyrosine side-chain has a role in determining stereospecificity as well as catalytic activity [327].

11.8 Conclusions

The early ideas about enzyme–substrate recognition by their association in a specific manner, such as the suggestion by Wurtz in 1880 that papain acts by becoming bound

to fibrin prior to its digestion, were generalized by Emil Fischer in 1898 and detailed descriptions of probable subsite interactions with polymeric substrates were codified by Schechter and Berger in 1967–1970. The key to the later and ongoing advances in understanding the recognition process and its intimate involvement in catalysis described in this chapter has its foundation in the now generally accepted concept of complementarity in the reaction transition state as against in the ground-state (Michaelis) complex. This change was suggested by J.B.S. Haldane in 1930 and elaborated by Linus Pauling in 1946. The advances here described were produced principally by application of combinations of reaction kinetics, crystallography, various spectroscopies, and structural variation in both substrate and enzyme, the latter provided both by protein engineering and by the use of natural variants. Future progress should benefit substantially from advances in understanding electrostatic effects, the roles of individual water molecules, and the dynamics of the protein and bound substrate provided by both computational and time-resolved structural studies.

Acknowledgments

We thank the Science and Engineering Research Council and the Biotechnology and Biological Sciences Research Council for support over many years, and Professor Mike Page for valued discussion.

References

1 Roberts, S.M.(ed.) (1989) *Molecular Recognition: Chemical and Biochemical Problems*, Special Publication No. 78, Royal Society of Chemistry, London.

2 Fischer, E. (1894) *Berichte der Deutschen Chemischen Gesellschaft*, **27**, 2985.

3 Haldane, J.B.S. (1930) *Enzymes*, Longmans Green, London, p. 182.

4 Pauling, L. (1946) *Chemical & Engineering News*, **24**, 1375.

5 Fersht, A. (1999) *Structure and Mechanism in Protein Science. A Guide to Enzyme Catalysis and Protein Folding*, Freeman, New York.

6 Wharton, C.W. and Eisenthal, R. (1981) *Molecular Enzymology*, Blackie, London.

7 Page, M.I. and Williams, A.(eds) (1987) *Enzyme Mechanisms*, Royal Society of Chemistry, London.

8 Page, M.I. (1987) In *Enzyme Mechanisms* (eds M.I. Page and A. Williams), Royal Society of Chemistry, London, p. 1.

9 Jencks, W.P. (1969) *Catalysis in Chemistry and Enzymology*, McGraw-Hill, New York.

10 Wolfenden, R. and Frick, L. (1987) In *Enzyme Mechanisms* (eds M.I. Page and A. Williams), Royal Society of Chemistry, London, p. 97.

11 Gandour, R.D. and Schowen, R.L.(eds) (1978) *Transition States of Biochemical Processes*, Plenum Press, New York.

12 Sinnot, M.(ed) (1998) *Comprehensive Biological Catalysis. A Mechanistic Reference*, Academic Press, London.

13 Resmini, M., Ostler, E.L., Brocklehurst, K. and Gallacher, G. (2004) In *Catalytic Antibodies* (ed. E. Keinan), VCH-Wiley Verlag GmbH, Weinheim, p. 468.

14 Klotz, I.M. (1987) In *Enzyme Mechanisms* (eds M.I. Page and A. Williams), Royal Society of Chemistry, London, p. 14.

15 Maddock, S.C., Pasetto, P. and Resmini, M. (2004) *Chemical Communications*, 536.

16 Wolfenden, R. (1978) In *Transition States of Biochemical Processes* (eds R.D. Gandour and R.L. Schowen), Plenum Press, New York, p. 55.

17 Leatherbarrow, R.J., Fersht, A.R. and Winter, G. (1985) *Proceedings of the National Academy of Sciences of the United States of America*, **82**, 7840.

18 Lowe, G. and Yuthavong, Y. (1971) *The Biochemical Journal*, **124**, 107.

19 Brocklehurst, K., Kowlessur, D., Patel, G., Templeton, W., Quigley, K., Thomas, E.W., Wharton, C.W., Willenbrock, F. and Szawelski, R. (1988) *The Biochemical Journal*, **250**, 761.

20 Brocklehurst, K., Watts, A.B., Patel, M., Verma, C. and Thomas, E.W. (1998) In *Comprehensive Biological Catalysis. A Mechanistic Reference*, vol. **I** (ed. M. Sinnot), Academic Press, London, p. 381.

21 Craik, C.S., Largman, C., Fletcher, T., Roczniak, S., Barr, P.J., Fletterick, R. and Rutter, W.J. (1985) *Science*, **228**, 291.

22 Estell, D.A., Graycar, T.P., Miller, J.V., Powers, D.B., Burnier, J.P., Ng, P.G. and Wells, J.A. (1986) *Science*, **233**, 659.

23 Wells, J.A., Powers, D.B., Bott, R.R., Graycar, T.P. and Estell, D.A. (1987) *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 1219.

24 Eyring, H. (1935) *Chemical Reviews*, **17**, 65.

25 Ingold, C.K. (1962) *The Transition State*, Special Publication No. 16, The Chemical Society, London, p. 118.

26 Thornton, E.K. and Thornton, E.R. (1978) In *Transition States of Biochemical Processes* (eds R.D. Gandour and R.L. Schowen), Plenum Press, New York, p. 3.

27 Wolfenden, R. (1969) *Nature*, **223**, 704.

28 Schowen, R.L. (1978) In *Transition States of Biochemical Processes* (eds R.D. Gandour and R.L. Schowen), Plenum Press, New York, p. 77.

29 Page, M.I. (1984) In *The Chemistry of Enzyme Action* (ed. M.I. Page), Elsevier, Amsterdam, p. 1.

30 Albery, W.J. and Knowles, J.R. (1977) *Angewandte Chemie (International Edition in English)*, **16**, 285.

31 Bender, M.L., Kezdy, F.J. and Gunter, C.R. (1964) *Journal of the American Chemical Society*, **86**, 3714.

32 Jencks, W.J. (1966) In *Current Aspects of Biochemical Energetics* (eds N.O. Kaplan and E.P. Kennedy), Academic Press, New York, p. 273.

33 Lienhard, G.E. (1973) *Science*, **180**, 149.

34 Knowles, J.R. (1987) *Science*, **236**, 1252.

35 Polanyi, M. (1921) *Zeitschrift fur Electrochemie*, **27**, 142.

36 Hammett, L.P. (1970) In *Physical Organic Chemistry*, 2nd edn, McGraw-Hill, New York, p. 117.

37 Wolfenden, R. (1974) *Molecular and Cellular Biochemistry*, **3**, 207.

38 Brocklehurst, K. (1977) *The Biochemical Journal*, **163**, 111.

39 Jencks, W.P. (1975) *Advances in Enzymology*, **43**, 219.

40 Jencks, W.P. (1987) *Cold Spring Harbour Symposia on Quantitative Biology*, **52**, 65.

41 Brocklehurst, K. (1982) *Methods in Enzymology*, **87**, 427.

42 Brocklehurst, K. (1994) *Protein Engineering*, **7**, 291.

43 Gul, S., Mellor, G.W., Thomas, E.W. and Brocklehurst, K. (2006) *The Biochemical Journal*, **396**, 17.

44 Hussain, S., Khan, A. and Brocklehurst, K. (2002) *Recent Research Developments in Biochemistry*, **3**, 653.

45 Gul, S., Hussain, S., Thomas, M.P., Resmini, M., Verma, C.S., Thomas, E.W. and Brocklehurst, K. (2008) *Biochemistry*, **47**, 2025.

46 Gul, S., Sreedharan, S.K. and Brocklehurst, K. (1998) *Enzyme Assays: Essential Data*, John Wiley & Sons, Ltd, Chichester.

47 Neurath, H. (1985) *Federation Proceedings*, **44**, 2907.

48 Perona, J.J. and Craik, C.S. (1995) *Protein Science*, **4**, 337.

49 Wharton, C.W. (1998) In *Comprehensive Biological Catalysis. A Mechanistic Reference*, vol. **I** (ed. M. Sinnot), Academic Press, London, p. 345.

50 Polanyi, J.C. and Zewail, A.H. (1995) *Accounts of Chemical Research*, **28**, 119.

51 Huber, R. and Bode, W. (1978) *Accounts of Chemical Research*, **11**, 114.

52 Davie, E.W., Fujikawa, K. and Kisiel, W. (1991) *Biochemistry*, **30**, 10363.

53 Smith, C.L. and DeLotto, R. (1994) *Nature*, **368**, 548.

54 Singer, P.T., Smalas, A., Carty, R.P., Mangel, W.F. and Sweet, R.M. (1993) *Science*, **259**, 669.

55 Farber, G.K. (1997) *Current Biology*, **7**, R352.

56 Moffat, K. (1997) In *Macromolecular Crystallography* (eds C.W. Carter and R.M. Sweet), Academic Press, p. 433.

57 White, A.J., Drabble, K. and Wharton, C.W. (1995) *The Biochemical Journal*, **306**, 843.

58 Matthews, B.W. (1977) In *The Proteins*, vol. **3** (eds H. Neurath and R.L. Hill), Academic Press, New York, p. 404.

59 Liao, D.I., Breddam, K., Sweet, R.M., Bullock, T. and Remington, S.J. (1992) *Biochemistry*, **31**, 9796.

60 Ollis, D.L., Cheah, E., Cygler, M., Dykstra, B., Frolow, F., Franken, S.M., Harel, H., Remington, S.J., Silman, I., Schrag, J., Sussman, J.L., Verschueren, K.H.G. and Goldman, A. (1992) *Protein Engineering*, **5**, 197.

61 Blow, D.M. (1976) *Accounts of Chemical Research*, **9**, 145.

62 Matthews, D.A., Alden, R.A., Birktoft, J.J., Freer, S.T. and Kraut, J. (1977) *The Journal of Biological Chemistry*, **252**, 8875.

63 Alber, T., Petsko, G.A. and Tsernoglou, D. (1976) *Nature*, **263**, 297.

64 Kossiakoff, A.A. and Spencer, S.A. (1981) *Biochemistry*, **20**, 6462.

65 Frey, P.A., Whitt, S.A. and Tobin, J.B. (1994) *Science*, **264**, 1927.

66 Ash, E.L., Sudmeier, J.L., DeFabo, E.C. and Bachovchin, W.W. (1997) *Science*, **278**, 1128.

67 Djurdjevic-Pahl, A., Hewage, C. and Malthouse, J.P. (2002) *The Biochemical Journal*, **362**, 339.

68 Wellner, N. and Zundel, G. (1994) *Journal of Molecular Structure*, **317**, 249.

69 Carey, P.R. and Tonge, P.J. (1995) *Accounts of Chemical Research*, **28**, 8.

70 Fastrez, J. and Fersht, A.R. (1973) *Biochemistry*, **12**, 1067.

71 Kahne, D. and Still, W.C. (1988) *Journal of the American Chemical Society*, **110**, 7529.

72 McPhalen, C.A. and James, M.N.G. (1988) *Biochemistry*, **27**, 6582.

73 Stein, R.L. and Strimpler, A.M. (1987) *Journal of the American Chemical Society*, **109**, 4387.

74 Warshel, A., Papazyan, A. and Kollman, P.A. (1995) *Science*, **269**, 102.

75 Fersht, A.R. and Sperling, J. (1973) *Journal of Molecular Biology*, **74**, 137.

76 Robillard, G. and Shulman, R.G. (1972) *Journal of Molecular Biology*, **71**, 507.

77 Gerlt, J.A. and Gassman, P.G. (1993) *Biochemistry*, **32**, 11943.

78 Braxton, S. and Wells, J.A. (1991) *The Journal of Biological Chemistry*, **266**, 11797.

79 Loewenthal, R., Sancho, J., Reinikainen, T. and Fersht, A.R. (1993) *Journal of Molecular Biology*, **232**, 574.

80 Jackson, S.E. and Fersht, A.R. (1993) *Biochemistry*, **32**, 13909.

81 Plou, F.J., Kowlessur, D., Malthouse, J.P.G., Mellor, G.W., Hartshorn, M.J., Pinitglang, S., Patel, H., Topham, C.M., Thomas, E.W., Verma, C. and Brocklehurst, K. (1996) *Journal of Molecular Biology*, **257**, 1088.

82 Pinitglang, S., Watts, A.B., Patel, M., Reid, J.D., Noble, M.A., Gul, S., Bokth, A., Naeem, A., Patel, H., Thomas, E.W., Sreedharan, S.K., Verma, C. and Brocklehurst, K. (1997) *Biochemistry*, **36**, 9968.

83 Carter, P. and Wells, J.A. (1990) *Proteins: Structure, Function, and Genetics*, **7**, 335.

84 Robertus, J.D., Alden, R.A., Birktoft, J.J., Kraut, J., Powers, J.C. and Wilcox, P.E. (1972) *Biochemistry*, **11**, 2439.

85 Robertus, J.D., Kraut, J., Alden, R.A. and Birktoft, J.J. (1972) *Biochemistry*, **11**, 4293.

86 Henderson, R. (1970) *Journal of Molecular Biology*, **54**, 341.

87 Brady, K., Wei, A., Ringe, D. and Abeles, R.H. (1990) *Biochemistry*, **29**, 7600.

88 Malthouse, J.P.G. (2007) *Biochemical Society Transactions*, **35**, 566.

89 O'Connell, T.P. and Malthouse, J.P.G. (1995) *The Biochemical Journal*, **307**, 353.

90 O'Sullivan, D.B., O'Connell, T.P., Mahon, M.M., Koenig, A., Milne, J.J., Fitzpatrick, T.B. and Malthouse, J.P.G. (1999) *Biochemistry*, **38**, 618.

91 Djurdjevic, A., Hewage, C. and Malthouse, J.P.G. (2002) *The Biochemical Journal*, **362**, 339.

92 Wright, C.S., Alden, R.A. and Kraut, J. (1969) *Nature*, **221**, 235.

93 Bode, W., Papamokos, E., Musil, D., Seemuller, U. and Fritz, H. (1986) *The EMBO Journal*, **5**, 813.

94 Gros, P., Betzel, C., Dauter, Z., Wilson, K.S. and Hol, W.G.J. (1989) *Journal of Molecular Biology*, **210**, 347.

95 Betzel, C., Pal, G.P. and Saenger, W. (1988) *European Journal of Biochemistry*, **178**, 155.

96 Betzel, C., Klupsch, S., Papendorf, G., Hastrup, S., Branner, S. and Wilson, K.S. (1992) *Journal of Molecular Biology*, **223**, 427.

97 van der Laan, J.M., Teplyakov, A.V., Kelders, H., Kalk, K.H., Misset, O., Mulleners, L.J.S.M. and Dijkstra, B.W. (1992) *Protein Engineering*, **5**, 405.

98 Poulos, T.L., Alden, R.A., Freer, S.T., Birktoft, J.J. and Kraut, J. (1976) *The Journal of Biological Chemistry*, **251**, 1097.

99 Takeuchi, Y., Noguchi, S., Satow, Y., Kojima, S., Kumagai, I., Miura, K., Nakamura, K.T. and Mitsui, Y. (1991) *Protein Engineering*, **4**, 501.

100 Takeuchi, Y., Satow, Y., Nakamura, K.T. and Mitsui, Y. (1991) *Journal of Molecular Biology*, **221**, 309.

101 Betzel, C., Singh, T.P., Visanji, M., Peters, K., Fittkau, S., Saenger, W. and Wilson, K.S. (1993) *The Journal of Biological Chemistry*, **268**, 15854.

102 Markland, F.S. and Smith, E.L. (1971) In *The Enzymes*, vol. **3** (ed. P.D. Boyer), Academic Press, New York, p. 516.

103 Wells, J.A., Cunningham, B.C., Graycar, T.P. and Estell, D.A. (1987) *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 5167.

104 Grøn, H., Meldal, M. and Breddam, K. (1992) *Biochemistry*, **31**, 6011.

105 Grøn, H. and Breddam, K. (1992) *Biochemistry*, **31**, 8967.

106 Rheinnecker, M., Baker, G., Eder, J. and Fersht, A.R. (1993) *Biochemistry*, **32**, 1199.

107 Matthews, B.W., Sigler, P.B., Henderson, R. and Blow, D.M. (1967) *Nature*, **214**, 652.

108 Watson, H.C., Shotton, D.M., Cox, J.M. and Muirhead, H. (1970) *Nature*, **225**, 806.

109 Ruhlmann, A., Kukla, D., Schwager, P., Bartels, K. and Huber, R. (1973) *Journal of Molecular Biology*, **77**, 417.

110 Sprang, S., Standing, T., Fletterick, R.J., Stroud, R.M., Finer-Moore, J., Xuong, N.H., Hamlin, R., Rutter, W.J. and Craik, C.S. (1987) *Science*, **437**, 905.

111 Read, R.J. and James, M.N.G. (1988) *Journal of Molecular Biology*, **200**, 523.

112 Fujinaga, M. and James, M.N.G. (1987) *Journal of Molecular Biology*, **195**, 373.

113 Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G. and Bartunik, H. (1983) *Journal of Molecular Biology*, **164**, 237.

114 Remington, S.J., Woodbury, R.G., Reynolds, R.A., Matthews, B.W. and Neurath, H. (1988) *Biochemistry*, **27**, 8097.

115 Navia, M.A., McKeever, B.M., Springer, J.P., Lin, T.Y., Williams, H.R., Fluder, E.M., Dorn, C.P. and Hoogsteen, K. (1989) *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 7.

116 Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S.R. and Hofsteenge, J. (1989) *The EMBO Journal*, **8**, 3467.

117 Padmanabhan, K., Padmanabhan, K.P., Tulinsky, A., Park, C.H., Bode, W., Huber, R., Blankenship, D.T., Cardin, A.D. and Kisiel, W. (1993) *Journal of Molecular Biology*, **232**, 947.

118 Narajana, S.V.L., Carson, M., El-Kabbani, O., Kilpatrick, J.M., Moore, D., Chen, X., Bugg, C.E., Volanakis, J.E. and DeLucas, L.J. (1994) *Journal of Molecular Biology*, **235**, 695.

119 Delbaere, L.T.J., Brayer, G.D. and James, M.N.G. (1979) *Canadian Journal of Biochemistry*, **57**, 135.

120 Moult, J., Sussman, F. and James, M.N.G. (1985) *Journal of Molecular Biology*, **182**, 555.

121 Nienaber, V.L., Breddam, K. and Birktoft, J.J. (1993) *Biochemistry*, **32**, 11469.

122 Brayer, G.D., Delbaere, L.T.J. and James, M.N.G. (1979) *Journal of Molecular Biology*, **131**, 743.

123 Greer, J. (1990) *Proteins: Structure, Function, and Genetics*, **7**, 317.

124 Lottenberg, R., Christensen, U., Jackson, C.M. and Coleman, P.L. (1981) *Methods in Enzymology*, **80**, 341.

125 Knight, C.G. (1995) *Methods in Enzymology*, **248**, 18.

126 Thomas, D.A., Francis, P., Smith, C., Ratcliffe, S., Ede, N.J., Kay, C., Wayne, G., Martin, S.L., Moore, K., Amour, A. and Hooper, N.M. (2006) *Proteomics*, **6**, 2112.

127 Tawa, P., Tam, J., Cassady, R., Nicholson, D.W. and Xanthoudakis, S. (2001) *Cell Death and Differentiation*, **8**, 30.

128 Meng, Q.C. and Oparil, S. (1996) *Journal of Chromatography. A*, **743**, 105.

129 Todd, M.J. and Gomez, J. (2001) *Analytical Biochemistry*, **296**, 179.

130 Horton, J.D., Cohen, J.C. and Hobbs, H.H. (2007) *Trends in Biochemical Sciences*, **32**, 71.

131 Kwon, H.J., Lagace, T.A., McNutt, M.C., Horton, J.D. and Deisenhofer, J. (2008) *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 1820.

132 Nassoury, N., Blasiole, D.A., Oler, A.T., Benjannet, S., Hamelin, J., Poupon, V., McPherson, P.S., Attie, A.D., Prat, A. and Seidah, N.G. (2007) *Traffic*, **8**, 718.

133 Brocklehurst, K., Willenbrock, F. and Salih, E. (1987) In *Hydrolytic Enzymes* (eds A. Neuberger and K. Brocklehurst), Elsevier, Amsterdam, p. 39.

134 Vines, S.H. (1905) *Annals of Botany*, **19**, 149.

135 Wurtz, A. and Bouchet, E. (1879) *Comptes Rendus de l'Academie des Sciences*, **89**, 425.

136 Berti, P.J. and Storer, A.C. (1995) *Journal of Molecular Biology*, **246**, 273.

137 Barrett, A.J., Rawlings, N.D. and Woessner, J.F. Jr(eds) (2004) *Handbook of Proteolytic Enzymes*, Elsevier, London.

138 Baines, B.S. and Brocklehurst, K. (1979) *The Biochemical Journal*, **177**, 541.

139 Baines, B.S. and Brocklehurst, K. (1982) *Journal of Protein Chemistry*, **1**, 119.

140 Thomas, M.P., Topham, C.M., Kowlessur, D., Mellor, G.W., Thomas, E.W., Whitford, D. and Brocklehurst, K. (1994) *The Biochemical Journal*, **300**, 805.

141 Topham, C.M., Salih, E., Frazao, C., Kowlessur, D., Overington, J.P., Thomas, M., Brocklehurst, S.M., Patel, M., Thomas, E.W. and Brocklehurst, K. (1991) *The Biochemical Journal*, **280**, 79.

142 Brocklehurst, K., Baines, B.S. and Malthouse, J.P.G. (1981) *The Biochemical Journal*, **197**, 739.

143 Malthouse, J.P.G. and Brocklehurst, K. (1976) *The Biochemical Journal*, **159**, 221.

144 Willenbrock, F. and Brocklehurst, K. (1985) *The Biochemical Journal*, **227**, 511.

145 O'Farrell, P.A. and Joshua-Tor, L. (2007) *The Biochemical Journal*, **401**, 421.

146 Brocklehurst, K., Carlsson, J., Kierstan, M.P.J. and Crook, E.M. (1973) *The Biochemical Journal*, **133**, 573.

147 Brocklehurst, K., Carlsson, J., Kierstan, M.P.J. and Crook, E.M. (1974) *Methods in Enzymology*, **34**, 531.

148 Brocklehurst, K., Courey, A.J., Gul, S., Lin, S.-H. and Moritz, R.L. (2004) In *Purifying*

Proteins for Proteomics (ed. R.J. Simpson), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 221.

149 Matthews, D.A., Smith, W.W., Ferre, R.A., Condon, B., Budahal, G., Sisson, W., Villafranca, J.E., Janson, C.A., McElroy, H.E., Gribskov, C.L. and Worland, S. (1994) *Cell*, **77**, 761.

150 Ding, J., McGrath, W.J., Sweet, R.M. and Mangel, W.F. (1996) *The EMBO Journal*, **15**, 1778.

151 Wilson, K.P., Black, J.-A.F., Thomson, J.A., Kim, E.E., Griffiths, J.P., Navia, M.A., Murcko, M.A., Chambers, S.P., Aldope, R.A., Raybuck, B.A. and Livingston, D.J. (1994) *Nature*, **370**, 270.

152 Lowe, G. (1976) *Tetrahedron*, **32**, 291.

153 Willenbrock, F. and Brocklehurst, K. (1984) *The Biochemical Journal*, **222**, 805.

154 Willenbrock, F. and Brocklehurst, K. (1985) *The Biochemical Journal*, **227**, 521.

155 Brocklehurst, K., Kowlessur, D., O'Driscoll, M., Patel, G., Quenby, S., Salih, E., Templeton, W., Thomas, E.W. and Willenbrock, F. (1987) *The Biochemical Journal*, **244**, 173.

156 Brocklehurst, K., Brocklehurst, S.M., Kowlessur, D., O'Driscoll, M., Patel, G., Salih, E., Templeton, W., Thomas, E., Topham, C.M. and Willenbrock, F. (1988) *The Biochemical Journal*, **256**, 543.

157 Kowlessur, D., Topham, C.M., Thomas, E.W., O'Driscoll, M., Templeton, W. and Brocklehurst, K. (1989) *The Biochemical Journal*, **258**, 755.

158 Patel, M., Kayani, I.S., Mellor, G.W., Sreedharan, S., Templeton, W., Thomas, E.W., Thomas, M. and Brocklehurst, K. (1992) *The Biochemical Journal*, **281**, 553.

159 Noble, M.A., Gul, S., Verma, C.S. and Brocklehurst, K. (2000) *The Biochemical Journal*, **351**, 723.

160 Mellor, G.W., Patel, M., Thomas, E.W. and Brocklehurst, K. (1993) *The Biochemical Journal*, **294**, 201.

161 Watts, A., Hafeez, A., Gul, S., Verma, C., Thomas, E.W. and Brocklehurst, K. (1997) *Biochemical Society Transactions*, **25**, 84.

162 Reid, J.D., Hussain, S., Sreedharan, S.K., Bailey, T.S.F., Pinitglang, S., Thomas, E.W., Verma, C.S. and Brocklehurst, K. (2001) *The Biochemical Journal*, **357**, 343.

163 Reid, J.D., Hussain, S., Bailey, T.S.F., Sonkaria, S., Sreedharan, S.K., Thomas, E.W., Resmini, M. and Brocklehurst, K. (2004) *The Biochemical Journal*, **378**, 699.

164 Fruton, J.S. (1982) *Advances in Enzymology and Related Areas of Molecular Biology*, **53**, 239.

165 Schechter, I. and Berger, A. (1967) *Biochemical and Biophysical Research Communications*, **27**, 157.

166 Schechter, I. and Berger, A. (1968) *Biochemical and Biophysical Research Communications*, **32**, 898.

167 Berger, A. and Schechter, I. (1970) *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, **257**, 249.

168 Drenth, J., Kalk, K.H. and Swen, H.M. (1976) *Biochemistry*, **15**, 3731.

169 Kraut, J. (1977) *Annual Review of Biochemistry*, **46**, 331.

170 Huber, C.P., Ozaki, Y., Pliura, D.H., Storer, A.C. and Carey, P.R. (1982) *Biochemistry*, **21**, 3109.

171 Storer, A.C., Lee, H. and Carey, P.R. (1983) *Biochemistry*, **22**, 4789.

172 Berti, P.J., Faerman, C.H. and Storer, A.C. (1991) *Biochemistry*, **30**, 1394.

173 Patel, M., Kayani, I.S., Templeton, W., Mellor, G.W., Thomas, E.W. and Brocklehurst, K. (1992) *The Biochemical Journal*, **287**, 881.

174 Lowe, G. and Yuthavong, Y. (1971) *The Biochemical Journal*, **124**, 117.

175 Templeton, W., Kowlessur, D., Thomas, E.W., Topham, C.M. and Brocklehurst, K. (1990) *The Biochemical Journal*, **266**, 645.

176 Kowlessur, D., Thomas, E.W., Topham, C.M., Templeton, W. and Brocklehurst, K. (1990) *The Biochemical Journal*, **266**, 653.

177 Boland, M.J. and Hardman, M.J. (1973) *European Journal of Biochemistry*, **36**, 575.

178 Baker, E.N., Boland, M.J., Calder, P.C. and Hardman, M.J. (1980) *Biochimica et Biophysica Acta*, **616**, 30.

179 Kamphuis, I.G., Kalk, K.H., Swarte, M.B.A. and Drenth, J. (1984) *Journal of Molecular Biology*, **179**, 233.

180 Baker, E.N. and Dodson, E.J. (1980) *Acta Crystallographica*, **A36**, 559.

181 Taralp, A., Kaplan, H., Sytwu, I.-I., Vlattas, I., Bohacek, R., Knap, A.K., Hirama, T., Huber, C.P. and Hasnain, S. (1995) *The Journal of Biological Chemistry*, **270**, 18036.

182 Thomas, M.P., Verma, C., Boyd, S.M. and Brocklehurst, K. (1995) *The Biochemical Journal*, **306**, 39.

183 O'Hara, B.P., Hemmings, A.M., Buttle, D.J. and Pearl, L.H. (1995) *Biochemistry*, **34**, 13190.

184 Takahashi, T., Dehdarani, A.H., Yonezawa, S. and Tang, J. (1986) *The Journal of Biological Chemistry*, **261**, 9375.

185 Brömme, D., Bescherer, K., Kirschke amd, H. and Fittkau, S. (1987) *The Biochemical Journal*, **245**, 381.

186 Kirschke, H., Wikstrom, P. and Shaw, E. (1988) *FEBS Letters*, **228**, 128.

187 Cole, P.W., Murakami, K. and Inagami, T. (1971) *Biochemistry*, **10**, 4246.

188 Sreedharan, S.K., Patel, H., Smith, S., Gharbia, S.E., Shah, H.N. and Brocklehurst, K. (1993) *Biochemical Society Transactions*, **21**, 218.

189 Alecio, M.R., Dann, M.L. and Lowe, G. (1974) *The Biochemical Journal*, **141**, 495.

190 Baker, E.N. and Drenth, J. (1987) In *Biological Macromolecules and Assemblies*, vol **3** (eds F. Jurnak and A. McPherson), John Wiley & Sons, Inc., New York, p. 313.

191 García-Echeverría, C. and Rich, D.H. (1992) *FEBS Letters*, **297**, 100.

192 Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. and Turk, V. (1990) *The EMBO Journal*, **9**, 1939.

193 Kalnitsky, G., Chatterjee, R., Singh, H., Lones, M. and Paszkowski, A. (1983) In *Proteinase Inhibitors in Medical and Biological Aspects* (eds N. Katunuma, H. Umezawa and H. Holzer), Springer, Berlin, p. 263.

194 Brocklehurst, K., Kowlessur, D., Patel, G., Templeton, W., Quigley, K., Thomas, E.W., Wharton, C.W., Willenbrock, F. and Szawelski, R.J. (1988) *The Biochemical Journal*, **250**, 761.

195 Kowlessur, D., O'Driscoll, M., Topham, C.M., Templeton, W., Thomas, E.W. and Brocklehurst, K. (1989) *The Biochemical Journal*, **259**, 443.

196 Patel, M., Thomas, M.P., Noble, M.A., Gul, S., Thomas, E.W. and Brocklehurst, K. (1994) *Biochemical Society Transactions*, **22**, 214.

197 Gul, S., Pinitglang, S., Thomas, E.W., Verma, C. and Brocklehurst, K. (1998) *Biochemical Society Transactions*, **26**, S171.

198 Pinitglang, S., Noble, M., Verma, C., Thomas, E.W. and Brocklehurst, K. (1996) *Biochemical Society Transactions*, **24**, 468.

199 Davies, G., Sinnott, M.L. and Withers, S.G. (1998) In *Comprehensive Biological Catalysis. A Mechanistic Reference*, vol. **I** (ed. M. Sinnot), Academic Press, London, p. 119.

200 Brown, A. and Bruice, T.C. (1973) *Journal of the American Chemical Society*, **95**, 1593.

201 Sinnott, M. L. (1987) In *Enzyme Mechanisms* (eds M.I. Page and A. Williams), Royal Society of Chemistry, London, p. 259.

202 Sinnott, M.L. (1990) *Chemical Reviews*, **90**, 1171.

203 Davies, G.J. and Henrissat, B. (1995) *Structure*, **3**, 853.

204 McCarter, J.D. and Withers, S.G. (1994) *Current Opinion in Structural Biology*, **4**, 885.

205 Davies, G.J., Tolley, S.P., Henrissat, B., Hjort, C. and Schulein, M. (1995) *Biochemistry*, **34**, 16210.

206 McIntosh, L.P., Hand, G., Johnson, P.E., Joshi, M.D., Korner, M., Plesniak, L.A., Ziser, L., Wakarchuk, W.W. and Withers, S.G. (1996) *Biochemistry*, **35**, 9958.

207 Walaszek, Z., Horton, D. and Ekiel, I. (1982) *Carbohydrate Research*, **106**, 193.

208 Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C.T., Phillips, D.C. and Sarma, V.R. (1965) *Nature*, **206**, 757.

209 Henrissat, B. and Bairoch, A. (1996) *The Biochemical Journal*, **316**, 695.

210 Jacobson, R.H., Zhang, X.J., DuBose, R.F. and Matthews, B.W. (1994) *Nature*, **369**, 761.

211 Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K.S. and Vorgias, C.E. (1996) *Nature Structural Biology*, **3**, 638.

212 Henrissat, B. and Davies, G. (1997) *Current Opinion in Structural Biology*, **7**, 637.

213 Koshland, D.E. (1953) *Biological Reviews of the Cambridge Philosophical Society*, **28**, 416.

214 Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P. and Davies, G. (1995) *Proceedings of the National Academy of Sciences of the United States of America*, **362**, 281.

215 Jenkins, J., Lo Leggio, L., Harris, G.W. and Pickersgill, R.W. (1995) *FEBS Letts*, **362**, 281.

216 Burmeister, W.P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S. and Henrissat, B. (1997) *Structure*, **5**, 663.

217 Jain, S., Drendel, W.B., Chen, Z.-W., Matthews, F.S., Sly, W.S. and Grubb, J.H. (1996) *Nature Structural Biology*, **3**, 375.

218 Buschiazzo, A., Amaya, M.F., Cremona, M.L., Frasch, A.C. and Alzari, P.M. (2002) *Molecular Cell*, **10**, 757.

219 Watts, A.G., Damager, I., Amaya, M.L., Buschiazzo, A., Alzari, P., Frasch, A.C. and Withers, S.G. (2003) *Journal of the American Chemical Society*, **125**, 7532.

220 Yip, V.L.Y., Varrot, A., Davies, G.J., Rajan, S.S., Yang, X., Thompson, J., Anderson, W.F. and Withers, S.G. (2004) *Journal of the American Chemical Society*, **126**, 8354.

221 Vasella, A., Davies, G.J. and Bohm, M. (2002) *Current Opinion in Chemical Biology*, **6**, 619.

222 Money, V.A., Smith, N.L., Scaffidi, A., Stick, R.V., Gilbert, H.J. and Davies, G.J. (2006) *Angewandte Chemie International Edition*, **45**, 5136.

223 Hengge, A.V. (1998) In *Comprehensive Biological Catalysis. A Mechanistic Reference* (ed. M. Sinnot), Academic Press, London, p. 517.

224 Hirsch, A.K.H., Fischer, F.R. and Diederich, F. (2007) *Angewandte Chemie International Edition*, **46**, 338.

225 Thaimattam, R., Banerjee, R., Miglani, R. and Iqbal, J. (2007) *Current Pharmaceutical Design*, **13**, 2751.

226 Gold, M.G., Barford, D. and Komander, D. (2006) *Current Opinion in Structural Biology*, **16**, 693.

227 Rockey, W.M. and Elcock, A.H. (2006) *Current Protein & Peptide Science*, **7**, 437.

228 Knighton, D.R., Zheng, J., Ten Eyck, L.F., Ashford, V.A., Xuong, N.-H., Taylor, S.S. and Sowadski, J.M. (1991) *Science*, **253**, 407.

229 Knighton, D.R., Bell, S.M., Zheng, J., Ten Eyck, L.F., Xuong, N.H., Taylor, S.S. and Sowadski, J.M. (1993) *Acta Crystallographica. Section D, Biological Crystallography*, **49**, 357.

230 Bossemeyer, D., Engh, R.A., Kinzel, V., Ponstingl, H. and Huber, R. (1993) *The EMBO Journal*, **12**, 849.

231 De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgan, D.O. and Kim, S.H. (1993) *Nature*, **363**, 595.

232 Zhang, F., Strand, A., Robbins, D., Cobb, M.H. and Goldsmith, E.J. (1994) *Nature*, **367**, 704.

233 Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J. and Pavletich, N.P. (1995) *Nature*, **376**, 313.

234 Xu, W., Harrison, S.C. and Eck, M.J. (1997) *Nature*, **385**, 595.

235 Lowe, E.D., Noble, M.E.M., Skamnaki, V.T., Oikonomakos, N.G., Owen, D.J. and Johnson, L.N. (1997) *The EMBO Journal*, **16**, 6646.

236 Fabian, M.A., Biggs, W.H., Treiber, D.K., Atteridge, C.E., Azimioara, M.D., Benedetti, M.G., Carter, T.A., Ciceri, P., Edeen, P.T., Floyd, M., Ford, J.M., Galvin, M., Gerlach, J.L., Grotzfeld, R.M., Herrgard, S., Insko, D.E., Insko, M.A., Lai, A.G., Lélias, J.-M., Mehta, S.A., Milanov, Z.V., Velasco, A.M., Wodicka, L.M., Patel, H.K., Zarrinkar, P.P. and Lockhart, D.J. (2005) *Nature Biotechnology*, **23**, 329.

237 Hunter, T. (1982) *The Journal of Biological Chemistry*, **257**, 4843.

238 Foulkes, J.G., Chow, M., Gorka, C., Frackelton, A.R. Jr, and Baltimore, D. (1985) *The Journal of Biological Chemistry*, **260**, 8070.

239 Kemp, B.E., Bylund, D.B., Huang, T.-S. and Krebs, E.G. (1975) *Proceedings of the National Academy of Sciences of the United States of America*, **72**, 3448.

240 Daile, P., Carnegie, P.R. and Young, J.D. (1975) *Nature*, **257**, 416.

241 Pearson, R.B. and Kemp, B.E. (1991) *Methods in Enzymology*, **200**, 62.

242 Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M.F., Piwnica-Worms, H. and Cantley, L.C. (1994) *Current Biology*, **4**, 973.

243 Hutti, J.E., Jarrell, E.T., Chang, J.D., Abbott, D.W., Storz, P., Toker, A., Cantley, L.C. and Turk, B.E. (2004) *Nature Methods*, **1**, 27.

244 Pinna, L.A. and Ruzzene, M. (1996) *Biochimica et Biophysica Acta*, **1314**, 191.

245 Songyang, Z., Lu, K.P., Kwon, Y.T., Tsai, L.H., Filhol, O., Cochet, C., Brickey, D.A., Soderling, T.R., Bartleson, C., Graves, D.J., DeMaggio, A.J., Hoekstra, M.F., Blenis, J., Hunter, T. and Cantley, L.C. (1996) *Molecular and Cellular Biology*, **16**, 6486.

246 Fiol, C.J., Wang, A., Roeske, R.W. and Roach, P.J. (1990) *The Journal of Biological Chemistry*, **265**, 6061.

247 Songyang, Z., Carraway, K.L., Eck, M.J., Harrison, S.C., Feldman, R.A., Mohammadi, M., Schlessinger, J., Hubbard, S.R., Smith, D.P., Eng, C., Lorenzo, M.J., Ponder, B.A.J., Mayer, B.J. and Cantley, L.C. (1995) *Nature*, **373**, 536.

248 Obata, T., Yaffe, M.B., Leparc, G.G., Piro, E.T., Maegawa, H., Kashiwagi, A., Kikkawa, R. and Cantley, L.C. (2000) *The Journal of Biological Chemistry*, **275**, 36108.

249 Ubersax, J.A., Woodbury, E.L., Quang, P.N., Paraz, M., Blethrow, J.D., Shah, K., Shokat, K.M. and Morgan, D.O. (2003) *Nature*, **425**, 859.

250 Williams, D.D., Marin, O., Pinna, L.A. and Proud, C.G. (1999) *FEBS Letters*, **448**, 86.

251 Biondi, R.M., Cheung, P.C.F., Casamayor, A., Deak, M., Currie, R.A. and Alessi, D.R. (2000) *The EMBO Journal*, **19**, 979.

252 Adams, P.D., Sellers, W.R., Sharma, S.K., Wu, A.D., Nalin, C.M. and Kaelin, W.G. Jr (1996) *Molecular and Cellular Biology*, **16**, 6623.

253 Schulman, B.A., Lindstrom, D.L. and Harlow, E. (1998) *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 10453.

254 Miller, M.E. and Cross, F.R. (2001) *Journal of Cell Science*, **114**, 1811.

255 Cheng, K.-Y., Noble, M.E.M., Skamnaki, V., Brown, N.R., Lowe, E.D., Kontogiannis, L., Shen, K., Cole, P.A., Siligardi, G. and Johnson, L.N. (2006) *The Journal of Biological Chemistry*, **281**, 23167.

256 Archambault, V., Buchler, N.E., Wilmes, G.M., Jacobson, M.D. and Cross, F.R. (2005) *Cell Cycle*, **4**, 125.

257 Brown, N.R., Noble, M.E.M., Endicott, J.A. and Johnson, L.N. (1999) *Nature Cell Biology*, **1**, 438.

258 Leng, X., Noble, M., Adams, P.D., Qin, J. and Harper, J.W. (2002) *Molecular and Cellular Biology*, **22**, 2242.

259 Takeda, D.Y., Wohlschlegel, J.A. and Dutta, A. (2001) *The Journal of Biological Chemistry*, **276**, 1993.

260 Chen, Y.-N.P., Sharma, S.K., Ramsey, T.M., Jiang, L., Martin, M.S., Baker, K., Adams, P.D., Bair, K.W. and Kaelin, W.G. Jr (1999) *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 4325.

261 Lowe, E.D., Tews, I., Cheng, K.Y., Brown, N.R., Gul, S., Noble, M.E., Gamblin, S.J. and Johnson, L.N. (2002) *Biochemistry*, **41**, 15625.

262 Davies, S.P., Reddy, H., Caivano, M. and Cohen, P. (2000) *The Biochemical Journal*, **351**, 95.

263 Knight, Z.A. and Shokat, K.M. (2005) *Chemistry & Biology*, **12**, 621.

264 Carnegie, P.R., Kemp, B.E., Dunkley, P.R. and Murray, A.W. (1973) *The Biochemical Journal*, **135**, 569.
265 Harauz, G., Ishiyama, N. and Bates, I.R. (2000) *Molecular and Cellular Biochemistry*, **209**, 155.
266 Turner, R.S., Kemp, B.E., Su, H. and Kuo, J.F. (1985) *The Journal of Biological Chemistry*, **260**, 11503.
267 Wise, B.C., Glass, D.B., Chou, C.H.J., Raynor, R.L., Katoh, N., Schatzman, R.C., Turner, R.S., Kibler, R.F. and Kuo, J.F. (1982) *The Journal of Biological Chemistry*, **257**, 8489.
268 Fersht, A.R. and Kaethner, M.M. (1976) *Biochemistry*, **15**, 818.
269 First, E.A. (1998) In *Comprehensive Biological Catalysis. A Mechanistic Reference*, vol. **I** (ed. M. Sinnot), Academic Press, London, p. 574.
270 Fersht, A.R. and Dingwall, C. (1979) *Biochemistry*, **18**, 2627.
271 Loftfield, R.B. and Eigner, E.A. (1966) *Biochimica et Biophysica Acta*, **130**, 426.
272 Tsui, W.C. and Fersht, A.R. (1981) *Nucleic Acids Research*, **9**, 4627.
273 Fersht, A.R. and Kaethner, M.M. (1976) *Biochemistry*, **15**, 3342.
274 Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y. and Winter, G. (1985) *Nature*, **314**, 235.
275 Fersht, A.R. (1987) *Biochemistry*, **26**, 8031.
276 Cusack, S., Yaremchuk, A. and Tukalo, M. (1996) *The EMBO Journal*, **15**, 2834.
277 Ghosh, G., Pelka, H., Schulman, L.H. and Brunie, S. (1991) *Biochemistry*, **30**, 9569.
278 Loftfield, R.B. and Vanderjagt, D. (1972) *The Biochemical Journal*, **128**, 1353.
279 Baldwin, A.N. and Berg, P. (1966) *The Journal of Biological Chemistry*, **241**, 831.
280 Freist, W. (1989) *Biochemistry*, **28**, 6787.
281 Jakubowski, H. and Goldman, E. (1992) *Microbiological Reviews*, **56**, 412.
282 Jakubowski, H. (1994) *Annals of the New York Academy of Sciences*, **745**, 4.
283 Fersht, A.R., Knill-Jones, J.W. and Tsui, W.C. (1982) *Journal of Molecular Biology*, **156**, 37.
284 Fersht, A.R. (1977) *Biochemistry*, **16**, 1025.
285 Freist, W. and Cramer, F. (1983) *European Journal of Biochemistry*, **131**, 65.
286 Nureki, O., Vassylyev, D.G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T.L., Schimmel, P. and Yokoyama, S. (1998) *Science*, **280**, 578.
287 Lin, L., Hale, S.P. and Schimmel, P. (1996) *Nature*, **384**, 33.
288 Rould, M.A., Perona, J.J. and Steitz, T.A. (1991) *Nature*, **352**, 213.
289 Biou, V., Yaremchuk, A., Tukalo, M. and Cusack, S. (1994) *Science*, **263**, 1404.
290 Knowles, J.R., Katze, J.R., Konigsberg, W. and Söll, D. (1970) *The Journal of Biological Chemistry*, **245**, 1407.
291 Rigler, R., Pachmann, U., Hirsch, R. and Zachau, H.G. (1976) *European Journal of Biochemistry*, **65**, 307.
292 Schulman, L.H. (1991) *Progress in Nucleic Acid Research and Molecular Biology*, **41**, 23.
293 Sherman, J.M., Thomann, H.-U. and Söll, D. (1996) *Journal of Molecular Biology*, **256**, 818.
294 Yan, W., Augustine, J. and Francklyn, C. (1996) *Biochemistry*, **35**, 6559.
295 Ibba, M., Curnow, A.W. and Söll, D. (1997) *Trends in Biochemical Sciences*, **22**, 39.
296 Fourmy, D., Mechulam, Y. and Blanquet, S. (1995) *Biochemistry*, **34**, 15681.
297 Bhat, T.N., Blow, D.M., Brick, P. and Nyborg, J. (1982) *Journal of Molecular Biology*, **158**, 699.
298 Winter, G., Fersht, A.R., Wilkinson, A.J., Zoller, M. and Smith, M. (1982) *Nature*, **299**, 756.
299 Brick, P. and Blow, D.M. (1987) *Journal of Molecular Biology*, **194**, 287.
300 Brick, P., Bhat, T.N. and Blow, D.M. (1989) *Journal of Molecular Biology*, **208**, 83.
301 Fersht, A.R. (1975) *Biochemistry*, **14**, 5.
302 Fersht, A.R., Ashford, J.S., Bruton, C.J., Jakes, R., Koch, G.L.E. and Hartley, B.S. (1975) *Biochemistry*, **14**, 1.
303 Fersht, A.R. and Jakes, R. (1975) *Biochemistry*, **14**, 3350.

304 Jakes, R. and Fersht, A.R. (1975) *Biochemistry*, **14**, 3344.
305 Wells, T.N.C. and Fersht, A.R. (1986) *Biochemistry*, **25**, 1881.
306 Fersht, A.R., Matouschek, A. and Serrano, L. (1992) *Journal of Molecular Biology*, **224**, 771.
307 Avis, J.M. and Fersht, A.R. (1993) *Biochemistry*, **32**, 5321.
308 de Prat Gay, G., Duckworth, H.W. and Fersht, A.R. (1993) *FEBS Letters*, **318**, 167.
309 Fersht, A.R. (1987) *Biochemistry*, **26**, 8031.
310 Fersht, A.R. and Winter, G.P. (1985) *Ciba Foundation Symposium*, **111**, 204.
311 First, E.A. and Fersht, A.R. (1995) *Biochemistry*, **34**, 5030.
312 Fersht, A.R., Knill-Jones, J.W., Bedouelle, H. and Winter, G. (1988) *Biochemistry*, **27**, 1581.
313 Fersht, A.R. and Dingwall, C. (1979) *Biochemistry*, **18**, 1245.
314 Fersht, A.R., Shindler, J.S. and Tsui, W.-C. (1980) *Biochemistry*, **19**, 5520.
315 Quinn, D.M. and Feaster, S.R. (1998) In *Comprehensive Biological Catalysis. A Mechanistic Reference*, vol. **I** (ed. M. Sinnot), Academic Press, London, p. 455.
316 Grochulski, P., Li, Y., Schrag, J.D. and Cygler, M. (1994) *Protein Science*, **3**, 82.
317 Verger, R. and De Haas, G.H. (1976) *Annual Review of Biophysics and Bioengineering*, **5**, 77.
318 Grochulski, P., Li, Y., Schrag, J.D., Bouthillier, F., Smith, P., Harrison, D., Rubin, B. and Cygler, M. (1993) *The Journal of Biological Chemistry*, **268**, 12843.
319 Grochulski, P., Bouthillier, F., Kazlauskas, R.J., Serreqi, A.N., Schrag, J.D., Ziomek, E. and Cygler, M. (1994) *Biochemistry*, **33**, 3494.
320 Myers-Payne, S.C., Hui, D.Y. and Brockman, H.L. (1995) *Biochemistry*, **34**, 3942.
321 Feaster, S.R., Quinn, D.M. and Barrett, B.L. (1997) *Protein Science*, **6**, 73.
322 Winkler, F.K., D'Arcy, A. and Hunziker, W. (1990) *Nature*, **343**, 771.
323 Egloff, M.-P., Marguet, F., Buono, G., Verger, R., Camibillau, C. and van Tilbeurgh, H. (1995) *Biochemistry*, **34**, 2751.
324 Egloff, M.P., Sarda, L., Verger, R., Cambillau, C. and van Tilbeurgh, H. (1995) *Protein Science*, **4**, 44.
325 Scott, D.L., White, S.P., Otwinowski, Z., Yuan, W., Gelb, M. and Sigler, P.B. (1990) *Science*, **250**, 1541.
326 White, S.P., Scott, D.I., Otwinowski, Z., Gelb, M.H. and Sigler, P.B. (1990) *Science*, **250**, 1560.
327 Kuipers, O.P., Dekker, N., Verheij, H.M. and de Haas, G.H. (1990) *Biochemistry*, **29**, 6094.
328 Kumar, M., Sekharudu, C., Ramakrishnan, B., Dupureur, C.M., Zhu, H., Tsai, M.-D. and Sundaralingam, M. (1994) *Protein Science*, **3**, 2082.
329 Carman, G.M., Deems, R.A. and Dennis, E.A. (1995) *The Journal of Biological Chemistry*, **270**, 18711.
330 Gelb, M.H., Jain, M.K., Hanel, A.M. and Berg, O.G. (1995) *Annual Review of Biochemistry*, **64**, 653.

12
Protein Recognition

Robyn E. Mansfield, Arwen J. Cross, Jacqueline M. Matthews, and Joel P. Mackay

12.1
General Introduction

Proteins function by making and breaking interactions with other molecules. The identities of these molecules cover the full spectrum of substances that one might find in an organism, illustrating the astounding structural and functional diversity that can be obtained from different combinations of roughly 20 amino acids. Thus, proteins are able to form functional interactions with partners ranging from metal ions and small diatomic molecules, such as oxygen and nitric oxide, through to typical organic molecules and macromolecules, such as nucleic acid polymers, polysaccharides, and of course other proteins. Given the enormous breadth of these interactions and the functions that they bring about, we have chosen in this chapter to focus on capturing a few of the general principles underlying protein recognition. As other chapters in this volume deal in depth with enzymes and enzyme–substrate interactions, we focus here on protein–protein and protein–DNA recognition.

First, we discuss the general physicochemical nature of the interfaces made between proteins and their partners, the strength of these interactions, and methods for measuring their strength. Next, we describe two areas of protein recognition that have recently received much attention: coupled folding and binding, in which at least one of the partners in a protein interaction is disordered in the absence of its partner, and the regulation of protein interactions by post-translational modification (PTM) – a phenomenon that underlies diverse biological processes such as signaling and gene regulation. Finally, we look at progress that has been made in exploiting our understanding of protein recognition to either inhibit protein–protein interactions or to engineer new interactions for research or therapeutic purposes.

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

12.2
Nature of Protein Interfaces

12.2.1
General Characteristics of Binding Sites

A number of published studies have sought to define the characteristics of protein interaction interfaces, in an effort to determine whether all interfaces share some underlying physicochemical properties. Such properties would be of enormous value in predicting the functional surfaces of a protein. In an examination of 75 protein–protein interfaces from the Protein Data Bank (PDB), Lo Conte *et al.* determined that the average protein–protein interface (i.e., the protein surface area that becomes solvent-inaccessible upon binding) is 1600 ± 400 Å^2 [1]. The study also showed that a poor correlation exists between the size of an interface and the molecular weight of the complex. However, there are trends relating the size of the interface to the nature of the interaction. For example, smaller than average interfaces tend to be associated with transient interactions and larger interfaces are more likely to be associated with large conformational changes upon binding [1]. Large conformational changes include domain movements, loop rearrangements, and also disorder–order transitions (see Section 12.5). One example is the complex formed between the plasma protease thrombin and hirudin, an anticlotting agent from medicinal leech, which buries 3300 Å^2 [2, 3]. Hirudin has a globular domain that binds to the active site of thrombin and a C-terminal tail that fills a groove on its surface. The tail is disordered in free hirudin and it undergoes a disorder–order transition upon binding [3, 4].

In contrast, average-sized interfaces most commonly involve only small changes to surface loops and side-chain positions. Protein–nucleic acid interactions are an exception to this trend, more commonly involving significant conformational changes than protein–protein interactions of comparable size [4–6]. It should be noted that the average size of protein–DNA interfaces closely mirrors that of protein–protein complexes (on a per-monomer basis) [7], although many of the former interactions involve multiple domains, through either dimerization or the repetition of modules within a single monomer. As a consequence, protein–DNA interfaces are, on average, twice as large as protein–protein interfaces, covering 2200–3100 Å^2 [2]. It is thought that a large interface area is required for sequence-specific recognition, so that enough nucleotide bases are involved in the contact. In support of this, non-sequence-specific DNase I utilizes a 1600 Å^2 interface, whereas the sequence-specific *Eco*RV restriction enzyme uses one almost 3 times larger [2]. It is common for DNA-binding proteins to contact DNA via a number of domains connected by flexible linkers. This is also a feature of RNA recognition [8]. The use of multiple domains enables recognition of a larger number of nucleotides and an increased binding affinity, without restricting the conformational flexibility of the DNA.

The chemical properties of protein–protein interfaces depend on whether the complex is obligate or transient. For example, interface hydrophobicity has been

reported to be intermediate between that of the protein core and the surface, with more permanent complexes being more hydrophobic than transient ones [9]. Separate studies conclude that obligate homodimer interfaces are even more hydrophobic, resembling protein cores [10]. If only transient interfaces are considered, the hydrophobicity of the interface appears to be similar overall to the remaining protein surface [1]. Protein–nucleic acid interactions tend to be less hydrophobic than protein–protein interactions, as expected given the charged nature of nucleic acids [5].

Some studies suggest that the propensity for each amino acid to appear at an interface is similar to the general propensity for that residue to appear on any part of a protein surface [11], while others suggest that residue propensities are a meaningful parameter for defining surfaces [1, 9]. Again, the apparent discrepancy between these conclusions probably arises from the choice of proteins in these studies. Interface amino acid propensities are more similar to the protein core in obligatory complexes and to the surface in transient complexes [12], whereas there are more polar residues at protein–nucleic acid interfaces than at protein–protein interfaces [7].

The number of hydrogen bonds in a protein–protein interface is roughly correlated with its size [1] and polar groups that become buried at interfaces almost always form intermolecular hydrogen bonds [13]. Protein–nucleic acid interactions tend to have more hydrogen bonds than protein–protein interfaces, with protein DNA interfaces having one hydrogen bond for every 125 $Å^2$ and protein–protein interfaces one every 170 $Å^2$. Further, two-thirds of protein–protein hydrogen bonds involve the backbone, whereas 40% of protein–DNA hydrogen bonds involve lysine and arginine side-chains [5].

Unlike enzyme active sites, protein–protein interfaces are relatively planar and transient interactions have even more planar interfaces than those of obligate complexes [9]. Interfaces have no preference for particular secondary structure elements and are usually made up from several segments of the primary sequence rather than a continuous stretch [9].

Water is present at both protein–protein and protein–nucleic acid interfaces. In some structures, a ring of water exists around a dry central interface. In others, the entire interface is "wet" with water packing cavities and these ordered water molecules can be packed as tightly as the protein core [1]. Interface water is important for hydrogen bonding, with water-mediated hydrogen bonds being as common as direct ones [14]. Protein–DNA interfaces tend to contain more buried water than protein–protein interfaces [7].

12.2.2
Modularity and Promiscuity in Protein Interactions

Whereas most protein–protein complexes for which detailed structural data are available are simple dimers involving a single interface, protein–nucleic acid interactions often use modularity to increase the specificity and/or affinity of the interaction [7]. Linkers between domains determine the distance between recognized

nucleotides and may additionally interact with the nucleic acid backbone [15]. It is notable that the individual modules in protein–nucleic acid interfaces tend to form interfaces of approximately the standard size; it is possible that the standard-sized interface exists because a certain size is required to exclude bulk solvent from the interface, keeping only ordered water [11]. Some clearly modular protein–protein interactions are also known. For example, tandemly arrayed LIM domain and fibronectin type 1 modules both recognize extended peptide motifs with high affinity [16, 17].

Analysis of large-scale protein interaction studies (e.g., [18, 19]) reveals that some proteins have few interaction partners, whereas others act as so-called hubs, making contacts with many interaction partners [20]. Hub proteins could either bind multiple partners using separate interaction domains or surfaces, or by having an adaptable interface. Interaction studies show that protein connectivity is not correlated with number of domains or with the size of the interface. This suggests that having an adaptable face, rather than multiple interaction faces, is a common way for a hub protein to interact with several partners [21].

12.2.3 Hotspots at Interfaces

A large number of mutational studies have made it clear that not all residues at an interface contribute equally to the interaction affinity. In their seminal work on the interaction between human growth hormone and its receptor, DeLano *et al*. introduced the term "hotspot" to describe interface residues that, when mutated to alanine, result in a drop in binding affinity of 10-fold or higher [22]. Hotspots are usually identified by alanine-scanning mutagenesis, whereby each residue at an interface is mutated to alanine in turn. Alanine is used since it eliminates the side-chain beyond the β-carbon without changing the main-chain conformation and is commonly found in both buried and exposed parts of proteins [23].

A follow-up study that scanned with a more diverse set of residues, rather than just alanine, showed some unexpected patterns at the interface between human growth hormone and its receptor [24]. Surprisingly, conserved residues were not unusually sensitive to mutation and chemically conservative mutations were often not tolerated. In contrast, more radical mutations, other than changes to proline and cysteine, were frequently tolerated. Hotspots previously identified by alanine scanning had the strictest requirements for substitution, with the wild-type amino acid being most favorable [24]. This conclusion is supported by a meta-analysis of alanine-scanning data from the literature, which concluded that evolutionarily conserved residues at interfaces are more likely to be hotspots [25].

Hotspots are enriched in tyrosine, tryptophan, and arginine [11], and it is possible that these residues are favored because they can form more than one type of interaction; for example, tyrosine can participate in both π-interactions and hydrogen bonds. It has also been suggested that hotspots tend to be located in clusters known as hot regions [21]. An analysis of 75 protein–protein complexes revealed that hotspot residues within hot regions are tightly packed, so that mutating a single residue is

likely to disrupt its neighbors, whereas the hot regions are independent of one another. Thus, hotspot residues appear to be preorganized into the conformation required for binding and often reside at the bottom of pockets, becoming buried upon binding [21]. In contrast, the position of hot regions is less well conserved between structurally related complexes, perhaps allowing for some adaptability at the interface.

In protein–protein interfaces, hotspots are more concentrated at the center of the interface. Interactions are more resilient to mutations in residues on the rim of the interface, since such changes often maintain solvent exclusion from the center of the interface [11]. This "o-ring" hypothesis, deriving from an analysis of protein complex structure, states that an interface is made up of hydrophobic residues surrounded by a polar ring; this arrangement gives rise to a dry environment at the center of the interface that is favorable for hydrophobic interactions and direct hydrogen bonds. It has also been proposed that amino acid propensities in the core but not the rim differ from the remaining protein surface [12]. Core residues are also less resilient to mutation and more hydrophobic [26], even though not all core residues are hotspots.

In summary, interaction interfaces do have some properties in common, but remain difficult to predict accurately. Efforts to improve such predictions are ongoing, and include the biennial CAPRI (Critical Assessment of Prediction of Interactions) project, in which groups attempt to predict the structures of protein–protein complexes for which structural coordinates have not yet been released [27].

12.3
Affinity of Protein Interactions

12.3.1
Introduction

Protein interactions are hugely diverse and cover a wide range of affinities. Early studies focused on tight complexes and it was widely held that weaker interactions were unlikely to be physiologically relevant, because of the low concentrations at which proteins are generally observed in cells. However, it has become clear in recent times that different affinities are required to perform a wide range of biological functions and, in particular, that weak interactions are responsible for many important processes *in vivo*. Here, we discuss the correlation between affinity and function.

The affinity between two components in a binary, noncovalent complex is most commonly described by the equilibrium dissociation constant, K_d, although its inverse, the equilibrium association constant, K_a, is also widely used. In simple cases, K_d is equal to the ratio k_d/k_a, where k_a and k_d are the rate constants for association ("on-rate") and dissociation ("off-rate"), respectively [4]. Of these two parameters, the dissociation rate constant varies the most between complexes (in the

range 10^6 to 10^{-7} s^{-1}), and hence has the greatest effect on K_d. The rate of association is determined in part by the rate of diffusion of the components, which is often quoted at 10^9 M^{-1} s^{-1} [28]. However, taking into account the fact that only a small patch on each protein is reactive, Berg has calculated that an on-rate of 10^6 M^{-1} s^{-1} is more appropriate [29]. Measured values for k_a generally fall in the range 10^5–10^8 M^{-1} s^{-1} (implying a K_d range of $\sim 10^{-3}$–10^{-14} M). Note that on-rates that exceed the diffusion-imposed limit calculated by Berg can arise from the existence of long-range electrostatic attraction, or steering. Recent work, including elegant nuclear magnetic resonance (NMR) studies centered on paramagnetic relaxation enhancement effects [30], indicates that the association phase in the formation of macromolecular complexes might be a two-step process involving the initial formation of a nonspecific encounter complex and subsequent reorganization into the final specific complex. It is thought that the formation of encounter complexes revolves around long-range electrostatic interactions that assist in bringing the two molecules together and reducing the dimensionality of the search process [28, 31–33].

12.3.2 "Irreversible" Interactions

For some biological functions, strong, irreversible complex formation is essential, and such complexes can be divided into obligate (in which the protomers are not functional outside the context of the complex) and nonobligate assemblies. Multisubunit enzymes [34], virus capsids, and muscle fibers [1] are examples of obligate complexes. The protomers of such complexes are often expressed simultaneously; for example, the nonidentical subunits of cathepsin D are encoded by genes under the same promoter [35], allowing the protomers to cofold and bind simultaneously [9, 34].

High-affinity, nonobligate complexes are also extremely stable once formed; however, the components are often expressed separately and are independently stable [36]. Complexes in this category include those between barnase (an extracellular ribonuclease of *Bacillus amyloliquefaciens*) and its polypeptide inhibitor barstar [37], and between colicin-specific immunity protein and DNase [38], which are among the tightest known protein–protein interactions (both with dissociation constants of 10^{-14} M). Many broad classes of protein interactions exhibit picomolar to nanomolar affinities with slow off-rates, including antigen–antibody, protein–DNA, and enzyme–inhibitor complexes.

12.3.3 Regulatory Interactions

Proteins involved in regulatory functions, such as intracellular signaling and the regulation of gene expression, often need to make and break interactions in a reversible manner. This outcome can be achieved through the use of low- to medium-affinity interactions, which are made and broken on short (millisecond to second) time scales, and thereby allow alternative binding partners to compete [36]. Such a situation allows effective regulation by local environmental factors, such as the

concentrations of potential binding partners and the introduction of covalent modifications into one or both partners (see Section 12.5).

Strong interactions can also be used in regulatory roles if an appropriate molecular trigger, such as the binding of nucleotide or metal-ion cofactors, can disrupt binding. The heterotrimeric G-protein, for example, forms a stable trimer with GDP bound, but dissociates into the G_α and $G_{\beta\gamma}$ subunits upon binding of GTP. This GTP/GDP exchange results in a 1000-fold change in affinity between the two components [36] and acts as a switch mechanism in the control of signal transduction.

12.3.4
Ultra-Weak Interactions

Recently, a new class of short-lived "ultra-weak" protein–protein complexes, prominently featuring electron transport proteins, has come to light. Biological energy transduction is performed by chains of redox proteins that combine to provide a path for the controlled flow of electrons. In these processes, soluble redox proteins facilitate electron transport between membrane-bound protein complexes. For example, plastocyanin and cytochrome c_6 function as electron shuttles between cytochrome *f* and photosystem I in the photosynthetic redox chain, and high-turnover conditions are necessary to sustain a continuous electron current. To achieve this, electron transfer occurs between partners in low-affinity, transient complexes that exist for only milliseconds or less [36, 39, 40].

Specificity is an important consideration for low-affinity, transient interactions. As redox proteins are required to transfer electrons to and from multiple partners with comparable affinities [39], it is actually important for these interactions to be relatively nonspecific. For example, affinities for the interaction of *Phormidium* cytochrome *f* with both physiological and nonphysiological partners have been determined, and range from 10^2 to $10^4\ M^{-1}$. It is notable that the tightest complex was formed with the nonphysiological partner, yeast cytochrome *c* [41].

As might be expected, transient electron-transfer complexes (e.g., plastocyanin–-cytochrome *f* [42, 43]) often interact via a small interface of approximately 600 $Å^2$, and consistent with the o-ring hypothesis this interface consists of a hydrophobic core and a polar periphery [39]. Despite the small interface, it is thought that electrostatic steering plays a role in establishing relatively rapid on-rates; for example, the association rate constants for the plastocyanin–cytochrome *f* complex ($K_d =$ 1 mM [44]) and the barnase–barstar complex ($K_d = 10$ fM [37]) are comparable ($\sim 10^8\ M^{-1}\ s^{-1}$ [39]).

A further similarity between ultra-weak complexes is low geometric complementarity and poor packing of side-chains at the interface [45]. If the side-chains are not closely packed, the interface remains permeable to water molecules which can "rewet" the binding site and disrupt the complex. Furthermore, van der Waals contacts across the interface cannot be maximized and so their contribution to the binding affinity will not be optimal. Together, these features promote fast dissociation and enable transiency in the interactions [39]. Indeed, ultra-weak complexes might, in at least some instances, resemble the encounter complexes

predicted for strong interactions, insofar as they comprise a dynamic ensemble of orientations rather than a single well-defined state [46]. NMR data collected on the transient complex of plastocyanin–cytochrome *f* can not be satisfied by a single structure, but only by a family of structures displaying significant variability, indicating that the complex is rather dynamic [46].

12.4 Measuring Protein Interactions

12.4.1 Introduction

There are numerous methods available for detecting or characterizing protein interactions. All of these methods have advantages and disadvantages, and deliver a different range of parameters, ranging from simply the detection of some sort of association (not necessarily direct), to a full detailed thermodynamic analysis. In the last few years several reviews [47] and large volumes have been written that provide a much more complete coverage of these methods (e.g., [48]). Here, we summarize a selection of these techniques and provide references from which a more detailed exploration can be instigated.

12.4.2 Discovering/Establishing Protein Interactions

The initial discovery of protein interactions is often made from some type of screening strategy. Two of the most common strategies are yeast two-hybrid (Y2H) and pulldown experiments. The Y2H assay is one of many protein complementation approaches that can detect interactions [49], in this case in the context of a yeast nucleus. The two domains of a yeast transcription factor are separately fused to the proteins in question and form a functional unit only when the partner proteins interact. This is detected through the use of a reporter gene that selects, for example, for yeast growth. Y2H can be used to screen for new partners (by screening a "bait" protein against a library of "prey" proteins) or to analyze a known interaction in detail.

Pulldown experiments are also frequently used both to discover new interactions and to dissect existing interactions. Most commonly, one protein is expressed as a glutathione-*S*-transferase (GST) fusion, immobilized on glutathione beads and treated with the other protein(s), either purified or in a cell lysate; sodium dodecylsulfate–polyacrylamide gel electrophoresis is used to detect captured proteins (e.g., [50]). Care must be taken in the interpretation of these experiments, however, because they are prone to false positives. Coimmunoprecipitation experiments similarly use antibodies to pull protein complexes out of cell lysates [51].

Electrophoretic mobility shift assay (EMSA, or gel shift) experiments use non-denaturing electrophoresis to compare the hydrodynamic and charge properties of DNA with or without bound proteins. DNA carries a negative charge and rapidly

migrates to the anode under electrophoresis. A protein bound to DNA will have a larger size and a different shape, and will (usually) be less electronegative and thus migrate more slowly on the gel than the free DNA. Protein–protein interactions can be detected in a similar way on native gels. EMSAs can also be used to estimate dissociation constants for tight interactions (below $\sim 10^{-7}$ M) by carrying out titrations, but the same approach is not common for protein–protein interactions.

Confocal microscopy combined with Förster resonance energy transfer (FRET) experiments [52] can be used to assess direct protein–protein interactions inside living cells. The genes for target proteins are modified so that the proteins are each tagged with different variants of Green Fluorescent Protein. The excitation and emission wavelengths of the fluorescent proteins are chosen such that the emission spectrum of one can excite the fluorescence of the other (the FRET effect). Only when the target proteins interact will the proteins be brought sufficiently close to generate a FRET signal within the cell.

12.4.3 Determining Interaction Stoichiometry

A number of experimental approaches allow an estimate to be made of complex stoichiometry, but are less well suited for the determination of dissociation constants.

In size-exclusion chromatography (SEC; or gel filtration), molecules are separated on the basis of their hydrodynamic properties. Provided a complex is of sufficient stability (K_d below $\sim 10^{-7}$ M), an interaction can be detected as a new peak in the chromatogram of a mixture of the protein and its partner, compared with chromatograms of the isolated species. The relative stoichiometry of the complex can be estimated by applying different ratios of the two components to the column and determining the ratio where only complex is formed (e.g. [53]). Rough estimates of K_d can sometimes be made by applying samples of different concentrations; the concentration where $\sim$50% of the protein is in the complex form corresponds approximately to the K_d.

Classical light (or Rayleigh) scattering involves the scattering of light from particles that are small compared to the wavelength of the incident light. For most soluble proteins and protein complexes, this scattering is independent of shape. Multi-angle laser light scattering (MALLS) measures Rayleigh scattering at multiple wavelengths, which allows the weight average molecular weight of proteins to be determined. MALLS is often used in-line with SEC to give good estimates of the molecular weight of eluting species [54]. Small angle X-ray scattering and small angle neutron scattering are other solution methods that can be used to estimate the molecular weight of proteins and protein complexes (e.g., [55]). These simple but powerful approaches can additionally provide accurate low-resolution structural models [55].

Using the relatively soft ionization techniques of electrospray and matrix-assisted laser desorption/ionization, mass spectrometry (MS) can be used to ionize intact noncovalent macromolecular protein complexes and thereby determine their stoichiometry [56]. This approach is quite new but has wide applicability because of the very sensitive nature of the instrument.

The analytical ultracentrifuge is a well established tool for the measurement of complex stoichiometry. Samples are subjected to a centrifugal field, and the distribution of the solute as it experiences buoyant, frictional, and centrifugal forces is observed, either as a function over time (sedimentation velocity) or once it has reached equilibrium (sedimentation equilibrium). Measurements are made using absorption, interference, or fluorescence optics, depending on the instrument. For heterologous protein complex formation, velocity experiments are probably most useful for detecting the presence of an interaction or estimating the molecular weight of a stable complex. Equilibrium experiments, although longer to run (days versus hours), can provide a very good estimate of shape-independent molecular weight without reference to molecular weight calibration standards. Dissociation constants can also be obtained for both homo- or heteromeric complexes, although the analysis can be very complicated in the latter case [57].

12.4.4 Measuring Affinities

In many cases, it is of interest to measure the affinity of an interaction. There are an increasing number of techniques available that are sufficiently sensitive to be practical, given the typical yields available for recombinant proteins.

Fluorescence anisotropy is related to the size (and shape) of a molecule. Thus, it can used to measure the K_d of a protein interaction by monitoring the formation of a protein complex that is significantly larger than the uncomplexed molecule [58]. In some cases intrinsic tryptophan fluorescence can be monitored; however, proteins or their ligands can be labeled with a suitable fluorescent tag.

There are many different forms of enzyme-linked immunosorbent assay (ELISA) that can be used to detect protein interactions. Since ELISAs involve numerous washing steps, the approach is best suited to strong interactions [59].

Isothermal titration calorimetry (ITC) measures the uptake or evolution of heat as one molecule (the "ligand") is titrated into its partner (or vice versa) [60]. A binding isotherm of the evolved heat as a function of the concentration of added ligand is generated, which can be fitted to derive K_d (for interactions in the range $\sim 10^{-4}$–10^{-8} M), the stoichiometry, and the ΔH of the interaction; ΔS and ΔG can also be calculated from the data. ITC requires carefully matched buffers (to avoid complications from heats of dilution or heats of ionization) and relatively large amounts of soluble protein (usually milligrams of both components).

Surface plasmon resonance is one of the few methodologies that can determine the on- and off-rates for an interaction [61]. A protein is immobilized onto a functionalized gold chip, using one of a range of different chemistries. The chip contains fluid channels though which the ligand is passed and binding over time (k_{on}) is detected by changes in the refractive index of the surface that occur upon binding. After equilibrium is reached, buffer is passed through the channel and the dissociation (k_{off}) process is measured. This technique has been multiplexed and automated, allowing examination of multiple interactions/mutants/ligands without user

intervention: quantitative analysis is generally straightforward, but becomes very difficult in cases other than simple 1 : 1 binding.

NMR spectroscopy can be used in a number of different ways to provide both structural and quantitative information on protein interactions. Chemical shift perturbation assays [62] are the most commonly employed approach: ^{15}N-heteronuclear single quantum coherence (HSQC) spectra of a ^{15}N-labeled recombinant protein are recorded during the course of the addition of an unlabeled partner. As the backbone amide of each residue gives rise to a single signal in the spectrum, those peaks that correspond to residues at the binding interface tend to shift, providing a rapid assessment of the binding interaction surface. In favorable cases, titration data can also be used to calculate a dissociation constant for the interaction. Other approaches, including the use of lanthanide tags [63] and nitroxide spin labels [30], have become popular for the rapid mapping of protein interaction interfaces.

12.4.5
Modulation of Binding Affinity

It is worth noting that binding affinities measured for purified components in dilute aqueous solution can differ from the effective affinity *in vivo* through differences either in the effective local concentration of the two components or the physicochemical environment. In vivo thus, the available amounts of each species can be modulated by changes in gene expression, protein degradation, secretion or localization. Proteins may be coexpressed or colocalized within a compartment, or may require directed diffusion or transport. Anchoring of one or both proteins in a membrane or other structural complex can also help to increase the effective local concentration [36].

The local physicochemical environment includes the concentration of unrelated macromolecules, ions, or other effecter molecules (e.g., ATP, Ca^{2+}), as well as changes in pH and temperature [36]. Molecular crowding (also known as the excluded volume effect), which reduces the effective volume available to the species in question, will increase the effective association constant; this effect is thought to be important in the crowded cellular milieu, although the magnitude of the effect is currently unclear [64].

12.5
Coupled Folding and Binding

12.5.1
Introduction

Recent work has revealed that many protein interactions involve one or more proteins that are either partially or totally disordered in the absence of their partner [65]. Here, we discuss the role of unstructured regions and proteins in protein–protein and protein–nucleic acid interactions.

12.5.2 Characteristics of Intrinsically Unstructured Proteins

There is no single line that separates structured and intrinsically disordered proteins. At one end of the continuum are folded domains, followed by multidomain proteins interspersed with disordered regions, then compact but disordered molten globules and finally extended unstructured states [66]. Most intrinsically unstructured proteins do not lie entirely to the latter end of the spectrum, but rather contain some residual structure [67], which most likely provides initiation points for folding and/or binding under the appropriate conditions.

Overall, unstructured regions are sufficiently different from structured regions in their properties that disorder can be predicted quite reliably from sequence data alone. Many programs that predict disorder, such as PONDR [68] and DISOPRED [69], use algorithms that are trained on proteins from the PDB, under the assumption that missing electron density indicates a disordered sequence. Some of the parameters that distinguish unstructured proteins are low sequence complexity and biased amino acid composition. Residues that are over-represented in disordered regions are glutamine, serine, proline, glutamic acid, lysine, glycine, alanine, and arginine [66, 70], whereas cysteine, valine, leucine, isoleucine, methionine, phenylalanine, tyrosine, and tryptophan correlate more strongly with folded domains. Thus, disordered regions have a large proportion of charged and polar residues at the expense of the bulky hydrophobic residues normally found in hydrophobic cores [71]. Some particular low-complexity motifs are observed in a substantial number of proteins and appear to have specific functions; for example, the PEST motif (which is rich in proline, glutamic acid, serine, threonine, and aspartic acid) appears to reduce the half-life of the parent protein. The motif is recognized by proteases in the caspase-, calpain-, and ubiquitin-mediated degradation pathways [70].

A slightly different class of disorder arises in short, linear motifs that have been identified from sequence analysis in eukaryotic genomes. Two studies that point out the importance of these regions in binding sites find that they have similar properties. Molecular Recognition Features [72] and Eukaryotic Linear Motifs [73] are both characterized as short disordered regions that have more hydrophobic residues than other disordered regions. These sequences are thought to play an important role in recognition and can perhaps be considered to be "disordered scaffolds" that present a set of binding residues in a manner that can adapt to binding partners with a range of structures.

12.5.3 Advantages of Disorder for Protein Recognition

The proportion of disordered proteins increases strikingly with organism complexity. Only 2% of archaean and 4% of eubacterial proteins have disordered regions longer than 30 residues, while 33% of eukaryotic proteins have this level of disorder [69]. Examining the gene ontology for disordered proteins suggests that they are often

involved in functions that are expanded in eukaryotes. Biosynthesis and metabolism rarely utilize unstructured proteins, whereas such proteins often function in signaling, transcriptional regulation and cytoskeletal organization.

Alternative splicing, which is more common in higher organisms, might also have a bearing on the distribution of disordered regions in proteins. A study by Romero *et al*. using 46 alternatively spliced genes found that 81% were either fully or partially disordered [74]. Further, of the five known pairs of structures of alternatively spliced protein isoforms, the regions associated with alternative splicing were missing from electron density maps in three cases and are likely to be disordered (the other two cases involve only very short alternatively spliced sections). A correlation between alternative splicing and disorder makes sense; such elements can perhaps be more readily spliced in and out without affecting the global fold of the protein.

With a few exceptions, disordered regions appear to be mutated faster than ordered ones; a study of disorder in 26 protein families found that in 19 families disordered sections of the protein chain diverged significantly faster than the ordered regions [75]. This phenomenon presumably occurs at least in part because there is no need to maintain a stable structure and could potentially allow for the rapid diversification of a process or pathway (e.g., neuronal development) during evolution.

An unfolded protein can also present a large interface with a relatively short sequence, compared to what is possible with a typical globular protein. For example, in the complex formed between CREB-binding protein (CBP) and activator for thyroid hormone and retinoid receptors (ACTR) [76], both proteins are disordered in the absence of the other, but fold together upon complex formation to form an extensive interaction surface (Figure 12.1). This arrangement represents remarkable economy, and it has been suggested that using unfolded proteins allows the cell to reduce the total amount of sequence and protein bulk required for its functions [77].

Intrinsically unstructured proteins tend to form interactions which are of low affinity, but high specificity. When a disordered protein folds on binding there is a high entropic cost associated with becoming ordered and this must be compensated for enthalpically. The requirement for favorable enthalpy changes means that specificity is favored [66]. Thus, highly specific interactions can be created that are still readily reversible.

Disorder can also favor the addition of PTMs by allowing modifying enzymes access to the protein. For example, phosphorylation occurs predominantly at sites that are intrinsically disordered [78]. Similarly, histone tails, which are disordered, are subject to a variety of PTMs [66].

Finally, it has been suggested that coupled folding and binding can speed up the search for a binding partner. Whereas a folded protein has to approach its binding partner to within a few Ångstroms at approximately the right orientation in order to bind, disorder potentially allows more freedom. A completely unfolded protein can approach within its radius of gyration to make contact with a binding partner. These contacts will be very weak, since in the unfolded state the binding residues cannot act cooperatively; however, their strength will increase on folding. The process

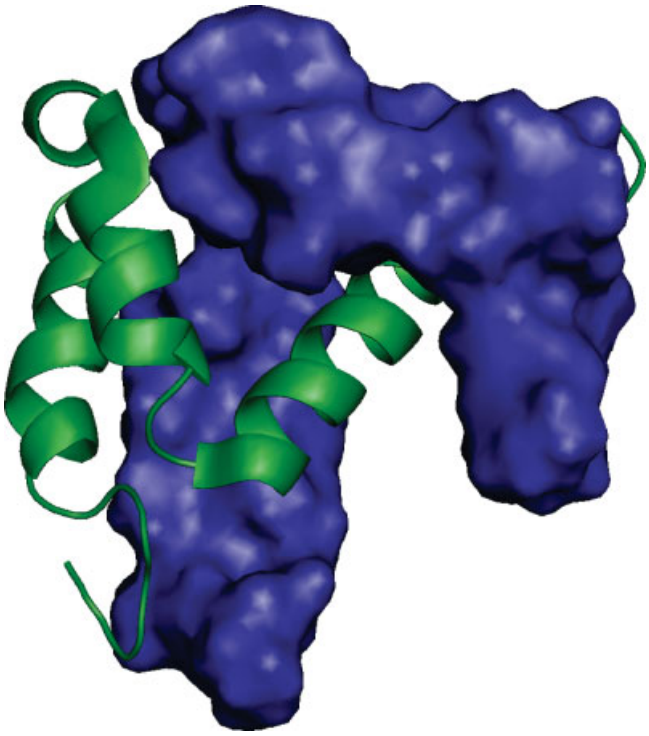


Figure 12.1 Structure of the complex formed between CBP (blue surface) and ACTR (green ribbon), illustrating the highly intertwined nature of the fold. Both domains are poorly ordered in the absence of their partner. PDB ID: 1kbl.

of a protein casting its chain out and "reeling in its catch" by folding, the fly-casting mechanism [79], proposes that binding is accelerated for components that are at low concentration; for example, a transcription factor could find a low-copy-number gene more quickly by being at least partially unfolded.

12.5.4 Diversity in Coupled Folding and Binding

The coupling of folding and binding can occur to many different degrees. At one extreme, the entire sequence of one or both partners undergoes a disorder to order transition, whereas at the other extreme a small flexible linker between domains can become ordered upon complex formation. As noted above, the interaction between CBP and ACTR affords a dramatic example of the phenomenon – the interaction domains of both proteins are disordered in isolation and fold together to form an intricate interaction surface. In contrast, the interaction between the transcriptional regulator TFIIIA and its DNA target involves the ordering only of flexible linkers between three zinc finger domains. This ordering is functionally important, however;

the ordered linkers pack against the adjacent zinc fingers, providing additional stability to the complex [80]. A third class of interactions involves short, disordered peptide motifs that recognize ordered domains. For example, as noted in Section 12.2.2, both LIM domains and fibronectin type 1 modules recognize short, disordered peptides that build on existing secondary structure upon formation of a complex.

12.6 Regulation of Interactions by PTMs

12.6.1 Introduction

Proteins can undergo a multitude of reversible covalent modifications (PTMs) and these modifications can potentially affect most or all of the properties of a protein, including stability and subcellular localization. For example, glycosylation of extracellular proteins is thought to improve stability against proteolysis [81, 82], while the addition of the small protein SUMO to lysine side-chains of target proteins can direct the SUMOylated protein to the cytoplasm [83]. PTMs are also an integral aspect of protein–protein recognition and provide a dynamic means for regulating cellular interactions. There are in general three aspects to the action of PTMs in protein recognition: the specific peptide sequences onto which modifications are introduced, the recognition domains that recognize peptides bearing specific modifications, and the enzymes that introduce and remove PTMs. All of these provide opportunities for regulation; for example, activation or inactivation of the modifying enzymes will have downstream effects on the network of interactions that are regulated by that enzyme.

12.6.2 Types of PTMs

The list of known protein PTMs is extensive and still growing. In Table 12.1, we have listed a number of PTMs that have been demonstrated to function by regulating

Table 12.1 PTMs and the domains that recognize them.

Modification	Residues affected	Recognition domains	References
Phosphorylation	Tyr, Thr, Ser	SH2, PTB, C2, MH2, WD40, WW, BRCT, FF, SRI, CID	[149]
Methylation	Lys, Arg	chomodomain, PHD, WD40, TUDOR, MBT	[150, 151]
Acetylation	Lys, Ser, Thr	bromodomains	[152, 153]
Hydroxylation	Pro, Asp, Glu		[154, 155]
Ubiquitination	Lys	UBM, UBA	[156]
SUMOylation	Lys	SIM	[157]

protein interactions. Probably the most extensively studied modification in this regard is the phosphorylation of serine, threonine, tyrosine, and histidine that underpins the transduction of extracellular signals to the cytoplasm and ultimately to the nucleus. PTMs of histones, which include lysine, arginine, serine, and threonine derivatives, have also received a great deal of attention in recent times for their role in regulating gene expression.

12.6.3 A Case Study – Histone Modifications

Over 60 modifications (or marks) to histone proteins, mostly localized to their disordered N-terminal tails, have been cataloged and the functional consequences of these modifications are only partially understood. Overall, it is clear that these modifications, which generally appear to be reversible, are responsible for determining the transcriptional status of nearby genes and that this regulation is achieved by the recruitment of other proteins to specific marks. The acetylation and methylation of lysine side-chains has perhaps received the most attention. At least 11 lysines on H3 alone can be methylated or acetylated, and different levels of methylation (mono-, di-, and tri-) have also been observed. One of the best established examples is the trimethylation of K9 of histone H3 (H3K9Me3); this mark can be introduced by one of many lysine methyltransferase enzymes (most prominently Suv39h1 [84]) and is recognized by heterochromatin protein 1 (HP1) through an N-terminal chromodomain on the latter protein [85]. The presence of HP1 is associated with transcriptional repression and the formation of heterochromatin [86]. Demethylase enzymes that can act on H3K9Me3, including JMJD2A [87], have also been described, suggesting that heterochromatin formation involves (among other things) a dynamic balance between methylation at H3K9, recognition of the H3K9Me3 mark, and demethylation. Interplay clearly also exists with other marks; for example, phosphorylation of the adjacent residue S10 on H3 reduces the affinity of HP1 for H3K9Me3 by two orders of magnitude [88]. It is also notable that, although the chromodomains from several other proteins also recognize methylated histone tails, some members of the family appear to play roles in DNA or RNA recognition [89].

In contrast, the presence of a trimethylated derivative of H3K4 is associated with actively transcribed genes. This modification can be introduced by enzymes such as Set1 [90] and is recognized by domains from several different fold families, including a tandem chromodomain in the chromatin remodeling enzyme CHD1 [91], PHD domains in BPTF [92] and ING2 [93], and the β-propeller WD40 repeat in WDR5 [94].

An analysis of the architectures of proteins that recognize histone marks reveals that many harbor multiple PTM-recognition domains separated by flexible linkers (a feature that they share with many other proteins involved in transcriptional regulation). For example, Brd2 and TAFII250 each contain double bromodomains and bind H3 and H4 tails that bear two acetylated lysine marks, respectively [95, 96]. Similarly, Polybromo (also known as BAF180) contains six bromodomains [97], although the contributions made by each of these domains to Polybromo activity

is not fully resolved. This modular structure allows recognition of specific combinations of modifications. Many other modifications and the associated proteins that mediate the addition, removal, and recognition of these marks have been described over the last 10 years; it remains the case, however, that the biology underlying these recognition events still remains largely obscure.

12.7
Engineering and Inhibiting Protein–Protein Interactions

12.7.1
Introduction

Owing to the clear importance of protein interactions in biology, a great deal of effort has gone both into developing novel protein interactions and into inhibiting existing interactions. This section summarizes current efforts to design new proteins that are capable of binding to a chosen target with high affinity and specificity. One of the main reasons for creating such new proteins is to block an existing and undesirable interaction in a medical context. We provide several examples where this type of strategy has already moved from the laboratory to a clinical setting and also briefly discuss the increasing use of small molecules to inhibit protein interactions.

12.7.2
Engineering Proteins with a Specific Binding Functionality

The field of protein design has taken enormous strides. Computational strategies have allowed stable proteins to be designed from scratch [98] and existing proteins to be engineered to have dramatically improved thermal stability [99] or altered enzymatic activity [100]. To date, efforts to design proteins with a tailored binding function have focused on modifying the surface of an existing protein fold that acts effectively as a scaffold to present a novel binding surface. Combinatorial approaches such as phage or ribosome display are employed so that a large fraction of sequence space can be searched.

Many different protein folds have been used as scaffolds in this manner, including thioredoxin, lipocalins, and protein A. The best established class of "designer" binding proteins is the monoclonal antibodies and the vast number of clinical trials currently in progress for drugs based on these molecules is a testament to the feasibility of the approach. The complementarity determining regions from antibodies raised in animals are grafted onto human scaffolds to reduce immunogenicity, and then optimized using display technologies [101]. Herceptin, a therapeutic antibody that targets the Her2 receptor and is frequently overexpressed in breast cancers, was created in this manner. As well as inhibiting the interaction between Her2 and its signaling ligands, the antibody activates cellular cytotoxicity pathways [102], inducing natural killer cells (which bind the Fc fragment of the antibody) to kill the target cell.

The development of nonantibody scaffolds for protein design has emerged as a powerful technology and the observation that many of the more promising scaffolds have been taken up by biotechnology/biotech start-up companies is an indication of the maturity of the field [103, 104]. Suitable scaffolds consist of a compact structural core, together with a loop or surface that can be randomized by mutagenesis while maintaining the structural integrity of the core. They must be small, highly stable and able to display specificity towards different types of targets. One category of scaffolds that has received much attention is the repeat proteins, including ankyrin (ANK, known commercially as DARPins) and tetratricopeptide (TPR) repeats [105, 106]. As the name suggests, these proteins consist of consecutive repeats of a basic structural motif, usually 20–40 amino acids in length, which assemble to form elongated domains with a continuous hydrophobic core. In nature, repeat proteins are found in all intracellular compartments and mediate a diverse range of protein interactions [107, 108]. A number of residues in each repeated module are essential for the fold, while other residues are responsible for mediating interactions and can be randomized to create a combinatorial library of potential binding proteins.

The 33-residue ANK repeat consists of a β-turn followed by two antiparallel α-helices and a short loop [109]. Plückthun *et al.* have created combinatorial libraries containing two or three ANK repeats in which six positions per repeat were randomized, and used ribosome display to select molecules with novel binding specificities in the nanomolar range for both the *E. coli* maltose-binding protein and a number of eukaryotic kinases [110, 111]. TPR repeat proteins display 34-residue repeats that form helix-turn-helix motifs and stack in parallel with adjacent units to form a right-handed superhelix [112]. Cortajarena *et al.* used a three-TPR repeat scaffold (CTPR3) to redesign a binding domain that targets the C-terminal peptide of the heat-shock protein HSP90 (a natural target for many TPR-containing proteins), as proof of principle [113]. Manipulation of residues in direct contact with the peptide produced variants with high affinity and specificity for the peptide.

Rather than starting with a structured scaffold, libraries of phage-displayed peptides can be used to find inhibitors. Hits from such screens can subsequently be incorporated into antibody or other scaffolds or used as a basis to develop peptidomimetics. Small peptides have the advantage of often being membrane-permeable, allowing intracellular targets to be addressed. For example, modified peptides that bind to X-linked inhibitor of apoptosis (XIAP, an intracellular target that inhibits caspase-9; see below) can activate apoptosis in cancer cell lines.

A common approach to improve peptide inhibitors in cases where the target is extracellular is to constrain the peptide library through a disulfide bond, thereby reducing the conformational space available to the peptide. This strategy was employed successfully to find an inhibitor for vascular endothelial growth factor [114] that bound with low micromolar affinity. It is notable that the selected peptides bound in a position that overlaps the binding sites for natural receptors and neutralizing antibodies.

Peptides can also be stabilized against degradation through the introduction of synthetic modifications. For example, "hydrocarbon staples" between side-chains can encourage helix formation. A peptide that mimics a single helix from the

pro-apoptotic BID protein can be used to activate apoptosis since it will bind to BCL-X_L and block its antiapoptotic actions. A hydrocarbon stapled peptide was more helical than an unmodified one and had a 6-fold higher binding affinity [115]. Such staples also increase lipophilicity which can help peptides to be transported across the cell membrane. Stabilization of peptides to degradation can also be achieved by introducing capping groups. Benzyl caps were used to stabilize tripeptides that bind XIAP [116]. The peptides bind to the same groove as the natural binding partner Smac/DIABLO and trigger apoptosis in the same manner by preventing XIAP from inhibiting caspase-9. Finally, it is worth noting that although peptides used in phage display normally lack PTMs, *in vitro* tyrosine phosphorylation has been used to find substrate-mimicking peptides that recognize SH2 domains [117].

12.7.3
Optimizing Protein Interactions

Although combinatorial design and selection is currently the dominant method for engineering proteins with novel binding abilities, considerable progress has been made in the more complex areas of rational and computational design. While there are currently no examples of high-affinity binding sites that have been designed from scratch [118], the affinity of natural binding sites has been successfully increased by rational design. For example, examination of the crystal structure of CD4 in complex with its substrate gp120 led to the prediction of three separate point mutations in CD4 that would create a tighter binder. Each change individually increased the binding affinity, and the triple mutant demonstrated a 40- to 80-fold increase in affinity for the substrate [119].

Several computational algorithms have been developed to aid in the rational manipulation of protein–protein interactions [120]. Some programs, including EGAD [121] and FOLDEF [122], focus on the stability of proteins or protein complexes, and try to predict the change in binding free energy that would result from interface mutations [121, 123]. In contrast, PARE (Predicting Association Rate Enhancement) calculates the electrostatic attraction between two proteins and can be used to predict residue substitutions, both at the binding interface and the periphery, which will increase the rate of association without affecting the dissociation rate [124]. This method was successfully used to gain a 250-fold enhancement in the binding affinity between TEM1 β-lactamase and its protein inhibitor BLIP.

12.7.4
Engineering DNA-Binding Proteins

Just as the design of protein–binding modules is highly desirable, the ability to construct tailored DNA-binding domains has many practical applications, particularly in the areas of gene regulation and gene repair. Most efforts to date have focused on classical zinc fingers, the most abundant DNA-binding domains in eukaryotic genomes. These domains recognize DNA in a modular fashion: each domain in a tandem array can recognize 3–4 base pairs and a small set of amino acids determines

DNA sequence specificity. Pioneering work in several laboratories has yielded domains that can selectively target many of the 64 DNA triplets [125–129], and several companies (e.g., Sangamo and ToolGen) have been founded on the basis of this work, with the goal of creating designer DNA-binding proteins. Publicly available tools for the design of such proteins have also recently been developed: Sander *et al.*, in collaboration with the Zinc Finger Consortium, have created a web-based tool ZiFiT (Zinc Finger Targeter), which is available at http://bindr.gdcb.iastate.edu/ZiFiT [130]. ZiFiT will analyze a user-provided target DNA sequence, determine 18-base-pair motifs that are most suitable for recognition by zinc fingers in the database (a sequence of this length is predicted to occur only once in the human genome by chance [131]), and output the sequence of six zinc finger proteins that are predicted to bind to the chosen DNA sequences.

The ability to generate a protein that can specifically target a unique site in the genome is immensely powerful, as it opens up the possibility for delivering functional modules, such as transcriptional activators or repressors, methylases, nucleases, and integrases, to sites of interest [132, 133]. Nucleases (DNA-cleaving enzymes) have received much attention due to their potential for use in gene therapy; targeting such proteins to the site at which integration of a new gene is required can potentially improve the rate of recombination of the therapeutic gene. In a prominent recent example, Urnov *et al.* designed a zinc finger nuclease targeted towards the interleukin receptor gene *IL2Rγ*, mutations in which are responsible for causing severe combined immune deficiency. Treatment of cultured human T cells using this designed protein resulted in gene correction in a substantially increased percentage of cells compared to controls lacking the zinc finger nuclease [134].

12.7.5 Searching for Small-Molecule Inhibitors of Protein Interactions

Protein–protein interfaces are, on the whole, smoother than enzyme active sites, meaning distinct pockets for the binding of small-molecule drugs are less common. As a consequence, the prevailing dogma has been that the inhibition of protein interactions by small molecules is not a viable drug design strategy [135]. However, a number of recent successes have caused this view to be revised [136].

In a manner that is analogous to the directed phage display libraries described above, the search for small molecule inhibitors generally starts with libraries comprising natural products and/or synthetic chemicals. Chemical libraries are often designed with the goal of achieving maximum chemical diversity. However, this approach may not help maximize the number of hits found since successful drugs are not evenly distributed in chemical space but rather are clustered because of the requirements for absorption, distribution, metabolism, and excretion, combined with issues of toxicity [137] and bioavailability [138].

Libraries of small molecules can be screened in different ways to find drugs. Traditional high-throughput screening was used to obtain inhibitors of the MDM2: p53 interaction. MDM2 binds to the tumor suppressor p53 with high affinity and inhibits its activity as a transcriptional activator. Small molecules of the nutlin

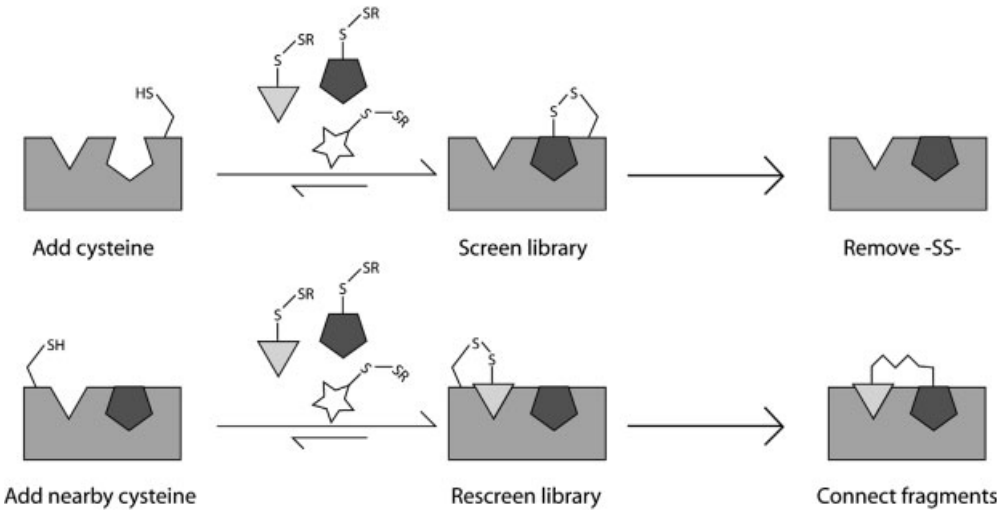


Figure 12.2 Schematic illustration of the tethering strategy for lead discovery. A cysteine is first introduced in the vicinity of the target site and then a library of small (~250 Da) fragments, derivatized with a flexibly linked thiol, is screened against the target. Retained fragments are identified by MS and resynthesized without the thiol group. The process is repeated using additional cysteine mutants until several suitable fragments are obtained. These can be combined by introducing a suitable linker to create a tightly binding compound.

family target the p53 binding site on MDM2, allowing activation of the p53 pathway and subsequent apoptosis of the target cell. An X-ray structure of the MDM2: peptide complex reveals that functional groups of the inhibitor reside in pockets normally occupied by side-chains from p53 [139].

An increasingly common approach is to use fragment based discovery, in which libraries of small fragments are first screened for binding to the target protein. The binding of these fragments is generally rather weak and is improved in subsequent steps by elaboration or by combination with other fragments that target nearby sites. Several approaches exist to identify fragments that target the desired site. Tethering, developed at Sunesis Pharmaceuticals, involves covalently linking the library to the protein through an engineered cysteine in order to amplify weak binding (Figure 12.2). Structure–activity relationships by NMR, pioneered by Steven Fesik at Abbott Laboratories, entail recording ^{15}N-HSQCs of the protein in the presence of small pools of compounds from a library [140]. In this manner, not only can binders be identified, but the surface of the protein to which they bind can be delineated. Two optimized fragments that bind to nearby patches of the target protein can then be selected and used to make linked compounds. This process can increase the affinity of the inhibitor dramatically; for example, Shuker *et al.* linked two fragments with millimolar affinities for FK506-binding protein to create an inhibitor with nanomolar affinity [140]. A very similar approach was used to create a small molecule with a nanomolar affinity for the antiapoptotic protein BCL-X_L [141], with the idea of enhancing apoptosis in cancer cells. The inhibitor, ABT-737, closely

mimics a natural inhibitor of BCL-X_L, BID, in that a carboxyl group mimics an aspartic acid, a fluorophenyl group replaced a leucine in a hydrophobic pocket, and another hydrophobic pocket normally occupied by isoleucine was occupied by a naphthalene derivative.

12.7.6 Flexibility and Allosteric Inhibitors

As noted above, the large planar surfaces of protein–protein interfaces include clusters of residues that are hotspots for binding their partners [22]. Indeed, sometimes multiple partners can bind to different combinations of hotspots at the same interface site. For example, β-lactamase-inhibitory protein binds to several different β-lactamases with different residues becoming important depending on which partner is bound [142]. In a similar vein, inhibitors do not need to bind at an interface in exactly the same way as the native binding partner in order to block binding, although they naturally tend to bind to the same face of the target protein. The hinge region of the Fc fragment of immunoglobulin G is the consensus site for binding by phage-selected peptides as well as for several natural ligands. The interactions with the peptides are very similar to the natural ligands in spite of their different structures, revealing that the hinge is a favorable location on the protein surface for binding. It is fairly hydrophobic and solvent accessible, but the same could be said for other sites on the protein. The key to its promiscuous binding properties is probably its ability to change conformations to suit different binding partners [143].

The range of dissimilar ligands that can bind a particular site depends on the presence of hotspots and the range of conformational change which is possible [144]. Even small movements allow ligands to bind in different ways, and ligand binding can shift the equilibrium in favor of a certain conformation. Mutations can also shift the populations of conformers and prevent an inhibitor from binding with high affinity. Competitive inhibitors of HIV protease bind at the dimer interface, and mutations at the interface can cause rotation between the domains that increases the off-rate for the inhibitor and lead to resistance to the drug [145].

Although the obvious target for an inhibitor of a protein interaction is the interaction surface, allosteric inhibition is also possible if a conformational change is required for ligand binding and an allosteric inhibitor can be found to bind to a hinge region involved in the conformational change. Lymphocyte function-associated antigen-1 (LFA-1) interacts with intracellular adhesion molecules like intercellular adhesive molecule-1 (ICAM-1) to promote T cell activation in the inflammatory response; the interaction occurs only when the former protein is in the "active" conformation. Both hydantoin-based molecules [146] and the cholesterol-lowering drug lovastatin [147] act as allosteric inhibitors, binding to locations remote from the ICAM-1-binding site [148] and thereby preventing the formation of the active conformation of LFA-1. The existence of such inhibitors is a reminder of the role of protein flexibility in recognition processes, and a reason why carrying out *in silico*

screens based on a crystal structure of the target protein and a virtual compound library will not necessarily find the best hits; the best site for an inhibitor to bind may not be visible in the structure.

12.8 Conclusions

The interaction of proteins with other biomacromolecules continues to be an area of enormous activity. In general, applications can only follow after a sound understanding of a system has been attained, and the technical difficulties associated with the detailed analysis of protein–protein and protein–DNA interactions has meant that applications are only now beginning to blossom. The spectacular recent success of designer zinc fingers is an excellent example. Just in the last few years, we have seen the first demonstrations that protein interactions are a plausible target for drug design and it is likely that this area will flourish in coming years, taking advantage of both our growing understanding of the principles underlying protein interactions, and the elegant molecular and structural tools that have recently been developed.

References

1 Lo Conte, L., Chothia, C. and Janin, J. (1999) *Journal of Molecular Biology*, **285**, 2177–2198.

2 Janin, J. (1995) *Proteins: Structure, Function, and Genetics*, **21**, 30–39.

3 Rydel, T.J., Tulinsky, A., Bode, W. and Huber, R. (1991) *Journal of Molecular Biology*, **221**, 583–601.

4 Janin, J. (1995) *Biochimie*, **77**, 497–505.

5 Nadassy, K., Wodak, S.J. and Janin, J. (1999) *Biochemistry*, **38**, 1999–2017.

6 Spolar, R.S. and Record, M.T. Jr (1994) *Science*, **263**, 777–784.

7 Jones, S., van Heyningen, P., Berman, H.M. and Thornton, J.M. (1999) *Journal of Molecular Biology*, **287**, 877–896.

8 Shamoo, Y., Abdul-Manan, N. and Williams, K.R. (1995) *Nucleic Acids Research*, **23**, 725–728.

9 Jones, S. and Thornton, J.M. (1996) *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 13–20.

10 Tsai, C.-J., Lin, S.L., Wolfson, H.J. and Nussinov, R. (1997) *Protein Science: A Publication of the Protein Society*, **6**, 53–64.

11 Bogan, A.A. and Thorn, K.S. (1998) *Journal of Molecular Biology*, **280**, 1–9.

12 Chakrabarti, P. and Janin, J. (2002) *Proteins: Structure, Function, and Genetics*, **47**, 334–343.

13 Janin, J. (2000) Kinetics and thermodynamics of protein–protein interactions, in *Protein–Protein Recognition* (ed. C. Kleanthous), Oxford University Press, Oxford, pp. 1–32.

14 Janin, J. (1999) *Structure*, **7**, R277–R279.

15 Lunde, B.M., Moore, C. and Varani, G. (2007) *Nature Reviews. Molecular Cell Biology*, **8**, 479–490.

16 Bhati, M., Lee, C., Nancarrow, A.L., Lee, M., Craig, V.J., Bach, I., Guss, J.M., Mackay, J.P. and Matthews, J.M. (2008) *The EMBO Journal*, **27**, 2018–2029.

17 Potts, J.R. and Campbell, I.D. (1996) *Matrix Biology*, **15**, 313–320 discussion 321.

18 Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, Y., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S. and Rothberg, J.M. (2000) *Nature*, **403**, 623–627.

19 Gavin, A.-C., Bosche, M., Krause, R., Grandi, P., Marzloch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.-A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. and Superti-Furga, G. (2002) *Nature*, **415**, 141–147.

20 Barabasi, A.-L. and Oltvai, Z.N. (2004) *Nature Reviews. Genetics*, **5**, 101–113.

21 Keskin, O., Ma, B., Rogale, K., Gunasekaran, K. and Nussinov, R. (2005) *Physical Biology*, **2**, S24–S35.

22 DeLano, W.L. (2002) *Current Opinion in Structural Biology*, **12**, 14–20.

23 Cunningham, B.C. and Wells, J.A. (1989) *Science*, **244**, 1081–1085.

24 Pal, G., Kouadio, J.-L., Artis, D.R., Kossiakoff, A.A. and Sidhu, S.S. (2006) *The Journal of Biological Chemistry*, **281**, 22378–22385.

25 Hu, Z., Ma, B., Wolfson, H. and Nussinov, R. (2000) *Proteins*, **39**, 331–342.

26 Guharoy, M. and Chakrabarti, P. (2005) *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 15447–15452.

27 Lensink, M.F., Mendez, R. and Wodak, S.J. (2007) *Proteins*, **69**, 704–718.

28 Zhou, H.-X. (2001) *Biopolymers*, **59**, 427–433.

29 Berg, O.G. (1985) *Biophysical Journal*, **47**, 1–14.

30 Clore, G.M., Tang, C. and Iwahara, J. (2007) *Current Opinion in Structural Biology*, **17**, 603–616.

31 Schreiber, G. (2002) *Current Opinion in Structural Biology*, **12**, 41–47.

32 Qin, S. and Zhou, H.-X. (2008) *The Journal of Physical Chemistry. B*, **112**, 5955–5960.

33 Shaul, Y. and Schreiber, G. (2005) *Proteins*, **60**, 341–352.

34 Mintseris, J. and Weng, Z. (2005) *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 10930–10935.

35 Faust, P.L., Kornfeld, S. and Chirgwin, J.M. (1985) *Proceedings of the National Academy of Sciences of the United States of America*, **82**, 4910–4914.

36 Nooren, I.M.A. and Thornton, J.M. (2003) *The EMBO Journal*, **22**, 3486–3492.

37 Schreiber, G. and Fersht, A.R. (1993) *Biochemistry*, **32**, 5145–5150.

38 Li, W., Keeble, A.H., Giffard, C., James, R., Moore, G.R. and Kleanthous, C. (2004) *Journal of Molecular Biology*, **337**, 743–759.

39 Crowley, P.B. and Ubbink, M. (2003) *Accounts of Chemical Research*, **36**, 723–730.

40 Bahadur, R.P., Chakrabarti, P., Rodier, F. and Janin, J. (2004) *Journal of Molecular Biology*, **336**, 943–955.

41 Crowley, P.B., Rabe, K.S., Worrall, J.A.R., Canters, G.W. and Ubbink, M. (2002) *ChemBioChem*, **3**, 526–533.

42 Ubbink, M., Ejdeback, M., Karlsson, B.G. and Bendall, D.S. (1998) *Structure*, **6**, 323–335.

43 Crowley, P.B., Otting, G., Schlarb-Ridley, B.G., Canters, G.W. and Ubbink, M. (2001) *Journal of the American Chemical Society*, **123**, 10444–10453.

44 Kannt, A., Young, S. and Bendall, D.S. (1996) *Biochimica e t Biophysica Acta – Bioenergetics*, **1277**, 115–126.

45 Crowley, P.B. and Carrondo, M.A. (2004) *Proteins*, **55**, 603–612.

46 Hulsker, R., Baranova, M.V., Bullerjahn, G.S. and Ubbink, M. (2008)

Journal of the American Chemical Society, **130**, 1985–1991.
47 Berggard, T., Linse, S. and James, P. (2007) *Proteomics*, **7**, 2833–2842.
48 Fu, H. (2004) *Protein–Protein Interactions; Methods and Applications*, Humana Press, New York.
49 Coates, P.J. and Hall, P.A. (2003) *The Journal of Pathology*, **199**, 4–7.
50 Brymora, A., Cousin, M.A., Roufogalis, B.D. and Robinson, P.J. (2001) *Analytical Biochemistry*, **295**, 119–122.
51 Kaboord, B. and Perr, M. (2008) *Methods in Molecular Biology*, **424**, 349–364.
52 Piston, D.W. and Kremers, G.-J. (2007) *Trends in Biochemical Sciences*, **32**, 407–414.
53 Cunningham, B.C., Ultsch, M., De Vos, A.M., Mulkerrin, M.G., Clauser, K.R. and Wells, J.A. (1991) *Science*, **254**, 821–825.
54 Wen, J., Arakawa, T. and Philo, J.S. (1996) *Analytical Biochemistry*, **240**, 155–166.
55 Neylon, C. (2008) *European Biophysics Journal*, **37**, 531–541.
56 Sharon, M. and Robinson, C.V. (2007) *Annual Review of Biochemistry*, **76**, 167–193.
57 Howlett, G.J., Minton, A.P. and Rivas, G. (2006) *Current Opinion in Chemical Biology*, **10**, 430–436.
58 Heyduk, T., Ma, Y., Tang, H. and Ebright, R.H. (1996) *Methods in Enzymology*, **274**, 492–503.
59 Crowther, J.R. (1995) *ELISA: Theory and Practice*, Humana Press, New York.
60 Velazquez-Campoy, A., Leavitt, S.A. and Freire, E. (2004) *Methods in Molecular Biology*, **261**, 35–54.
61 Huber, W. and Mueller, F. (2006) *Current Pharmaceutical Design*, **12**, 3999–4021.
62 Zuiderweg, E.R. (2002) *Biochemistry*, **41**, 1–7.
63 Pintacuda, G., John, M., Su, X.-C. and Otting, G. (2007) *Accounts of Chemical Research*, **40**, 206–212.
64 Minton, A.P. (2001) *The Journal of Biological Chemistry*, **276**, 10577–10580.
65 Dyson, H.J. and Wright, P.E. (2002) *Current Opinion in Structural Biology*, **12**, 54–60.
66 Dyson, H.J. and Wright, P.E. (2005) *Nature Reviews. Molecular Cell Biology*, **6**, 197–208.
67 Dyson, H.J. and Wright, P.E. (2002) *Advances in Protein Chemistry*, **62**, 311–340.
68 Romero, P., Obradovic, Z. and Dunker, A.K. (2004) *Applied Bioinformatics*, **3**, 105–113.
69 Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. and Jones, D.T. (2004) *Journal of Molecular Biology*, **337**, 635–645.
70 Le Gall, T., Romero, P.R., Cortese, M.S., Uversky, V.N. and Dunker, A.K. (2007) *Journal of Biomolecular Structure & Dynamics*, **24**, 325–341.
71 Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) *Proteins*, **41**, 415–427.
72 Vacic, V., Oldfield, C.J., Mohan, A., Radivojac, P., Cortese, M.S., Uversky, V.N. and Dunker, A.K. (2007) *Journal of Proteome Research*, **6**, 2351–2366.
73 Fuxreiter, M., Tompa, P. and Simon, I. (2007) *Bioinformatics*, **23**, 950–956.
74 Romero, P.R., Zaidi, S., Fang, Y.Y., Uversky, V.N., Radivojac, P., Oldfield, C.J., Cortese, M.S., Sickmeier, M., LeGall, T., Obradovic, Z. and Dunker, A.K. (2006) *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 8390–8395.
75 Brown, C.J., Takayama, S., Campen, A.M., Vise, P., Marshall, T.W., Oldfield, C.J., Williams, C.J. and Dunker, A.K. (2002) *Journal of Molecular Evolution*, **55**, 104–110.
76 Demarest, S.J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H.J., Evans, R.M. and Wright, P.E. (2002) *Nature*, **415**, 549–553.
77 Gunasekaran, K., Tsai, C.-J., Kumar, S., Zanuy, D. and Nussinov, R. (2003) *Trends in Biochemical Sciences*, **28**, 81–85.
78 Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G.,

Obradovic, Z. and Dunker, A.K. (2004) *Nucleic Acids Research*, **32**, 1037–1049.
79 Shoemaker, B.A., Portman, J.J. and Wolynes, P.G. (2000) *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 8868–8873.
80 Wuttke, D.S., Foster, M.P., Case, D.A., Gottesfeld, J.M. and Wright, P.E. (1997) *Journal of Molecular Biology*, **273**, 183–206.
81 Dwek, R.A. (1996) *Chemical Reviews*, **96**, 683–720.
82 Shental-Bechor, D. and Levy, Y. (2008) *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 8256–8261.
83 Morita, Y., Kanei-Ishii, C., Nomura, T. and Ishii, S. (2005) *Molecular Biology of the Cell*, **16**, 5433–5444.
84 Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. and Jenuwein, T. (2001) *Nature*, **410**, 116–120.
85 Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V. and Laue, E.D. (2002) *Nature*, **416**, 103–107.
86 Lomberk, G., Wallrath, L.L. and Urrutia, R. (2006) *Genome Biology*, **7**, 228.
87 Klose, R.J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J. and Zhang, Y. (2006) *Nature*, **442**, 312–316.
88 Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H. and Allis, C.D. (2005) *Nature*, **438**, 1116–1122.
89 Brehm, A., Tufteland, K.R., Aasland, R. and Becker, P.B. (2004) *BioEssays*, **26**, 133–140.
90 Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y.R., Winston, F. and Allis, C.D. (2001) *Genes and Development*, **15**, 3286–3295.
91 Flanagan, J.F., Mi, L.-Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F. and Khorasanizadeh, S. (2005) *Nature*, **438**, 1181–1185.
92 Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., Wu, C. and Allis, C.D. (2006) *Nature*, **442**, 86–90.
93 Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M.R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B.R., Ayer, D.E., Kutateladze, T.G., Shi, Y., Cote, J., Chua, K.F. and Gozani, O. (2006) *Nature*, **442**, 96–99.
94 Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H. and Allis, C.D. (2005) *Cell*, **121**, 859–872.
95 Jacobson, R.H., Ladurner, A.G., King, D.S. and Tjian, R. (2000) *Science*, **288**, 1422–1425.
96 Kanno, T., Kanno, Y., Siegel, R.M., Jang, M.K., Lenardo, M.J. and Ozato, K. (2004) *Molecular Cell*, **13**, 33–43.
97 Xue, Y., Canman, J.C., Lee, C.S., Nie, Z., Yang, D., Moreno, G.T., Young, M.K., Salmon, E.D. and Wang, W. (2000) *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 13015–13020.
98 Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L. and Baker, D. (2003) *Science*, **302**, 1364–1368.
99 Malakauskas, S.M. and Mayo, S.L. (1998) *Nature Structural Biology*, **5**, 470–475.
100 Toscano, M.D., Woycechowsky, K.J. and Hilvert, D. (2007) *Angewandte Chemie International Edition*, **46**, 3212–3236.
101 Stockwin, L.H. and Holmes, S. (2003) *Expert Opinion on Biological Therapy*, **3**, 1133–1152.
102 Nahta, R. and Esteva, F.J. (2006) *Cancer Letters*, **232**, 123–138.
103 Skerra, A. (2007) *Current Opinion in Biotechnology*, **18**, 295–304.
104 Binz, H.K., Amstutz, P. and Pluckthun, A. (2005) *Nature Biotechnology*, **23**, 1257–1268.
105 Main, E.R., Jackson, S.E. and Regan, L. (2003) *Current Opinion in Structural Biology*, **13**, 482–489.

106 Main, E.R., Lowe, A.R., Mochrie, S.G., Jackson, S.E. and Regan, L. (2005) *Current Opinion in Structural Biology*, **15**, 464–471.

107 Andrade, M.A., Perez-Iratxeta, C. and Ponting, C.P. (2001) *Journal of Structural Biology*, **134**, 117–131.

108 Bork, P. (1993) *Proteins*, **17**, 363–374.

109 Sedgwick, S.G. and Smerdon, S.J. (1999) *Trends in Biochemical Sciences*, **24**, 311–316.

110 Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grutter, M.G. and Pluckthun, A. (2004) *Nature Biotechnology*, **22**, 575–582.

111 Amstutz, P., Binz, H.K., Parizek, P., Stumpp, M.T., Kohl, A., Grutter, M.G., Forrer, P. and Pluckthun, A. (2005) *The Journal of Biological Chemistry*, **280**, 24715–24722.

112 Das, A.K., Cohen, P.W. and Barford, D. (1998) *The EMBO Journal*, **17**, 1192–1199.

113 Cortajarena, A.L., Kajander, T., Pan, W., Cocco, M.J. and Regan, L. (2004) *Protein Engineering, Design and Selection*, **17**, 399–409.

114 Sidhu, S.S., Fairbrother, W.J. and Deshayes, K. (2003) *ChemBioChem*, **4**, 14–25.

115 Walensky, L.D., Kung, A.L., Escher, I., Malia, T.J., Barbuto, S., Wright, R.D., Wagner, G., Verdine, G.L. and Korsmeyer, S.J. (2004) *Science*, **305**, 1466–1470.

116 Oost, T.K., Sun, C., Armstrong, R.C., Al-Assaad, A.-S., Betz, S.F., Deckwerth, T.L., Ding, H., Elmore, S.W., Meadows, R.P., Olejniczak, E.T., Oleksijew, A., Oltersdorf, T., Rosenberg, S.H., Shoemaker, A.R., Tomaselli, K.J., Zou, H. and Fesik, S.W. (2004) *Journal of Medicinal Chemistry*, **47**, 4417–4426.

117 Gram, H., Schmitz, R., Zuber, J.F. and Baumann, G. (1997) *European Journal of Biochemistry*, **246**, 633–637.

118 Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H. and Schreiber, G. (2007) *Current Opinion in Structural Biology*, **17**, 67–76.

119 Martin, L., Stricher, F., Misse, D., Sironi, F., Pugniere, M., Barthe, P., Prado-Gotor, R., Freulon, I., Magne, X., Roumestand, C., Menez, A., Lusso, P., Veas, F. and Vita, C. (2003) *Nature Biotechnology*, **21**, 71–76.

120 Kortemme, T. and Baker, D. (2004) *Current Opinion in Chemical Biology*, **8**, 91–97.

121 Guerois, R., Nielsen, J.E. and Serrano, L. (2002) *Journal of Molecular Biology*, **320**, 369–387.

122 Pokala, N. and Handel, T.M. (2005) *Journal of Molecular Biology*, **347**, 203–227.

123 Reynolds, K.A., Thomson, J.M., Corbett, K.D., Bethel, C.R., Berger, J.M., Kirsch, J.F., Bonomo, R.A. and Handel, T.M. (2006) *The Journal of Biological Chemistry*, **281**, 26745–26753.

124 Selzer, T., Albeck, S. and Schreiber, G. (2000) *Nature Structural Biology*, **7**, 537–541.

125 Bae, K.-H., Kwon, Y.D., Shin, H.-C., Hwang, M.-S., Ryu, E.-H., Park, K.-S., Yang, H.-Y., Lee, D.-k., Lee, Y., Park, J., Kwon, H.S., Kim, H.-W., Yeh, B.-I., Lee, H.-W., Sohn, S.H., Yoon, J., Seol, W. and Kim, J.-S. (2003) *Nature Biotechnology*, **21**, 275–280.

126 Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. and Barbas, C.F. 3rd (2001) *The Journal of Biological Chemistry*, **276**, 29466–29478.

127 Dreier, B., Fuller, R.P., Segal, D.J., Lund, C.V., Blancafort, P., Huber, A., Koksch, B. and Barbas, C.F. 3rd (2005) *The Journal of Biological Chemistry*, **280**, 35588–35597.

128 Liu, Q., Xia, Z., Zhong, X. and Case, C.C. (2002) *The Journal of Biological Chemistry*, **277**, 3850–3856.

129 Segal, D.J., Dreier, B., Beerli, R.R. and Barbas, C.F. 3rd (1999) *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 2758–2763.

130 Sander, J.D., Zaback, P., Joung, J.K., Voytas, D.F. and Dobbs, D. (2007) *Nucleic Acids Research*, **35**, W599–W605.

131 Mandell, J.G. and Barbas, C.F. 3rd (2006) *Nucleic Acids Research*, **34**, W516–W523.

132 Beerli, R.R. and Barbas, C.F. 3rd (2002) *Nature Biotechnology*, **20**, 135–141.

133 Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M.H. and Chandrasegaran, S. (2005) *Nucleic Acids Research*, **33**, 5978–5990.

134 Urnov, F.D., Miller, J.C., Lee, Y.-L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D. and Holmes, M.C. (2005) *Nature*, **435**, 646–651.

135 Toogood, P.L. (2002) *Journal of Medicinal Chemistry*, **45**, 1543–1558.

136 Wells, J.A. and McClendon, C.L. (2007) *Nature*, **450**, 1001–1009.

137 Lipinski, C.A. (2000) *Journal of Pharmacological and Toxicological Methods*, **44**, 235–249.

138 Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (1997) *Advanced Drug Delivery Reviews*, **23**, 3–25.

139 Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N. and Liu, E.A. (2004) *Science*, **303**, 844–848.

140 Shuker, S.B., Hajduk, P.J., Meadows, R.P. and Fesik, S.W. (1996) *Science*, **274**, 1531–1534.

141 Oltersdorf, T., Elmore, S.W., Shoemaker, A.R., Armstrong, R.C., Augeri, D.J., Belli, B.A., Bruncko, M., Deckwerth, T.L., Dinges, J., Hajduk, P.J., Joseph, M.K., Kitada, S., Korsmeyer, S.J., Kunzer, A.R., Letai, A., Li, C., Mitten, M.J., Nettesheim, D.G., Ng, S., Nimmer, P.M., O'Connor, J.M., Oleksijew, A., Petros, A.M., Reed, J.C., Shen, W., Tahir, S.K., Thompson, C.B., Tomaselli, K.J., Wang, B., Wendt, M.D., Zhang, H., Fesik, S.W. and Rosenberg, S.H. (2005) *Nature*, **435**, 677–681.

142 Zhang, Z. and Palzkill, T. (2004) *The Journal of Biological Chemistry*, **279**, 42860–42866.

143 DeLano, W.L., Ultsch, M.H., de Vos, A.M. and Wells, J.A. (2000) *Science*, **287**, 1279–1283.

144 Ma, B., Shatsky, M., Wolfson, H.J. and Nussinov, R. (2002) *Protein Science*, **11**, 184–197.

145 Rose, R.B., Craik, C.S. and Stroud, R.M. (1998) *Biochemistry*, **37**, 2607–2621.

146 Last-Barney, K., Davidson, W., Cardozo, M., Frye, L.L., Grygon, C.A., Hopkins, J.L., Jeanfavre, D.D., Pav, S., Qian, C., Stevenson, J.M., Tong, L., Zindell, R. and Kelly, T.A. (2001) *Journal of the American Chemical Society*, **123**, 5643–5650.

147 Kallen, J., Welzenbach, K., Ramage, P., Geyl, D., Kriwacki, R., Legge, G., Cottens, S., Weitz-Schmidt, G. and Hommel, U. (1999) *Journal of Molecular Biology*, **292**, 1–9.

148 Arkin, M.R. and Wells, J.A. (2004) *Nature Reviews. Drug Discovery*, **3**, 301–317.

149 Seet, B.T., Dikic, I., Zhou, M.M. and Pawson, T. (2006) *Nature Reviews. Molecular Cell Biology*, **7**, 473–483.

150 Pahlich, S., Zakaryan, R.P. and Gehring, H. (2006) *Biochimica et Biophysica Acta*, **1764**, 1890–1903.

151 Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D. and Patel, D.J. (2007) *Nature Structural & Molecular Biology*, **14**, 1025–1040.

152 Mukherjee, S., Hao, Y.H. and Orth, K. (2007) *Trends in Biochemical Sciences*, **32**, 210–216.

153 Polevoda, B. and Sherman, F. (2002) *Genome Biology*, **3**, reviews0006.

154 Kaelin, W.G. (2005) *Annual Review of Biochemistry*, **74**, 115–128.

155 Walsh, G. and Jefferis, R. (2006) *Nature Biotechnology*, **24**, 1241–1252.

156 Hicke, L. (2001) *Nature Reviews. Molecular Cell Biology*, **2**, 195–201.

157 Guo, B., Yang, S.H., Witty, J. and Sharrocks, A.D. (2007) *Biochemical Society Transactions*, **35**, 1414–1418.

13
Mammalian Peptide Hormones: Biosynthesis and Inhibition

Karen Brand and Annette G. Beck-Sickinger

13.1
Introduction

Endocrine hormones are molecules that are formed in endocrine cells and transfer information from one cell to another via the bloodstream (Figure 13.1). Moreover, there are tissue hormones that are produced in tissues and act there in an autocrine or paracrine manner (Figure 13.1).

The chemical composition of hormones is diverse [1, 2], including:

- Peptide and proteohormones
- Amino acid-derived hormones, including catecholamines and thyroid hormones
- Cholesterol-derived hormones, including steroid hormones and vitamin D
- Fatty acid-derived hormones, including eicosanoids and retinoids
- Nucleotides and small inorganic molecules such as nitrogen monoxide.

Peptide hormones consist of up to 50 amino acids, while larger polymers that show a distinct tertiary structure are usually referred to as proteohormones.

Both hormonal and neurotransmitter functions have been described for many neuropeptides, including neuropeptide Y (NPY) [3], oxytocin [4], and cholecystokinin (CCK) [5].

While lipid-derived hormones can pass the cell membrane, peptide hormones generally bind to their receptors on the cell surface and thereby initiate an intracellular signaling cascade.

The main class of receptors that can be activated by hormones are G-protein-coupled receptors (GPCRs). Rashid *et al*. reviewed the signaling diversity of peptidergic GPCRs, including different G-proteins, and presented some mechanisms for how the receptors mediate different responses to endogenous ligands [6].

In addition to the classical cell surface receptor-mediated activity, several peptide hormones can also act as intracrines, which means they act inside the cell either after internalization by their target cells or retention in their cells of synthesis [7, 8]. There is compelling evidence that GPCRs can regulate genomic activity not only by the classical second messenger pathway, but also after endocytosis of the ligand–receptor

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

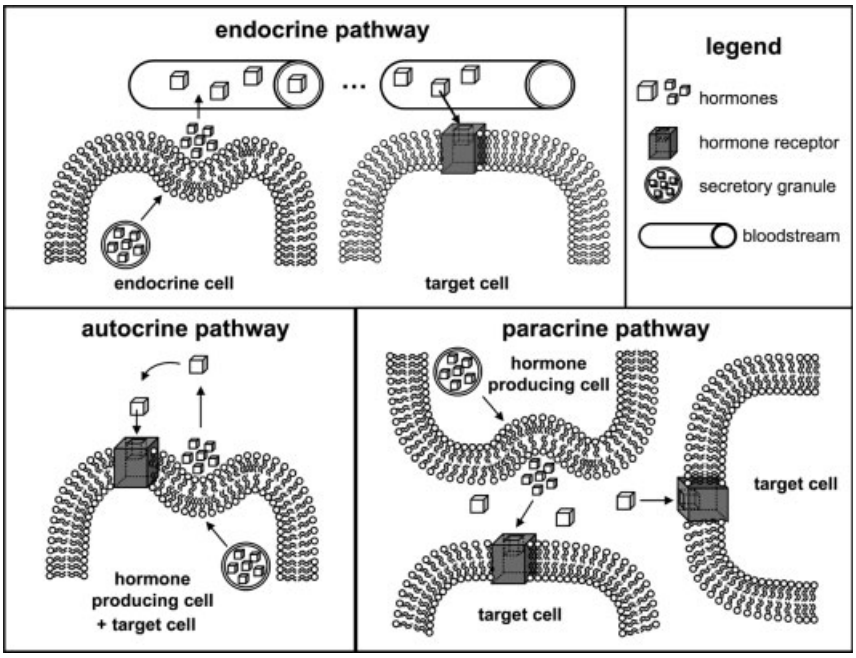


Figure 13.1 Endocrine, autocrine, and paracrine pathways of hormone activity.

complex and translocation to the nucleus or after activation of nuclearly located GPCR by a nonsecreted ligand [9]. A recent list of intracrines, including the hormones angiotensin [10], oxytocin [11] and NPY [12], is given by Re *et al.* [13].

13.2
Mammalian Peptide Hormones

Peptide hormones show a large variety of sites of formation, sites of action as well as a lot of different effects.

The hypothalamus represents one possible site of formation, where releasing factors [called liberins, such as thyroliberin (TRH) or corticoliberin according to an International Union of Pure and Applied Chemistry recommendation [14] and release-inhibiting factors (called statins, such as melanostatin or somatostatin [14]) are produced. These hormones regulate the release of other hormones that are formed in the pituitary and are designated by the ending "-tropin" (e.g., thyrotropin, corticotropin, melanotropin, or somatotropin). A huge number of peptide hormones can be formed in the gut [e.g., NPY, CCK, ghrelin, motilin, glucagon-like peptide (GLP) 1]. Moreover, peptide hormones can be produced in the pancreas (e.g., insulin, glucagon, pancreatic polypeptide), liver (e.g., angiotensin II), heart (e.g., atrial

natriuretic peptide), thyroid (e.g., calcitonin), and placenta (e.g., relaxin). One peptide hormone is not necessarily formed in only one tissue. For instance, gastrointestinal hormones like NPY [15], CCK [16], and ghrelin [17] are also formed in the brain.

One example that shows the diversity of functions of a single peptide hormone is NPY. It is the most abundant peptide in the central nervous system (CNS) of mammals [18], but it is also widespread within the peripheral nervous system [19]. It stimulates feeding [20], is involved in the regulation of blood pressure [21], but also in epilepsy [22], in memory retention [23], and in the regulation of the circadian rhythm [24]. NPY belongs to the NPY family, which also includes peptide YY and the pancreatic polypeptide. All three members are 36 amino acids in length and show a high sequence identity among different species [25].

Table 13.1 gives an overview of the diversity of peptide hormones. It contains representatives of peptide hormones that act in different parts of the body and representatives that cause different actions, which are arranged in peptide families consisting of hormones with similar sequences. More information on mammalian peptide hormones and peptide hormones of other classes can be found in Kastin [26].

13.3
Biosynthesis of Peptide Hormones

This section gives a schematic overview of the biosynthesis of peptide hormones (Figure 13.2). The single steps are described in more detail in the following subsections (Sections 13.3.1–13.3.4).

Peptide hormones are produced as inactive precursors that require proteolytic processing to liberate the biologically active peptides. Since the precursors can undergo a variety of modifications, these are additionally listed in Table 13.2.

13.3.1
Production and Maturation of Prohormones before Entering the Secretory Pathway

Peptide hormones are produced by transcription and splicing of the resulting heterogeneous nuclear RNA followed by translation into the amino acid sequence, which frequently results in inactive precursors that require proteolytic processing to liberate the biologically active peptides. The precursors carry a signal sequence on their N-terminus, which directs the ribosomes to the rough endoplasmic reticulum (ER) during synthesis by binding to signal recognition particles in a GTP-dependent process [67]. The growing peptide chain is transferred into the ER, in which the signal peptide is split off by a signal peptidase. The resulting prohormone is folded and initial post-translational modifications (PTMs) can occur, like disulfide bond formation [68] and glycosylation. It has been shown that some substrates of prohormone convertases can be cleaved already at this stage [69, 70], but more frequently prohormones are processed in the trans-Golgi network or in a later step of the secretory pathway (see Section 13.3.2).

Vesicles containing the folded prohormones are transferred from the ER to the Golgi apparatus. Phosphorylation of peptide hormone precursors takes place in the

Table 13.1 Overview of some mammalian peptide hormones.

Hormone	Abbreviation	Length (human)	Modification[a]	Main site of formation	Physiological action
Angiotensin					
angiotensin II	Ang II	8		brain, heart, liver	increases blood pressure
angiotensin IV	Ang IV	6		brain	enhances memory retention, long-term potentiation, increases blood pressure [27]
Calcitonin peptide family					
amylin/islet amyloid polypeptide	IAPP	37	disulfide bond between Cys2 and Cys7	pancreas	role in glucose homeostasis [28], inhibitor of bone resorption [29]
calcitonin	CT	32	disulfide bond between Cys1 and Cys7; $CONH_2$	thyroid	inhibitor of bone formation [29]
calcitonin gene-related peptide	CGRP	37	disulfide bond between Cys2 and Cys7	CNS	potent endogenous vasodilator, contributes to migraine pathogenesis [30]
Cholecystokinin/gastrin family					
cholecystokinin	CCK	different forms: 58, 33, 22, 8	Tyr (SO_4) within C-terminal bioactive heptasequence	gut, brain	regulates gall bladder contraction, pancreatic enzyme secretion, intestinal motility, satiety signaling, inhibits gastric acid secretion, neurotransmitter in brain and periphery [31]
gastrin		different forms: 71, **34**, **17**, 14, 6	Tyr (SO_4) within C-terminal bioactive hexasequence, $CONH_2$	gut, brain	regulates gastric acid secretion and mucosal growth [31]
Glucagon family					
glucagons	GG	29		pancreas	increase of blood glucose level
glucagon-like peptide 1	GLP1	37	Gly37 can be transformed to $CONH_2$	gut	stimulation of insulin secretion, inhibition of gastric emptying, acid secretion, food intake and glucagon secretion [32]

Guanylin family					
guanylin		15	disulfide bond between Cys4 and Cys12 and between Cys7 and Cys15	gut	both are involved in fluid and electrolyte homeostasis by stimulating secretion of chloride and hydrogencarbonate and inhibiting sodium resorption, which leads to water secretion; leading to natriuresis, kaliuresis, and diuresis in the kidney [33]
uroguanylin		16	disulfide bond between Cys4 and Cys12 and between Cys7 and Cys15	gut, stomach, kidney, lung, pancreas	
Insulin superfamily					
insulin	INS	A: 21	A and B chain connected via disulfide bonds between $Cys7^{(A)}$ and $Cys7^{(B)}$ and $Cys20^{(A)}$, and $Cys19^{(B)}$	pancreas	decreases blood glucose level, enhances cellular uptake of monosaccharides, fatty acids and amino acids, enhances glycogen synthesis [34]
		B: 30	A chain contains disulfide bond between Cys6 and Cys11		
relaxin 2	RLX	A: 24	A and B chain connected via disulfide bonds between $Cys11^{(A)}$ and $Cys11^{(B)}$ and $Cys24^{(A)}$, and $Cys23^{(B)}$	ovary, uterus, placenta, prostate	numerous pregnancy-specific actions like preparation for lactation, regulation of blood pressure, plasma osmolality, angiogenesis, collagen turnover, renal function [35]
		B: 29	A chain contains disulfide bond between Cys10 and Cys15		
Motilin/ghrelin family					
ghrelin		28	Ser3 octanoylated, des-acyl-ghrelin also occurs	gut, brain	stimulates gastric acid secretion, stimulates appetite, decreases blood pressure, stimulates release of several other hormones (growth hormone, ACTH, cortisol, prolactin) [36]

(*Continued*)

Table 13.1 (*Continued*)

Hormone	Abbreviation	Length (human)	Modification[a]	Main site of formation	Physiological action
motilin		22		gut	regulates motility of digestive tract [37]
Natriuretic peptide family					
atrial natriuretic peptide	ANP	28	disulfide bond between Cys7 and Cys23	heart	natriuretic, diuretic, vasorelaxant, antimitogenic effects [38]
brain-type natriuretic peptide	BNP	32	disulfide bond between Cys10 and Cys26	heart	natriuretic, diuretic, vasorelaxant, antimitogenic effects [38], protects heart against myocyte damages and fibrosis [39]
C-type natriuretic peptide	CNP	22	disulfide bond between Cys6 and Cys22	predominantly localized in endothelial cells and CNS	natriuretic, diuretic, vasorelaxant, antimitogenic effects [38], modulation of skeletal growth and bone development [40]
NPY family					
neuropeptide tyrosine	NPY	36	$CONH_2$	brain, gut	orexigenic, regulation of blood pressure, memory retention, seizure [41], role in alcohol intake and addiction [42]
peptide tyrosine tyrosine	PYY	36	$CONH_2$	gut, brain	anorexigenic, regulates glucose homeostasis [43]
pancreatic polypeptide	PP	36	$CONH_2$	pancreas, brain	anorexigenic, inhibition of pancreatic secretion and gallbladder motility [44]
Opiod family					
adrenocorticotropic hormone	ACTH	39	phosphorylated at Ser31	brain	stimulates steroid synthesis [45]
melanocyte-stimulating hormone	MSH	α: 13	α: Ac-NH; $CONH_2$	brain	feeding, energy metabolism, inflammation
		α: 22 γ: 12			α γ: increase blood pressure and heart rate γ: sodium metabolism [46]

Orexin family					
orexin A/ hypocretin 1		33	N-terminal pGlu; disulfide bonds between Cys6 and Cys12 and Cys7 and Cys14; $CONH_2$	brain	both: regulators of feeding behavior, sleep/wake states, involved in modulation of brain reward function, critical role in development of addiction processes [47, 48]
orexin B/ hypocretin 2		28	$CONH_2$	brain	
Somatostatin					
somatostatin	SST	different forms: 28, 14	cyclization of the last 12 amino acids by disulfide bond	brain, gut, pancreas	inhibits release of other hormones including growth hormone, acts as neurotransmitter [49]
Vasopressin/oxytocin family					
oxytocin	Oxt	9	disulfide bond between Cys1 and Cys6, $CONH_2$	brain	important role in reproductive physiology, including maternal behavior, social recognition, and bonding [50]
(arginine) vasopressin/ antidiuretic hormone	Avp	9	disulfide bond between Cys1 and Cys6, $CONH_2$	brain	antidiuretic effect, regulation of blood pressure, facilitates aggression, influences pair bonding [51]

[a]N-terminal acetylation is denoted by Ac-NH; C-terminal amidation is denoted by $CONH_2$.

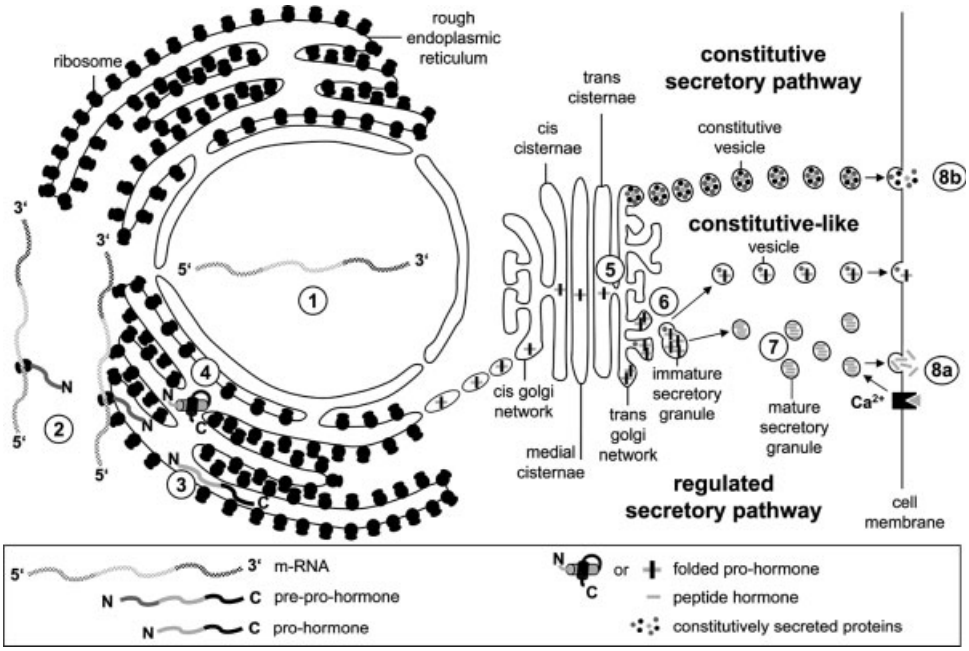


Figure 13.2 Schematic overview on the biosynthesis of peptide hormones. They are formed as preprohormones by transcription (1) and translation (2). After completion of the signal peptide (2) the growing peptide chain is transferred to the ER, where the signal peptide is split off and the prohormone is formed (3). Within the ER, folding of the prohormone takes place and first PTMs like disulfide bond formation or glycosylation can occur (4). The folded prohormone is transferred to the Golgi, where further PTMs like phosphorylation and sulfation can occur (5). By specific sorting mechanisms the products released by the Golgi are transferred to immature secretory granules or constitutive vesicles (6). Furthermore, from immature secretory granules constitutive-like vesicles can be formed (6). Within the secretory granules prohormone processing and, in most cases, further modifications result in the formation of the bioactive peptide (7) followed by its exocytosis after stimulation (8a). On the contrary, the content of constitutive vesicles is continuously released without any stimulus (8b).

trans cisternae of the Golgi [71]. Later, in the trans-Golgi network, peptide hormone precursors can be sulfated by a tyrosylprotein sulfotransferase that catalyzes the transfer of sulfate from 3′-phosphoadenosine-5′-phosphosulfate to specific tyrosine residues of a protein [62].

The prohormones leave the trans-Golgi network within secretory vesicles, which can occur by two main pathways [72].

13.3.2 Secretory Pathways

Two general pathways of secretion can be distinguished (Figure 13.2). All eukaryotic cells have the constitutive secretory pathway, which is characterized by an immediate release of vesicle cargo without any stimulus.

Table 13.2 PTMs apart from proteolytic processing.

Modification	Compartment	Enzyme	Examples	Site of modification	Prediction	Detection
Disulfide bond formation	ER	protein disulfide isomerase (EC 5.3.4.1)	insulin, calcitonin	two Cys residues are transformed to cystine	[52]	[53, 54]
N- and *O*-Linked glycosylation	*N*: ER	glycosyl-transferases (EC 2.4.X.X)	*N* and *O*: POMC	*N*: oligosaccharide precursor added to Asn (at Asn–X–Ser/Thr or rarely Asn–X–Cys)	*N*: NetNGlyc [55]	[57]
	O: Golgi			*O*: stepwise addition of sugars to branched starting structure that have been added to Ser/Thr	*O*: NetOGlyc [56]	
Phosphorylation	trans cisternae of Golgi stack	protein kinases (EC 2.7.X.X)	ACTH	mostly on Ser, also Thr, Tyr	"GANNPhos" [58], "DISPHOS" [59]	[60, 61]
O-Sulfation	trans-Golgi network	tyrosylprotein sulfotransferase (EC 2.8.2.20)	CCK, gastrin	tyrosine residues in exposed protein domains containing acid amino acids [62]	"Sulfinator" [63]	[64, 65]
Removal of basic residues (Section 13.3.4.1)	secretory granules	carboxypeptidase E, aminopeptidase B	TRH	C- or N-terminus of intermediate		
C-Terminal amidation (Section 13.3.4.2)	secretory granules	peptidylglycine α-amidating mono-oxygenase (EC 1.14.17.3)	calcitonin, NPY, oxytocin	C-terminal glycine residue is split off and delivers NH_2-group		[66]
Acylation (Section 13.3.4.3)		acyl transferases (EC 2.3.X.X)	α-MSH, ghrelin			
N-Terminal pyroglutamylation (Section 13.3.4.4)	secretory granules	glutaminyl cyclase (EC 2.3.2.5)	orexin, TRH	N-terminal Gln or Glu is transferred into pGlu		

In addition to this pathway, endocrine, neuronal, and exocrine cells use the regulated secretory pathway [72], which enables large amounts of so-called regulated secretory proteins to be stored in secretory granules, until their release is induced by an external stimulation. Therefore, secretory granules serve as storage organelles for rapid release of bioactive peptides. However, only a small fraction of the multitude of existing secretory granules is immediately releasable [73]. The observed slower secretion following the rapid release has been explained by the existence of a large depot pool of granules that have to mature to releasable granules [74]. Moreover, recent studies have shown that granules do not necessarily release all their cargo at once, but cells might control the rate of release either by modulating the properties of the fusion pore (designated as "kiss and run" [75] or "cavicapture" [76] exocytosis) or by storing the cargo in states that disperse slowly in the extracellular space (designated as solubility-limited exocytosis) (for a review, see [77]).

The denotation "granule" results from the granular appearance of these organelles, which is due to an electron-dense core formed by aggregated protein. As depicted in Figure 13.2, secretory granules are initially formed as immature granules. Different models for sorting of regulated secretory proteins to these granules are discussed [78].

According to the "sorting for entry" hypothesis, the trans-Golgi network acts as the primary operator for protein sorting [78]. Trans-Golgi network sorting occurs due to the intrinsic ability of regulated secretory proteins to aggregate, due to a sorting signal on the prohormones themselves, or due to trans-Golgi network sorting receptors like carboxypeptidase E [79].

In the "sorting by retention" hypothesis molecules enter the immature granules without being sorted and a subset of proteins is discharged from there while regulated secretory proteins remain since their condensation limits their ability to escape [78]. Proteins that are eliminated from immature granules via formation of vesicles end up in the so-called constitutive-like secretory pathway [78, 80]. Proteins like prohormones that remain within secretory granules are further processed.

As pointed out by Dikeakos and Reudelhuber [81], several sorting signals can contribute to the overall efficiency of sorting or retention; thus, both above-mentioned hypotheses are not mutually exclusive.

13.3.3
Prohormone Cleavage

Since it is not the precursor as a whole but only part of it that is biologically active or becomes biologically active after further modification, removal of flanking sequences is necessary. Within their precursor, peptide hormone sequences are frequently flanked by two neighboring basic residues (Figure 13.3), most often Lys–Arg, but Arg–Arg, Lys–Lys, and Arg–Lys are also possible [82]. In some cases even single arginine residues occur [83].

Different processing enzymes responsible for prohormone processing at these sites have been found (Table 13.3). Interestingly, endopeptidases of four different classes can participate in proprotein processing of some prohormones.

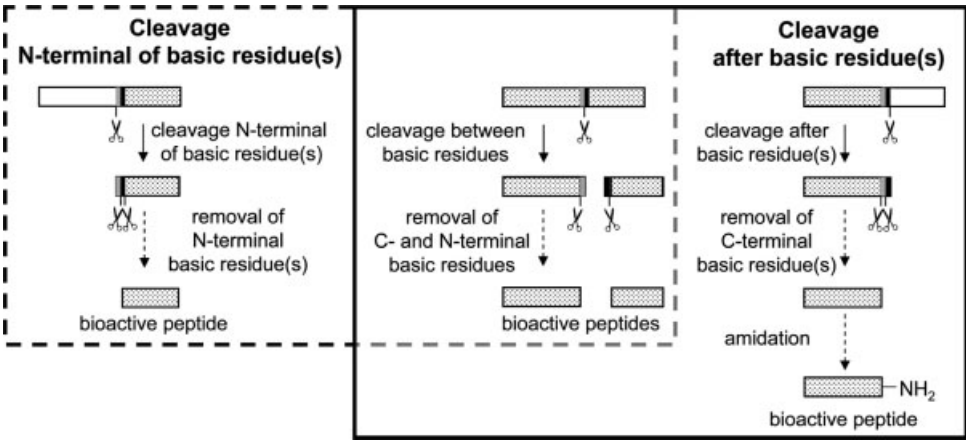


Figure 13.3 Different cleavage patterns of prohormones. The figure includes only cleavages at basic residues, since these are the most common. Prohormones often contain more than one of those cleavage sites (see Section 13.3.3.2).

Two distinct processing pathways can be distinguished (Figure 13.3):

- Cleavage after basic residues, mostly followed by the removal of the C-terminal basic residues.
- Cleavage at the N-terminal site of basic residues, mostly followed by the removal of the N-terminal basic residues.

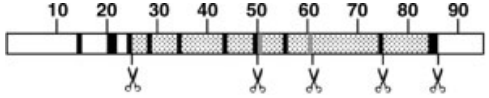
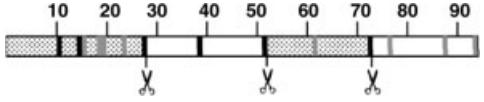
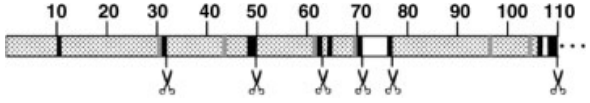
In both cases the cleavage can also take place in between a basic pair which can be followed by removal of basic residues at the C- and N-terminus of the products, respectively.

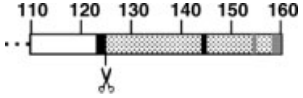
Proprotein convertases cleaving after basic residues are the major endoproteolytic processing enzymes of the secretory pathway [93]. As can be seen from Table 13.4, a

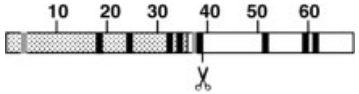
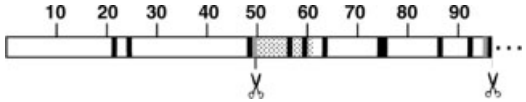
Table 13.3 Proprotein convertases and other enzymes involved in prohormone cleavage.

Serine endopeptidases (EC 3.4.21.X)	**Cysteine endopeptidases (EC 3.4.22.X)**	**Aspartic endopeptidases (EC 3.4.23.X)**	**Metalloendopeptidases (EC 3.4.24.X)**
Proprotein convertases (see Section 13.3.3)	cathepsin L (prohormone thiol protease complex) [85, 86]	renin (angiotensinogenase)	pro-oxytocin/ neurophysin-converting enzyme (magnolysin) [90]
PC1, PC2, PC5/6A, furin, PACE4		pro-opiomelano-cortin-converting enzyme [87]	nardilysin [91]
SKI-1 [84]		chromaffin granule 70-kDa aspartic protein-ase [88, 89]	ECE-2 [92]

Table 13.4 Processing scheme of different prohormones.

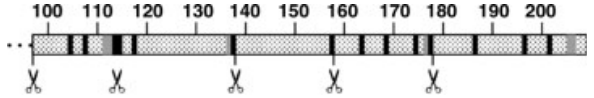
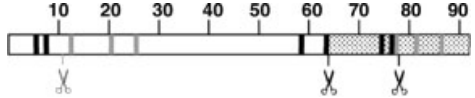
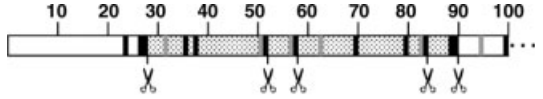
Precursor	Schematic structure	Denotation of the cleavage product(s)	Position of the product(s) within the precursor	Cleavage site	Cleaving enzyme(s)
Human pro-CCK		all derivatives		R85R86↓	intestine: PC1 [101]
		CCK-8	76–83	R75↓	intestine: PC1 [101]
		CCK-22	62–83	K61↓	intestine: PC2 [101]
		CCK-33	51–83	R50↓K51	intestine: PC1 [101]
		CCK-58	26–83	R25↓	intestine: PC1 [101]
Mouse proghrelin		ghrelin	1–28	R28↓	probably PC1 [102]
		obestatin	53–74	R52↓	probably PC1 [102]
				R74↓	
Mouse proglucagon		glicentin	1–69	K70R71↓	intestine: PC1, . . . [103]
					pancreas: PC2 [104]

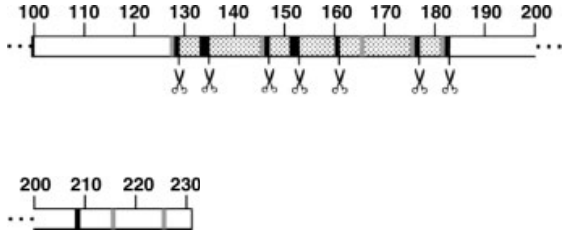


	glicentin-related pancreatic polypeptide	1–30	K31R32↓	intestine: PC1 [103]
				pancreas: PC2 [104]
	oxyntomodulin	33–69	K70R71↓	intestine: PC1, . . . [103]
	glucagon	33–61	K62R63↓	pancreas: PC2 [104]
	miniglucagon	51–61	R49R50↓	miniglucagon-generating endopeptidase [91]
	GLP1	78–107	R77↓	intestine: PC1 [103]
			R109R110↓	
	GLP2	126–158	R124R125↓	intestine: PC1 [103]
Rat pro-NPY		NPY	1–36	K38R39↓	PC1 > PC2 [105]
Rat pro-opiomelanocortin		γ-MSH	51–61	R49K50↓	

(*Continued*)

Table 13.4 (*Continued*)

Precursor	Schematic structure	Denotation of the cleavage product(s)	Position of the product(s) within the precursor	Cleavage site	Cleaving enzyme(s)
		adrenocorticotropic hormone	98–136	K96R97↓	m + c: PC1 [106]
				K137R138↓	m + c: PC1 [106]
		α-MSH	98–110	within KKRR115	m: PC2 [106]
		CLIP (corticotrophin-like intermediate lobe)	116–136	KKRR115	m: PC2 [106]
		β-lipotropic hormone	139–209	K137R138↓	m + c: PC1 [106]
		γ-lipotropic hormone	139–176	K177R178↓	m + (c): PC2 [106]
		β-MSH	159–176	K158↓	PC2? [106]
		β-endorphin	179–209	K177R178↓	m + (c): PC2 [106]
Rat prosomatostatin		pSST 1–10 (Antrin)	1–10	L11↓	SKI-1 [84]
		SST-14	79–92	R77K78↓	PC2, (PC1) [107]
		SST-28	65–92	R64↓	PACE4, furin [107]
Rat pro-TRH		pFT22	29–50	R27R28↓	PC1 > PC2 [108]



		K51R52↓	PC1 > PC2 [108]
TRH	53–55	K57R58↓	PC2 > PC1 [108]
pEH24	59–82	K83R84↓	PC1 > PC2 [108]
TRH	85–87	R89R90↓	PC1 > PC2 [108]
TRH	130–132	K128R129↓	PC1 > PC2 [108]
		R134R135↓	PC1 > PC2 [108]
pST10	136–145	K146R147↓	PC1 [108]
TRH	148–150	R152R153↓	PC1 [108]
pFE22	154–175	K176R177↓	PC1 [108]
pSE14	162–175	R161↓	PC2 [108]
TRH	178–180	K182R183↓	PC1 [108]

m: melanotropes, c: corticotropes.

lot of precursors are cleaved by enzymes of this family. For pro-oxytocin a special converting enzyme has been described, a metalloendopeptidase called pro-oxytocin/neurophysin-converting enzyme [90], that also cleaves after basic residues and has been used for several investigations concerning the importance of the three-dimensional structure of the substrate for cleavability [94, 95].

Cathepsin L, the catalytic subunit of the prohormone thiol protease complex [85, 86], and the metallopeptidase nardilysin [91] cleave N-terminally of basic residues.

Furthermore, two members of the proprotein convertase family that cleave after R-X1-hydrophobic-X2 (SKI-1/S1P [96, 97]) and V-F-A-Q (NARC-1/PCSK9 [98]) have been discovered. SKI-1 is a membrane-bound subtilase. As reviewed by Seidah *et al.*, SKI-1 is a key enzyme in the regulation of lipid metabolism and cholesterol homeostasis [99], but it has also been shown to be involved in the processing of prosomatostatin [84].

The metallopeptidase endothelin-converting enzyme (ECE)-2 has also been proposed to be involved in the nonclassical processing of bioactive peptides at nonbasic sites, including conversion of big-endothelin-1 into endothelin-1 [92].

Prohormone cleavage is not necessarily performed by one enzyme exclusively. Cleavage products have been found even though the enzyme believed to be responsible for the cleavage has been knocked out [100]. This indicates that converting enzymes are, at least in part, able to substitute each other.

13.3.3.1 Basic Amino Acid-Specific Members of the Proprotein Convertase Family

Proprotein convertase represent the major endoproteolytic processing enzymes [93]. They are serine endoproteases, belong to the family of subtilases and are involved in the processing of many prohormones. The proprotein convertase family includes seven mammalian basic amino acid-specific members (Table 13.5) that are structurally related to yeast kexin (EC 3.4.21.61).

Based on the crystal structures of furin and kexin, analysis of the proteolytic domain explains the preference for basic residues, in particular arginine, at the P_1 position [109] (see Section 13.3.3.3 for nomenclature).

PC1, PC2, and partly PC5/6A (an isoform of PC5/6) are targeted to dense core secretory granules of endocrine and neuroendocrine cells [110], and are therefore mainly responsible for the processing of prohormones. Moreover, PC5/6A can be constitutively secreted into the extracellular milieu like PC4 and paired amino acid-converting enzyme (PACE) 4 [99], and its cysteine-rich domain can serve as a cell surface anchor and allows processing of membrane-bound precursors on the cell surface [111]. The other three basic amino acid specific members (furin, PC7, and PC5/6B) are type I membrane-bound proteases [99].

The prohormone convertases themselves are produced as inactive precursors requiring proteolytic processing. Proprotein convertases consist of a signal peptide, a pro-domain, a catalytic domain containing the catalytic triad Ser–His–Asp and an oxyanion hole, a P-domain, and a C-terminal segment, which differs most between the members of the family (Figure 13.4).

As an example, the processing of PC1 is illustrated in Figure 13.5. The convertases are routed to the ER and thereby to the secretory pathway by their signal peptide. The

Table 13.5 Proprotein convertases.

Denotations	Soluble/membrane protein	Distribution	Function
PC1/PC3/PC1/3/subtilisin-like proprotein convertase 3/ neuroendocrine convertase 1	contains an amphipathic helix, not clear whether soluble [112] or transmembrane protein [113]	mainly in secretory granules of endocrine and neuroendocrine cells [110]	cleavage of precursors entering the regulated secretory pathway
PC2/subtilisin-like proprotein convertase 2/neuroendocrine convertase 2	contains an amphipathic helix, not clear whether soluble [109] or transmembrane protein [114]	mainly in secretory granules of endocrine and neuroendocrine cells [110]	cleavage of precursors entering the regulated secretory pathway
PC4/subtilisin-like proprotein convertase 5	soluble [109]	testicular germ cells, macrophage like cells of the ovary [115]	role in sperm fertilization [116, 117], and early embryonic development [116]
PC5A/PC6A/PC5/6A/subtilisin-like proprotein convertase 6A	soluble [118]; Cys-rich domain serves as cell surface anchor [111]	secretory granules of endocrine and neuroendocrine cells [110] and cell surface [111]	cleavage of precursors entering the regulated secretory pathway and of membrane-bound precursors on cell surface
PC5B/PC6B/PC5/6B/subtilisin-like proprotein convertase 6B	transmembrane protein, soluble after shedding [118]	cycles between cell surface via endosomes back to trans-Golgi network [99]	
PC7/PC8/subtilisin-like proprotein convertase 7/lymphoma proprotein convertase	transmembrane protein	widespread [119] cycles between cell surface via endosomes back to the trans-Golgi network [99]	growth factors, may be involved in processing of nonessential substrates since null mice show no apparent phenotype [120]
PACE4/subtilisin-like proprotein convertase 4	soluble; Cys-rich domain serves as cell surface anchor [111]	cell surface [111]	processing of matrix metalloproteinases [121], transforming growth factor-β-related factors [122]
Furin/PACE/subtilisin-like proprotein convertase 1	transmembrane protein [123], soluble after shedding [124]	widespread; cycles between cell surface via endosomes back to trans-Golgi-network [125–127]	major processing enzyme of the constitutive secretory pathway (processing of growth factors, hormones, receptors, plasma proteins, matrix metalloproteinases, viral envelope glycoproteins, bacterial exotoxins) [128]

Convertases important for prohormone processing are given in italics.

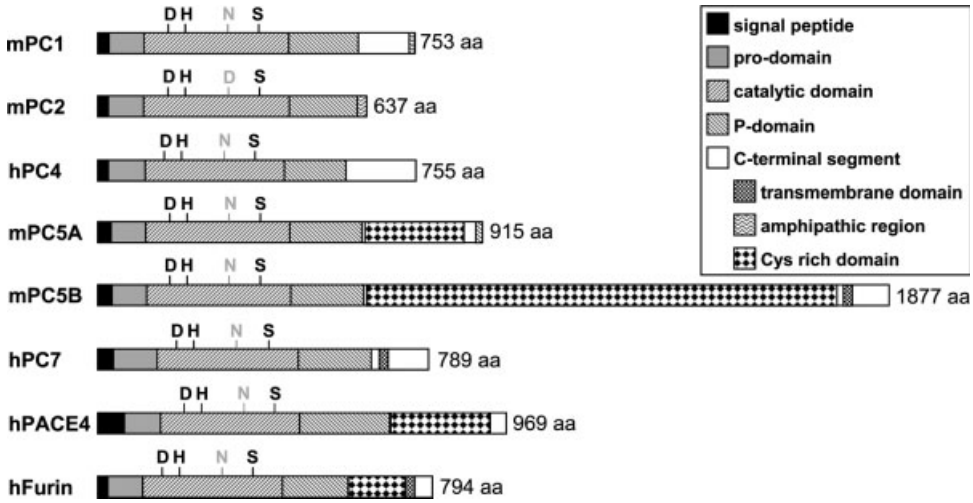


Figure 13.4 Schematic domain structure of the basic amino acid-specific members of the proprotein convertases. PC5 occurs in two alternatively spliced isoforms: soluble PC5A and membrane-bound PC5B. The residues aspartic acid, histidine and serine of the catalytic triad and the asparagine (in the case of PC2, aspartic acid) of the oxanion hole are indicated. The C-terminal segment can include a transmembrane domain, an amphipathic region or a cysteine-rich domain.

cleavage of the signal peptide results in the pro form of the convertase, which is processed further. While maturation of pro-PC2 occurs within a late compartment of the secretory pathway [129], the pro domain of pro-PC1 is autocatalytically split off in the ER [130]. Nevertheless, it remains associated with PC1 as a kind of autoinhibitor [131, 132]. This protein complex reaches the trans-Golgi network and the full-length PC1 is transported further into secretory granules. There, the C-terminal segment is split off, which results in the fully active convertase that is stored in secretory granules and serves for the processing of several precursors. It has

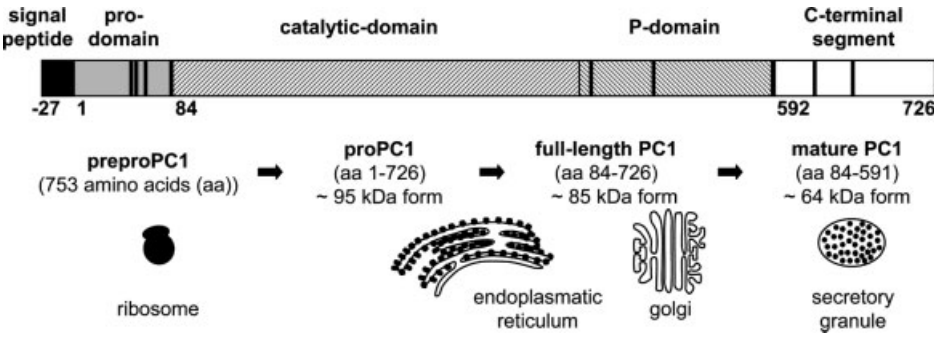


Figure 13.5 Schematic domain structure of PC1 and its processing during the regulated secretory pathway. Dibasic sequences within the pro-PC1 sequence are depicted as black bars.

been shown *in vitro* that the C-terminal segment itself increases PC1 activity at nanomolar concentrations, while it decreases enzyme activity at micromolar concentrations [133].

Several studies have shown that removal of the C-terminal segments of PC1 [134], PC2 [114], and PC5/6A [118] prevent their sorting to large dense core secretory vesicles. The C-terminal segments of these convertases even have the capacity to redirect a constitutively secreted protein to granule-containing cytoplasmic extensions [110].

13.3.3.2 Different Biologically Active Peptides from one Precursor

While some precursors like pro-NPY contain only one peptide hormone, there are a lot of precursors whose processing results in the formation of several biologically active peptides as can be taken from Table 13.4. These include pro-TRH, which contains several other products beside TRH [108]. Moreover, this precursor contains the same sequence (Gln–His–Pro–Gly) several times, which allows the simultaneous formation of several copies of the biologically active TRH (pGlu–His–Pro-NH_2).

Since different cell types produce different amounts of the respective processing enzymes, prohormone precursors may undergo tissue-specific processing, as for instance described for proglucagon [103] and pro-opiomelanocortin [106]. While in the pancreas, proglucagon is processed into glicentin-related pancreatic polypeptide, glucagon, and the major proglucagon fragment, in intestine, glicentin, GLP1, and GLP2 are formed [103]. Through differential processing and degradation, different levels of peptides can be achieved although they are derived from the same precursor.

13.3.3.3 Nomenclature at the Cleavage Site

The general nomenclature used today to describe the cleavage site of a substrate and the corresponding positions in the active site of the enzyme was developed by Schechter and Berger [135]. Accordingly, the cleavage site is located between residue P_1 and residue P_1'. Residues N-terminal to the cleavage site are numbered ascendingly P_i, and on the C-terminal site ascending numbers P_i' are used. The enzyme positions located next to P_i/P_i' are denoted by S_i/S_i' (Figure 13.6).

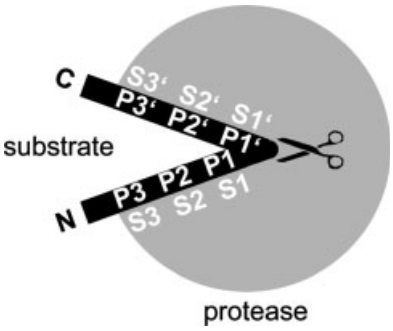


Figure 13.6 Nomenclature at the cleavage site.

13.3.3.4 Prediction of Cleavage Sites – Discovery of New Bioactive Peptides

Prediction of prohormone cleavage sites can be helpful to discover new bioactive peptides from genomic data or protein sequence databases.

Since not all single or paired basic amino acid residues are actually cleaved by proprotein convertases, prediction of a cleavage site is not easily possible just from the sequence.

In 2004, Duckert *et al.* developed a neural network-based method (ProP) using the sequence patterns of experimentally verified cleavage sites of furin or prohormone convertases in general to predict whether a given site is a potential cleavage site [136]. The network was trained on viral and eukaryotic proteins obtained from the Swiss-Prot database. Since the amount of experimentally verified sites was limited for convertases other than furin, only a furin-specific neural network could be trained. Cleavage sites for the other proprotein convertases are included in a general approach. However, since cleavage specificity differs between the members of the proprotein convertase family and PC1 and PC2 sites are under-represented in the data set prediction of cleavage sites in PC1 or PC2 substrates is suboptimal in the ProP program.

In 2006, Amare *et al.* developed a binary logistic regression model trained on mammalian prohormones and compared their model with previously described programs. They could show that an improved prediction of mammalian processing sites was achieved [82]. NeuroPred is available online at http://neuroproteomics.scs.uiuc.edu/neuropred.html [137]. Apart from the prediction of cleavage sites, the program also provides the molecular mass of the predicted peptides, including possible PTMs. Accordingly, the program is well suited to confirm new neuropeptides in mass spectrometry-based neuropeptidomic studies.

Another bioinformatic search tool to find new bioactive peptides has been developed by Mirabeau *et al.* [138]. Their approach is based on a hidden Markov model formalism that uses several peptide hormone sequence features to estimate the likelihood that the protein contains a processed and secreted peptide [138]. The suitability of their tool is demonstrated by an application to an alignment of mammalian proteomes where 90% of the top 300 proteins contained known peptide hormones [138]. Furthermore, the authors could identify two novel candidate peptide hormones [138].

13.3.4 Further PTMs

13.3.4.1 Removal of Basic Amino Acids

After endoproteolytic cleavage, in most cases, basic or dibasic extensions are split off; however, peptides can also be bioactive with basic extensions. It has been shown that removal can lead to alteration of receptor specificity, for instance enkephalins change their selectivity for opiate receptors [139]. The pentapeptides without C-terminal basic residues show selectivity for δ-receptors, whereas extended peptides do not discriminate between μ- and δ-receptors, but show a marked increase in affinity for κ-receptors [139].

The removal of the remaining basic residues is done depending on the enzyme(s) involved in prohormone processing. While metallocarboxypeptidases (EC 3.4.17.X), mainly carboxypeptidase E (EC 3.4.17.10), remove basic residue extensions from the C-terminus of peptide intermediates [140], aminopeptidase B (EC 3.4.11.6) removes such extensions from the N-terminus [141, 142].

For many years carboxypeptidase E, also known as carboxypeptidase H, was thought to be the only carboxypeptidase involved in peptide processing in secretory vesicles of mammalian cells [140]. However, a defect of carboxypeptidase E does not completely eliminate removal of dibasic residues, which suggests that one or more additional carboxypeptidases are involved in peptide hormone processing [143]. Carboxypeptidase D (EC 3.4.17.22) has been discovered in a search for carboxypeptidase E-like enzymes [144]. However, it has been shown that this enzyme is enriched in the trans-Golgi network [145], cycles to the cell surface, and returns to the trans-Golgi network, and functions primarily in the processing within the trans-Golgi network and/or immature secretory granules [146].

Further enzymes that are involved in the specific removal of basic residues from peptides and proteins include carboxypeptidase M (EC 3.4.17.12), which is bound to extracellular membranes, and carboxypeptidase N (EC 3.4.17.3), which occurs in plasma.

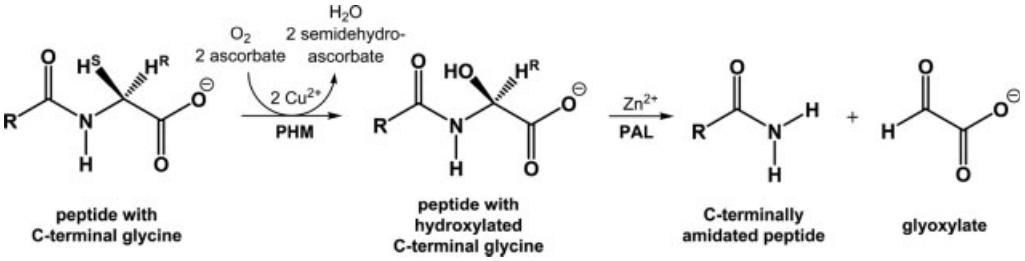
13.3.4.2 **C-Terminal Amidation**

Numerous peptide hormones require C-terminal amidation to reach their full biological activity [147].

Amidation is performed by the bifunctional enzyme peptidylglycine α-amidating mono-oxygenase (PAM), which consists of the copper- and ascorbate-dependent peptidylglycine α-hydroxylating monooxygenase (PHM) and peptidyl-α-hydroxyglycine α-amidating lyase (PAL) [148, 149]. As the name suggests, amidation occurs with the C-terminal glycine residue as the amide donor. The amidation reaction is depicted in Scheme 13.1.

However, the glycine-extended hormones may also show biological function. It has been shown that glycine-extended gastrin is biologically active and, interestingly, it comprises a different set of activities compared to amidated gastrin [150].

Amidated and glycine-extended hormones can also show more similar biological effects as in the case of amidated and glycine-extended GLP1 [151]. However, it has



Scheme 13.1 C-terminal amidation reaction catalyzed by the bifunctional PAM consisting of PHM and PAL.

been shown that amidated GLP1 has a slightly longer C-terminal plasma half-life than the glycine-extended GLP1 [152]. Therefore, a function of N-terminal amidation might also be the improvement of hormone stability.

13.3.4.3 **Acylation**

While N-terminal acetylation of proteins is a frequently occurring cotranslational modification, peptide hormone acylation occurs post-translationally (reviewed in [153]) and is a rare modification within native peptide hormones. It has been shown that N-terminal acetylation can influence peptide hormone activity. Administration of α-melanocortin-stimulating hormone (MSH), an *N*-acetylated bioactive peptide resulting from pro-opiomelanocortin processing, or administration of desacetyl-α-MSH result in quantitatively different or functionally antagonistic effects [154]. The differences in potency of the two peptides might be due to the increased stability of the acetylated form [154].

Apart from N-terminal acetylation, side-chain acylation can also be important. In the case of ghrelin it has been shown that octanoylation at Ser3 is essential for its biological activity [155].

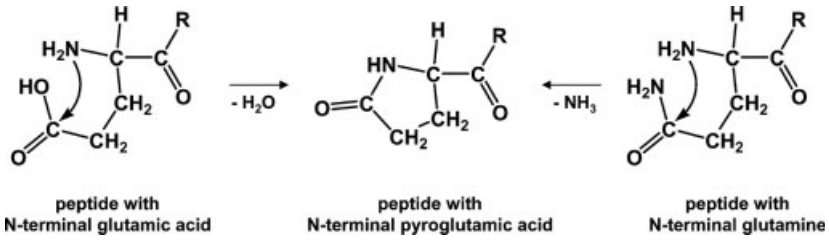
13.3.4.4 **Pyroglutamylation**

N-Terminal pyroglutamate occurs in several peptide hormones including orexin [156], gonadoliberin [157] or TRH [157]. Glutaminyl cyclase (EC 2.3.2.5) was identified as being be responsible for the formation of N-terminal pyroglutamic acid either from N-terminal glutamine at basic pH or from N-terminal glutamate at acidic pH [158–160], that probably occurs like peptide amidation in secretory granules [161] (Scheme 13.2).

Selective removal of the pyroglutamate residue, which is required for the primary structure determination by Edman degradation, can be achieved by mild hydrolysis with high concentrations of aqueous methanesulfonic acid at low temperature [162].

13.3.4.5 **N-Terminal Truncation**

N-Terminal truncation is a postsecretory processing event performed by dipeptidyl peptidase IV (EC 3.4.14.5) or dipeptidyl peptidase IV-like enzymes [163–165]. Dipeptidyl peptidase IV, also called CD26, is a member of the prolyl oligopeptidase



Scheme 13.2 N-terminal pyroglutamate formation catalyzed by glutaminyl cyclase using peptides with N-terminal glutamate or glutamine.

family of serine proteases that removes preferentially Xaa–Pro dipeptides from the N-terminus of substrates.

The N-terminal truncation can lead to an altered receptor specificity as has been shown for NPY as well as for the pancreatic polypeptide [166, 167]. Both truncated peptides bind preferentially to the Y_2 receptor subtype, while their full-length counterparts show equal affinity for both the Y_1 receptor subtype and the Y_2 receptor subtype. Since this shift in signaling potential might participate in induction of neovascularization in developing tumors, N-terminal truncation might be important for cancer [165].

In contrast to this, dipeptidyl peptidase IV-like enzyme activity can also result in the formation of inactive peptides like in the case of the GLP1 [168] and glucose-dependent insulinotropic polypeptide [169].

13.4 Inhibition of Biosynthesis

As hormones can be involved in pathological processes, selective inhibition of their action can have favorable therapeutic effects. Several approaches may serve for the development of such inhibition. The formation of a bioactive peptide can be avoided by interfering with its biosynthesis. Moreover, the hormone–receptor interaction can be destroyed by blocking the receptors or the hormonal effect can be reduced by accelerating peptide hormone degradation.

13.4.1 Readout Systems to Investigate Cleavage by Proteases

Readout systems to observe the cleavage reaction are essential for scanning combinatorial libraries for inhibitors and for inhibitor development or improvement. Some general possibilities are depicted in Figure 13.7.

Cushman and Cheung described a spectrophotometric assay for the determination of the amount of hippuric acid formed, which has been used for the development of the inhibitor captopril [170]. However, before the determination, the hippuric acid formed has to be extracted. This separation step can be omitted using chromogenic or fluorogenic precursors that are transferred to chromophores or fluorophores after cleavage, including coumarin-labeled peptide substrates (Figure 13.7, *1*). For example, pGlu-RTKR-(4-methylcoumaryl-7-amide) is a substrate that is often used to test proprotein convertase preparations or inhibitors [133, 171, 172]. *N*-Acyl-aminocoumarins with diverse nonpeptidic, low-molecular-weight *N*-acyl groups have been used for the discovery of small-molecule substrates of serine and cysteine proteases that can be used for the development of small-molecule inhibitors [173–175].

Apart from 7-amino-4-methylcoumarin, the bifunctional leaving group 7-amino-4-carbamoyl-methylcoumarin can be used, which delivers comparable kinetic data, but permits reduction of enzyme and substrate concentration due to higher quantum yield [176]. Moreover, the bifunctionality of this compound allows efficient

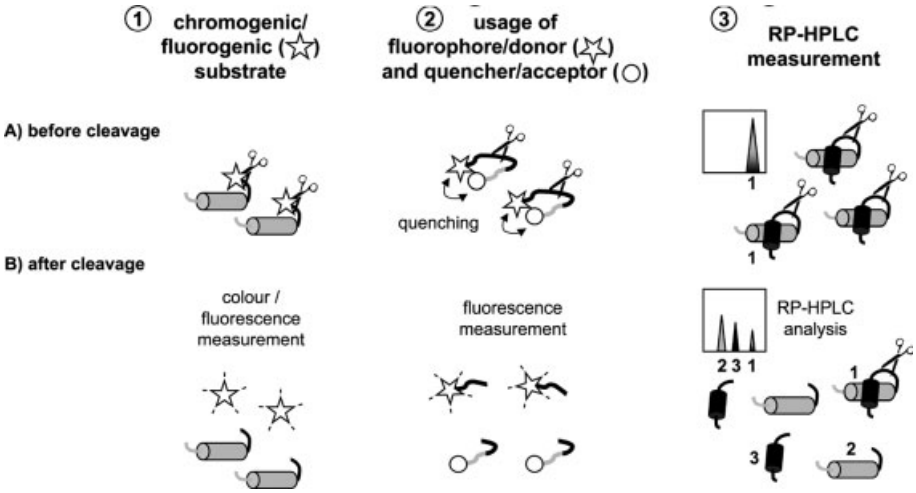


Figure 13.7 Different methods to investigate cleavage by proteases. The schemes illustrate in all three cases the situation before the cleavage (top) and the situation after the cleavage (bottom).

production of substrates and substrate libraries by Fmoc-based solid-phase peptide synthesis [176].

However, one major disadvantage of coumarin substrates is that the coumarin has to be attached through an amide bond at the scissile bond, since only formation of the free amino group by protease cleavage results in the greatly enhanced fluorescence of coumarin. Investigation is therefore restricted to the nonprime site (for nomenclature, see Section 13.3.3.3).

For screening of libraries containing potential inhibitors, interference of the fluorophore with organic compounds can lead to selection for or against candidates with particular spectral properties. Since many organic compounds absorb in the ultraviolet region, they can interfere with coumarin-based fluorescence assays. Although outside the ultraviolet region there is less interference with organic compounds; also red-shifted fluorescent dyes, such as rhodamine, will result in a selection for or against candidates with particular spectral properties [177]. Therefore Grant *et al.* developed a dual-substrate assay and combined aminomethylcoumarin- and rhodamine-based fluorogenic substrates in a single screening [177]. Since interference of the same compound at two separate wavelengths is less likely, selection for or against candidates with particular spectral properties can be reduced [177].

Moreover, substrates with a fluorescence donor and acceptor that are situated within the range of the Förster radius or internally quenched substrates can be used to investigate prohormone cleavage (Figure 13.7, *2*), Table 13.6).

In both cases, fluorescence of the donor or fluorophore is detectable after separation of quencher and fluorophore mediated by cleavage of the precursor. This strategy can be employed for the investigation of both the nonprime and the prime sites (for nomenclature, see Section 13.3.3.3) for their influence on the cleavability.

Table 13.6 Different donor/acceptor and fluorophore/quencher pairs.

Donor/fluorophore	Acceptor/ quencher	Practical aspects	References
o-Aminobenzoic acid; synonymous anthranilic acid (Abz)	3-nitrotyrosine	can be prepared by solid-phase peptide synthesis complete quenching with more than 20 Å	[178, 179]
Edans	4-dimethylamino-azobenzene-4-carboxylic acid (Dabcyl)	relatively long excited state lifetime	[180]
Edans	4-dimethylamino-azobenzene-4′-sulfonyl (Dabsyl)	efficient energy transfer	[181]
N-(2-Aminoethyl)-4-amino-3,6-disulfo-1,8-naphthalimide (Lucifer Yellow)	4-dimethylamino-azobenzene-4′-sulfonyl (Dabsyl)	substrates investigated can be coupled to beads → easy detection of fluorescent bead after cleavage extracts of cells expressing recombinant protease used	[182]
2,3-Diazabicyclo [2.2.2]oct-2-ene (Dbo)	intrinsic Trp/Tyr	easy incorporation of Dbo through Asn side-chain long fluorescence lifetime low background fluorescence due to nanosecond time resolved fluorescence detection increased differentiation between cleaved and uncleaved substrate	[183]

Substrates identified in this way can be used as inhibition indicators, as has been done by Meldal using a one-bead, two-compound assay for solid-phase screening [184]. Accordingly, beads are used as an assay container in which the putative inhibitor competes with the identified substrate [184]. Strong inhibition leads to an uncleaved substrate that remains quenched, while weak inhibition results in substrate cleavage which leads to fluorescence and can be used for sorting [184].

Another method that does not depend on modification is the usage of reverse-phase high-performance liquid chromatography (RP-HPLC) analysis to separate the released peptides from the substrate and quantify their amount (Figure 13.7, *3*). Of course, this is considerably less sensitive than fluorescence analysis.

13.4.2 Rational Design of Inhibitors of the Angiotensin-Converting Enzyme

Design from enzyme-related information has been used for the design of inhibitors of the angiotensin-converting enzyme (ACE), a component of the renin–angiotensin

system, which represents a hormonal cascade that controls cardiovascular, renal and adrenal functions (reviewed in [185]). In addition to the "classical," peripheral renin–angiotensin system, a complete brain renin–angiotensin system exists that comprises all necessary precursors and enzymes required for the formation and metabolism of the biologically active forms of angiotensin (reviewed in [186, 187]).

The renin–angiotensin system and the development of the ACE inhibitor captopril [188, 189] and further inhibitors are reviewed in the following since this the development of this inhibitor represents a good example for rational inhibitor design from enzyme-related information.

The "classical" formation of the peptide hormone angiotensin II is a two-step process (Figure 13.8). In the first step the precursor angiotensinogen is cleaved by the aspartate protease renin (EC 3.4.23.15), which results in the formation of the

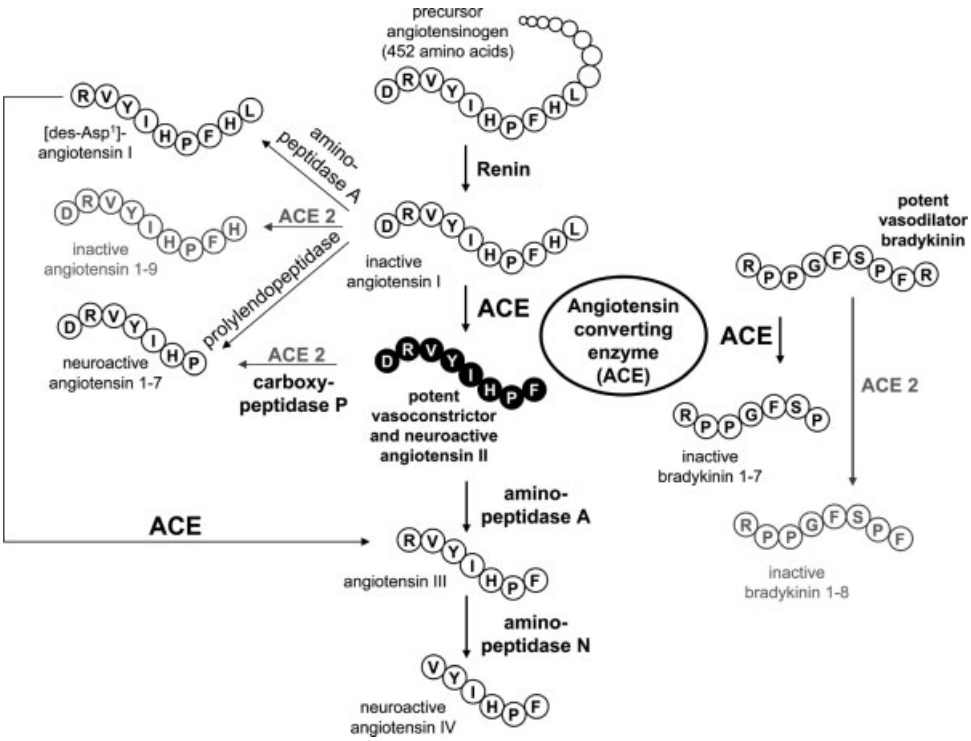


Figure 13.8 Schematic view of the biosynthesis of different angiotensin analogs and of the bradykinin inactivation. The main focal point is on the formation of the potent vasoconstrictor and neuroactive angiotensin II and the inactivation of the potent vasodilator bradykinin by ACE that can be inhibited by ACE inhibitors. The whole network shows the complexity of the system. For clarity reasons direct angiotensinogen conversion to angiotensin II by cathepsin G or tonin is not depicted. The scheme includes both angiotensin analogs of the "classical" pathway, the peripheral renin–angiotensin system, and neuroactive angiotensins occurring within the renin–angiotensin system of the CNS. Additionally, currently observed ACE2 action is depicted in gray.

biologically inactive decapeptide angiotensin I. Angiotensin I can be further processed by ACE to the vasoconstrictory octapeptide angiotensin II by removal of the two C-terminal residues (Figure 13.8). ACE (EC 3.4.15.1), a chloride-dependent zinc glycoprotein that belongs to the peptidyl dipeptidase family, also inactivates the vasodilatory nonapeptide bradykinin (Figure 13.8) which enhances the vasoconstrictory effect. Thus, inhibitors of ACE have the potential to decrease blood pressure by preventing production of angiotensin II and by preventing the inactivation of bradykinin. Recently, another enzyme has been shown to be involved in angiotensin processing. The so-called ACE2 hydrolyzes angiotensin I to angiotensin 1–9, angiotensin II to inactive angiotensin 1–7, and bradykinin to the inactive metabolite bradykinin 1–8 (Figure 13.8).

Within the brain, neuroactive angiotensin II can also be formed directly from angiotensinogen by the serine endopeptidases cathepsin G or tonin [190]. The membrane-bound zinc-dependent aminopeptidase A converts angiotensin II to angiotensin III by splitting off the N-terminal acidic amino acid aspartate [191]. Neuroactive angiotensin IV is formed from angiotensin III by the membrane-bound zinc-dependent aminopeptidase N [191]. A further neuroactive fragment, angiotensin 1–7, can either be formed by prolylendopeptidases from angiotensin I or by carboxypeptidase P action from angiotensin II [187].

The first ACE inhibitors were found in the venom of the Brazilian viper (*Bothrops jararaca*) [192], including the pentapeptide bradykinin-potentiating peptide (BPP) 5a (pGlu-Lys-Trp-Ala-Pro) and the nonapeptide SQ 20 881 (pGlu-Trp-Pro-Arg-Pro-Gln-Ile-Pro-Pro), which is used clinically under the generic name teprotide. It has been shown that the most favorable inhibitory peptides show the C-terminal sequence Ala–Pro [188]. However, BPP5a was rapidly hydrolyzed. Teprotide shows higher stability due to four proline residues, but it is not orally available.

Byers and Wolfenden demonstrated that L-benzylsuccinic acid was a potent competitive inhibitor of the well-characterized carboxypeptidase A [193]. Cushman *et al.* [194] synthesized succinyl-L-proline as a potential inhibitor of ACE (Figure 13.9, *1*), and further improved this compound by rational design assuming mechanistic homology and including the knowledge from the venom peptides (Figure 13.9, *2*). The inhibitory activity could be enhanced by addition of a methyl group (Figure 13.9, *3*) in analogy to the methyl group of alanine in the BPP5a [188, 194]. Replacement of the putative zinc-binding 4-carboxylfunction by a sulfhydryl group led to the extremely potent and specific competitive inhibitor captopril (Figure 13.9, *4*) [194].

Many inhibitory substances followed, including enalaprilat [195] and lisinopril [196]. These and a variety of further ACE inhibitors with their respective structures have been reviewed by Nemec *et al.* [189].

Recently gained insights enabled the development of even further inhibitors. Two isoforms of ACE have been described that are transcribed from the same gene in a tissue-specific manner [197]: the somatic form, which can be found in a variety of tissues and consists of two highly homologous parts (designated as the N- and C-domains), and the testicular form, which is exclusively expressed in germinal cells and is identical to the C-domain of the somatic form except for the first 36 residues [198]. Each domain contains the characteristic HEXXH zinc-binding motif

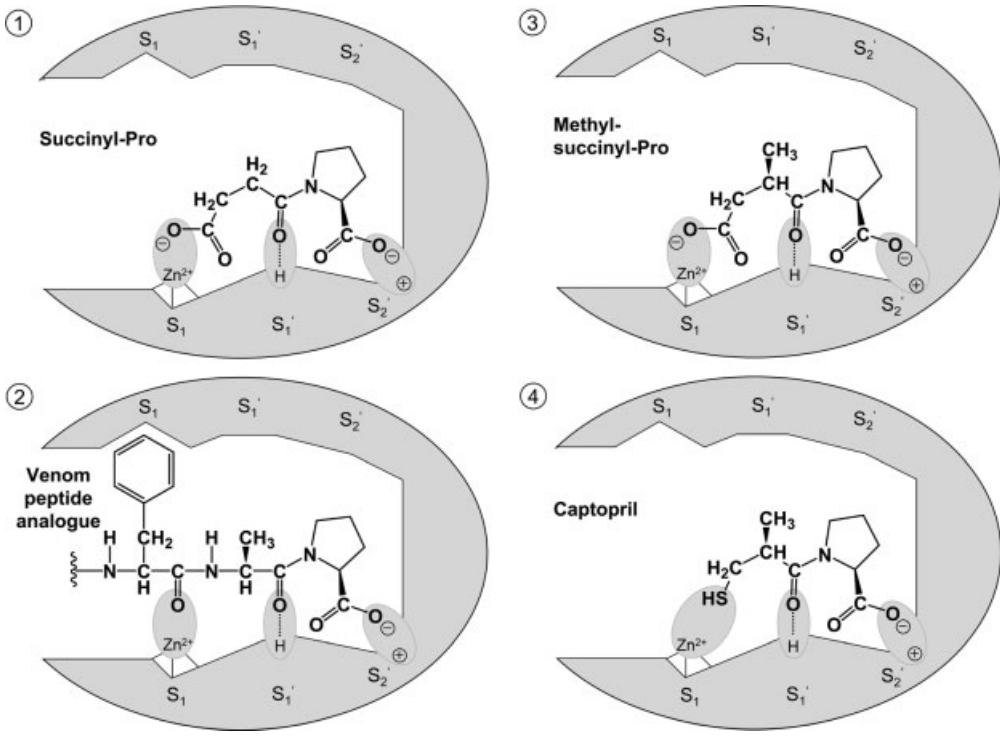
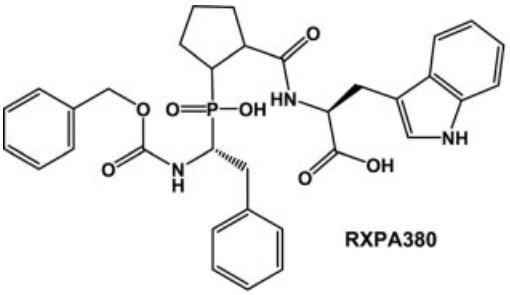


Figure 13.9 Hypothetical binding of competitive inhibitors of ACE according to [188] that illustrate the development of the ACE inhibitor captopril.

of zinc peptidases [199]. Both domains hydrolyze angiotensin I and bradykinin at comparable rates, but with different chloride concentration requirements [200]. The recently published crystal structures of the N-domain of human somatic ACE [201] and testicular ACE [202] provided the possibility for the design of "domain-specific" ACE inhibitors. Interestingly, analysis of the three-dimensional structure of ACE shows that it bears little similarity to that of carboxypeptidase A that had been used for inhibitor design [202].

While the earlier inhibitors captopril, enalaprilat, and lisinopril show comparable selectivity for both active sites [203], RXPA380 (Figure 13.10) was reported as the first highly selective inhibitor of the C-domain [204] and RXP407 (Figure 13.10) as the first highly selective inhibitor of the N-domain of somatic ACE [205]. The cocrystallization study of RXPA380 with ACE could be used to explain the high affinity for the C-domain by the interaction of bulky moieties with residues unique to the C-domain and provides the first structural data for the design of domain-specific pharmacophores [206].

Interestingly, ACE exhibits endopeptidase activity for substrates with amidated C-termini, including gonadoliberin, and for substrates with N-terminal pyroglutamic acid, including also gonadoliberin [208]. On the basis of their respective crystal



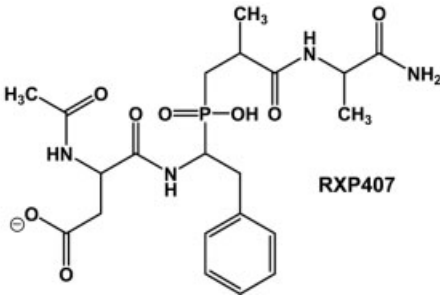


Figure 13.10 Structures of domain-selective inhibitors of ACE: RXPA380 [207], the first highly selective inhibitor of the C-domain of ACE, and RXP407 [204], the first highly selective inhibitor of the N-domain of ACE.

structures, the N- and C-domains of ACE have been studied in complex with gonadoliberin by molecular modeling, which offers new insights into subsites that are distant from the binding site of small molecules and provides an extension of the pharmacophore model used for structure-based drug design up to the S_7 subsite of ACE [209].

Apart from ACE inhibition only, new strategies have been developed to combine this inhibition with intervention of other signaling pathways that contribute to the control of blood pressure. After the development of dual vasopeptidase inhibition, which is based on inhibition of ACE and neutral endopeptidase (EC 3.4.24.11), triple vasopeptidase inhibition has been developed including inhibition of ACE, neutral endopeptidase, and ECE-1 (EC 3.4.24.71). One example of such triple inhibitors is CGS 35 601 [210]. Through inhibition of ACE, triple vasopeptidase inhibitors block the conversion of angiotensin I to the potent vasoconstrictor angiotensin II as well as the degradation of the vasodilator bradykinin. Through the inhibition of neutral endopeptidase, the degradation of peptidic vasodilators, including bradykinin, natriuretic peptides, and adrenomedullin, and peptidic vasoconstrictors, such as endothelins, is prevented, and through the inhibition of ECE-1 the conversion of big endothelin-1 into the potent vasoconstrictor endothelin-1 is suppressed.

13.4.3
Proprotein Convertase Inhibitors

Proprotein convertase inhibitors generally target the catalytic site, which is highly specific for the cleavage after at least one basic residue. Several proprotein convertase inhibitors have been detected, including endogenous protein inhibitors and derived inhibitors thereof (Section 13.4.3.1), as well as peptide inhibitors (Section 13.4.3.2) and peptide-derived inhibitors (Section 13.4.3.3). However, frequently inhibitors are not significantly specific for one proprotein convertase. This section mainly focuses on inhibitors of PC1, PC2, and PC5A, which are involved in the regulated secretory pathway as well as on inhibitors of the widespread proprotein convertase furin.

Table 13.7 Endogenous protein inhibitor and derived peptide K_i values (nM).

Protein inhibitor/derived peptide	PC1	PC2	Furin	References
C-Terminally extended pro-domain of PC1	6 ± 0.6	10000 ± 1000	10 ± 1	[131]
PC1 pro-domain (39–62)	15 400	ND	700	[214]
PC1 (50–83)	700	ND	4800	[214]
Little PEN-LEN rSAAS (221–254)	119 ± 1.5	ND	$19\,300 \pm 1800$	[215]
SAAS (235–244)	9 ± 0.5	ND	261 ± 44	[215]
SAAS (235–246)	51 ± 3.8	ND	$39\,400 \pm 4300$	[215]
7B2	no inhibition	6.7	ND	[216]
Barley serine protease inhibitor 2-derived peptide	3900	ND	300	[217]
SPN4A	ND	3.5	0.013	[218]
CRES	no inhibition	25 ± 3.1	no inhibition	[219]

Inhibition constants for the main prohormone processing enzymes PC1 and PC2 as well as for the widespread convertase furin are compared. In most cases pGlu-RTKR-MCA has been used as substrate; for the 7B2 investigations, carbobenzoxy-RTKR-MCA has been used.ND, not determined.

13.4.3.1 Endogenous Protein Inhibitors and Derived Inhibitors

Endogenous protein inhibitors may serve as templates for the design of inhibitory peptides. Several pro-domains of prohormone convertases have been shown to inhibit convertase activity, including the pro-domain of PC1 [131] and furin [211, 212]. However, it has been shown that these pro-domains show only a partial specificity towards their cognate enzyme [212, 213]. For instance, the pro-domain of PC1 is a slow-tight binding inhibitor of both PC1 and furin with K_i in the nanomolar range [131] (Table 13.7). Therefore, the usage of pro-domains as specific inhibitors would require modification to get more specific compounds since proprotein convertases are often coexpressed within cells and poorly selective inhibitors will affect multiple cellular functions.

Apart from the pro-domain, the neuroendocrine proteins proSAAS and 7B2 serve as endogenous inhibitors of PC1 [220] and PC2 [216], respectively. The PC1 inhibitory region of proSAAS was mapped to an 8–12 residue region near the C-terminus containing a Lys–Arg sequence [221]. The 27-kDa 7B2, but not the C-terminally truncated 21-kDa 7B2, has been shown to be a tight binding inhibitor of PC2, while PC1 *in vitro* cleaves the 27-kDa 7B2 and is not inhibited [216].

Moreover, barley serine proteinase inhibitor-2, a known subtilisin inhibitor, has been altered by specific mutations to achieve potent inhibition of PC1 and furin [217]. Since the inhibitors developed were slowly cleaved by PC1 and furin, the P_1–P'_1 amide bond has been replaced by a nonpeptidic aminomethylene bond [217]. Interestingly, this change has resulted in a potent furin inhibitor that was not cleaved; however, PC1 was no longer inhibited at the same concentrations [217].

Richer *et al*. used homology-search programs for the screening of serine protease inhibitor (also known as serpin) genes in eukaryotes and identified the gene *spn4* in *Drosophila melanogaster* that encodes the most potent and effective endogenous

serpin [218]. SPN4A, which contains an Arg–Arg–Lys–Arg sequence in its active-site loop, has been shown to inactivate human furin and *Drosophila* PC2 by a slow-binding mechanism and formation of a sodium decylsulfate-stable complex with both enzymes [218]. Removal of an alanine residue from the active-site loop of the serpin results in a reduction of inhibition potency against PC2, while inhibition potency against furin remains unchanged [218].

The cystatin-related epididymal spermatogenic (CRES) protein is related to the family 2 cystatins of the cystatin superfamily of cysteine protease inhibitors, but it lacks sequences important for cysteine protease inhibition [219]. However, CRES has been shown *in vitro* to inhibit PC2, but not furin and PC1 [219].

13.4.3.2 **Peptide Inhibitors**

Apart from peptides derived from endogenous inhibitors (see Section 13.4.3.1), positional scanning of synthetic peptide combinatorial libraries has been used to identify potential inhibitory peptides for PC1 [222, 223] and PC2 [222] as well as PC5A [224], and to get information about active-site determinants. Accordingly, S_5, S_4, and S_3 subsites of PC1 appear to resemble those of PC2, while the S_6 subsite of PC1 appears to be more selective than that of PC2 [222]. Nevertheless, the most potent inhibitor identified in the study by Lindberg *et al.* (Ac-Leu-Leu-Arg-Val-Lys-Arg-NH_2) proved to be the most potent inhibitor of both convertases determined by observing the rate of hydrolysis of pGlu-Arg-Thr-Lys-Arg-MCA (Table 13.8) [222]. Interestingly, this hexapeptide sequence was later identified within the endogenous inhibitor proSAAS [220].

Using positional-scanning synthetic peptide combinatorial libraries not only nona-L-arginine has been identified as a potent inhibitor of furin, but also nona-D-arginine, which is more stable against proteolytic degradation [223, 224]. Introducing this poly-D-arginine sequence into engineered polypeptides might achieve greater specificity than polyarginine alone while retaining high potency [225].

13.4.3.3 **Peptide-Derived Inhibitors**

Acyl-peptidyl-chloromethyl ketones, including decanoyl-Arg–Val–Lys–Arg-chloromethyl ketone, are often used as small-molecule inhibitors of proprotein convertases, especially furin [226–229]. They form irreversible complexes with the active site of the convertase through their peptidyl group and are able to enter cells by the acyl moiety

Table 13.8 Peptide inhibitor K_i values (nM).

Peptide sequence of the inhibitor	PC1	PC2	PC5A	Furin	References
Ac-LLRVKR-NH_2	3.2 ± 1.0	360 ± 50	ND	1400 ± 230	[222]
Nona-L-arginine	12000 ± 2500	no inhibition	150	42 ± 3	[223, 224]
Nona-D-arginine	ND	ND	19	1.3	[224]

Inhibition constants for the main prohormone processing enzymes PC1, PC2, and PC5A as well as for the widespread convertase furin are compared. In all cases pGlu-RTKR-MCA has been used as substrate.ND, not determined.

within their structure [230]. However, their use *in vivo* is limited due to their toxicity. They can be used as active-site titrants to determine the catalytically active enzyme concentration [231] or pH optima of different convertase processing forms [232]. Moreover, the crystal structure of furin with decanoyl-Arg–Val–Lys–Arg-chloromethylketone has led to the identification of the active amino acids that form the S_1–S_4 pocket [233]. Modeling and docking analysis has led to the extrapolation of further subsites and has shown high similarities between the different convertases [109]. Therefore it is unlikely that short-peptide inhibitors might be specific for individual proprotein convertases. However, differences have been recognized for the S_6 subsite, which indicates that short peptide inhibitors will require at least six residues [109, 225].

13.4.3.4 Are there Conformational Requirements for Substrates?

Are there conformational requirements for prohormone convertase substrates that can be used for inhibitor design? Although there are several preferences for prohormone cleavage sites [83, 234–236] that can be visualized well by a sequence logo [237, 238] (an example is shown in Figure 13.11), examination of sequences around cleaved sites finally could not reveal any consensus primary sequence [239] that is cleaved in all cases. Thus, the question arises, what is the reason for a prohormone convertase to cleave at some basic sites and to leave a lot of other basic sites intact?

It can be assumed that a specific secondary structure of the substrate is a necessity for proprotein convertase activity. Consequently, the secondary structure of substrates and uncleaved peptides with dibasic sequences should be investigated experimentally. If no experimental information is available, structure predictions could also be helpful.

Secondary structure predictions of the regions around basic amino acids in different hormone precursors reveal that cleaved sites are likely in or next to flexible regions with high β-turn [239] or ω-loop [240] formation probability. In addition to these theoretical investigations, the influence of the structure of the substrate on prohormone processing has been investigated in several experimental studies [241, 242]. Brakch *et al.* have demonstrated on peptide analogs of the pro-oxytocin/neurophysin processing domain that processing at dibasic sites is associated with a β-turn structure, the presence of which was confirmed by circular dichroism [95, 241]

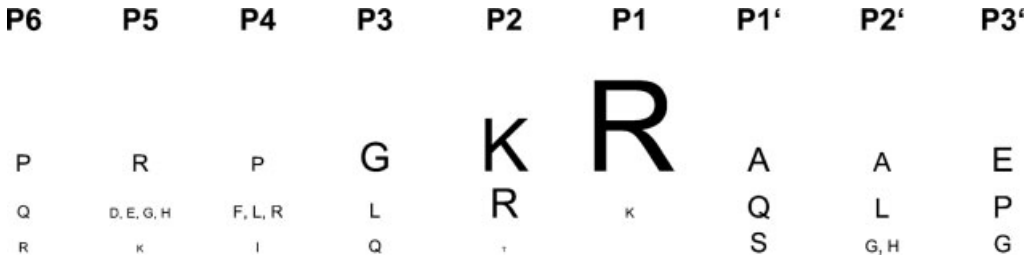


Figure 13.11 Sequence logo for substrate cleavage sites cleaved by PC1. The figure is generated from 42 cleavage sites within 14 prohormones that are cleaved by PC1.

as well as Fourier transform infrared spectroscopy [95, 243] and nuclear magnetic resonance spectroscopy [95]. Moreover, it has been demonstrated that the β-turn structure is an interchangeable motif [241, 244]. Later, this group showed that introduction of a β-turn breaker into the proposed turn region abolishes the enzyme–substrate interaction [242]. Moreover, they state that the flexibility of the peptide substrate is necessary for the interaction [242, 245].

A general computer analysis of superimposed crystal structures including 95 small-molecule inhibitors bound to 17 serine proteases reveals that inhibitors, including substrate analogs, commonly bind in an extended β-strand conformation at the active site [246]. From the analysis of crystal structure data of furin and molecular modeling, Henrich *et al.* assumed that many substrates might bind their cognate convertase in an overall extended conformation [109].

Stabilization of defined secondary structures can help to find out whether special conformational requirements of substrates are necessary for enzymatic activity like cleavage by prohormone processing enzymes [247]. Such requirements could be used as restraints in the design of inhibitors. For instance, conformational restriction by cyclization has been used to stabilize secondary structures to obtain an insight into conformational requirements for ligands necessary for receptor binding that can be used for drug design [248]. However, to what extent the substrate conformation influences its cleavability remains an open question.

13.5 Conclusions

Peptide hormones represent a large group of hormones and are involved in a variety of physiological functions. Their biosynthesis, starting from precursor sequences, may be regulated at several steps including cleavage by proprotein convertases, which can lead to different products depending on the cleavage sites actually used, and other PTMs that may change hormone activity and may even lead to alteration of receptor specificity. In the case of angiotensin II, inhibition of ACE, which is responsible for angiotensin II formation, has led to clinically relevant pharmacology. In addition to prohormone processing, mainly by PC1 and PC2, proprotein convertases including furin are also involved in processing of bacteria and viruses. However, selective inhibition of proprotein convertases is still challenging.

References

1 Greenspan, F.S. and Gardner, D.G.(eds) (2004) *Basic & Clinical Endocrinology*, 7th edn, McGraw-Hill, New York.

2 Krauss, G. (2008) *Biochemistry of Signal Transduction and Regulation*, 4th edn, Wiley-VCH Verlag GmbH, Weinheim.

3 O'Donohue, T.L., Chronwall, B.M., Pruss, R.M., Mezey, E., Kiss, J.Z., Eiden, L.E., Massari, V.J., Tessel, R.E., Pickel, V.M., DiMaggio, D.A., Hotchkiss, A.J., Crowley, W.R. and Zukowska-Grojec, Z. (1985) *Peptides*, **6**, 755–768.

4 Kiss, A. and Mikkelsen, J.D. (2005) *Endocrine Regulations*, **39**, 97–105.

5 Williams, J.A. (1982) *Biomedical Research*, **3**, 107–121.

6 Rashid, A.J., O'Dowd, B.F. and George, S.R. (2004) *Endocrinology*, **145**, 2645–2652.

7 Re, R. (1999) *Hypertension*, **34**, 534–538.

8 Re, R.N. and Cook, J.L. (2006) *Regulatory Peptides*, **133**, 1–9.

9 Gobeil, F., Fortier, A., Zhu, T., Bossolasco, M., Leduc, M., Grandbois, M., Heveker, N., Bkaily, G., Chemtob, S. and Barbaz, D. (2006) *Canadian Journal of Physiology and Pharmacology*, **84**, 287–297.

10 Re, R.N. (1989) *Journal of Molecular and Cellular Cardiology*, **21** (Suppl. 5), 63–69.

11 Kinsey, C.G., Bussolati, G., Bosco, M., Kimura, T., Pizzorno, M.C., Chernin, M.I., Cassoni, P. and Novak, J.F. (2007) *Journal of Cellular and Molecular Medicine*, **11**, 96–110.

12 Jacques, D., Sader, S., Perreault, C., Fournier, A., Pelletier, G., Beck-Sickinger, A.G. and Descorbeth, M. (2003) *Canadian Journal of Physiology and Pharmacology*, **81**, 288–300.

13 Re, R.N. and Cook, J.L. (2008) *Journal of Clinical Pharmacology*, **48**, 344–350.

14 IUPAC–IUB Commission on Biochemical Nomenclature (1975) *The Journal of Biological Chemistry*, **250**, 3215–3216.

15 Tatemoto, K. (1982) *Proceedings of the National Academy of Sciences of the United States of America*, **79**, 5485–5489.

16 Goltermann, N.R., Rehfeld, J.F. and Roigaard-Petersen, H. (1980) *The Journal of Biological Chemistry*, **255**, 6181–6185.

17 Cowley, M.A., Smith, R.G., Diano, S., Tschop, M., Pronchuk, N., Grove, K.L., Strasburger, C.J., Bidlingmaier, M., Esterman, M., Heiman, M.L., Garcia-Segura, L.M., Nillni, E.A., Mendez, P., Low, M.J., Sotonyi, P., Friedman, J.M., Liu, H., Pinto, S., Colmers, W.F., Cone, R.D. and Horvath, T.L. (2003) *Neuron*, **37**, 649–661.

18 Gray, T.S. and Morley, J.E. (1986) *Life Sciences*, **38**, 389–401.

19 Dumont, Y., Martel, J.C., Fournier, A., St-Pierre, S. and Quirion, R. (1992) *Progress in Neurobiology*, **38**, 125–167.

20 Gehlert, D.R. (1999) *Neuropeptides*, **33**, 329–338.

21 Malmstrom, R.E. (2002) *European Journal of Pharmacology*, **447**, 11–30.

22 Vezzani, A., Sperk, G. and Colmers, W.F. (1999) *Trends in Neurosciences*, **22**, 25–30.

23 Flood, J.F., Baker, M.L., Hernandez, E.N. and Morley, J.E. (1989) *Brain Research*, **503**, 73–82.

24 Albers, H.E. and Ferris, C.F. (1984) *Neuroscience Letters*, **50**, 163–168.

25 Cabrele, C. and Beck-Sickinger, A.G. (2000) *Journal of Peptide Science*, **6**, 97–122.

26 Kastin, A.J.(ed.) (2006) *Handbook of Biologically Active Peptides*, Elsevier, Amsterdam.

27 Stragier, B., De Bundel, D., Sarre, S., Smolders, I., Vauquelin, G., Dupont, A., Michotte, Y. and Vanderheyden, P. (2008) *Heart Failure Reviews*, **13**, 321–337.

28 Martin, C. (2006) *Diabetes Educator*, **32** (Suppl. 3), 101S–104S.

29 Lerner, U.H. (2006) *Journal of Musculoskeletal & Neuronal Interactions*, **6**, 87–95.

30 Benemei, S., Nicoletti, P., Capone, J.A. and Geppetti, P. (2007) *Neurological Sciences*, **28** (Suppl. 2), S89–S93.

31 Rehfeld, J.F., Friis-Hansen, L., Goetze, J.P. and Hansen, T.V.O. (2007) *Current Topics in Medicinal Chemistry*, **7**, 1154–1165.

32 Drucker, D.J. (2002) *Gastroenterology*, **122**, 531–544.

33 Sindic, A. and Schlatter, E. (2006) *Journal of the American Society of Nephrology*, **17**, 607–616.

34 Kasik, J.W., Lu, C. and Menon, R.K. (2000) *Pediatric Diabetes*, **1**, 169–177.

35 Dschietzig, T., Bartsch, C., Baumann, G. and Stangl, K. (2006) *Pharmacology & Therapeutics*, **112**, 38–56.

36 Kojima, M. and Kangawa, K. (2005) *Physiological Reviews*, **85**, 495–522.

37 Poitras, P. and Peeters, T.L. (2008) *Current Opinion in Endocrinology, Diabetes & Obesity*, **15**, 54–57.

38 Pandey, K.N. (2005) *Peptides*, **26**, 901–932.

39 Tamura, N., Ogawa, Y., Chusho, H., Nakamura, K., Nakao, K., Suda, M., Kasahara, M., Hashimoto, R., Katsuura, G., Mukoyama, M., Itoh, H., Saito, Y., Tanaka, I., Otani, H., Katsuki, M. and Nakao, K. (2000) *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 4239–4244.

40 Matsukawa, N., Grzesik, W.J., Takahashi, N., Pandey, K.N., Pang, S., Yamauchi, M. and Smithies, O. (1999) *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 7403–7408.

41 Pedrazzini, T., Pralong, F. and Grouzmann, E. (2003) *Cellular and Molecular Life Sciences*, **60**, 350–377.

42 Thorsell, A. (2007) *Peptides*, **28**, 480–483.

43 Boey, D., Sainsbury, A. and Herzog, H. (2007) *Peptides*, **28**, 390–395.

44 Kojima, S., Ueno, N., Asakawa, A., Sagiyama, K., Naruo, T., Mizuno, S. and Inui, A. (2007) *Peptides*, **28**, 459–463.

45 Konig, S., Luger, T.A. and Scholzen, T.E. (2006) *Experimental Dermatology*, **15**, 751–761.

46 Humphreys, M.H. (2007) *Current Opinion in Nephrology and Hypertension*, **16**, 32–38.

47 Boutrel, B. and de Lecea, L. (2008) *Physiology & Behavior*, **93**, 947–951.

48 Ohno, K. and Sakurai, T. (2008) *Frontiers in Neuroendocrinology*, **29**, 70–87.

49 Patel, Y.C. (1999) *Frontiers in Neuroendocrinology*, **20**, 157–198.

50 Lee, H.-J., Caldwell, H.K., Macbeth, A.H., Tolu, S.G. and Young, W.S. 3rd (2008) *Endocrinology*, **49**, 3256–3263.

51 Caldwell, H.K., Lee, H.-J., Macbeth, A.H. and Young, W.S. 3rd (2008) *Progress in Neurobiology*, **84**, 1–24.

52 Lu, C.-H., Chen, Y.-C., Yu, C.-S. and Hwang, J.-K. (2007) *Proteins*, **67**, 262–270.

53 Zhang, M. and Kaltashov, I.A. (2006) *Analytical Chemistry*, **78**, 4820–4829.

54 Tang, H.-Y. and Speicher, D.W. (2004) *Current Protocols in Protein Science*, Unit 11.11. DOI: .

55 Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S. and Brunak, S. (2004) *Proteomics*, **4**, 1633–1649.

56 Julenius, K., Molgaard, A., Gupta, R. and Brunak, S. (2005) *Glycobiology*, **15**, 153–164.

57 Wuhrer, M., Catalina, M.I., Deelder, A.M. and Hokke, C.H. (2007) *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, **849**, 115–128.

58 Tang, Y.-R., Chen, Y.-Z., Canchaya, C.A. and Zhang, Z. (2007) *Protein Engineering, Design and Selection*, **20**, 405–412.

59 Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z. and Dunker, A.K. (2004) *Nucleic Acids Research*, **32**, 1037–1049.

60 Gafken, P.R. and Lampe, P.D. (2006) *Cell Communication & Adhesion*, **13**, 249–262.

61 Raeck, C. and Berger, S. (2007) *Analytical and Bioanalytical Chemistry*, **389**, 2161–2165.

62 Huttner, W.B. (1988) *Annual Review of Physiology*, **50**, 363–376.

63 Monigatti, F., Gasteiger, E., Bairoch, A. and Jung, E. (2002) *Bioinformatics*, **18**, 769–770.

64 Monigatti, F., Hekking, B. and Steen, H. (2006) *Biochimica et Biophysica Acta*, **1764**, 1904–1913.

65 Yu, Y., Hoffhines, A.J., Moore, K.L. and Leary, J.A. (2007) *Nature Methods*, **4**, 583–588.

66 Tatemoto, K. and Mutt, V. (1978) *Proceedings of the National Academy of Sciences of the United States of America*, **75**, 4115–4119.

67 Bacher, G., Lutcke, H., Jungnickel, B., Rapoport, T.A. and Dobberstein, B. (1996) *Nature*, **381**, 248–251.

68 Sevier, C.S. and Kaiser, C.A. (2002) *Nature Reviews. Molecular Cell Biology*, **3**, 836–847.

69 Salvas, A., Benjannet, S., Reudelhuber, T.L., Chretien, M. and Seidah, N.G. (2005) *FEBS Letters*, **579**, 5621–5625.

70 Bass, J., Turck, C., Rouard, M. and Steiner, D.F. (2000) *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 11905–11909.

71 Rosa, P., Mantovani, S., Rosboch, R. and Huttner, W.B. (1992) *The Journal of Biological Chemistry*, **267**, 12227–12232.

72 Kelly, R.B. (1985) *Science*, **230**, 25–32.

73 Trifaro, J.-M., Glavinovic, M. and Rose, S.D. (1997) *Neurochemical Research*, **22**, 831–841.

74 Neher, E. and Zucker, R.S. (1993) *Neuron*, **10**, 21–30.

75 Fesce, R., Grohovaz, F., Valtorta, F. and Meldolesi, J. (1994) *Trends in Cell Biology*, **4**, 1–4.

76 Henkel, A.W. and Almers, W. (1996) *Current Opinion in Neurobiology*, **6**, 350–357.

77 Michael, D.J., Cai, H., Xiong, W., Ouyang, J. and Chow, R.H. (2006) *Trends in Endocrinology and Metabolism*, **17**, 408–415.

78 Arvan, P. and Castle, D. (1998) *The Biochemical Journal*, **332**, 593–610.

79 Cool, D.R., Normant, E., Shen, F.-S., Chen, H.-C., Pannell, L., Zhang, Y. and Loh, Y.P. (1997) *Cell*, **88**, 73–83.

80 Kuliawat, R. and Arvan, P. (1992) *The Journal of Cell Biology*, **118**, 521–529.

81 Dikeakos, J.D. and Reudelhuber, T.L. (2007) *The Journal of Cell Biology*, **177**, 191–196.

82 Amare, A., Hummon, A.B., Southey, B.R., Zimmerman, T.A., Rodriguez-Zas, S.L. and Sweedler, J.V. (2006) *Journal of Proteome Research*, **5**, 1162–1167.

83 Nakayama, K., Watanabe, T., Nakagawa, T., Kim, W.S., Nagahama, M., Hosaka, M., Hatsuzawa, K., Kondoh-Hashiba, K. and Murakami, K. (1992) *The Journal of Biological Chemistry*, **267**, 16335–16340.

84 Mouchantaf, R., Watt, H.L., Sulea, T., Seidah, N.G., Alturaihi, H., Patel, Y.C. and Kumar, U. (2004) *Regulatory Peptides*, **120**, 133–140.

85 Hook, V., Yasothornsrikul, S., Greenbaum, D., Medzihradszky, K.F., Troutner, K., Toneff, T., Bundey, R., Logrinova, A., Reinheckel, T., Peters, C. and Bogyo, M. (2004) *Biological Chemistry*, **385**, 473–480.

86 Hook, V., Funkelstein, L., Lu, D., Bark, S., Wegrzyn, J. and Hwang, S.-R. (2008) *Annual Review of Pharmacology and Toxicology*, **48**, 393–423.

87 Loh, Y.P., Parish, D.C. and Tuteja, R. (1985) *The Journal of Biological Chemistry*, **260**, 7194–7205.

88 Azaryan, A.V., Schiller, M.R. and Hook, V.Y.H. (1995) *Biochemical and Biophysical Research Communications*, **215**, 937–944.

89 Hook, V.Y., Schiller, M.R. and Azaryan, A.V. (1996) *Archives of Biochemistry and Biophysics*, **328**, 107–114.

90 Plevrakis, I., Clamagirand, C., Creminon, C., Brakch, N., Rholam, M. and Cohen, P. (1989) *Biochemistry*, **28**, 2705–2710.

91 Fontes, G., Lajoix, A.-D., Bergeron, F., Cadel, S., Prat, A., Foulon, T., Gross, R., Dalle, S., Le-Nguyen, D., Tribillac, F. and Bataille, D. (2005) *Endocrinology*, **146**, 702–712.

92 Mzhavia, N., Pan, H., Che, F.Y., Fricker, L.D. and Devi, L.A. (2003) *The Journal of Biological Chemistry*, **278**, 14704–14711.

93 Steiner, D.F. (1998) *Current Opinion in Chemical Biology*, **2**, 31–39.

94 Brakch, N., Boussetta, H., Rholam, M. and Cohen, P. (1989) *The Journal of Biological Chemistry*, **264**, 15912–15916.

95 Paolillo, L., Simonetti, M., Brakch, N., D'Auria, G., Saviano, M., Dettin, M., Rholam, M., Scatturin, A., Di Bello, C. and Cohen, P. (1992) *The EMBO Journal*, **11**, 2399–2405.

96 Seidah, N.G., Mowla, S.J., Hamelin, J., Mamarbachi, A.M., Benjannet, S., Toure, B.B., Basak, A., Munzer, J.S., Marcinkiewicz, J., Zhong, M., Barale, J.-C., Lazure, C., Murphy, R.A., Chretien, M. and Marcinkiewicz, M. (1999) *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 1321–1326.

97 Sakai, J., Rawson, R.B., Espenshade, P.J., Cheng, D., Seegmiller, A.C., Goldstein, J.L. and Brown, M.S. (1998) *Molecular Cell*, **2**, 505–514.

98 Seidah, N.G., Benjannet, S., Wickham, L., Marcinkiewicz, J., Jasmin, S.B., Stifani, S., Basak, A., Prat, A. and Chretien, M. (2003) *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 928–933.

99 Seidah, N.G., Khatib, A.M. and Prat, A. (2006) *Biological Chemistry*, **387**, 871–877.

100 Pan, H., Nanno, D., Che, F.-Y., Zhu, X., Salton, S.R., Steiner, D.F., Fricker, L.D. and Devi, L.A. (2005) *Biochemistry*, **44**, 4939–4948.

101 Rehfeld, J.F. (2006) *Journal of Molecular Medicine*, **84**, 544–550.

102 Zhu, X., Cao, Y., Voogd, K. and Steiner, D.F. (2006) *The Journal of Biological Chemistry*, **281**, 38867–38870.

103 Ugleholdt, R., Zhu, X., Deacon, C.F., Orskov, C., Steiner, D.F. and Holst, J.J. (2004) *Endocrinology*, **145**, 1349–1355.

104 Furuta, M., Zhou, A., Webb, G., Carroll, R., Ravazzola, M., Orci, L. and Steiner, D.F. (2001) *The Journal of Biological Chemistry*, **276**, 27197–27202.

105 Brakch, N., Rist, B., Beck-Sickinger, A.G., Goenaga, J., Wittek, R., Burger, E., Brunner, H.R. and Grouzmann, E. (1997) *Biochemistry*, **36**, 16309–16320.

106 Tanaka, S. (2003) *Zoological Science*, **20**, 1183–1198.

107 Brakch, N., Galanopoulou, A.S., Patel, Y.C., Boileau, G. and Seidah, N.G. (1995) *FEBS Letters*, **362**, 143–146.

108 Nillni, E.A. and Sevarino, K.A. (1999) *Endocrine Reviews*, **20**, 599–648.

109 Henrich, S., Lindberg, I., Bode, W. and Than, M.E. (2005) *Journal of Molecular Biology*, **345**, 211–227.

110 Dikeakos, J.D., Mercure, C., Lacombe, M.-J., Seidah, N.G. and Reudelhuber, T.L. (2007) *FEBS Journal*, **274**, 4094–4102.

111 Nour, N., Mayer, G., Mort, J.S., Salvas, A., Mbikay, M., Morrison, C.J., Overall, C.M. and Seidah, N.G. (2005) *Molecular Biology of the Cell*, **16**, 5215–5226.

112 Stettler, H., Suri, G. and Spiess, M. (2005) *Biochemistry*, **44**, 5339–5345.

113 Arnaoutova, I., Smith, A.M., Coates, L.C., Sharpe, J.C., Dhanvantari, S., Snell, C.R., Birch, N.P. and Loh, Y.P. (2003) *Biochemistry*, **42**, 10445–10455.

114 Assadi, M., Sharpe, J.C., Snell, C. and Loh, Y.P. (2004) *Biochemistry*, **43**, 7798–7807.

115 Tadros, H., Chretien, M. and Mbikay, M. (2001) *Journal of Reproductive Immunology*, **49**, 133–152.

116 Mbikay, M., Tadros, H., Ishida, N., Lerner, C.P., De Lamirande, E., Chen, A., El-Alfy, M., Clermont, Y., Seidah, N.G., Chretien, M., Gagnon, C. and Simpson, E.M. (1997) *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 6842–6846.

117 Gyamera-Acheampong, C., Tantibhedhyangkul, J., Weerachatyanukul, W., Tadros, H., Xu, H., van de Loo, J.-W., Pelletier, R.-M., Tanphaichitr, N. and Mbikay, M. (2006) *Biology of Reproduction*, **74**, 666–673.

118 De Bie, I., Marcinkiewicz, M., Malide, D., Lazure, C., Nakayama, K., Bendayan, M. and Seidah, N.G. (1996) *The Journal of Cell Biology*, **135**, 1261–1275.

119 Seidah, N.G. and Chretien, M. (1997) *Current Opinion in Biotechnology*, **8**, 602–607.

120 Scamuffa, N., Calvo, F., Chretien, M., Seidah, N.G. and Khatib, A.-M. (2006) *The FASEB Journal*, **20**, 1954–1963.

121 Bassi, D.E., Mahloogi, H. and Klein-Szanto, A.J.P. (2000) *Molecular Carcinogenesis*, **28**, 63–69.

122 Constam, D.B. and Robertson, E.J. (2000) *Genes and Development*, **14**, 1146–1155.

123 Misumi, Y., Oda, K., Fujiwara, T., Takami, N., Tashiro, K. and Ikehara, Y. (1991) *The Journal of Biological Chemistry*, **266**, 16954–16959.

124 Thimon, V., Belghazi, M., Dacheux, J.-L. and Gatti, J.-L. (2006) *Reproduction*, **132**, 899–908.

125 Molloy, S.S., Thomas, L., VanSlyke, J.K., Stenberg, P.E. and Thomas, G. (1994) *The EMBO Journal*, **13**, 18–33.

126 Chapman, R.E. and Munro, S. (1994) *The EMBO Journal*, **13**, 2305–2312.

127 Liu, G., Thomas, L., Warren, R.A., Enns, C.A., Cunningham, C.C., Hartwig, J.H. and Thomas, G. (1997) *The Journal of Cell Biology*, **139**, 1719–1733.

128 Nakayama, K. (1997) *The Biochemical Journal*, **327**, 625–635.

129 Shennan, K.I., Taylor, N.A., Jermany, J.L., Matthews, G. and Docherty, K. (1995) *The Journal of Biological Chemistry*, **270**, 1402–1407.

130 Zhou, Y. and Lindberg, I. (1993) *The Journal of Biological Chemistry*, **268**, 5615–5623.

131 Boudreault, A., Gauthier, D. and Lazure, C. (1998) *The Journal of Biological Chemistry*, **273**, 31574–31580.

132 Tangrea, M.A., Bryan, P.N., Sari, N. and Orban, J. (2002) *Journal of Molecular Biology*, **320**, 801–812.

133 Rabah, N., Gauthier, D., Dikeakos, J.D., Reudelhuber, T.L. and Lazure, C. (2007) *FEBS Journal*, **274**, 3482–3491.

134 Jutras, I., Seidah, N.G. and Reudelhuber, T.L. (2000) *The Journal of Biological Chemistry*, **275**, 40337–40343.

135 Schechter, I. and Berger, A. (1967) *Biochemical and Biophysical Research Communications*, **27**, 157–162.

136 Duckert, P., Brunak, S. and Blom, N. (2004) *Protein Engineering, Design and Selection*, **17**, 107–112.

137 Southey, B.R., Amare, A., Zimmerman, T.A., Rodriguez-Zas, S.L. and Sweedler, J.V. (2006) *Nucleic Acids Research*, **34**, W267–W272.

138 Mirabeau, O., Perlas, E., Severini, C., Audero, E., Gascuel, O., Possenti, R., Birney, E., Rosenthal, N. and Gross, C. (2007) *Genome Research*, **17**, 320–327.

139 Magnan, J., Paterson, S.J. and Kosterlitz, H.W. (1982) *Life Sciences*, **31**, 1359–1361.

140 Fricker, L.D. (1988) *Annual Review of Physiology*, **50**, 309–321.

141 Foulon, T., Cadel, S. and Cohen, P. (1999) *The International Journal of Biochemistry & Cell Biology*, **31**, 747–750.

142 Hwang, S.-R., O'Neill, A., Bark, S., Foulon, T. and Hook, V. (2007) *Journal of Neurochemistry*, **100**, 1340–1350.

143 Fricker, L.D., Berman, Y.L., Leiter, E.H. and Devi, L.A. (1996) *The Journal of Biological Chemistry*, **271**, 30619–30624.

144 Song, L. and Fricker, L.D. (1995) *The Journal of Biological Chemistry*, **270**, 25007–25013.

145 Song, L. and Fricker, L.D. (1996) *The Journal of Biological Chemistry*, **271**, 28884–28889.

146 Kalinina, E., Varlamov, O. and Fricker, L.D. (2002) *Journal of Cellular Biochemistry*, **85**, 101–111.

147 Merkler, D.J. (1994) *Enzyme and Microbial Technology*, **16**, 450–456.

148 Bolkenius, F.N. and Ganzhorn, A.J. (1998) *General Pharmacology*, **31**, 655–659.

149 Eipper, B.A., Milgram, S.L., Husten, E.J., Yun, H.Y. and Mains, R.E. (1993) *Protein Science*, **2**, 489–497.

150 Dockray, G.J., Varro, A., Dimaline, R. and Wang, T. (2001) *Annual Review of Physiology*, **63**, 119–139.

151 Orskov, C., Wettergren, A. and Holst, J.J. (1993) *Diabetes*, **42**, 658–661.

152 Wettergren, A., Pridal, L., Wojdemann, M. and Holst, J.J. (1998) *Regulatory Peptides*, **77**, 83–87.

153 Polevoda, B. and Sherman, F. (2002) *Genome Biology*, **3**, reviews0006.

154 Wilkinson, C.W. (2006) *Peptides*, **27**, 453–471.

155 Kojima, M., Hosoda, H., Date, Y., Nakazato, M., Matsuo, H. and Kangawa, K. (1999) *Nature*, **402**, 656–660.

156 Sakurai, T., Amemiya, A., Ishii, M., Matsuzaki, I., Chemelli, R.M., Tanaka, H., Williams, S.C., Richardson, J.A., Kozlowski, G.P., Wilson, S., Arch, J.R.S., Buckingham, R.E., Haynes, A.C., Carr, S.A., Annan, R.S., McNulty, D.E., Liu, W.-S., Terrett, J.A., Elshourbagy, N.A., Bergsma, D.J. and Yanagisawa, M. (1998) *Cell*, **92**, 573–585.

157 Awade, A.C., Cleuziat, P., Gonzales, T. and Robert-Baudouy, J. (1994) *Proteins*, **20**, 34–51.

158 Schilling, S., Hoffmann, T., Manhart, S., Hoffmann, M. and Demuth, H.-U. (2004) *FEBS Letters*, **563**, 191–196.

159 Busby, W.H. Jr, Quackenbush, G.E., Humm, J., Youngblood, W.W. and Kizer, J.S. (1987) *The Journal of Biological Chemistry*, **262**, 8532–8536.

160 Fischer, W.H. and Spiess, J. (1987) *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 3628–3632.

161 Bockers, T.M., Kreutz, M.R. and Pohl, T. (1995) *Journal of Neuroendocrinology*, **7**, 445–453.

162 Kobayashi, J., Ohki, K., Okimura, K., Hashimoto, T. and Sakura, N. (2006) *Chemical & Pharmaceutical Bulletin*, **54**, 827–831.

163 Rosenblum, J.S. and Kozarich, J.W. (2003) *Current Opinion in Chemical Biology*, **7**, 496–504.

164 Frerker, N., Wagner, L., Wolf, R., Heiser, U., Hoffmann, T., Rahfeld, J.-U., Schade, J., Karl, T., Naim, H.Y., Alfalah, M., Demuth, H.-U. and von Horsten, S. (2007) *Peptides*, **28**, 257–268.

165 Busek, P., Malik, R. and Sedo, A. (2004) *The International Journal of Biochemistry & Cell Biology*, **36**, 408–421.

166 Grandt, D., Teyssen, S., Schimiczek, M., Reeve, J.R. Jr, Feth, F., Rascher, W., Hirche, H., Singer, M.V., Layer, P., Goebell, H., Ho, F.J. and Eysselein, V.E. (1992) *Biochemical and Biophysical Research Communications*, **186**, 1299–1306.

167 Mentlein, R., Dahms, P., Grandt, D. and Kruger, R. (1993) *Regulatory Peptides*, **49**, 133–144.

168 Mentlein, R., Gallwitz, B. and Schmidt, W.E. (1993) *European Journal of Biochemistry*, **214**, 829–835.

169 Deacon, C.F. (2004) *Hormone and Metabolic Research*, **36**, 761–765.

170 Cushman, D.W. and Cheung, H.S. (1971) *Biochemical Pharmacology*, **20**, 1637–1648.

171 Jean, F., Boudreault, A., Basak, A., Seidah, N.G. and Lazure, C. (1995) *The Journal of Biological Chemistry*, **270**, 19225–19231.

172 Bruzzaniti, A. and Mains, R.E. (2002) *Peptides*, **23**, 863–875.

173 Wood, W.J.L., Patterson, A.W., Tsuruoka, H., Jain, R.K. and Ellman, J.A. (2005) *Journal of the American Chemical Society*, **127**, 15521–15527.

174 Salisbury, C.M. and Ellman, J.A. (2006) *ChemBioChem*, **7**, 1034–1037.

175 Patterson, A.W., Wood, W.J.L. and Ellman, J.A. (2007) *Nature Protocols*, **2**, 424–433.

176 Harris, J.L., Backes, B.J., Leonetti, F., Mahrus, S., Ellman, J.A. and Craik, C.S. (2000) *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 7754–7759.

177 Grant, S.K., Sklar, J.G. and Cummings, R.T. (2002) *Journal of Biomolecular Screening*, **7**, 531–540.

178 Grahn, S., Kurth, T., Ullmann, D. and Jakubke, H.-D. (1999) *Biochimica et Biophysica Acta*, **1431**, 329–337.

179 Meldal, M. and Breddam, K. (1991) *Analytical Biochemistry*, **195**, 141–147.

180 Wang, G.T., Chung, C.C., Holzman, T.F. and Krafft, G.A. (1993) *Analytical Biochemistry*, **210**, 351–359.

181 Elston, C., Wallach, J. and Saulnier, J. (2007) *Analytical Biochemistry*, **368**, 87–94.

182 Rosse, G., Kueng, E., Page, M.G.P., Schauer-Vukasinovic, V., Giller, T., Lahm, H.-W., Hunziker, P. and Schlatter, D. (2000) *Journal of Combinatorial Chemistry*, **2**, 461–466.

183 Hennig, A., Florea, M., Roth, D., Enderle, T. and Nau, W.M. (2007) *Analytical Biochemistry*, **360**, 255–265.

184 Meldal, M. (2002) *Biopolymers*, **66**, 93–100.

185 Carey, R.M. and Siragy, H.M. (2003) *Endocrine Reviews*, **24**, 261–271.

186 Wright, J.W. and Harding, J.W. (1994) *Neuroscience and Biobehavioral Reviews*, **18**, 21–53.

187 von Bohlen und Halbach, O. and Albrecht, D. (2006) *Cell and Tissue Research*, **326**, 599–616.

188 Cushman, D.W., Cheung, H.S., Sabo, E.F. and Ondetti, M.A. (1982) *The American Journal of Cardiology*, **49**, 1390–1394.

189 Nemec, K. and Schubert-Zsilavecz, M. (2003) *Pharmazie in Unserer Zeit*, **32**, 11–16.

190 Lippoldt, A., Paul, M., Fuxe, K. and Ganten, D. (1995) *Clinical and Experimental Hypertension*, **17**, 251–266.

191 Zini, S., Fournie-Zaluski, M.-C., Chauvel, E., Roques, B.P., Corvol, P. and Llorens-Cortes, C. (1996) *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 11968–11973.

192 Ferreira, S.H., Bartelt, D.C. and Greene, L.J. (1970) *Biochemistry*, **9**, 2583–2593.

193 Byers, L.D. and Wolfenden, R. (1972) *The Journal of Biological Chemistry*, **247**, 606–608.

194 Cushman, D.W., Cheung, H.S., Sabo, E.F. and Ondetti, M.A. (1977) *Biochemistry*, **16**, 5484–5491.

195 Ferguson, R.K., Vlasses, P.H., Irvin, J.D., Swanson, B.N. and Lee, R.B. (1982) *Journal of Clinical Pharmacology*, **22**, 281–289.

196 Karlberg, B.E., Fyhrquist, F., Gronhagen-Riska, C., Tikkanen, I. and Ohman, K.P. (1984) *Scandinavian Journal of Urology and Nephrology (Supplement)*, **79**, 103–106.

197 Hubert, C., Houot, A.M., Corvol, P. and Soubrier, F. (1991) *The Journal of Biological Chemistry*, **266**, 15377–15383.

198 Ehlers, M.R.W., Fox, E.A., Strydom, D.J. and Riordan, J.F. (1989) *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 7741–7745.

199 Soubrier, F., Alhenc-Gelas, F., Hubert, C., Allegrini, J., John, M., Tregear, G. and Corvol, P. (1988) *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 9386–9390.

200 Jaspard, E., Wei, L. and Alhenc-Gelas, F. (1993) *The Journal of Biological Chemistry*, **268**, 9496–9503.

201 Corradi, H.R., Schwager, S.L.U., Nchinda, A.T., Sturrock, E.D. and Acharya, K.R. (2006) *Journal of Molecular Biology*, **357**, 964–974.

202 Natesh, R., Schwager, S.L.U., Sturrock, E.D. and Acharya, K.R. (2003) *Nature*, **421**, 551–554.

203 Wei, L., Clauser, E., Alhenc-Gelas, F. and Corvol, P. (1992) *The Journal of Biological Chemistry*, **267**, 13398–13405.

204 Georgiadis, D., Beau, F., Czarny, B., Cotton, J., Yiotakis, A. and Dive, V. (2003) *Circulation Research*, **93**, 148–154.

205 Dive, V., Cotton, J., Yiotakis, A., Michaud, A., Vassiliou, S., Jiracek, J., Vazeux, G., Chauvet, M.-T., Cuniasse, P. and Corvol, P. (1999) *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 4330–4335.

206 Corradi, H.R., Chitapi, I., Sewell, B.T., Georgiadis, D., Dive, V., Sturrock, E.D. and Acharya, K.R. (2007) *Biochemistry*, **46**, 5473–5478.

207 Georgiadis, D., Cuniasse, P., Cotton, J., Yiotakis, A. and Dive, V. (2004) *Biochemistry*, **43**, 8048–8054.

208 Skidgel, R.A. and Erdos, E.G. (1985) *Proceedings of the National Academy of Sciences of the United States of America*, **82**, 1025–1029.

209 Papakyriakou, A., Spyroulias, G.A., Sturrock, E.D., Manessi-Zoupa, E. and Cordopatis, P. (2007) *Biochemistry*, **46**, 8753–8765.

210 Battistini, B., Daull, P. and Jeng, A.Y. (2005) *Cardiovascular Drug Reviews*, **23**, 317–330.

211 Anderson, E.D., VanSlyke, J.K., Thulin, C.D., Jean, F. and Thomas, G. (1997) *The EMBO Journal*, **16**, 1508–1518.

212 Zhong, M., Munzer, J.S., Basak, A., Benjannet, S., Mowla, S.J., Decroly, E., Chretien, M. and Seidah, N.G. (1999) *The Journal of Biological Chemistry*, **274**, 33913–33920.

213 Fugere, M., Limperis, P.C., Beaulieu-Audy, V., Gagnon, F., Lavigne, P., Klarskov, K., Leduc, R. and Day, R. (2002) *The Journal of Biological Chemistry*, **277**, 7648–7656.

214 Basak, A. and Lazure, C. (2003) *The Biochemical Journal*, **373**, 231–239.

215 Basak, A., Koch, P., Dupelle, M., Fricker, L.D., Devi, L.A., Chretien, M. and Seidah, N.G. (2001) *The Journal of Biological Chemistry*, **276**, 32720–32728.

216 Martens, G.J., Braks, J.A., Eib, D.W., Zhou, Y. and Lindberg, I. (1994) *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 5784–5787.

217 Villemure, M., Fournier, A., Gauthier, D., Rabah, N., Wilkes, B.C. and Lazure, C. (2003) *Biochemistry*, **42**, 9659–9668.

218 Richer, M.J., Keays, C.A., Waterhouse, J., Minhas, J., Hashimoto, C. and Jean, F. (2004) *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 10560–10565.

219 Cornwall, G.A., Cameron, A., Lindberg, I., Hardy, D.M., Cormier, N. and Hsia, N. (2003) *Endocrinology*, **144**, 901–908.

220 Fricker, L.D., McKinzie, A.A., Sun, J., Curran, E., Qian, Y., Yan, L., Patterson, S.D., Courchesne, P.L., Richards, B., Levin, N., Mzhavia, N., Devi, L.A. and Douglass, J. (2000) *The Journal of Neuroscience*, **20**, 639–648.

221 Qian, Y., Devi, L.A., Mzhavia, N., Munzer, S., Seidah, N.G. and Fricker, L.D. (2000) *The Journal of Biological Chemistry*, **275**, 23596–23601.

222 Apletalina, E., Appel, J., Lamango, N.S., Houghten, R.A. and Lindberg, I. (1998) *The Journal of Biological Chemistry*, **273**, 26589–26595.

223 Cameron, A., Appel, J., Houghten, R.A. and Lindberg, I. (2000) *The Journal of Biological Chemistry*, **275**, 36741–36749.

224 Fugere, M., Appel, J., Houghten, R.A., Lindberg, I. and Day, R. (2007) *Molecular Pharmacology*, **71**, 323–332.

225 Fugere, M. and Day, R. (2005) *Trends in Pharmacological Sciences*, **26**, 294–301.

226 Denault, J.-B., D'Orleans-Juste, P., Masaki, T. and Leduc, R. (1995) *Journal of Cardiovascular Pharmacology*, **26** (Suppl. 3), S47–S50.

227 Bassi, D.E., Mahloogi, H., Lopez De Cicco, R. and Klein-Szanto, A. (2003) *The American Journal of Pathology*, **162**, 439–447.

228 Wick, W., Wild-Bode, C., Frank, B. and Weller, M. (2004) *Journal of Neurochemistry*, **91**, 1275–1283.

229 Stawowy, P., Meyborg, H., Stibenz, D., Stawowy, N.B.P., Roser, M., Thanabalasingam, U., Veinot, J.P., Chretien, M., Seidah, N.G., Fleck, E. and Graf, K. (2005) *Circulation*, **111**, 2820–2827.

230 Angliker, H., Wikstrom, P., Shaw, E., Brenner, C. and Fuller, R.S. (1993) *The Biochemical Journal*, **293**, 75–81.

231 Fujiwara, K., Kitagawa, T. and Tsuru, D. (1981) *Biochimica et Biophysica Acta*, **658**, 10–16.

232 Boudreault, A., Gauthier, D., Rondeau, N., Savaria, D., Seidah, N.G., Chretien, M. and Lazure, C. (1998) *Protein Expression and Purification*, **14**, 353–366.

233 Henrich, S., Cameron, A., Bourenkov, G.P., Kiefersauer, R., Huber, R., Lindberg, I., Bode, W. and Than, M.E. (2003) *Nature Structural Biology*, **10**, 520–526.

234 Devi, L. (1991) *FEBS Letters*, **280**, 189–194.

235 Rholam, M., Brakch, N., Germain, D., Thomas, D.Y., Fahy, C., Boussetta, H., Boileau, G. and Cohen, P. (1995) *European Journal of Biochemistry*, **227**, 707–714.

236 Cameron, A., Apletalina, E.V. and Lindberg, I. (2002) In *The Enzymes, Vol. XXII: Co- and Posttranslational Proteolysis of Proteins*, 3rd edn (eds R.E. Dalbey and D.S. Sigman), Academic Press, New York.

237 Schneider, T.D. and Stephens, R.M. (1990) *Nucleic Acids Research*, **18**, 6097–6100.

238 Crooks, G.E., Hon, G., Chandonia, J.-M. and Brenner, S.E. (2004) *Genome Research*, **14**, 1188–1190.

239 Rholam, M., Nicolas, P. and Cohen, P. (1986) *FEBS Letters*, **207**, 1–6.

240 Bek, E. and Berry, R. (1990) *Biochemistry*, **29**, 178–183.

241 Brakch, N., Rholam, M., Boussetta, H. and Cohen, P. (1993) *Biochemistry*, **32**, 4925–4930.

242 Brakch, N., Rholam, M., Simonetti, M. and Cohen, P. (2000) *European Journal of Biochemistry*, **267**, 1626–1632.

243 Simonetti, M. and Di Bello, C. (2001) *Biopolymers*, **62**, 109–121.

244 Brakch, N., Boileau, G., Simonetti, M., Nault, C., Joseph-Bravo, P., Rholam, M. and Cohen, P. (1993) *European Journal of Biochemistry*, **216**, 39–47.

245 Glandieres, J.-M., Hertzog, M., Lazar, N., Brakch, N., Cohen, P., Alpert, B. and Rholam, M. (2002) *FEBS Letters*, **516**, 75–79.

246 Fairlie, D.P., Tyndall, J.D.A., Reid, R.C., Wong, A.K., Abbenante, G., Scanlon, M.J., March, D.R., Bergman, D.A., Chai, C.L.L. and Burkett, B.A. (2000) *Journal of Medicinal Chemistry*, **43**, 1271–1281.

247 Dettin, M., Falcigno, L., Campanile, T., Scarinci, C., D'Auria, G., Cusin, M., Paolillo, L. and Di Bello, C. (2001) *Journal of Peptide Science*, **7**, 358–373.

248 Hruby, V.J. (2002) *Nature Reviews. Drug Discovery*, **1**, 847–858.

14
Insect Peptide Hormones

R. Elwyn Isaac and Neil Audsley

14.1
Introduction

Peptides are ancient signaling molecules that appeared early in the evolution of the nervous and endocrine systems of metazoans, serving as both circulating hormones and as neurotransmitters/modulators within the central nervous system. Insects, with their complex life cycles and adaptation to a wide range of habitats, rely on peptide hormones to control growth and development, and to regulate physiological processes essential for survival in diverse and sometimes hostile environments. Peptide hormones control the timing of development in response to nutritional status, size, and environmental stimuli, such as temperature and day length. They are also key components of homeostatic mechanisms ensuring the efficient storage and supply of energy substrates, and maintaining water/ion balance through the control of the flow of urine by the renal tubules and water absorption by the gut. Other physiological processes regulated by peptide hormones include reproduction, behavior, and muscle activity. The study of insect hormones started in the early part of the twentieth century and has continued with increasing intensity to the present time (for recent reviews, see [1–5]). Part of the incentive to make progress in this field is the possibility of exploiting the peptidergic endocrine system to develop new and specific chemicals to disrupt insect development, physiology, and behavior, and thus provide novel strategies for controlling insect pest populations. The recent application of whole-genome sequencing to analyzing insect genomes and the ability of software programs to identify genes encoding peptide hormones has accelerated research output in this field. Parallel advances have been made in the qualitative analysis of insect peptides by soft-ionization mass spectrometry, which in combination with high-performance liquid chromatography and computer programs that predict the primary structure of preprohormones from DNA sequences, can provide a comprehensive profile of peptides present in neuroendocrine cells and neurohemal organs. It is now commonplace to conduct reverse endocrinology by chemically synthesizing peptides predicted from genome sequence data for

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

biological testing and by heterologous expression of peptide receptors belonging to the G-protein-coupled receptor (GPCR) class in cell lines for pharmacological studies [6]. Analysis of the genomic DNA sequence of insect species (*Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera* and *Tribolium castaneum*) predicts the number of genes coding for preprohormones to be between 30 and 40, depending on the species [7–10]. The number of GPCR genes of the peptide/protein receptor families is predicted to be between 35 and 50 in the aforementioned insects, whose genomes have been sequenced and analyzed [5, 7, 11]. Insect peptides can be grouped into families based on common structural features and biological activity [1]. Some peptide families are distantly related to vertebrate peptide hormones [e.g., substance P, cholecystokinin (CCK), insulin, corticotropin-releasing factor (CRF), calcitonin, and neuropeptide Y], whereas other families are only found in invertebrates (e.g., allatostatins) or are restricted to arthropods and insects [e.g., "CAPA", proctolin and adipokinetic hormone (AKH)].

Considering the large number of insect peptides, it is not appropriate in this chapter to provide a comprehensive review of all insect peptide hormones. Instead, we aim to cover recent developments in the field and to focus attention on informative examples of the following insect peptide modulators and hormones:

(i) **Proctolin** – a powerful myotropin neuromodulator and hormone, and the first insect bioactive peptide to be structurally characterized;
(ii) **Allatostatin (A-type)** – a large family of insect peptides that inhibit juvenile hormone (JH) biosynthesis and are myoinhibitory.
(iii) **CRF-related diuretic hormones (DHs)** – these are homologs of vertebrate CRF peptides (e.g., sauvigine, urocortin, and urotensin I).
(iv) **Sex peptide** – a peptide hormone that is delivered to the female in male seminal fluid, and that has profound effects on the behavior and physiology of the postmated female.

We aim to review interesting aspects of the chemistry and biochemistry of these peptides. Several excellent reviews on insect peptide receptors have been published elsewhere and therefore this aspect of insect peptide action is not covered here [3, 7, 11–13].

14.2 Structure and Biosynthesis of Insect Peptide Hormones

Peptides are structurally the most diverse of all signaling molecules utilized by the nervous and endocrine systems of insects. They can range in size from five to 300 amino acids and are synthesized as larger inactive precursors that need to be processed to generate biologically active peptides. The precursor polypeptides are made in neurosecretory cells of the nervous system, in cells of endocrine glands, and in endocrine cells of the midgut, and are converted to active peptides by the concerted action of specialized peptidases and amino acid modifying enzymes in the secretory pathway. There appears to be little difference between invertebrates and

vertebrates in the processing events that convert precursor polypeptides to the mature hormone within the secretory pathway of endocrine cells. Insect prohormones undergo limited proteolysis by subtilisin-like serine proteases, known as prohormone convertases, which recognize motifs comprising several basic residues [14–16]. The peptide intermediates formed are further processed by removal of C-terminal arginine and lysine residues by carboxypeptidase E-like enzymes to generate either the mature C-terminus or a peptide with a C-terminal glycine, which becomes a substrate for the neuropeptide α-amidating enzymes (peptidylglycine α-hydroxylating monooxygenase and peptidyl-α-hydroxylglycine α-amidating lyase) [17–19]. An uncommon modification is the sulfation of a tyrosine residue in insect sulfakinins [20–23], peptides that show some sequence similarity to the bioactive C-terminal regions of vertebrate gastrin and CCK-8, which also contain a sulfated tyrosine. This conjugation is important for bioactivity since nonsulfated sulfakinins are very weak agonists of the sulfakinin receptor [24].

Peptide hormones originating in neuroendocrine cells are often stored in neurohemal organs from where they can be efficiently released into the circulating hemolymph, which bathes peripheral target tissues [25]. Peptidases either in the insect hemolymph or on the surface of peripheral tissues will degrade circulating peptide hormones unless the peptide has some built in structural features to protect it from nonspecific degradation by peptidases, especially by exopeptidases that remove either N- or C- terminal amino acids. Thus, most insect peptides that operate in the hemolymph possess one or more structural features that reduce the rate of catabolic clearance from the circulation. The most common protective measure is the α-amidation of the C-terminal residue – a modification that is not only critical for receptor binding and activation, but which also renders peptides resistant to hydrolysis by carboxypeptidases [19]. Some insect peptides are protected from aminopeptidase attack either by a pyroglutamyl residue resulting from the cyclization of an N-terminal glutamine or by the presence of a proline close to the N-terminus. For example, all members of the AKH family of insect peptides have both an amidated C-terminal amino acid and an N-terminal pyroglutamate residue, both of which are critical for biological activity, and will provide metabolic protection and stability in the circulation appropriate for the homeostatic role of AKHs in regulating energy reserves [1, 26]. The frequency with which proline is found towards the N-terminus of insect peptides is probably due to the fact that this amino acid offers good protection from indiscriminate aminopeptidases attempting to cleave the first peptide bond [27]. Proline is unique amongst the naturally occurring amino acids in that the side-chain is attached to both the α-carbon and the nitrogen atom that forms the peptide bond. The resulting imino ring structure prevents free rotation around the αC−N bond and the conformational restraint placed on peptide bonds in the vicinity of the proline residue can confer resistance to nonspecialized peptidases [28]. For example, the majority of characterized insect tachykinins have proline at position 2, irrespective of the length of the peptide and insect order, and this proline is not critical for receptor activation [29].

Larger peptides often have additional strategies for protecting bioactivity, such as secondary structure and the presence of one or more disulfide bonds leading to

cyclic peptides or increased structural complexity from dimerization of two peptide chains. The size and shape of the peptide will often restrict entry to the enzyme active site, making the peptide less susceptible to proteolysis. Many insect peptides have a core C-terminal sequence required for eliciting a biological response and which can comprise fewer than six residues plus a C-terminal α-amide group (e.g., peptides of the tachykinin, FMRFamide, and allatostatin families), which means that the N-terminal region can vary in composition and in length without compromising fundamental signaling properties [1, 29–31]. It is not uncommon for a single prohormone to contain the sequences of several peptides that on processing will give rise to a set of bioactive peptides with identical or very similar C-terminal structure, but with a variable N-terminal region [32–34]. The variation in the primary sequence of these peptides is likely to result in differences in susceptibility to peptidases, while conserving signaling properties. For example, the prohormone of cockroach (*Diploptera punctata*) A-type allatostatins gives rise to a large family of peptides that have a conserved Y/FXFGLamide C-terminal sequence (see Section 14.5), but which can vary greatly in length and in amino acid residues beyond this core region [35, 36]. The release of cocktails of functionally homologous peptides with different N-terminal sequences and possibly different rates of metabolism might therefore provide some security in the uncertain environment of the hemolymph, where proteolytic activities will vary both quantitatively and qualitatively depending upon the physiological state and developmental stage of the insect.

14.3 Proctolin

Proctolin (Arg-Tyr-Leu-Pro-Thr) was the first insect signaling peptide to be isolated and chemically characterized. It was isolated as a potent myotropin of insect visceral muscle (gut and oviduct), but the pentapeptide also stimulates contractions of skeletal and heart muscles from a range of insect species [37, 38]. Proctolin has the properties of a neuromodulator or a cotransmitter that works in conjunction with a classical fast-acting neurotransmitter (e.g., glutamate in the case of muscle) [37, 39, 40]. Many of the proctolin-sensitive tissues are innervated by proctolin-containing motoneurons, but other responsive tissues might be stimulated by proctolin acting as a hormone. Proctolin is present in neurosecretory cells with endings either in recognized neurohemal regions or outside the blood-brain barrier [41–43]. These observations and the detection of the peptide in the hemolymph of two insect species (*Leucophea maderae* and *Locusta migratoria*) is consistent with a role as a circulating hormone [38], and might provide the primary route to responsive tissues that are not innervated (e.g., the Malpighian tubules of *L. migratoria*, and the heart of *Periplaneta americana* and *T. castaneum*), as well as an alternative mechanism for stimulating visceral muscle.

Structurally, proctolin stands out as an atypical insect peptide hormone. It is composed of only five amino acids, all of which are important for bioactivity and

appear to be invariant [44]. Apart from a proline residue at position four, which protects the peptide from hydrolysis by hemolymph peptidyl dipeptidase A [angiotensin-converting enzyme (ACE)] [45, 46], it is poorly protected from proteolysis, which probably explains its rapid catabolism in insect hemolymph. Another unusual structural feature is the presence of arginine at the N-terminus. As already mentioned, the mechanism for the prohormone processing involves cleavage of peptide bonds after a basic residue in the precursor, which explains why peptide hormones with N-terminal basic residues will rarely occur. Therefore, the presence of arginine, which is important for bioactivity, at the N-terminus of proctolin presents a biosynthetic challenge. The recent characterization of a preproproctolin gene (*proct*) from *Drosophila* has provided an explanation of how proctolin is processed from its precursor [47]. The predicted proctolin preproprotein is 140 amino acids in length and comprises a secretion signal peptide (amino acids 1–38) followed by a single copy of the proctolin sequence, which is C-terminally flanked by a 94–amino acid peptide (Figure 14.1). Therefore, it appears that successful processing of proctolin relies on a signal peptidase to generate the peptide with an N-terminal arginine. A second cleavage occurring at Arg44–Ser45 peptide bond by a prohormone convertase would generate an intermediate extended at the C-terminus with Arg44, which would be removed by a carboxypeptidase D processing enzyme. Precursors of insect neuropeptides and peptide hormones can contain multiple copies of an identical peptide or, in some instances, several peptides with distinct structures and biological activities [2, 3]. This can provide a mechanism either to generate a cocktail of peptides with different structures and bioactivities or to increase the molar efficiency of production of a single hormone by having multiple copies in one larger precursor. However, the strategy of using the signal peptidase to generate the N-terminus of proctolin means that only one copy of the mature peptide can be processed from the preproprotein. Since the use of basic residues as recognition motifs for prohormone

Figure 14.1 Alignment of the N-terminal region of the preproproctolin amino acid sequences from three *Drosophila* species, the beetle, *T. castaneum*, and the ticks, *R. microplus* and *I. scapularis* (compiled from [48, 142]). The mature proctolin pentapeptide is highlighted in red and the predicted cleavages by a signal peptidase and a prohormone convertase are indicated by complete and dashed arrows, respectively.

convertases is highly conserved, it is very likely that all proctolin preprohormones will have a similar basic protein structure of a signal peptide followed immediately by Arg-Tyr-Leu-Pro-Thr and then a prohormone convertase cleavage site. Putative proctolin genes and copy DNA (cDNA) clones have been sequenced from other drosophilid species, the beetle, *Tenebrio molitor*, and two ticks, *Rhipicephalus microplus* and *Ixodes scapularis* [10, 48], and as predicted the position of the proctolin pentapeptide sequence in the preprohormone and the Arg–Ser recognition site for cleavage by a prohormone convertase are conserved in all cases. Somewhat surprisingly, proctolin precursor genes have not as yet been found in the genomes of mosquitoes (*A. gambiae* and *Anopheles aegypti*). The hindgut and oviduct of *A. aegypti* is insensitive to proctolin even at concentrations up to 1 μM [49], which is consistent with the absence of the proctolin gene in mosquitoes.

14.4 Sex Peptide

Seminal fluid of insects contains numerous secreted products of the male accessory glands that are important for reproductive success [50, 51]. The most studied and well characterized of these is the sex peptide of *D. melanogaster* (Figure 14.2) [52, 53]. This 36 amino acid peptide is a multifunctional sex hormone synthesized in the male accessory gland of adult flies as a 55-amino-acid preprohormone. After processing, the mature peptide is secreted into the seminal fluid and on mating is transferred to the female where it elicits several physiological and dramatic behavioral responses. These postmating responses include increased rate of egg laying, reduced attractiveness to and rejection of courting males, better sperm storage, elevated JH biosynthesis, increased appetite, and stimulation of the innate immune system of the mated female [53, 54]. Sex peptide has several distinctive structural features: a tryptophan-rich N-terminal region followed by a hydroxyproline-rich domain and a C-terminal cyclic structure formed by the formation of an intradisulfide bridge between Cys24 and Cys36. Another post-translational modification is known to occur at Ile14, but the nature of this modification has not been resolved. Experiments that measured postmating responses to sex peptide fragments injected into females identified the

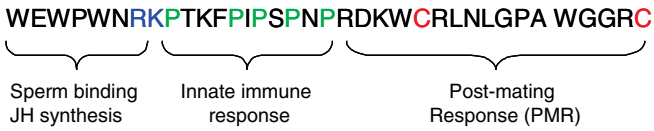


Figure 14.2 The different regions of the sex peptide of *D. melanogaster* and their associated biological activities. Color code: blue, the trypsin-like cleavage site involved in releasing sex peptide from sperm tails; green, hydroxyprolines that stimulate expression of immunity genes; red, cysteines that form a disulfide bridge [58, 143].

C-terminal portion as responsible for stimulating ovulation and altering sexual receptivity. Hydroxylation of the proline residues and modification of the Ile14 does not appear to be necessary for biological activity, but an intact disulfide bridge is critical [55].

Some of these responses can last for up to 1 week – an effect that is dependent upon transfer of sperm in addition to seminal fluid [53]. Elegant experiments using transgenic flies expressing modified forms of sex peptide have recently revealed the mechanism by which sex peptide extends its influence on mated female behavior. Sex peptide without the tryptophan-rich peptide sequence does not bind to sperm, whereas sex peptide lacking the trypsin recognition site remains permanently attached [56]. Thus, sex peptide binds to the surface of sperm tails through interaction with the N-terminal tryptophan-rich domain (Figure 14.2) and is gradually cleaved by proteolysis at a trypsin-like cleavage site (basic residues) to slowly release the active hormone from a pool of immobilized hormone.

Sex peptide also appears to have a role in activating the innate immune system of mated females [57]. This results in synthesis of a range of antimicrobial peptides that can protect against bacteria and fungi that might possibly be introduced during copulation. The central portion of the sex peptide, which includes five hydroxyproline residues, appears to be required for elicitation of this response and it has been proposed that this region can mimic the carbohydrate components of bacterial peptidoglycans that are known to induce the innate immune response in *Drosophila* [58].

How does sex peptide elicit the diverse responses in mated females? The fact that sex peptide injected into the abdomen of females or introduced by ectopic expression of the sex peptide gene in female fat body (a tissue that is functionally equivalent to the liver) elicits both the reluctance to mate and increased ovulation is consistent with sex peptide operating via the hemolymph [55, 59]. Labeled sex peptide binds to peripheral nerves, the subesophageal ganglion, the cervical connective as well as the thoracic ganglion and the genital tract [60, 61]. Thus, there is strong evidence that some of the sex peptide from male seminal fluid crosses from the reproductive tract of the mated female into the hemolymph and is transported to the central nervous system (CNS) where it triggers dramatic behavioral responses. The recent identification and molecular characterization of the sex peptide receptor from *D. melanogaster* was a major step forward in understanding the mechanism of action of sex peptide [62]. The receptor is a GPCR and is highly expressed in the female reproductive tract, especially the sperm storage organs and the lower oviduct. The receptor was also localized to surface regions of the brain, ventral regions of the subesophageal ganglion, and the cervical connective and nerve roots in the brain and ventral nerve cord. The receptor is therefore well placed on the surface of the CNS for interaction with the sex peptide being carried in the hemolymph [62]. Orthologs of the *D. melanogaster* sex peptide receptor are found in other insect species (*Drosophila pseudoobscura*, *A. aegypti*, *Bombyx mori*, *A. gambiae* and *T. castaneum*) suggesting the possibility of a general conserved endocrine mechanism amongst insects for altering behavior of mated females.

14.5
A-Type Allatostatins

The A-type allatostatins were first identified from the brains of the cockroach *D. punctata* [63, 64] based on their ability to inhibit JH synthesis in the corpora allata. They are characterized by a conserved C-terminal motif (Tyr/Phe-Xaa-Phe-Gly-Leu-NH_2, where Xaa = Asp, Asn, Gly, Ser, or Ala) that forms the active region of the peptide. As with many insect neuropeptides, allatostatins are pleiotropic, belong to families that are present in many different invertebrates, and are present in multiple forms which vary in their N-terminal sequences in an individual species. Homologs of this peptide have been identified in six different cockroach species [65–67]. In the cockroaches *D. punctata* and *P. americana*, the complimentary DNAs encoding the allatostatin precursors have been identified and share 71% sequence identity. The precursor of *D. punctata* produces 13 allatostatins, whereas 14 homologs (Peram-AST1–14) are derived from the precursor of *P. americana* (Tables 14.1 and 14.2 [65]). An additional allatostatin, $AST2^{11-18}$ (Leu-Pro-Val-Tyr-Asn-Phe-Gly-Leu-NH_2), has also recently been identified, derived from the cleavage of AST2 at its dibasic (Lys9–Arg10) processing site in both cockroach species (reviewed in [65, 68–70]. Although allatostatin-like peptides have been identified in many other insect orders (Diptera, Lepidotpera, and Orthoptera), they have only been shown to act as allatostatins in cockroaches [63, 64], crickets [71], and termites [72], and hence may not be directly involved in the regulation of JH biosynthesis in most other insect species.

Table 14.1 Structures of *D. punctata* allatostatins and their rank orders of effectiveness of the inhibition of JH biosynthesis and proctolin-induced contractions of the hindgut.

Peptide	Sequence	Rank order (JH biosynthesis)[a]	Rank order (gut motility)[b]
Dippu-AST 1	LYDFGLa	13	8
Dippu-AST2	AYSYVSEYKRLPVYNFGLa	12	10
Dippu-AST3	SKMYGFGLa	11	11
Dippu-AST4	DGRMYSFGLa	2	6
Dippu-AST5	DRLYSFGLa	3	9
Dippu-AST6	ARPYSFGLa	5	12
Dippu-AST7	APSGAQRLYGFGLa	1	4
Dippu-AST8	GGSLYSFGLa	10	2=
Dippu-AST9	GDGRLYAFGLa	6	1
Dippu-AST10	PVNSGRSSGRFNFNFGLa	7	7
Dippu-AST11	YPQEHRFSFGLa	4	5
Dippu-AST12	PFNFGLa	8	13
Dippu-AST13	IPMYDFGIa	9	2=

Lower case "a" denotes C-terminal amidation.
[a]From Tobe *et al.* [35].
[b]From Lange *et al.* [144].

Table 14.2 Structures of *P. americana* allatostatins and their rank orders of effectiveness of the inhibition of JH biosynthesis and receptor activation.

Peptide	Sequence	Rank order (JH biosynthesis)	Rank order (receptor activity)
Peram-AST1	LYDFGLa	10	2
Peram-AST2	AYSYVSEYKRLPVYNFGLa	1	7
Peram-AST2$^{11–18}$	LPVYNFGLa	9	4
Peram-AST3	SKMYGFGLa	12	9
Peram-AST4	SGNDRLYSFGLa	4	14
Peram-AST5	DRMYSFGLa	15	8
Peram-AST6	ARPYSFGLa	2	3
Peram-AST7	SPSGMQRLYGFGLa	3	12
Peram-AST8	GGSMYSFGLa	11	5
Peram-AST9	ADGRLYAFGLa	5	10
Peram-AST10	PVSSARSQTGSRFNFGLa	6	1
Peram-AST11	SPQGHRFSFGLa	8	6
Peram-AST12	SLHYAFGLa	7	11
Peram-AST13	PYNFGLa	13	12
Peram-AST14	IPMYDFGIa	14	15

Lower case "a" denotes C-terminal amidation.
From Gade *et al.* [32].

Allatostatins have also been implicated in the modulation of muscle activity in various tissues (foregut, hindgut, heart, oviduct, antennal pulsatile organ), inhibition of vitellogenesis, and modulation of midgut enzymes [36, 68]. The cellular localization of A-type allatostatins in the stomatogastric nervous system (including the frontal ganglion and recurrent nerve) and gut reflect their roles in the control of visceral muscle motility [36, 73–75], as does their presence in the antennal pulsatile organ [67] and oviduct [76]. Various A-type allatostatins have been shown to inhibit spontaneous contractions of the foregut in noctuid moth larvae [77, 78]. In the German cockroach, *Blattella germianica*, A-type allatostatins were shown to inhibit hindgut contractions, but had no effect on foregut peristalsis *in vitro* [79], whereas the converse was reported for *L. maderae* where foregut peristalsis was inhibited by A-type allatostatins in a dose-dependent manner, but they were inactive on the hindgut [80].

The conserved core pentapeptide (Tyr/Phe-Xaa-Phe-Gly-Leu-NH_2) is most likely the minimum requirement for biological activity and represents the portion of the peptide responsible for direct receptor interaction. Variation in the N-terminal amino acids is believed to be responsible for differences in the relative potencies of the various homologs as inhibitors of JH biosynthesis in adult female *D. punctata* [32, 81].

It is unclear why an insect requires multiple allatostatin homologs when they appear to have only one (cockroaches) or two (*D. melanogaster*) allatostatin receptors (see Section 14.2 for one possible explanation). Although structural information (from genomic and peptidomic data) of A-type allatostatins (and their receptors) is

available for many different species, the biological functions, their modes of action, and interactions of these peptides with each other and different signaling molecules (e.g., other allatoregulatory peptides) are still poorly understood.

14.6 CRF-Related Diuretic Hormones (DH)

Kataoka *et al.* identified the first insect DH from head extracts of the tobacco hornworm, *Manduca sexta* (Sphingidae), by its ability to promote post eclosion diuresis when injected into adults of the small white butterfly, *Pieris rapae* [82]. This 41-residue peptide, Manse-DH, is homologous to the vertebrate CRF peptides, which include suavigine, urocortin, and urotensin I. A cDNA encoding a 138-amino-acid precursor for Manse-DH was cloned from *M. sexta* consisting of the 41-amino-acid peptide bounded by dibasic amino acid cleavage sites and the C-terminal amidation signal. The sequence upstream of Manse-DH has limited homology to the ovine CRF precursor [83]. Similar peptides (at least 20) have since been structurally characterized from a variety of insect species from several orders (Orthoptera, Diptera, Coleoptera, Isoptera, Hymenotpera), including a second, smaller (30-residue) DH (Manse-DPII) from *M. sexta* (Table 14.3 [84]). Most insect CRF-related DHs have been characterized by peptide purification from tissue extracts utilizing multiple liquid chromatography steps and sequenced by Edman degradation (reviewed by Coast *et al.* [85]).

More recently, peptides have been identified through BLAST searching of genome databases [10, 86, 87]. Two DHs have also been identified from *T. molitor* (Tenmo-DH_{37} and Tenmo-DH_{41}), the only CRF-related DHs that are not amidated at their C-terminus (Table 14.3; [88, 89]). Two CRF-related DHs (Trica-DH_{37} and Trica-DH_{47}) were identified from the red flour beetle, *T. castaneum*, genome [10, 87]. The amino acid sequence of Trica-DH_{47} is identical to Tenmo-DH_{47}; however, the presence of a glycine residue preceding the C-terminal cleavage site implies Trico-DH_{47} is amidated [87], whereas Tenmo-DH_{47} is not [89]. There is a 73% sequence identity between Tenmo-DH_{37} and Trica-DH_{37} [10]. The two CRF-related DHs from *T. castaneum* are encoded by the same gene as a result of alternative splicing, in contrast to insects such as *D. melanogaster* and the honey bee, *A. mellifera*, where alternative splicing does not occur and only one CRF-related DH has been identified [10]. The sequences of DHs of closely related species are often identical, such as those from the house fly, *Musca domestica* (Musdo-DH), the stable fly, *Stomoxys calcitrans* (Stoca-DH) [90], *D. melanogaster* (Drome-DH) [86], and the two DHs characterized from the sphingid moth, *Hyles lineata* [91], differ from the *M. sexta* DHs by only one residue (reviewed in [85]). These insect CRF-related DHs promote diuresis through the stimulation of fluid secretion by the Malpighian tubules via cyclic AMP as their second messenger [92–96]. Insect CRF-related peptides (except Manse-DPII) have greater than 40% sequence identity with each other and 20–30% sequence identity with the vertebrate CRF peptides grouping them in the same superfamily [85]. Table 14.3 shows the sequence alignment of selected insect

Table 14.3 Sequences (single-letter amino acid code) of selected insect CRF-related DHs and sauvigine.

Species	Peptide	Peptide sequence					
Locusta migratoria	Locmi-DH	----MGMGPS	L**S**IVNPMDV**L**	RQRLLLEIAR	RRLRDAE-EQ	IKA**N**KDF**L**QQ	Ia
Manduca sexta	Manse-DH	------RMPS	L**S**IDLPMSV**L**	RQKLSLEKER	KVHALRA---	-AA**N**RNF**L**ND	Ia
Manduca sexta	Manse-DPII	---------S	F**S**VNPAVDI**L**	QHRYMEKVAQ	N---------	---**N**RNF**L**NR	Va
Tenebrio molitor	Tenmo-DH$_{47}$	AGALGESGAS	L**S**IVNSLDV**L**	RNRLLLEIAR	KKAKEGA---	NR-**N**RQI**L**LS	L-OH
Tenebrio molitor	Tenmo-DH	-------SPT	I**S**ITAPIDV**L**	RKTWEQERAR	KQMVR-----	---**N**REF**L**NS	LN-OH
Phyllomedusa sauvagei	sauvigine	------QGPP	I**S**IDLSLEL**L**	RKMIEIEKQE	KEKQQAA---	--N**N**RLL**L**DT	Ia

Dashes represent gaps inserted for alignment and lower case "a" denotes C-terminal amidation. Amino acids conserved in all peptides are given in bold.
From Coast *et al.* [85].

CRF-related DHs and sauvigine as a comparison. The alignments are as described by Coast *et al*. [85]. Four residues (shaded; Table 14.3), serine and leucine in the N-terminal region and asparagine and leucine in the C-terminal region, are conserved in all CRF-related peptides. In the vertebrate CRFs, the regions around these amino acids are also highly conserved (i.e., Pro–Pro–Ile–Ser, Leu–Leu–Arg, Asn–Arg and Leu–Leu). In the insect CRFs these regions, although not identical, have conservative substitutions retaining the properties of these areas and are believed to be important for biological activity [97]. Even though Manse-DPII is much shorter than the other insect CRF-related DHs, it shares significant sequence identity (47%) to Manse-DH and Coast [98] suggests Manse-DPII is most likely the product of a gene duplication event. Sequence alignments have been used to deduce phylogenetic and putative evolutionary relationships between CRF-related peptides [10, 98].

Using *in vitro* Malpighian tubule fluid secretion assays [97] and/or cyclic AMP production [99], the insect CRF-related peptides have been shown to cross-react in a heterospecific manner to varying degrees [100]. Sauvigine, bovine CRF and human CRF had small, but significant effects on fluid secretion by Malpighian tubules of *M. sexta*, whereas bovine CRF, human CRF, or sucker fish urotensin I had no effect on *L. migratoria* Malpighian tubules [101].

All insect CRF-related peptides, except *T. molitor* DHs, are C-terminally amidated. This amidation appears to be important for biological activity because Manse-DH acid is 1000 times less potent than the amidated peptide [82, 101]. The potency of *T. molitor*-DH to stimulate fluid secretion by Malpighian tubules is similar to other insect CRF-related DHs when assayed conspecifically (reviewed in [98]), but unlike other CRF-DHs, *T. molitor* DH has no effect on *M. sexta* Malpighian tubules. In contrast, Manse-DH is only 17 times less potent than Tenmo-DH at promoting cyclic AMP production in Malpighian tubules of *T. molitor* [88].

The N-terminal region of insect CRF-related DHs is important for receptor activation [100, 102, 103]. Compared to the intact Achdo-DH, deletion of amino acids from the N-terminus, up to residue 7, reduced the potency, but had little effect on diuretic activity of the peptide. However, further truncation (Achdo-DH$^{11–46}$) significantly affected biological activity and Achdo-DH$^{23–46}$ was completely inactive, implying the N-terminus is involved in signal transduction [100]. An α-helical CRF$^{9–41}$ analog had no intrinsic biological activity on fluid secretion by the Malpighian tubules of *A. domesticus*. However, this analog partially inhibited the stimulation of fluid secretion by Achdo-DH, presumably by binding to DH receptors on the tubules, suggesting that the C-terminal region of the peptide is important for receptor binding [95], although some interaction between the N- and C-terminals is most likely required for full biological activity [100].

14.7
Insect Peptide Hormones and Insect Control

There are several approaches that can be adopted to disrupt peptidergic signaling [104]. These include the inappropriate activation of peptide receptors using

superagonists, the blocking of signaling activity with receptor antagonists, and the disruption of signaling by inhibiting peptide processing and inactivation enzymes. Peptides themselves are poor candidates for insect control chemicals because of the problems that peptides will encounter in reaching target tissues. Entry through the integument is hindered by the hydrophobic nature of the insect cuticle, and uptake via ingestion can be thwarted by the peptidases of the gut, hemolymph, and other tissues, as well as the limited permeability of the gut epithelium. Nevertheless, progress has been made in the design of compounds that can penetrate the cuticle and in the development of ways to avoid or minimize degradation by peptidases.

Nachman *et al*. have been successful in the design of amphiphylic pseudomimics of pheromone biosynthesis-activating neuropeptide (PBAN) – a C-terminally amidated 33-amino-acid peptide that stimulates biosynthesis of the sex pheromone in moths [105]. PBAN belongs to a large family of insect peptides with a highly conserved C-terminal pentapeptide sequence (Phe-Xaa-Pro-Arg-Leu-NH_2, where X = Ser, Thr, Gly, or Val). The family includes the pyrokinins, myotropins, melanization and reddish coloration hormone, and diapause hormone [106]. The Phe-Xaa-Pro-Arg-Leu-NH_2 is the minimal structure for inducing pheromone biosynthesis and has been used as the basic peptide unit for the development of amphiphylic pseudopeptides with lipids attached to the N-terminus. The attachment of 6-phenylhexanoic acid to Phe-Thr-Pro-Arg-Leu-NH_2 resulted in a topically active stimulant of pheromone production [107]. Other modifications that improved cuticle penetration included replacing phenylalanine with hydrocinnamic acid and the conjugation of short-chain fatty acids (<12 carbons), cholic acid, 9-fluoreneacetic acid, 1-pyrenebutyric acid or 2-amino-7-bromofluorene to the N-terminal phenylalanine of the core pentapeptide sequence [108–111]. The latter three pseudopeptides were active as PBAN agonists for up to 24 h after a single topical application.

Substantial progress has also been made in the direction of enhancing the bioactivity of pseudopeptides by limiting susceptibility to inactivation by insect peptidases, such as aminopeptidase, the endopeptidase neprilysin (NEP), and the peptidyl dipeptidase ACE. Such an approach has been very successful in the design of novel biostable agonists of insect kinins. This large peptide family have a conserved C-terminal sequence of Phe-Xaa^1-Xaa^2-Trp-Gly-NH_2, where Xaa^1 = His, Asn, or Tyr; Xaa^2 = Ser, Pro, or Ala [1, 112]. Insect kinins were first isolated from insect tissues as powerful stimulators of the spontaneous contractions of insect gut, but have since been shown to have diuretic activity on isolated Malpighian (renal) tubules of several insect species and to limit weight gain by larvae of two major cotton pests (tobacco budworm, *Heliothis virescens*, and corn earworm, *Helicoverpa zea*) [113–116]. The helicokinin 2 analog (Val-Arg-Phe-Ser-Ser-Trp-Gly-NH_2) is cleaved at several positions by tissue peptidases, but can be stabilized by incorporation of the sterically hindered α,α-disubstituted amino acid, aminoisobutyric acid (Aib), and pGlu at the N-terminus [115]. Thus, pGlu-Arg-Phe-Ser-Aib-Trp-Gly-NH_2 is very resistant to tissue peptidases and when injected into 5-day-old larvae of *H. zea* there was a 50% reduction in larval weight compared to control insects [117].

Similar approaches have been used to enhance the stability and uptake of an allatostatin (Ala-Arg-Pro-Tyr-Ser-Phe-Gly-Leu-NH_2) in the gut of larval lepidopterans.

Allatostatins are rapidly degraded by endopeptidases that cleave the Phe6–Gly7 peptide bond. Replacement of the G residue with Fmoc-1-amino-cyclopropylcarboxylic acid offered protection against these enzymes. In addition, the incorporation of hexanoic acid at the N-terminus increased the transepithilial (midgut) transport of allatostatin analogs compared to the native (unmodified) peptide [118, 119].

Another successful approach has been to substitute one or more α-amino acid with a β-amino acid. This introduces a methylene group between the α-carbon and either the acid (β^3-amino acid) or amino group (β^2-homo-amino acid) [120]. These substitutions can maintain favorable interactions between the side-chains of the peptide ligand and its receptor, whilst providing resistance to peptidase attack at the adjacent peptide bond. One or two β^3-amino acids and β^2-homo-amino acids were introduced into the insect kinin Ac-Arg-Phe-Phe-Pro-Trp-Gly-NH_2 or the heptapeptide Ac-Arg-Phe-Phe-Phe-Pro-Trp-Gly-NH_2 instead of Phe^2, Phe^3, Pro^3, and Trp^5 [120]. The *N*-acetylated arginine prevented metabolic degradation by aminopeptidases and the positioning of the β-amino acids was designed to minimize hydrolysis by NEP and ACE, present either on the surface of insect tissues or in the hemolymph. Four of the eight analogs synthesized showed both a potent *in vitro* diuretic activity in cricket Malpighian tubule secretion assay and resistance to degradation by NEP and ACE [121].

Another approach to the disruption of peptidergic signaling has been the development of peptide antagonists by rational design. Alstein *et al.* have targeted the mating behavior of moths by developing PBAN antagonists that inhibit sex pheromone biosynthesis (for reviews, see [122, 123]). They substituted in turn each amino acid of the pyrokinin/PBAN peptide Arg-Tyr-Phe-Ser-Pro-Arg-Leu-NH_2 with D-phenylalanine and found a potent PBAN antagonist in Arg-Tyr-Phe-D-Phe-Pro-Arg-Leu-NH_2 [124]. This discovery of a lead antagonist led to the development of backbone cyclic antagonists that have several advantages over linear peptides for insect control [125, 126]. Importantly, cyclic peptide antagonists are less susceptible to degradation and have reduced polarity, which will improve the bioavailability of the peptides. In addition, they can display high selectivity and increased biological activity.

Alternatives to topical application for the delivery of peptide agonists to insect pests have been sought. Baculovirsuses infect insects and provide natural control of insect populations. However, their slow speed of kill limits their application to the control of major agricultural pests. Attempts to improve the speed by which baculovirsuses kill insects have included the development of transgenic viruses that overexpress insect peptide hormones [127]. A baculovirus genetically modified to express high levels of the DH of *M. sexta* improved the speed of kill of insect larvae, presumably by disrupting water balance [128]. A plant virus has also been used to deliver a peptide hormone, trypsin-modulating oostatic factor (TMOF), for insect control [129]. Tobacco leaves infected with transgenic tobacco mosaic virus expressing a mosquito TMOF peptide stunt the growth of feeding larvae of *H. virescens* by inhibiting trypsin and chymotrypsin activities in the insect midgut.

A different approach has also been successful in delivering an allatostatin peptide that when administered by injection inhibits feeding and weight gain, in larvae of

the tomato moth, *Lacanobia oleracea*, resulting in increased mortality [130]. The peptide is ineffective when administered orally as it is not protected from the digestive enzymes of the insect midgut, but when it is fused to the snowdrop lectin, *Galanthus nivalis* agglutinin, the hormone is protected and carried across the gut epithelium, by transcytosis, into the hemolymph, where it exerts its effects (suppresses feeding and growth, resulting in death) similar to those of the injected peptide [130, 131].

Hormone imbalance leading to disruption of growth and development of insect pests might also be achieved by inhibiting peptidases responsible for the synthesis and inactivation of peptide hormones. One such peptidase is the insect ACE, which is a soluble metallopeptidase found in reproductive tissues and in the hemolymph of insects [132]. It is both a dipeptidyl carboxypeptidase and an endopeptidase, and is capable of inactivating a variety of small- to medium-size peptide hormones by cleavage of C-terminal dipeptides and dipeptide amides [133]. ACE inhibitors stunt growth and development in the lepidopteran species, *H. virescens* [116] and *M. sexta* [134]. Although captopril, lisinopril, and enalapril on their own had no significant effect on larval growth of *H. virescens*, when ACE inhibitors were coinjected with the diuretic peptide helicokinin I, larval mortality reached 80% [116]. Fosinopril, another ACE inhibitor, abolished growth of fourth instar *M. sexta* larvae and resulted in high larval mortality [134]. Adult female mosquitoes (*Anopheles stephensi*) fed either captopril or another ACE inhibitor, lisinopril, in their blood meal, laid fewer eggs – a response that was dose-dependent [135, 136]. ACE inhibitors fed to adult male *A. stephensi* and *A. gambiae* in the glucose diet resulted in a dramatic loss of fecundity (60–80% reduction in the numbers of eggs) of mated females [134]. ACE inhibitors also have profound effects on reproduction in lepidopteran insects. Feeding captopril to adult female Egyptian cotton leafworm, *Spodoptera littoralis*, resulted in a 30% reduction in the numbers of eggs laid [137, 138]. ACE inhibitors injected into the last larval instar of the tomato moth, *Lacanobia oleracea*, reduce both oviposition and fertility [134, 139]. Fosinopril had an affect on spermiogenesis in *L. oleracea*. Male sixth instar larvae injected with fosinopril produced around half the normal number of eupyrene sperm bundles, which probably contributed to the low fecundity observed in females mated with inhibitor-treated male *L. oleracea*. Biochemical studies show that there are important differences between insect and mammalian ACEs that might be exploited to design selective inhibitors of value in controlling insect pest populations [132, 140, 141].

14.8 Conclusions

The recent advances made in the identification and characterization of insect peptide hormones, and the application of genomics, bioinformatics, and reverse endocrinology, is providing unparalleled information and resources for those scientists who are interested in exploring fundamental endocrine mechanisms

of how peptides control development, behavior, appetite, locomotor activity, and reproduction. These advances are also providing new opportunities for the development of insect control strategies that target the peptidergic endocrine system. Importantly, the genetics of *D. melanogaster* together with the emerging application of double-stranded RNA to selectively knock-down gene expression in a range of insect species can provide the necessary proof-of-principle that interfering with components of the peptide signaling system can have detrimental consequences for insect pests.

References

1 Gade, G. (1997) *Fortschritte der Chemie Organischer Naturstoffe*, **71**, 1–128.
2 Nässel, D.R. (2002) *Progress in Neurobiology*, **68**, 1–84.
3 Taghert, P.H. and Veenstra, J.A. (2003) *Advances in Genetics*, **49**, 1–65.
4 Ewer, J. (2005) *Hormones and Behavior*, **48**, 418–429.
5 Claeys, I., Poels, J., Simonet, G., Franssens, V., Van Loy, T., Van Hiel, M.B., Breugelmans, B. and Vanden Broeck, J. (2005) *Vitamins and Hormones*, **73**, 217–282.
6 Reumer, A., Van Loy, T., Clynen, E. and Schoofs, L. (2008) *General and Comparative Endocrinology*, **155**, 22–30.
7 Hewes, R.S. and Taghert, P.H. (2001) *Genome Research*, **11**, 1126–1142.
8 Riehle, M.A., Garczynski, S.F., Crim, J.W., Hill, C.A. and Brown, M.R. (2002) *Science*, **298**, 172–175.
9 Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman, G., Huybrechts, J., Eewing, M.A., Vierstraeet, E., Rodrigues-Zas, S.L., Schoofs, L., Robinson, G.E. and Sweedler, J.V. (2006) *Science*, **314**, 647–649.
10 Li, B., Predel, R., Neupert, S., Hauser, F., Tanaka, Y., Cazzamali, G., Williamson, M., Arakane, Y., Verleyen, P., Schoofs, L., Schachtner, J., Grimmelikhuijzen, C.J.P. and Park, Y. (2008) *Genome Research*, **18**, 113–122.
11 Hauser, F., Cazzamali, G., Williamson, M., Park, Y., Li, B., Tanaka, Y., Predel, R., Neupert, S., Schachtner, J., Verleyen, P. and Grimmelikhuijzen, C.J.P. (2008) *Frontiers in Neuroendocrinology*, **29**, 142–165.
12 Broeck, J.V. (2001) *Archives of Insect Biochemistry and Physiology*, **48**, 1–12.
13 Hauser, F., Cazzamali, G., Williamson, M., Blenau, W. and Grimmelikhuijzen, C.J.P. (2006) *Progress in Neurobiology*, **80**, 1–19.
14 Roebroek, A.J.M., Pauli, I.G.L., Zhang, Y. and Van de Ven, W.J.M. (1991) *FEBS Letters*, **289**, 133–137.
15 Roebroek, A.J.M., Creemers, J.W.M., Pauli, I.G.L., Kurzik-Dumke, U., Rentrop, M., Gateff, E.A.F., Leunissen, J.A.M. and Van de Ven, W.J.M. (1992) *The Journal of Biological Chemistry*, **267**, 17208–17215.
16 Veenstra, J.A. (2000) *Archives of Insect Biochemistry and Physiology*, **43**, 49–63.
17 Settle, S.H., Green, M.M. and Burtis, K.C. (1995) *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 9470–9474.
18 Jiang, N., Kolhekar, A.S., Jacobs, P.S., Mains, R.E., Eipper, B.A. and Taghert, P.H. (2000) *Developmental Biology*, **226**, 118–136.
19 Han, M., Park, D., Vanderzalm, P.J., Mains, R.E., Eipper, B.A. and Taghert, P.H. (2004) *Journal of Neurochemistry*, **90**, 129–141.
20 Nachman, R.J., Holman, G.M., Haddon, W.F. and Ling, N. (1986) *Science*, **234**, 71–73.
21 Nachman, R.J., Holman, G.M., Cook, B.J., Haddon, W.F. and Ling, N. (1986)

Biochemical and Biophysical Research Communications, **140**, 357–364.
22 Nichols, R., Schneuwly, S.A. and Dixon, J.E. (1988) *The Journal of Biological Chemistry*, **263**, 12167–12170.
23 Duve, H., Thorpe, A., Scott, A.G., Johnsen, A.H., Rehfeld, J.F., Hines, E. and East, P.D. (1995) *European Journal of Biochemistry*, **232**, 633–640.
24 Kubiak, T.M., Larsen, M.J., Burton, K.J., Bannow, C.A., Martin, R.A., Zantello, M.R. and Lowery, D.E. (2002) *Biochemical and Biophysical Research Communications*, **291**, 313–320.
25 Predel, R. (2001) *The Journal of Comparative Neurology*, **436**, 363–375.
26 Oudejans, R.C.H.M., Vroemen, S.F., Jansen, R.F.R. and Van der Horst, D.J. (1996) *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 8654–8659.
27 Vanhoof, G., Goossens, F., De Meester, I., Hendriks, D. and Scharpe, S. (1995) *The FASEB Journal*, **9**, 736–744.
28 Yaron, A. and Naider, F. (1993) *Critical Reviews in Biochemistry and Molecular Biology*, **28**, 31–81.
29 Nässel, D.R. (1999) *Peptides*, **20**, 141–158.
30 Thorpe, A., Johnsen, A.H., Rehfeld, J.F., East, P.D. and Duve, H. (1994) *Netherlands Journal of Zoology*, **45**, 251–259.
31 Bendena, W.G., Donly, B.C. and Tobe, S.S. (1999) *Annals of the New York Academy of Sciences*, **897**, 311–329.
32 Gade, G., Marco, H.G., Richter, D. and Weaver, R.J. (2008) *Journal of Insect Physiology*, **54**, 988–996.
33 Muren, J.E. and Nässel, D.R. (1996) *Regulatory Peptides*, **65**, 185–196.
34 Siviter, R.J., Coast, G.M., Winther, A.M.E., Nachman, R.J., Taylor, C.A.M., Shirras, A.D., Coates, D., Isaac, R.E. and Nässel, D.R. (2000) *The Journal of Biological Chemistry*, **275**, 23273–23280.
35 Tobe, S.S., Zhang, J.R., Bowser, P.R.F., Donly, B.C. and Bendena, W.G. (2000) *Journal of Insect Physiology*, **46**, 231–242.
36 Audsley, N., Matthews, H.J., Price, N.R. and Weaver, R.J. (2008) *Journal of Insect Physiology*, **54**, 969–980.
37 Orchard, I., Belanger, J.H. and Lange, A.B. (1989) *Journal of Neurobiology*, **20**, 470–496.
38 Lange, A.B. (2002) *Peptides*, **23**, 2063–2070.
39 Cook, B.J. and Holman, G.M. (1979) *Comparative Biochemistry and Physiology*, **64**, 21–28.
40 Adams, M.E. and O'Shea, M. (1983) *Science*, **221**, 286–289.
41 Nässel, D.R. and O'Shea, M. (1987) *The Journal of Comparative Neurology*, **265**, 437–454.
42 Nässel, D.R., Holmqvist, B.I. and Moverus, B.J. (1989) *The Journal of Comparative Neurology*, **283**, 450–463.
43 Davis, N.T., Velleman, S.G., Kingan, T.G. and Keshishian, H. (1989) *The Journal of Comparative Neurology*, **283**, 71–85.
44 Starratt, A.N. and Brown, B.E. (1979) *Biochemical and Biophysical Research Communications*, **90**, 1125–1130.
45 Lamango, N.S., Nachman, R.J., Hayes, T.K., Strey, A. and Isaac, R.E. (1997) *Peptides*, **18**, 47–52.
46 Siviter, R.J., Nachman, R.J., Dani, M.P., Keen, J.N., Shirras, A.D. and Isaac, R.E. (2002) *Peptides*, **23**, 2025–2034.
47 Taylor, C.A.M., Winther, A.M.E., Siviter, R.J., Shirras, A.D., Isaac, R.E. and Nässel, D.R. (2004) *Journal of Neurobiology*, **58**, 379–391.
48 Christie, A.E. (2008) *General and Comparative Endocrinology*, **157**, 174–185.
49 Messer, A.C. and Brown, M.R. (1995) *The Journal of Experimental Biology*, **198**, 2325–2336.
50 Chen, P.S. (1984) *Annual Review of Entomology*, **29**, 233–255.
51 Chapman, T. and Davies, S.J. (2004) *Peptides*, **25**, 1477–1490.
52 Chen, P.S., Stumm-Zollinger, E., Aigaki, T., Balmer, J., Bienz, M. and Bohlen, P. (1988) *Cell*, **54**, 291–298.
53 Kubli, E. (2003) *Cellular and Molecular Life Sciences*, **60**, 1689–1704.

54 Carvalho, G.B., Kapahi, P., Anderson, D.J. and Benzer, S. (2006) *Current Biology*, **16**, 692–696.

55 Schmidt, T., Coffat, Y., Klauser, S. and Kubli, E. (1993) *Journal of Insect Physiology*, **39**, 361–368.

56 Peng, J., Chen, S., Busser, S., Liu, H., Honegger, T. and Kubli, E. (2005) *Current Biology*, **15**, 207–213.

57 Peng, J., Zipperlen, P. and Kubli, E. (2005) *Current Biology*, **15**, 1690–1694.

58 Domanitskaya, E.V., Liu, H., Chen, S. and Kubli, E. (2007) *FEBS Journal*, **274**, 5659–5668.

59 Aigaki, T., Fleischmann, I., Chen, P.S. and Kubli, E. (1991) *Neuron*, **7**, 557–563.

60 Ottiger, M., Soller, M., Stocker, R.F. and Kubli, E. (2000) *Journal of Neurobiology*, **44**, 57–71.

61 Ding, Z., Haussmann, I., Ottiger, M. and Kubli, E. (2003) *Journal of Neurobiology*, **55**, 372–384.

62 Yapici, N., Kim, Y.-J., Ribeiro, C. and Dickson, B.J. (2008) *Nature*, **451**, 33–37.

63 Woodhead, A.P., Stay, B., Seidel, S.L., Khan, M.A. and Tobe, S.S. (1989) *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 5997–6001.

64 Pratt, G.E., Farnsworth, D.E., Siegel, N.R., Fok, K.F. and Feyereisen, R. (1989) *Biochemical and Biophysical Research Communications*, **163**, 1243–1247.

65 Weaver, R.J., Edwards, J.P., Bendena, W.G. and Tobe, S.S. (1998) In *Recent Advances in Arthropod Endocrinology* (eds G.M Coast and S.G Webster), Society for Experimental Biology Seminar Series No. 65, Cambridge University Press, Cambridge, pp. 3–32.

66 Belles, X., Graham, L.A., Bendena, W.G., Ding, Q., Edwards, J.P., Weaver, R.J. and Tobe, S.S. (1999) *Peptides*, **20**, 11–22.

67 Predel, R., Kellner, R., Rapus, J. and Gade, G. (1999) *Regulatory Peptides*, **82**, 81–89.

68 Gade, G. (2002) *Invertebrate Reproduction & Development*, **41**, 127–135.

69 Stay, B. and Tobe, S.S. (2007) *Annual Review of Entomology*, **52**, 277–299.

70 Hult, E.F., Weadick, C.J., Chang, B.S.W. and Tobe, S.S. (2008) *Journal of Insect Physiology*, **54**, 959–968.

71 Lorenz, M.W., Kellner, R. and Hoffmann, K.H. (1995) *Regulatory Peptides*, **57**, 227–236.

72 Yagi, K.J., Kwok, R., Chan, K.K., Setter, R.R., Myles, T.G., Tobe, S.S. and Stay, B. (2005) *Journal of Insect Physiology*, **51**, 357–365.

73 Duve, H., Johnsen, A.H., Maestro, J.-L., Scott, A.G., Winstanley, D., Davey, M., East, P.D. and Thorpe, A. (1997) *Peptides*, **18**, 1301–1309.

74 Duve, H., Johnsen, A.H., Maestro, J.-L., Scott, A.G., Jaros, P.P. and Thorpe, A. (1997) *European Journal of Biochemistry*, **250**, 727–734.

75 Duve, H., Audsley, N., Weaver, R. and Thorpe, A. (2003) *Cell and Tissue Research*, **314**, 281–295.

76 Garside, C.S., Koladich, P.M., Bendena, W.G. and Tobe, S.S. (2002) *Insect Biochemistry and Molecular Biology*, **32**, 1089–1099.

77 Duve, H., East, P.D. and Thorpe, A. (1999) *The Journal of Comparative Neurology*, **413**, 405–416.

78 Matthews, H.J., Audsley, N. and Weaver, R.J. (2007) *Journal of Insect Physiology*, **53**, 75–83.

79 Aguilar, R., Maestro, J.L., Vilaplana, L., Pascual, N., Piulachs, M.-D. and Belles, X. (2003) *Regulatory Peptides*, **115**, 171–177.

80 Duve, H., Johnsen, A.H., Scott, A.G. and Thorpe, A. (1995) *Regulatory Peptides*, **57**, 237–245.

81 Stay, B., Tobe, S.S. and Bendena, W.G. (1994) *Advances in Insect Physiology*, **25**, 267–337.

82 Kataoka, H., Troetschler, R.G., Li, J.P., Kramer, S.J., Carney, R.L. and Schooley, D.A. (1989) *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 2976–2980.

83 Digan, M.E., Roberts, D.N., Enderlin, F.E., Woodworth, A.R. and Kramer, S.J. (1992) *Proceedings of the National Academy*

of Sciences of the United States of America, **89**, 11074–11078.

84 Blackburn, M.B., Kingan, T.G., Bodnar, W., Shabanowitz, J., Hunt, D.F., Kempe, T., Wagner, R.M., Raina, A.K., Schnee, M.E. and Ma, M.C. (1991) *Biochemical and Biophysical Research Communications*, **181**, 927–932.

85 Coast, G.M., Orchard, I., Phillips, J.E. and Schooley, D.A. (2002) *Advances in Insect Physiology*, **29**, 279–409.

86 Vanden Broeck, J. (2001) *Peptides*, **22**, 241–254.

87 Weaver, R.J. and Audsley, N. (2008) *Peptides*, **29**, 168–178.

88 Furuya, K., Schegg, K.M., Wang, H., King, D.S. and Schooley, D.A. (1995) *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 12323–12327.

89 Furuya, K., Schegg, K.M. and Schooley, D.A. (1998) *Peptides*, **19**, 619–626.

90 Clottens, F.L., Holman, G.M., Coast, G.M., Totty, N.F., Hayes, T.K., Kay, I., Mallet, A.I., Wright, M.S., Chung, J.-S., Truong, O. and Bull, D.L. (1994) *Peptides*, **15**, 971–979.

91 Furuya, K., Harper, M.A., Schegg, K.M. and Schooley, D.A. (2000) *Insect Biochemistry and Molecular Biology*, **30**, 127–133.

92 Kay, I., Wheeler, C.H., Coast, G.M., Totty, N.F., Cusinato, O., Patel, M. and Goldsworthy, G.J. (1991) *Biological Chemistry Hoppe-Seyler*, **372**, 929–934.

93 Kay, I., Coast, G.M., Cusinato, O., Wheeler, C.H., Totty, N.F. and Goldsworthy, G.J. (1991) *Biological Chemistry Hoppe-Seyler*, **372**, 505–512.

94 Kay, I., Patel, M., Coast, G.M., Totty, N.F., Mallet, A.I. and Goldsworthy, G.J. (1992) *Regulatory Peptides*, **42**, 111–122.

95 Coast, G.M., Hayes, T.K., Kay, I. and Chung, J.S. (1992) *The Journal of Experimental Biology*, **162**, 331–338.

96 Audsley, N., Coast, G.M. and Schooley, D.A. (1993) *The Journal of Experimental Biology*, **178**, 231–243.

97 Coast, G.M., Rayne, R.C., Hayes, T.K., Mallet, A.I., Thompson, K.S.J. and Bacon, J.P. (1993) *The Journal of Experimental Biology*, **175**, 1–14.

98 Coast, G.M. (1998) In *Recent Advances in Arthropod Endocrinology* (eds G.M., Coast and S.G Webster), Society for Experimental Biology Seminar Series No. 65, Cambridge University Press, Cambridge, pp. 189–209.

99 Coast, G.M., Cusinato, O., Kay, I. and Goldsworthy, G.J. (1991) *Journal of Insect Physiology*, **37**, 563–573.

100 Coast, G.M. and Kay, I. (1994) *The Journal of Experimental Biology*, **187**, 275–343.

101 Audsley, N., Kay, I., Hayes, T.K. and Coast, G.M. (1995) *Comparative Biochemistry and Physiology. Part A, Physiology*, **110**, 87–93.

102 Reagan, J.D., Li, J.P., Carney, R.L. and Kramer, S.J. (1993) *Insect Biochemistry and Physiology*, **23**, 135–145.

103 Reagan, J.D. (1995) *Insect Biochemistry and Molecular Biology*, **25**, 535–539.

104 Gade, G. and Goldsworthy, G.J. (2003) *Pest Management Science*, **59**, 1063–1075.

105 Teal, P.E.A., Meredith, J.A. and Nachman, R.J. (1999) *Annals of the New York Academy of Sciences*, **897**, 348–360.

106 Altstein, M. (2004) *Peptides*, **25**, 1491–1501.

107 Abernathy, R.L., Teal, P.E.A., Meredith, J.A. and Nachman, R.J. (1996) *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 12621–12625.

108 Teal, P.E.A. and Nachman, R.J. (1997) *Regulatory Peptides*, **72**, 161–167.

109 Teal, P.E.A., Meredith, J.A. and Nachman, R.J. (1999) *Peptides*, **20**, 63–70.

110 Teal, P.E.A. and Nachman, R.J. (2002) *Peptides*, **23**, 801–806.

111 Nachman, R.J., Teal, P.E.A. and Strey, A. (2002) *Peptides*, **23**, 2035–2043.

112 Torfs, P., Nieto, J., Veelaert, D., Boon, D., Van de Water, G., Waelkens, E., Derua, R., Calderon, J., DeLoof, A. and Schoofs, L. (1999) *Annals of the New York Academy of Sciences*, **897**, 361–373.

113 Coast, G.M., Holman, G.M. and Nachman, R.J. (1990) *Journal of Insect Physiology*, **36**, 481–488.

114 Coast, G.M., Zabrocki, J. and Nachman, R.J. (2002) *Peptides*, **23**, 701–708.

115 Nachman, R.J., Strey, A., Isaac, E., Pryor, N., Lopez, J.D., Deng, J.-G. and Coast, G.M. (2002) *Peptides*, **23**, 735–745.

116 Seinsche, A., Dyker, H., Losel, P., Backhaus, D. and Scherkenbeck, J. (2000) *Journal of Insect Physiology*, **46**, 1423–1431.

117 Nachman, R.J., Tilley, J.W., Hayes, T.K., Holman, G.M. and Beier, R.C. (1994) *ACS Symposium Series*, **551**, 210–229.

118 Audsley, N., Matthews, J., Nachman, R.J. and Weaver, R.J. (2008) *Peptides*, **29**, 286–294.

119 Audsley, N., Matthews, J., Nachman, R. and Weaver, R.J. (2007) *General and Comparative Endocrinology*, **153**, 80–87.

120 Taneja-Bageshwar, S., Strey, A., Zubrzak, P., Williams, H., Reyes-Rangel, G., Juaristi, E., Pietrantonio, P. and Nachman, R.J. (2008) *Peptides*, **29**, 302–309.

121 Zubrzak, P., Williams, H., Coast, G.M., Isaac, R.E., Reyes-Rangel, G., Juaristi, E., Zabrocki, J. and Nachman, R.J. (2007) *Biopolymers*, **88**, 76–82.

122 Altstein, M. (2001) *Biopolymers*, **60**, 460–473.

123 Altstein, M. (2004) *Peptides*, **25**, 1373–1376.

124 Ben-Aziz, O., Zeltser, I. and Altstein, M. (2005) *Journal of Insect Physiology*, **51**, 305–314.

125 Ben-Aziz, O., Zeltser, I., Bhargava, K., Davidovitch, M. and Altstein, M. (2006) *Peptides*, **27**, 2147–2156.

126 Altstein, M., Ben-Aziz, O., Zeltser, I., Bhargava, K., Davidovitch, M., Strey, A., Pryor, N. and Nachman, R.J. (2007) *Peptides*, **28**, 574–584.

127 Inceoglu, A.B., Kamita, S.G., Hinton, A.C., Huang, Q., Severson, T.F., Kang, K.-D. and Hammock, B.D. (2001) *Pest Management Science*, **57**, 981–987.

128 Maeda, S. (1989) *Biochemical and Biophysical Research Communications*, **165**, 1177–1183.

129 Borovsky, D., Rabindran, S., Dawson, W.O., Powell, C.A., Ianotti, D.A., Morris, T.J., Shabanowitz, J., Hunt, D.F., DeBondt, H.L. and DeLoof, A. (2006) *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 18963–18968.

130 Fitches, E., Audsley, N., Gatehouse, J.A. and Edwards, J.P. (2002) *Insect Biochemistry and Molecular Biology*, **32**, 1653–1661.

131 Audsley, N., Weaver, R.J. and Edwards, J.P. (2001) *Physiological Entomology*, **26**, 181–188.

132 Isaac, R.E. (2004) In *Handbook of Proteolytic Enzymes*, vol. **1** (eds A.J. Barrett, N.D. Rawlings and J.F. Woessner), Elsevier Academic Press, Amsterdam, pp. 347–349.

133 Isaac, R.E., Lamango, N.S., Nachman, R.J., Strey, A. and Hayes, T.K. (1997) *Annals of the New York Academy of Sciences*, **814**, 339–341.

134 Isaac, R.E., Lamango, N.S., Ekbote, U., Taylor, C.A., Hurst, D., Weaver, R.J., Carhan, A., Burnham, S. and Shirras, A.D. (2007) *Peptides*, **28**, 153–162.

135 Ekbote, U., Looker, M. and Isaac, R.E. (2003) *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology*, **134**, 593–598.

136 Ekbote, U., Coates, D. and Isaac, R.E. (1999) *FEBS Letters*, **455**, 219–222.

137 Vercruysse, L., Gelman, D., Van De Velde, S., Raes, E., Hooghe, B., Vermeirssen, V., van Camp, J. and Smagghe, G. (1040) *Annals of the New York Academy of Sciences*, **2005**, 498–500.

138 Vercruysse, L., Gelman, D., Raes, E., Hooghe, B., Vermeirssen, V., Van Camp, J. and Smagghe, G. (2004) *Archives of Insect Biochemistry and Physiology*, **57**, 123–132.

139 Ekbote, U.V., Weaver, R.J. and Isaac, R.E. (2003) *Insect Biochemistry and Molecular Biology*, **33**, 989–998.

140 Houard, X., Williams, T.A., Michaud, A., Dani, P., Isaac, R.E., Shirras, A.D., Coates, D. and Corvol, P. (1998) *European Journal of Biochemistry*, **257**, 599–606.

141 Coates, D., Isaac, R.E., Cotton, J., Siviter, R., Williams, T.A., Shirras, A., Corvol, P. and Dive, V. (2000) *Biochemistry*, **39**, 8963–8969.

142 Isaac, R.E., Taylor, C.A., Hamasaka, Y., Nässel, D.R. and Shirras, A.D. (2004) *Invertebrate Neuroscience*, **5**, 51–64.

143 Kubli, E. (1992) *BioEssays*, **14**, 779–784.

144 Lange, A.B., Bendena, W.G. and Tobe, S.S. (1995) *Journal of Insect Physiology*, **41**, 581–588.

15
Plant Peptide Signals

Javier Narváez-Vásquez, Martha L. Orozco-Cárdenas, and Gregory Pearce

15.1
Introduction

Endogenous peptide signals play an important role in the regulation of plant growth, development, and reproduction, as well as mediators in the response of plants to environmental cues, including interactions with other organisms. The first peptide signal discovered in plants was systemin – a defense-related peptide isolated from tomato leaves in the laboratory of Professor Clarence A. Ryan [1]. At this time it was considered that plants only used small lipophilic molecules, such as auxins, cytokinins, gibberellins, abscisic acid, and ethylene, as regulators of plant growth and development [2, 3].

The absence of reliable bioassays and purification protocols, together with the low abundance, small size, and the presence of large families of functionally redundant peptide signals in plants, have made their identification a difficult task. To date, over a dozen families of peptide signals have been identified in plants using either biochemical or genetic screens (Table 15.1). However, the sequenced genomes of various plants, including *Arabidopsis*, rice, and poplar trees, contain hundreds of genes encoding potential peptide signal receptors [e.g., leucine-rich repeat receptor-like kinases (LRR-RLKs)] without known ligands (so-called "orphan receptors"). Through the use of bioinformatics, the development of new bioassays, and the refinement of high-performance liquid chromatography (HPLC) and mass spectrometry (MS) techniques, novel peptide families are being identified [4, 5].

Most plant regulatory peptides are extracellular signals derived from larger precursor proteins; they are active at very low concentrations and interact with cell membrane receptors to activate intracellular signal transduction events. Plant regulatory peptides seem to function at, or very close to, their site of synthesis in plants (e.g., paracrine hormones). For this chapter on plant regulatory peptides, the term peptide signals has been adopted to differentiate them from the classical endocrine peptide hormones known in animals [6].

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

Table 15.1 Plant peptides signals and their precursor proteins.

Peptide family name[a]	Precursor length (amino acids)	Processed peptides[b] (amino acids)	Signal peptide	Basic motifs[c]	PTMs[d]	Function	References
Systemin	200	18	no	no	no	defense	[1, 19]
HypSys	135–291	18–20	yes	no	O, G	defense	[31]
*At*Pep1	75–154	23	no	no	no	defense	[37]
CLV3/CLE	74–250	12	yes	yes	O	stem cell perpetuation and differentiation	[4, 47]
RALF	115	49	yes	yes	SS	cell division, root growth and development	[78]
ROT4/DVL	51–53	unknown	no	yes	no	polar cell proliferation	[84, 85]
CEP1	82–126	15	yes	no	O	root growth, elongation	[5]
PSK	77–87	5	yes	yes	-SO_3	cell proliferation, differentiation	[87]
PSY1	71–75	18	yes	no	O, -SO_3	cell proliferation, expansion	[104]
PLS	36	unknown	no	no	unknown	root growth, leaf vascularization	[107]
IDA/IDL	77–103	unknown	yes	yes	No	abscission	[111]
4k-P	119	37	yes	no	SS	cell proliferation, callus growth, defense	[115, 121]
SCR/SP11	74–81	50	yes	no	SS	self-incompatibility	[124, 125]

[a]See main text for abbreviations.
[b]Unknown, mature processed peptides have not been isolated from plant tissues yet.
[c]Potential mono/dibasic (prohormone convertase-like) processing sites flanking the bioactive peptide sequence.
[d]PTM of peptides or their protein precursors. O = proline hydroxylation; G = glycosylation; SS = disulfide bond; -SO_3 = sulfation.

This chapter presents a comprehensive review on peptide signals in plants, including the description of their biological activities, biochemical properties, gene family members and their regulation, and if known, their interactions with receptors and the signal transduction pathways they regulate. The different peptide families have been grouped according to their known or suspected function. The first group includes defense-related peptide signals, which includes systemin. The second larger group includes peptides involved in plant growth and development, through the regulation of cell division and differentiation, organ development, or abscission. The third group consists of peptides involved in the recognition of self (i.e., self-incompatibility), important in plant reproduction. Tables 15.1 and 15.2 present a summary of the main characteristics of the peptides included in this chapter. A methodological section is included to show the state-of-the-art techniques and strategies currently in use for the extraction and identification of bioactive peptides in plants. Conclusive remarks and perspectives for future research are presented in a final section that discusses several aspects that require further study, especially regarding the biogenesis of peptide signals in plants and the identification of their receptors, which for most peptide signals remain to be elucidated.

15.2 Defense-Related Peptides

Plants have evolved different families of functionally related peptide signals that serve to amplify the defense response against insect/herbivore and pathogen attacks through the activation of the jasmonic acid (JA) signaling pathway [7, 8]. This peptide-based signaling mechanism for plant defense is analogous to the activation of the inflammatory and acute-wound defense response mediated by peptide cytokines, such as the interleukins and tumor necrosis factor-α, through the eicosanoid signaling pathway in animals [9, 10].

15.2.1 Systemin

Systemin is an 18-amino-acid peptide involved in the activation and amplification of the defense response of plants against herbivores in some species of the Solanaceae [11]. At low nanomolar concentrations, systemin is very active in inducing defense proteinase inhibitors (PIs) in a plant bioassay (see Section 15.5.1) and also inhibits a proton pump ATPase in the cell membrane, causing a rapid increase in the pH of the extracellular medium of cell suspension cultures of *Solanum (Lycopersicon) peruvianum* (known as the alkalinization response, see Section 15.5.2). Systemin upregulates a whole array of defense-related genes, including several PIs and enzymes that cross-link proteins and degrade amino acids in the stomach of insect herbivores, diminishing food quality and digestibility [12]. More recently, it has been shown that systemin increases indirect defense mechanisms by inducing the release of plant volatiles that attract insects that parasitize attacking herbivores [13].

Table 15.2 Peptide sequences, conserved motifs, and receptors of plant peptide signals.

Peptide family	Representative peptide sequence with flanking amino acids[a]	Conserved motifs/structural feature	Receptors
Systemin	-REDL**AVQSKPPSKRDPPKMQTD**NNKL*	xxQxBPPxBBxPPBxQxx (palindromic sequence; polyproline II 3_1-helix structure) [1, 14]	SR-160 kDa, $K_d = 0.17$ nM [25]
HypSys	-GY**GRANLOOSOASSOOSKE**VSN-	(1) GA(Q/E)ARTLL (signal peptidase splice site) (2) -OOXO- or -OXOO- motifs (X = A, S, O, or T) (3) 3–4 hydroxy-Pro (O) with pentose attachments (6–17) [31] (4) conserved processing sites: NH_3-Y/F/N↓GR; xE/P/Q↓V/I/L/T-COOH [32]	unknown
*At*Pep1	-VTSR**ATKVKAKQRGKEKVSSGRPGQHN***	SSG(R/K)xGxxN (conserved C-terminal domain in *At*Pep1–7) [37]	(1) *At*PEPR1, LRR-RLK 170 kDa, $K_d = 0.25$ nM [41] (2) *At*PEPR2, LRR-RLK 170 kDa [42]
CLV3/CLE	-KGLGLHEEL**RTVOSGODPLHH**HVN*	KRxVPxGPNPLHNR (CLE motif) [4]	(1) CLV1, LRR-RLK 105 kDa, $K_d = 17.5$ nM [52] (2) CLV2, LRR-RLK, 80 kDa [53] (3) CRN, receptor kinase, 44 kDa [54]
RALF	-RRIL**ATKKYISYGALQKNSVPCSRRGSASY YNCKPGAQANPYSRGCSAITRCRS***	(1) RRIL↓ATK (potential processing dibasic site) (2) four conserved cysteines form disulfide bridges [78]	25 and 120 kDa, $K_d = 0.8$ nM [81]
ROT4/DVL1	MEMKRVMMSSAERSKEKKRSISRRLGKYM KEQK**GRIYIIRRCMVMLLCSHD***	KEQKGRIYIIRRCMVMLLCSHD (RTF domain) [85]	unknown
CEP1	-KEVIAHPT**DFROTNPGNSOGVGH**SNGRH*	DFxPxNPGxS(P/Q)G(V/I)xH [5]	unknown

PSK	RRSLVLHTD**YIYTQ**NHKP*	(1) Y(SO_3)IY(SO_3)TQN (disulfated pentapeptide) (2) RRSLVL↓HTDY (*At*SBT1.1 cleavage site) [96]	(1) PSKR1, LRR-RLK 120 kDa, $K_d = 1.4$ nM [102] (2) PSKR2, LRR-RLK 160 kDa, $K_d = 27$ nM [100, 104]
PSY	RSLLMVNVE**DYGDPSANPKHDPGVOOS**ATG QRVVGRG*	(1) DY(SO_3)xDP(S/T)ANxBHDPxxPxx (2) hydroxy-Pro at position 16 in *At*PSY1 has three pentoses(L-arabinose) attached [104]	*At*1g72300, LRR-RLK, 120 kDa [104]
PLS	MKPRLCFNFRRRSISPCYISISYLLVAKLFKLFKIH*	unknown	unknown
IDA/IDL	-ARIGATMEMKKNIKRLTFKNSHIFGYLPKGVP IPPSAPSKRHNSFVNSLPH*	(L/F)(P/S)Bxx(P/L)(I/V)PxSxBHN (conserved C-terminal domain) [111]	unknown
4k-P	-KIEA**ADCNGACSPFEVPPCRSRDCRCVPIGL FVGFCIHPTGLSSV**-	T-knot scaffold structure with three β-strands stabilized by three disulfide bridges between conserved cysteines [119]	23k-P, receptor-like glycoprotein, 43 kDa, $K_d = 18$ nM [117]
SCR/SP11	QELEANLMKR**C**TRGFRKLGK**C**TTLEEEK**C**KT LYPRGQ**C**T**C**SDSKMNTHS**C**D**C**KS**C**	protein folds into an α/β-barrel structure stabilized by four disulfide bridges between conserved cysteines [130]	SRK, 110 kDa, $K_d = 0.7$ nM [137]

[a]Sequences corresponding to isolated bioactive peptides are bold. O = hydroxy-Pro; B = basic residue;
* = C-terminus.

Systemin has a palindromic sequence centered around two sets of double prolines (Table 15.2). Circular dichroism experiments on synthetic systemin revealed a polyproline II 3_1-helix [14], a more extended structure than an α-helix, often associated with ligand–receptor interaction in animals [15, 16]. Alanine-substitution (alanine-scanning) and -deletion analysis revealed that replacement of most of the amino acids in systemin had little effect. However, a complete loss of activity was obtained when the penultimate threonine was replaced by alanine or when the C-terminal aspartic acid was removed [17]. Both of these analogs were competitive inhibitors of systemin in both the excised plant assay and the alkalinization assay. The C-terminal of the peptide is essential for activity, whereas the N-terminal may be important for binding the receptor [17, 18].

Systemin is derived from the C-terminus of a 200-amino-acid precursor called prosystemin [19]. Unlike most prohormone precursors in animals and yeast, prosystemin does not have a putative signal sequence, or endoplasmic reticulum (ER)/Golgi-dependent post-translational modifications (PTMs), suggesting that it is not synthesized through the secretory pathway, but on free ribosomes in the cytosol. Direct immunocytochemical evidence indicates that prosystemin is processed and compartmentalized in the cytosol of vascular phloem parenchyma cells [20]. The mechanism of systemin transport to the extracellular space is still under investigation.

The processing sites and the responsible enzyme(s) that release systemin from its protein precursor are currently unknown. Prosystemin does not have classical dibasic sites for processing by prohormone convertase-like enzymes [21], except for a dibasic KR sequence present in the middle of systemin (Table 15.2). In fact, a systemin-binding protease of the furin-like family of prohormone convertases is found in tomato cell membranes that cleaves systemin at this site [22]. The protease is apparently involved in systemin turnover, since the proteolytic products have significantly reduced biological activities [17]. Accordingly, substitution of these basic residues with chemically modified amino acids increases the half-life and activity of the modified systemin [23].

The importance of prosystemin as an integral component of defense signaling in tomato plants was demonstrated by production of transgenic plants constitutively expressing the prosystemin gene in sense and antisense orientation under the regulation of the cauliflower mosaic virus 35S promoter. The transgenic plants overexpressing the sense prosystemin gene constitutively accumulated large amounts of PIs among other defense compounds, and the plants were more resistant to herbivory by *Manduca sexta* larvae compared to wild-type plants [24]. Conversely, antisense prosystemin plants were consumed more readily by *M. sexta* larvae than the control plants [12].

Systemin interacts with a membrane receptor at the cell surface of *S. peruvianum* cells [25, 26]. The binding is very specific ($K_d = 0.17$ nM), saturable, and reversible [25]. A systemin-binding protein (160 kDa) was isolated from microsomal membranes of *S. peruvianum* cells, using an iodinated-systemin photoaffinity analog [26]. MS analysis of tryptic digests identified the binding protein as a LRR-RLK, that was proposed to be the same BRI1-like brassinolide receptor CURL3 [26, 27].

Recently, however, it has been shown that the systemin receptor and the brassinolide receptor are two different entities, and that BRI1 is not involved in systemic signaling [28, 29]. The search for the real systemin receptor is still on-going.

Upon systemin binding, a mitogen-activated protein (MAP) kinase is activated and a phospholipase cleaves linolenic acid (18 : 3 fatty acid) out of cell membranes, which, through the octadecanoid pathway, is converted to JA and conjugated to isoleucine. JA-Ile is a potent activator of defense-related genes, including the prosystemin gene itself [8]. A positive feedback loop is initiated, in which elevated levels of JA cause an increase in prosystemin synthesis, thereby producing more JA. Like the prosystemin gene, the genes that code for the octadecanoid pathway enzymes are upregulated in the vascular bundles of tomato leaves. After wounding by herbivore attacks, JA is rapidly produced and transported through the phloem to activate a systemic wound-defense response in tomato [8, 20].

15.2.2
Hydroxyproline-Rich Systemin Glycopeptides

Hydroxyproline-rich systemin (HypSys) glycopeptides are small glycopeptides that show little sequence homology with systemin, but are functionally similar. HypSys peptides are induced upon injury by herbivore attack, target to a membrane-bound receptor, and activate the JA signaling pathway. First isolated from tobacco leaves (*Nicotianum tabacum*, *Nt*) [30], HypSys peptides have now been isolated from tomato (*S. lycopersicum, Sl*) [11], petunia (*Petunia hybrida*, *Ph*) [31], and nightshade (*Solanum nigrum, Sn*) [32]) of the Solanaceae family, and recently from sweet potato (*Ipomoea batatas, Ib*) of the Convolvulaceae family [33]. The glycopeptides share common features including similar sizes of 18–20 amino acids, a central -OOXO- or -OXOO- motif (O = hydroxy-Pro; X = Ala, Ser, hydroxy-Pro, or Thr), an arginine near the N-terminal and usually a glutamine or glutamic acid as the C-terminal amino acid (Table 15.2). The central hydroxyproline region is the site of arabinosylation, and the exact location of the attachment sites and the lengths of the pentose chains are currently under investigation.

The HypSys glycopeptides from tobacco and tomato have been also shown to induce the expression of PIs in the plant bioassay and to induce the alkalinization response of cell cultures (see Section 15.5.2) [11, 30]. In petunia, supplying *Ph*HypSys to detached plants induced the expression of *DEFENSIN*, a gene involved in pathogen defense, but did not induce PIs [31]. Transgenic tobacco and tomato plants overexpressing homologous preproHypSys genes under the regulation of the 35S promoter, constitutively accumulated high levels of defense proteins, like the prosystemin sense plants [34, 36]. The transgenic tobacco plants were more resistant to insect attack [34].

*Nt*HypSys glycopeptide sequences were used to design primers and a copy DNA (cDNA) was isolated that contained two of the HypSys peptides – the first example of multiple bioactive peptides coded in a single gene in plants [30]. Using a similar approach, cDNAs have been found in other species that contain between two and six putative glycopeptide signals per gene. The sequences of the proproteins contain similar features including a signal sequence (in contrast to prosystemin), a conserved

10 amino acid propeptidase cleavage site, conserved bioactive glycopeptide processing sites, and conservation at the C-termini (Table 15.2).

All of the HypSys precursors contain a signal sequence, which along with the hydroxylation of prolines, and subsequent glycosylation, indicates that the precursor is synthesized through the secretory pathway. In tomato, the *Sl*proHypSys precursor was found localized in the cell wall and, like prosystemin, associated with the vasculature of the plant [35]. Recently, it was shown that systemin and *Sl*HypSys work cooperatively in tomato to upregulate the systemic defense response [36].

15.2.3
*Arabidopsis At*Pep1-Related Peptides

Following the development of the alkalinization assay (see Section 15.5.2), the search for bioactive peptides was extended to the genetic model plant, *Arabidopsis*. An alkalinizing peak was found with similar chromatographic characteristics to the systemins, but upon purification, a completely different peptide was identified. The peptide, consisting of 23 amino acids, was named *At*Pep1 (from *A. thaliana*) [37]. Like systemin, *At*Pep1 is derived from the C-terminus of a larger precursor protein (*PROPEP1*) of 92 amino acids in length that lacks a signal sequence. Also, similar to prosystemin, *PROPEP1* gene expression was upregulated by wounding and methyl jasmonate (MeJA). When synthetic *At*Pep1 was supplied to excised *Arabidopsis* leaves, there was an increased expression of the pathogen-related *DEFENSIN* gene and an increase in H_2O_2 production. Defensin and H_2O_2 are components of the innate immune response that are induced by exogenous pathogen-associated molecular pattern (PAMP) signals, such as flg22, a 22-amino-acid peptide derived from bacterial flagellin [38], and elf18, an 18-amino-acid peptide derived from an abundant bacterial elongation factor protein [39]. *At*Pep1 activated the same genes as the PAMPs and is the first example of an endogenous peptide elicitor of pathogen defense [37]. To further study the effects of *At*Pep1, *Arabidopsis* plants overexpressing *AtPROPEP1* were infected with the root pathogen *Pythium irregulare* and were found to have increased resistance to the pathogen compared to wild-type plants [37].

AtPROPEP1 is a member of a small gene family consisting of seven genes [40]. Although there is extreme heterogeneity among the *At*Pep1 paralogs, they all contain a conserved motif within the carboxy-region of the mature *At*Pep1-related peptides (Table 15.2). Unlike systemin, which is found only amongst close relatives of tomato, *At*Pep1 sequences have been identified in a number of plants including agriculturally important crops, such as maize, barley, wheat, rice, and soybean [37].

The *PROPEP*1-related genes exhibit differential expression to MeJA and methyl salicylate (MeSA) [40]. *PROPEP1* and *2* are strongly induced by MeJA and *PROPEP2* and *3* are induced by MeSA [40]. When six of the synthesized *At*Pep1 paralogs were fed to excised *Arabidopsis* leaves, only *At*Pep1 and 2 induced *DEFENSIN*, whereas all except *At*Pep4 induced expression of pathogenesis-elated protein-1. The *At*Pep1 family of peptide signals appear to be involved in an amplification loop, where wounding or pathogen attack activates the expression of the *PROPEP*1-related genes,

possibly acting through the SA and/or JA pathways to amplify the defense response against pathogen attack [10].

The receptor for *At*Pep1 (*At*PEPR1) was isolated from *Arabidopsis* suspension cells [41]. An [^{125}I]Tyr-*At*Pep1 analog was found first to bind to cell membranes in a strong and specific manner. Then, an [^{125}I]Cys-azido-labeled *At*Pep1 photoaffinity analog specifically interacted with a membrane-bound 170-kDa protein that was subsequently purified and identified as a member of a large LRR-RLK family. All seven of the *At*Peps were able to compete for receptor binding with the [^{125}I] Tyr-*At*Pep1 analog to varying degrees that corresponded to their abilities to alkalinize the *Arabidopsis* cell culture medium. Microsomal membranes from two T-DNA-insertional knockout mutants of the receptor were unable to interact with the azido-labeled *At*Pep1 analog. The *AtPEPR1* receptor gene was transformed into tobacco suspension cells, which normally do not respond to *At*Pep1, and the transgenic cell line gained the ability to alkalinize upon addition of the peptide, proving that *At*PEPR1 was a functional *At*Pep receptor [41]. More recently, an LRR-RLK was identified with 72% homology with *At*PEPR1 (called *At*PEPR2) [42]. *At*PEPR1 and 2 genes are similarly induced in response to wounding, MeJA, and *At*Pep1. The interaction of these two receptors with all seven of the *At*Peps is currently being investigated [42].

15.3 Peptides Involved in Growth and Development

15.3.1 CLAVATA3 and the CLE Peptide Family

The CLAVATA3/endosperm surrounding region (CLE) is so far the largest class of plant peptide signals that function in plant development by promoting or inhibiting cell division and differentiation of stem cells, in shoot and root meristems, and in vascular tissues [43, 44]. Bioactive CLE peptides are dodecapeptides derived by post-translational proteolysis from a conserved 14-amino-acid semiconsensus sequence called the CLE motif (Table 15.2), located at the C-terminus of larger precursor proteins [45, 46].

15.3.1.1 CLAVATA3 (CLV3)

CLAVATA3 (CLV3) was the founding member of the CLE peptide family in plants [47]. CLV3 is a 12-amino-acid peptide derived from the C-terminus of a small 96-amino-acid preproprotein, synthesized with a classical signal peptide through the secretory pathway. The peptide isolated from *Arabidopsis* contains two hydroxylated prolines (Pro3 and Pro6, Table 15.2), but not carbohydrate structures attached to them, as determined by matrix-assisted laser desorption/ionization-time-of-flight MS [48]. Proline hydroxylation is not required for full biological activity of the peptide.

The *CLV3* precursor gene is expressed in the upper layers of stem cells at the tip of the shoot-apical meristem (SAM) and the protein is localized in the cell walls of SAM

cells [49, 50]. Below the first three layers of stem cells, CLV3 causes SAM cells to differentiate into aerial organs. At the same time, CLV3 acts to maintain stem cells in an undifferentiated state in the SAM, which is essential for plant growth and organogenesis to continue [49].

CLV3 and other members of this peptide signaling pathway were identified by genetic screens in *Arabidopsis* (Figure 15.1a). Genetic and biochemical data support the hypothesis that CLV3 is the ligand for at least two different receptors: CLV1 [51, 52] and the CLV2/CORYNE (CRN) receptor complex [53, 54]. Mutations in the *CLV1, CLV2, CLV3* and *CRN* genes cause accumulation of stem cells, enlargement of the SAM, and the production of supernumerary floral organs and other structural defects (Figure 15.1b). In contrast, overexpression of CLV3 causes loss of stem cells and, in turn, developmental arrest [49].

CLV1 encodes a receptor kinase protein (~105 kDa) with LRR motifs, a transmembrane domain, and a serine–threonine protein kinase domain [55]. Ligand binding assay and photoaffinity labeling indicated that the CLV3 peptide directly binds the CLV1 extracellular domain with a dissociation constant (K_d) of 17.5 nM [52].

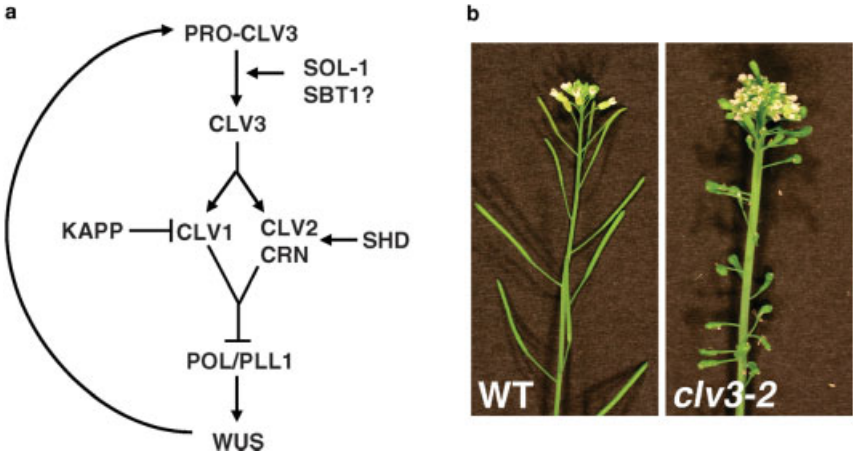


Figure 15.1 CLV3 signaling in *Arabidopsis*. (a) Current model of CLV3 signaling depicting the CLV–WUS feedback loop. CLV3 propeptide is secreted into the extracellular space (cell wall apoplast), where it is processed by the combined action of a subtilase-like serine protease (possibly SBT1) [96] and exopeptidases that may include the Zn^{2+}-carboxypeptidase SOL-1 [64]. CLV3 acts through at least two receptor systems, CLV1 and CLV2/CRN, to restrict stem cell fate in shoot and floral meristems [53, 54]. In the presence of CLV3, CLV1 is autophosphorylated and associates with the protein phosphatase KAPP, which can bind and autophosphorylate CLV1, acting as a negative regulator of the pathway [57, 58]. SHD is a HSP90-like protein, possibly involved in the assembly of a functional receptor complex [63]. The protein phosphatases POL1/PLL1 act downstream as positive regulators of WUS transcription that is downregulated by the CLV complex. WUS activity confers stem cell identity and positively regulates CLV3 expression. This CLV–WUS negative-feedback loop confers stem cell homeostasis on the shoot apical meristem. (b) Loss of the restrictive CLV3 signal in the the *clv3-2* mutant mutant causes stem cell overproliferation and the enlargement of shoot and floral meristems (Photo courtesy of Dr. Harley Smith).

The CLV1 ectodomain also interacts with other structurally related CLE peptides, but with distinct affinities depending on the specific amino acid sequence [52].

CLV2 is a receptor-like protein (~80 kDa) similar to CLV1, but lacks the intracellular kinase domain [53]. CLV2 appears to be involved in multiple developmental pathways, independent of CLV1 and CLV3. In addition to defects in meristem regulation, mutations at the *CLV2* locus result in defects in gynoecium, pedicel, and stamen development [53]. The broad expression pattern of *CLV2* in *Arabidopsis* also suggests that it may be involved in a wide-range of CLE-mediated signaling pathways [54, 56].

CRN is another receptor kinase that appears to act in conjunction with CLV2, and in parallel with CLV1, to perceive the CLV3 signal [54]. Since CRN protein lacks a distinct extracellular domain, it has been proposed that CRN and CLV2 interact via their transmembrane domains to establish a functional receptor. Mutations in *CRN* cause stem cell proliferation, similar to *clv1*, *clv2*, and *clv3* mutants. Like CLV2, CRN has additional functions during plant development, including floral organ formation [54].

Other intracellular signal transduction components that modulate CLV3 signaling include (Figure 15.1a): (i) a type 2C kinase-associated protein phosphatase (KAPP) that interacts directly and dephosphorylates the CLV1 kinase domain, and acts as a negative regulator of CLV3 signal transduction [57, 58]; (ii) another protein phosphatase, POLTERGEIST (POL) and its relative PLL1, provide partial, additive suppression of the *clv* stem cell accumulation defect [59–61]. *POL/PLL1* overexpression inhibits differentiation and induces stem cell accumulation, especially in a *clv* background, suggesting that POL/PLL1 are intermediates downstream of the CLV1 and CRN/CLV2 receptors [54, 61]; (iii) ROP, a Rho GTPase-like protein that is related to the Ras GTPase superfamily in animals [62]; (iv) SHEPHERD, a chaperone-like HSP90 protein that may be involved in folding or assembling of CLV3 signaling components including receptor complexes [63]; and, finally, but not least, (v) SOL-1, a Zn^{2+}-carboxypeptidase that can be involved in the processing of CLE peptides in root meristems [64].

One major target of the CLV3 signaling pathway is to downregulate the expression of WUS, a member of the WOX family of homeodomain transcription factors [65, 66]. *WUS* mRNA expression is confined to a group of cells in the center of the meristem. WUS acts non-cell-autonomously to confer stem cell identity on the overlying cells and promote CLV3 expression in those cells. Through this CLV3–WUS negative-feedback loop, stem cell homeostasis is maintained in the SAM. The cumulative evidence indicates that a balance between WUS and CLV3 is required both to regulate stem cell proliferation and to position the stem cell boundary in the SAM [66–68]. Thus, the CLV3 polypeptide is a key developmental mediator required to communicate cell fate information between the stem cells and their neighbors in the SAM.

15.3.1.2 **CLV3-Related Peptides**

The protein precursors of CLE peptides are encoded by a large and very diversified gene family in plants [4, 44]. Little sequence similarity exists upstream of the CLE motif. The genomes of *Arabidopsis* and rice, respectively, contain 37 and 47 putative

CLE precursor genes, indicating the diverse and possibly overlapping functions of CLE peptides in these plants. A recent bioinformatic analysis identified 179 members of the CLE protein family from various plant species, as well as the existence of multiple CLE domains in some *CLE* genes [4]. Based on sequence similarities in the primary CLE motif, proteins were clustered into 13 distinct groups that correlate with phylogenetic analysis and known biological functions of CLE signaling peptides based on *in vitro* bioassays and ectopic overexpression studies.

CLE genes are expressed in most plant organs, but some genes such as *CLV3* in *Arabidosis*, and its rice ortholog *FON4/FON2*, are only expressed in the stem cells of shoot and floral meristems [70, 71]. In maize [72], three highly homologous *CLE* genes are expressed in the ESR. Overlapping expression profiles of *CLE* genes in one or more tissues also suggest functional redundancy. In fact, ectopic expression of at least 12 *Arabidopsis CLE* genes in the SAM produced phenotypes similar to those of *CLV3* overexpression [73]. Expression of the same genes in the *CLV3* mutant reversed the mutant phenotype. Overexpression of other group of *CLE* genes caused phenotypes similar to, but milder than, those caused by *CLV3* overexpression [73].

The function of 26 chemically synthetic CLE peptides, which correspond to the predicted products of the 31 *Arabidopsis* CLE genes, was also investigated in *Arabidopsis* and rice [46, 74]. One group of synthetic CLE dodecapeptides inhibited root elongation in rice as well as in *Arabidopsis*. The same group reduced the size of the SAM in *Arabidopsis*, but not in rice, suggesting that the molecular mechanisms of shoot and apical meristem maintenance are highly conserved in different plant species, although some degree of divergence and specialization also exists [4, 44, 75].

Tracheary elements differentiation inhibitory factor (TDIF) is a recently isolated CLE dodecapeptide that suppresses xylem cell development at a concentration of 10^{-11} M and promotes cell division of *Zinnia elegans in vitro* cultured cells [46]. Similar to CLV3, Pro4 and Pro7 in TDIF are hydroxylated, but hydroxylation is not required for peptide activity. Alanine-substitution studies indicated that amino acids at positions 1, 3, 6, 8, 9, and 12 are essential for full TDIF activity [46].

Outside the plant kingdom, *CLE* genes have also been found in certain species of plant root endoparasite nematodes [69]. The soybean cyst nematode (*Heterodera glycines*) *HgSYV46* gene encodes a protein of 139 amino acids, with a predicted N-terminal signal peptide and the characteristic CLE domain at the C-terminus [76]. The nematode CLE peptide sequence is highly homologous to a group of CLEs predominantly expressed in *Arabidopsis* roots. It has been proposed that the peptides are secreted by the nematode at the early stages of colonization of plant roots, to reprogram cell division and differentiation, and establish specialized cyst feeding sites in the host [69]. This finding suggests an interesting coevolution of a plant parasite that mimics developmentally regulatory peptides for its own parasitic advantage [77].

CLE peptides should interact with membrane-bound, LRR-RLKs, similar to CLV1, or receptor complexes such as CRN/CLV2, to promote or inhibit cell proliferation and

differentiation (Figure 15.1a, Section 15.3.1.1). The different binding affinities of CLE peptides for the ectodomain of CLV1 [52] and related receptors may, together with the location and timing of expression of these components, determine the specificity of some of these functionally redundant peptide signals in plant tissues.

15.3.2 Rapid Alkalinization Factor Peptides

Rapid alkalinization factor (RALF) was first detected in tobacco leaves while purifying the defense glycopeptides *Nt*HypSys I and II, utilizing the alkalinization assay (see Section 15.5.2). The suspension cell response to the RALF HPLC peak was very rapid and had a greater magnitude compared to the HypSys glycopeptides (~5 versus 15 min for HypSys). Purification yielded a peptide 49 amino acids in length with a mass of 5.3 kDa and a half-maximal response of 1 nM in the alkalinization assay [78]. A MAP kinase activity was associated with RALF and was also more rapidly activated than MAP kinase activity induced by HypSys peptides. RALFs were isolated from tomato, alfalfa, and poplar with similar N-terminal sequences and masses. A cDNA coding for tobacco RALF revealed a precursor of 115 amino acids in length, containing a signal sequence of 25 amino acids with the mature RALF at the C-terminus (Table 15.2).

A mature tomato RALF peptide was synthesized and the reduced form was active at about one order of magnitude less than native RALF in the alkalinization assay. If the reduced form was treated with iodoacetamide to block disulfide bridge formation, RALF activity was completely abolished. Reduced RALF can apparently fold into the proper conformation for receptor recognition; however, when the disulfide bonds fail to form, there is no receptor recognition. Disulfide bridge formation under oxidizing conditions resulted in a mixture of products that could be separated by HPLC, revealing one peptide with full biological activity (see Section 15.5.2) [78].

Gene databank searches have revealed that RALF is ubiquitous in the plant kingdom, and RALFs have been identified in all tissues and organs of *Arabidopsis* [78] and poplar [79]. Treatment of *Arabidopsis* and tomato plants with RALF caused arrested root growth [78] and experiments done with *Nicotiana attenuata* root hairs reveal that RALF regulates apoplastic pH, affecting root tip growth and development [80]. These experiments and the ubiquitous nature of RALF reveal an essential role for RALF in growth and development in roots. The underlying mechanisms of RALF function and its role in other tissues remain to be explored.

Using techniques similar to those used to characterize the systemin receptor, [^{125}I] azido-labeled RALF bound specifically to two cell surface membrane proteins with masses of 25 and 120 kDa, which are believed to be part of a receptor complex [81]. In *Arabidopsis*, 34 genes were identified that code for RALF or RALF-like proteins [82], making loss of function mutational analysis of RALF extremely difficult. The complexity of expression and the large number of RALFs makes purification and cloning of the receptor essential to understanding the role of RALF, as this would allow researchers to perform loss of function mutational analyses.

Ectopic expression of RALF fused with Green Fluorescent Protein (GFP) in *Arabidopsis* showed the localization of the RALF–GFP fusion product in the apoplast [83]. This result, along with the presence of a signal sequence, and its interaction with the putative membrane-bound receptor, indicate that RALF is an extracellular signaling peptide.

15.3.3 Rotundifolia4 and Devil1

Rotundifolia4 (ROT4) and Devil1 (DVL1) are small homologous proteins (~6.2 kDa) that restrict polar cell proliferation during plant development. *ROT4* and *DVL1* genes were both independently discovered in *Arabidopsis* using gain-of-function genetic screens for genes affecting leaf and fruit development, respectively [84, 85]. Compared with wild-type plants, the dominant mutants had shorter, rounded leaves, petioles, floral organs, and stems, and siliques with horned tips. It was then observed that overexpression of *ROT4*/*DVL1* genes primarily limited the longitudinal growth of lateral organs with determinate development, such as leaves, sepals, and petals [84, 85].

ROT4 and *DVL1* genes belong to a multigene family in plants (22 members in *Arabidopsis*), weakly expressed in the shoot apex, young leaves, flowers, and roots [84, 85]. However, the expression patterns and levels differ among members of the family. When ectopically overexpressed, most members of the *ROT4/DVL1* gene family caused the same short-leaf phenotype, indicating functional redundancy [84, 86]. Reverse transcription-polymerase chain reaction (RT-PCR) analysis of genes involved in fruit development indicated that the expression of the MADS- box *FRUITFUL* (*FUL/AGL8*) gene is reduced 2- to 5- fold in DVL1 overexpressing lines [86].

The proteins encoded by *ROT4* (53 amino acids) and *DVL1* (51 amino acids) lack a predictable signal peptide sequence. Members of this peptide family share homology primarily at their C-terminal 29 amino acids (called the "RTF" domain, Table 15.2). Overexpression of the RTF domain alone was sufficient to confer the short leaf phenotype, indicating the importance of the C-terminal domain for protein function. In fact, amino acids Y37, I38, C42, and L46 in the RTF domain are absolutely conserved in all members and essential for the overexpression phenotype [86].

The ROT4 and DVL1 proteins may be processed to smaller peptides at mono- or dibasic sites present at the beginning of the RTF motif, since mutation of R23 in DVL1 was also required for the short-leaf overexpression phenotype [86]. Expression of GFP–DVL1 protein fusions suggested that the putative precursor protein may be processed in the cytosol, although the GFP–ROT4 fusion was found associated with the plasma membrane [86].

It is not clear yet if a processed ROT4/DVL1 peptide is also secreted to the extracellular space to interact with a receptor to initiate a signaling cascade that negatively regulates cell proliferation in developing organs. So far, the mature protein has not been biochemically purified or detected in plant tissues. It would be interesting to see if synthetic RTF peptides have any biological activity in limiting the growth of lateral organs when artificially supplied to plant seedlings.

15.3.4
C-Terminally Encoded Peptide 1

A C-terminally-encoded peptide 1 (CEP1) is a recently discovered 15-amino-acid bioactive peptide involved in root growth and development in *Arabidopsis* [5]. Using a bioinformatics approach, Ohyama *et al.* [5] identified over 100 open reading frames potentially encoding small, C-terminally located secreted bioactive peptides, including CEP1 (see Section 15.5.4). To test whether the CEP1 peptide was processed and secreted *in vivo*, the gene encoding the putative CEP1 precursor protein was overexpressed in *Arabidopsis* under the control of the strong constitutive 35S promoter. Transgenic plants were then grown *in vitro*, fully submerged in liquid medium, and the secreted CEP1 peptide was extracted from the culture medium with *o*-chlorophenol followed by acetone precipitation, and identified by liquid chromatography (LC)-MS/MS [5]. Phenotypically, the plants overexpressing CEP1 showed reduced primary root growth and lateral root elongation, resulting from a limited number of cells in the meristematic zone and reduced cell size in the elongating region. *Arabidopsis* seedlings grown in the presence of synthetic CEP1 also showed root growth inhibition in a dose-dependent manner ($\geq 10^{-7}$ M).

There are five genes encoding highly divergent CEP1 preproproteins of 82–126 amino acids, with a putative signal peptide and a conserved 17-amino-acid CEP1 domain at the C-terminus (Table 15.2). Two of the prolines (Pro4 and Pro11) in CEP1 are hydroxylated, but the hydroxyprolines are neither glycosylated nor required for biological activity [5]. RT-PCR analysis indicated that most CEP1-encoding genes are primarily expressed in roots, except *At2g35612*, which was only detected in flowers. Expression of a GUS reporter gene under the regulation of the CEP1 promoter was only observed in the SAM and lateral root primordia [5]. Despite no CEP1 receptor being identified yet, the evidence indicates that CEP1 is a new bioactive extracellular signal peptide in plants that regulates lateral root growth and development.

15.3.5
Tyrosine-Sulfated Peptides

15.3.5.1 Phytosulfokine

Plant suspension cell cultures secrete a "conditioning factor" into the liquid medium that helps maintain the mitotic activity and high rates of cell proliferation of freshly diluted (low-density) cultures. This factor was first purified by Matsubayashi and Sakagami [87] from *Asparagus officinalis* mesophyll cell cultures, and found to be a disulfated pentapeptide [Y(SO_3)IY(SO_3)TQ] called phytosulfokine (PSK).

Nanomolar quantities of PSK were capable of promoting cell division and further differentiation of asparagus suspension cultures [88], tracheary element differentiation of *Zinnia* mesophyll cells [89], and somatic embryogenesis of carrot cells [90]. PSK also promotes the formation of adventitious roots and buds in tissue cultures [91], and enhances pollen germination *in vitro* [92].

Peptides identical to PSK have been isolated from many different plant species (monocots and dicots), confirming an important conserved role for PSK in regulating

plant cell proliferation *in vivo*. Structure–function and deletion analyses using synthetic PSK indicated that the first three amino acids and the sulfate groups are essential for biological activity, which explains the strict conservation of the PSK sequence in plants [93].

PSK is synthesized as a preproprotein of around 80 amino acids that has a signal peptide at its N-terminus, suggesting that it is processed through the ER/Golgi secretory system. In fact, PSK was the first sulfated peptide found in plants and the presence of *O*-sulfated tyrosine in PSK led to the discovery of a Golgi-localized tyrosylprotein sulfotransferase (TPST) [94]. Acidic amino acid residues adjacent to the tyrosine residues of acceptor peptides were essential for the plant TPST-catalyzed sulfation reaction [94].

PSK is located at the C-terminus of its protein precursor, preceded by a highly conserved dibasic pair of amino acids located about eight residues upstream of PSK (Table 15.2). Mono- and dibasic pairs of amino acids are preferred processing sites by prohormone convertase-type enzymes in animals [95]. Recently, however, it has been reported that *Arabidopsis* prepro*At*PSK4 is proteolytically processed by *At*SBT1.1 – a subtilisin-like serine protease at a different cleavage site (RRSLVL↓HTDY), four amino acids downstream of the basic residues [96]. The resultant processed peptide had extended N- and C-terminal tails that should be trimmed by still unknown exopeptidases to generate the mature, fully active pentapeptide. The expression of the *AtSTBT1.1* gene and the release of PSK4 were activated after dissection and transfer of root segments into culture medium for cell proliferation and callus formation *in vitro* [96]. This is the first prohormone processing enzyme identified in plants.

The PSK precursor is encoded by a small, six-member gene family in *Arabidopsis* [97]. The genes are expressed in most plant tissues and at all developmental stages. Ectopic expression and antisense inhibition of the rice PSK gene in transgenic cell lines reconfirmed its role in modulating cell proliferation [98]. The PSK overexpressing cell lines grew faster and the antisense lines slower than the wild-type. However, overexpression of the PSK gene in transgenic plants does not cause a visible effect on plant growth or development [99].

PSK action is mediated by its interaction with a membrane receptor at the cell surface. Using [^{3}H]PSK and [^{35}S]PSK, high- and low-affinity binding sites ($K_d = 1.4$ and 27 nM, respectively) were detected in plasma membrane fractions of rice. The binding was saturable and reversible, with around 10^4 binding sites per rice cell [100]. Two proteins of 120 and 160 kDa were identified in SDS-gels after photoaffinity-labeling using a [^{125}I]PSK analog in the rice plasma membranes [101]. Subsequently, a 120-kDa PSK receptor protein (PSKR1) was isolated from microsomal fractions of carrot cells by detergent solubilization and affinity chromatography, and its gene was cloned [102]. PSKR1 belongs to a family of serine/threonine receptor kinases with extracellular LRRs, similar to other LRR-RLKs in *Arabidopsis*, such as the brassinolide and the CLV1 receptors [102].

The function of PSKR1 as the receptor for PSK was reconfirmed by overexpressing the *PSKR1* cDNA in carrot cells. The transgenic cells exhibited a significant increase in PSK binding, and showed higher growth rates than the wild-type cell line [102].

Arabidopsis plants overexpressing *AtPSKR1* also showed larger leaves and a delayed senescence compared with the wild-type [103].

15.3.5.2 Plant Peptides Containing Sulfated Tyrosine 1

A recent search for secreted tyrosine-sulfated peptides in plant cell culture medium using LC-MS led to the discovery of a new 18-amino-acid tyrosine sulfated glycopeptide in *Arabidopsis* (see Section 15.5.3) [104]. The peptide named PSY1 (plant peptide containing sulfated tyrosine 1) promotes cell proliferation and plant growth, similar to PSK. PSY1 is derived from the C-terminus of a 75-amino-acid precursor protein, which has a typical secretion signal sequence at its N-terminus. Two of the prolines (Pro16 and 17) in PSY1 are hydroxylated, and the peptide is decorated with a chain of three L-arabinoses attached to the hydroxy-Pro16 (Table 15.2). The *Arabidopsis* genome has three genes encoding homologs of the PSY1 precursor protein. In all PSY1 precursors, the tyrosine-sulfated glycopeptide is flanked by the conserved sequence RSLLM, which resembles the cleavage site for *At*proPSK processing by *At*SBT1.1 [96]. This suggests that both PSK and PSY1 may be proteolytically released by the same subtilase-like processing peptidase.

Histochemical analysis of transgenic plants carrying a PSY1 promoter–GUS reporter gene indicated that PSY1 is expressed in various tissues in *Arabidopsis*, with higher levels of expression in the SAM, root elongation zone, and the leaf margins [104]. In addition, the PSY1 gene was upregulated after wounding of leaves. Transgenic *Arabidopsis* seedlings overexpressing the PSY1 gene developed longer roots and larger cotyledons than the wild-type control, due to an increase in cell size [104]. In addition, germination of *Arabidopsis* seeds in medium supplemented with natural PSY1 enhanced the growth of seedlings and promoted the proliferation of cultured cells in a dose-dependent manner. *Arabidopsis* PSY1 also promoted cell proliferation of asparagus mesophyll cells, indicating a conserved signaling pathway for the regulation of cell proliferation and expansion in plants.

The *Arabidopsis* genome contains two *PSKR1* paralog genes, *At5g53890* and *At1g72300*, that share 48.6 and 43.6% sequence identity. Genetic analysis using loss-of-function mutants and transgenic lines overexpressing each of these genes indicated that *At*5g53890 is also involved in PSK perception and *At*1g72300 is involved in perception of PSY1. The cumulative results suggest that PSK and PSY1 are two genetically different, but functionally redundant peptide signals that appear to activate the same signal transduction pathway through interaction with different receptors. This poses an interesting question about the evolution of sulfated peptide signals in plants. The signaling pathway downstream of PSKR1, PSKR2, and *At*1g72300, as well as the mechanism of action and interplay of PSK with other hormones in regulating plant growth and development remain to be elucidated [105].

15.3.6 Polaris

Polaris (PLS) peptide (36 amino acids) modulates cell division and longitudinal cell expansion during root growth and leaf vascular development in *Arabidopsis*

plants [106]. PLS is encoded by a small gene of about 600 base pairs, genetically identified with the use of a promoter trap strategy [107]. *PLS* promoter trap insertional mutant showed a distinct phenotype consisting of shorter roots and reduced vascularization of rosette leaves [108]. *PLS* is a single-copy gene in *Arabidopsis*, expressed in the basal polar axis of developing embryos and in the root tips of seedlings. *PLS* mRNA is also detected in young leaves, and transcription is induced by auxin and repressed by ethylene – two well-known plant growth regulators [109]. PLS does not contain a predictable signal peptide (Table 15.2) and its subcellular location is still unknown. Since the PLS peptide has not been purified from plant tissues, it is not clear if PLS is post-translationally modified or processed. The specific cellular target(s) and mode of action of PLS are still unknown. It has been suggested that PLS may promote polar transport and homeostasis of auxin to counteract the inhibitory action of ethylene on cell division/expansion at the root tips [109].

15.3.7
Inflorescence Deficient in Abscission

Abscission is the shedding of plant organs that are no longer playing an important function during the life cycle of a plant. Organ abscission can also be a protective response against harsh environmental conditions, such as cold or biotic stress. Abscission of plants organs results from the disassembly of components of the cell wall by enzymes such as cellulases (endo-β1–4-glucanases), polygalacturonases, and pectinases [110]. The final separation step(s) in the shedding of floral organs in *Arabidopsis* is controlled by the gene for inflorescence deficient in abscission (*IDA*) [111, 112]. Ectopic expression of the *IDA* gene causes the premature abscission of floral organs [113], whereas floral abscission is completely blocked in *IDA* mutants [111].

The *IDA* gene is expressed in a very distinct layer of cells defined as the abscission zone, located at the base of filaments, petals, and sepals [112]. The *IDA* gene encodes for a protein of 77 amino acids, which has a putative N-terminal secretion signal peptide. Inflorescence deficient in abscission-like (IDL) proteins have been identified in *Arabidopsis* and multiple plant species [111]. RT-PCR analysis indicated that the IDL encoding genes are expressed in different organs and tissues, suggesting that they may be involved in other environmental or developmental processes in addition to organ abscission.

All IDL proteins have a signal peptide, elevated p*I* values, and a highly conserved C-terminal domain flanked by basic amino acids (Table 15.2) [111]. IDL proteins may be processed to release putative peptide ligands from the C-terminus, similar to CLE peptide signals. A potential receptor for IDA is HAESA, a plasma membrane-associated LRR-RLK involved in the control of floral organ abscission [114]. There are four genes in *Arabidopsis* that are closely related to HAESA. Like IDA, HAESA is only expressed in the abscission zone of floral organs [112]. Mature IDL peptides have not been isolated from plant tissues yet, which is a prerequisite to study the biogenesis of IDL peptides and the potential interaction with HAESA receptors.

15.3.7.1 4-kDa Peptide

The 4-kDa peptide (4k-P, also known as leginsulin) is a 37-amino-acid peptide ligand isolated from roots of soybean seedlings [115], based on its ability to bind to a 43-kDa receptor-like glycoprotein (43k-P), a seed basic 7S globulin [116, 117]. 43k-P also binds insulin and insulin-like growth factors, although there is no sequence similarity between 4k-P and insulin or insulin-like factors [118]. However, 43k-P has partial sequence homology with the human insulin-like growth factor-binding protein, and shares some functional and structural similarities with the insulin receptor [118]. 43k-P consists of four pairs of α- and β- subunits linked by six disulfide bridges, with four glycosylation sites with asparagine-linked glycans, a putative transmembrane domain, and a conserved ATP-binding site. Like the rat insulin receptor, 43k-P has protein tyrosine kinase activity, which is stimulated upon binding of 4k-P [115, 118]. The disulfide bonds and glycosyl groups are required for 4k-P binding to 43k-P [117].

4k-P is synthesized as a larger precursor protein that contains a putative N-terminal signal peptide, followed by the 4k-P peptide, a linker peptide, and a 6-kDa peptide. A C-terminal glycine seems to be processed from mature 4k-P [115]. 4k-P has a molecular mass of 3.92 kDa, with six highly conserved cysteines that form three disulfide bonds. The peptide adopts a T-knot scaffold structure, with three β-strands stabilized by the disulfide bridges [119].

Both the mature 4k-P and the 43k-P have been localized in the cell wall and the plasma membranes of soybean cotyledons, where they can interact [120]. The 4k-P has been identified only in legumes, although 43k-P homologs are found in other plant families [116, 117]. The biological activity of 4k-P has been associated with the regulation of *in vitro* cell proliferation and callus growth [119], as well as in plant defense against insects [121]. Overexpression of the soybean 4k-P gene in carrot and bird's-foot trefoil cells enhanced cell proliferation, callus growth, and differentiation in this nonlegume species [119], suggesting the existence of similar 4k-P/43k-P signaling components in other plants.

15.4 Peptides Involved in Self-Recognition

Plants have developed self-recognition mechanisms by which genes expressed in the pollen and the stigma determine whether there is a compatible or incompatible interaction for fertilization [122]. In the Brassicaceae family of plants, multiallelic S-locus genes (with over 60 different alleles) are physically linked genes involved in self-incompatibility [123].

15.4.1 S-Locus Cysteine Rich Peptides

The S-locus cysteine-rich peptide (SCR/SP11) gene is expressed in anther tapetum cells and pollen grains, and encodes small hydrophilic, cysteine-rich polypeptides

(<10 kDa) with a putative signal peptide for secretion [124, 125]. The mature SCR proteins (50–59 amino acids) diffuse through the pollen cell wall and interact with S-locus receptor kinases (SRKs) located in the surface of epidermal (papilla) cells of the stigma [126, 127]. Upon SCR binding of SRK, a signal transduction cascade is initiated that leads to self-incompatibility. Thus, SCR and SRK act as a ligand–receptor pair during pollen self-recognition.

SCR proteins are highly variable, except for eight cysteine residues (designated C1 to C8), which are highly conserved in most SCR peptides [128]. Structural analysis by nuclear magnetic resonance and computer modeling indicates that SCRs adopt a similar three-dimensional structure despite their extensive sequence divergence [129]. The protein folds into a compact α/β-barrel structure extensively stabilized by four intramolecular disulfide bridges that link C2 to C5, C3 to C6, C4 to C7, and C1 to C8 (Table 15.2) [130].

SRK is an integral plasma membrane protein, with a ligand-binding extracellular domain and a cytoplasmic serine/threonine kinase domain at the C-terminus [131]. In the papilla cells, the cytoplasmic tail of SRK is known to interact with Armadillo repeat-containing protein 1 (ARC1) [132] and M-locus protein kinase [133]. Both proteins are required for SRK signaling, leading to the inhibition of pollen germination in the stigma. ARC1 is a U-box protein with E3 ubiquitin ligase activity, suggesting the involvement of protein degradation by the ubiquitin/proteasome system during the self-incompatibility response. In addition, two thioredoxin *h*-like proteins are thought to act as negative regulators of SRK signaling by promoting the formation of inactive SRK homodimers [134, 135].

Another S-locus glycoprotein (*SLG*) gene involved in self-incompatibility is coordinately expressed with *SRK* in the stigma epidermis, and encodes a secreted protein highly homologous to the ectodomain region of SRK [136–138]. Although the role of SLG in the self-incompatibility reaction is not clear yet, SLG appears to enhance the interaction of SCRs with their SRK receptors [139].

The switch to self-fertility is often accompanied by inactivation or structural remodeling of S-locus genes in some species of the Brassicaceae [140, 141]. For instance, *A. thaliana* is a self-compatible species, which contains only short truncated or rearranged remnants of *SRK* and *SCR* sequences. On the other hand, the self-incompatible *Arabidopsis lyrata* contains fully functional *SRK* and *SCR* alleles [142]. The evolution of self and nonself mechanisms is influenced by selective pressures, and may involve different ligand–receptor proteins in flowering plants [122]. Small cysteine-rich peptide signals, produced in the stigma, and LRR-LRKs specifically expressed in pollen tubes, seem to play a role in pollen self-compatibility in tomato [143].

15.5
Methods in Plant Regulatory Peptide Research

Peptide signaling in plants is a relatively new field, initiated by the discovery of the defense peptide systemin in 1991 [1]. Since then, new peptide signals have been

found that regulate growth, development, and reproduction [99, 144, 145]. In this section, we will summarize the events leading to the discovery of systemin along with the technical aspects of the isolation of peptides from plants.

15.5.1 Discovery of Systemin

In 1972, Green and Ryan described a new phenomenon in plant science. When a leaf on a tomato plant was wounded either mechanically or by insects, defense PI proteins were produced, not only around the wound site, but in other leaves of the plant [146]. This led to the hypothesis that a chemical (wound) signal traveled through the plant to cause systemic production of the PIs. A search for this mobile signal, termed proteinase inhibitor-inducing factor (PIIF), was initiated that would eventually lead to the discovery of the first peptide hormone, systemin.

Initially, cell wall components were thought to be the most likely candidates for PIIF, as polysaccharides from fungal cell walls were capable of inducing a defense response in plants. Oligosaccharides from plant cell walls might then be released from wound sites and transported via the phloem as signals for the production of PIs. Thus, work on the wound response centered on the isolation of an abundant carbohydrate component of cell walls (i.e., pectin molecules of diverse size and linkages). Organic solvent extraction, anion-exchange chromatography, and gel-filtration chromatography were utilized to fractionate the pectic fragments. These fractions were then tested for their abilities to induce inhibitors using an excised tomato plant assay as follows. Young (2-week-old) plants were excised with a sharp razor blade and supplied with solutions containing the carbohydrates through their cut stems for 30 min. The excised plants were placed in vials containing water in a closed plexiglass container and allowed to incubate for 24 h in a growth chamber. Thereafter, leaf juice was expressed with a mortar and pestle and placed in wells on an immunodiffusion plate containing antibody to the inhibitor [147]. After 24 h incubation, the plate was developed and the amount of inhibitor produced by the fractions was compared to buffer supplied controls. Polygalacturonic acid oligomers with degrees of polymerization ranging from 2 to 20 were isolated and found to be very good inducers of PIs [148]. These pectic fragments were good candidates for mobile signals because they induced high levels of inhibitor, were extremely abundant, and they were present at the wound site where they could be released by pectic enzymes. However, radiolabeled pectic fragments showed no ability to move when placed on the wound site of a tomato leaf [149]. Thus, pectic fragments appear to play a role in localized defense responses only.

While pectic fragments are acidic and will strongly bind to anion-exchange resin, an inhibitor inducing activity was also found in the nonbinding material. In sharp contrast to the pectic fragments, which induce inhibitors when supplied at 0.5 mg/ml, this inducing compound was extremely potent, active at levels that could not be weighed.

After the crude extraction, four HPLC steps were required to isolate the inhibitor inducing substance:

(i) Preparative reverse-phase C_{18} HPLC in 0.1% trifluoroacetic acid (TFA) with an acetonitrile gradient.
(ii) Strong cation-exchange HPLC in potassium phosphate, pH 3 with a potassium chloride gradient.
(iii) Analytical C_{18} HPLC buffered at pH 6 with potassium phosphate using an acetonitrile gradient.
(iv) Analytical C_{18} HPLC in 0.1% TFA using a methanol gradient.

For each HPLC purification step, the fractions were checked for biological activity with the excised plant assay and pooled fractions were applied to the next column. Owing to the inherent variability of using whole plants, activity of each fraction was checked with a minimum of four plants and repeated 3 times. With a 3-day turnaround required for the assay, the research was tedious and time consuming.

A peptide was first suspected as the bioactive compound when a loss of activity was observed upon treatment with proteases. Sequence analysis revealed an 18-amino-acid peptide that was subsequently synthesized with full biological activity and the peptide was termed "systemin" [1] (see Section 15.2.1). The peptide was capable of inducing a response when as little as 40 fmol was supplied to a tomato plant. It was estimated that a 60-lb (27-kg) preparation of tomato leaves yielded about 1 μg of the systemin peptide. After purification and sequencing of systemin, degenerated primers were designed to isolate the cDNA encoding the systemin precursor protein [19].

15.5.2
Identification of Novel Peptide Signals using the Cell Alkalinization Assay

Upon isolation of tomato systemin, similar bioactive peaks were found in tobacco plants, another member of the Solanaceae family, which also possesses wound-inducible inhibitors. Partial sequence analysis was achieved, but the tobacco plant bioassay was much more difficult, as there was more variability among the plants and the antiserum for the immunodiffusion assay was weak. A paper from the laboratory of Thomas Boller [150] had concurrently shown that systemin had the ability to interact with *S. peruvianum* suspension cells to give a rapid increase in the extracellular pH of the media. Tobacco (*N. tabacum*) cells were cultured and a strong alkalinizing effect was observed with the tobacco bioactive fractions. A bioassay based on this alkalinization response was developed as shown in Figure 15.2. Owing to the uniformity of the suspension cells and the decreased time to obtain results, the tobacco peptides were quickly purified and characterized as 18-amino-acid glycopeptides containing hydroxyproline (see Section 15.2.2) [31].

Utilizing the same extraction protocol described above and the alkalinization assay, bioactive peaks were observed in a number of solanaceous species utilizing both tobacco and tomato suspension cells (Figure 15.3). HypSys peptides were purified from tomato leaves (*S. lycopersicum*) [11]; and after suspension cells were produced

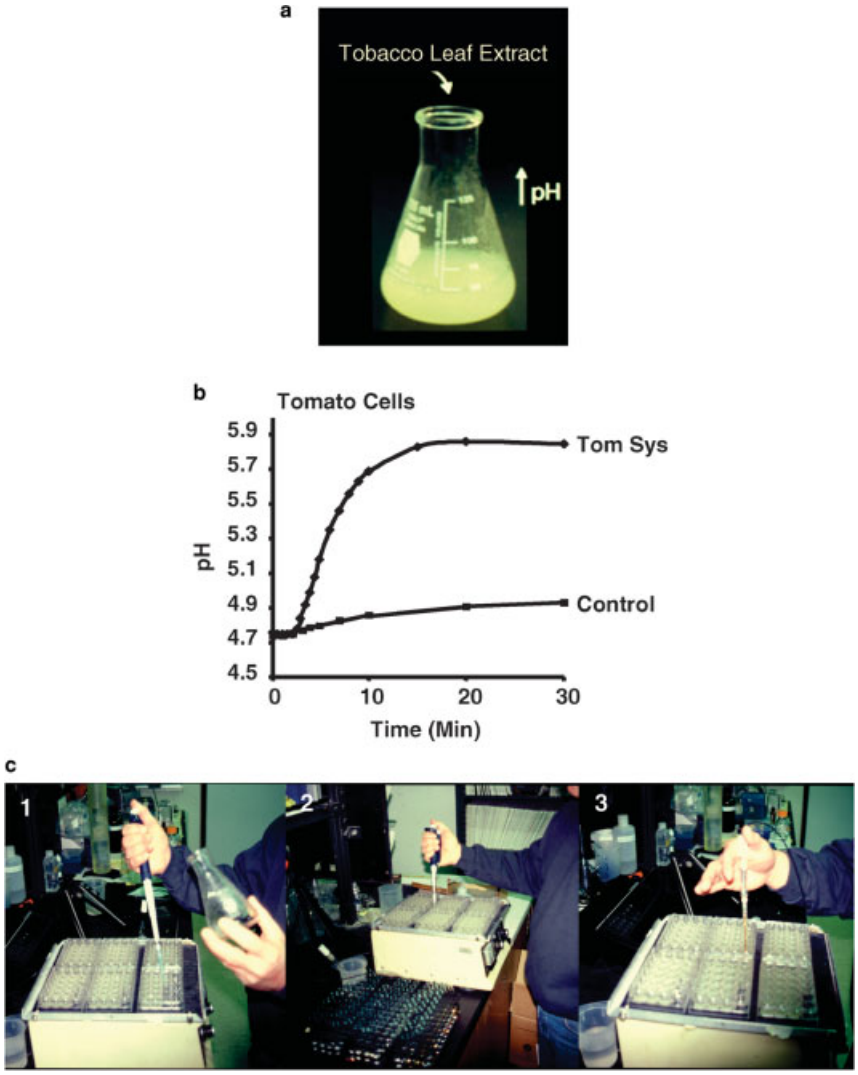


Figure 15.2 Alkalinization bioassay. (a) Suspension cells of the desired species are grown in 125-ml flasks (45 ml) on an orbital shaker at 160 rpm. Every 7 days, 2–5 ml of cells are transferred to fresh media. When the cells are growing in log phase, usually 3–6 day, they can be utilized in the alkalinization assay. (b) As an example of how suspension cells respond to a bioactive peptide, tomato systemin (1 nM) or water were added to 1-ml aliquots of tomato suspension cells and the pH was checked at various time intervals. (c) For the analysis of tobacco bioactive peptides, (1) aliquots of tobacco suspension cells (1 ml) are added to 24-well culture cluster plates and shaken at 160 rpm for 1 h; (2) small aliquots (1–10 µl) of extracted tobacco peptide HPLC fractions are added to the wells; (3) after 20 min, the pH of the cell media is measured and recorded [30].

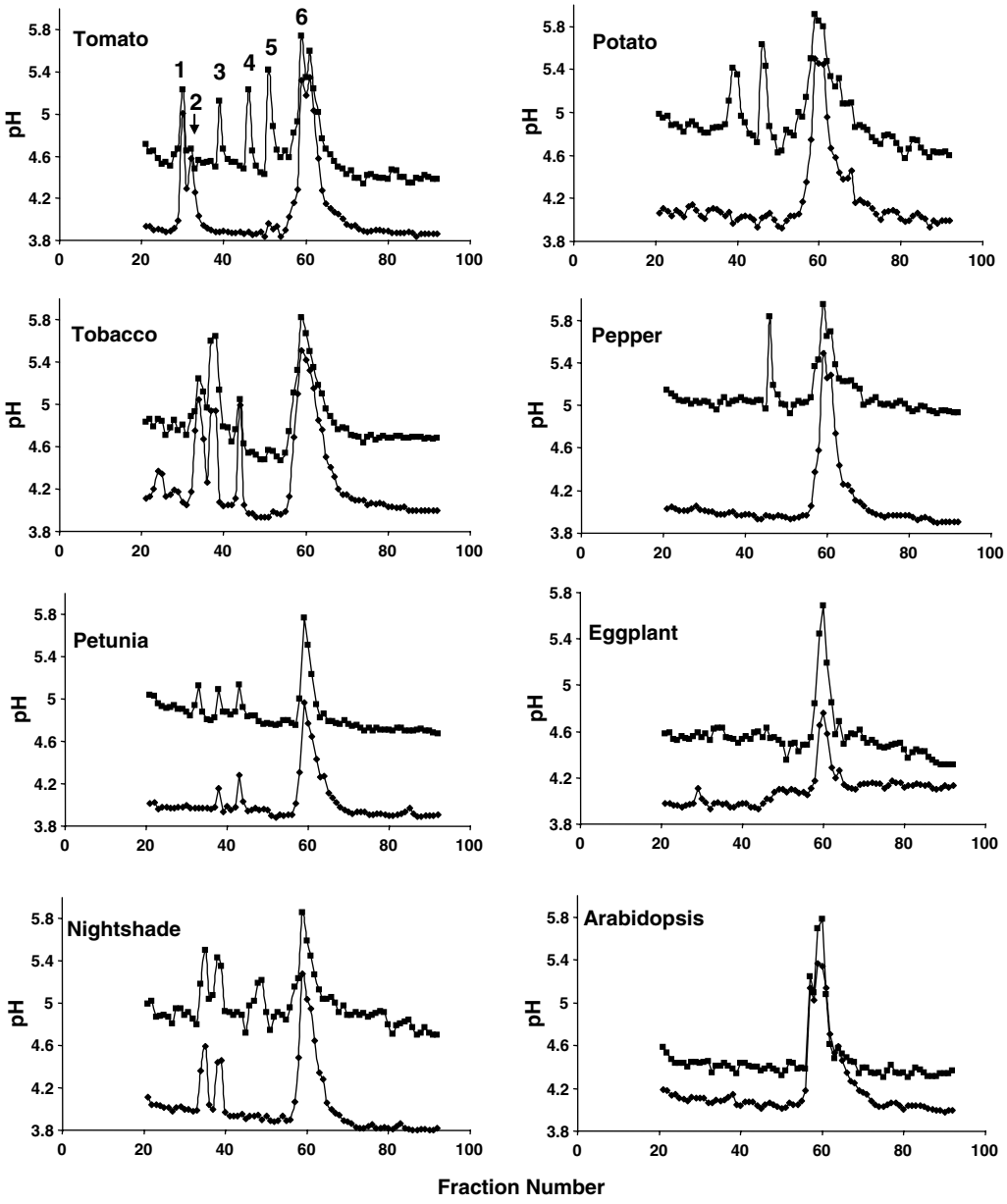


Figure 15.3 Alkalinization activities of HPLC fractions from different Solanaceae species. In each case 25 mg of crude extract was applied to the C_{18} reverse-phase column and fractions were eluted with a 0–40% acetonitrile/0.1% TFA gradient over 90 min. The flow rate was 2 ml/min and 1-min fractions were collected. Aliquots of 10 µl were added to either 1 ml of tomato or tobacco cells and the pH was recorded after 30 min. The squares denote the tomato cell assay and the diamonds denote the tobacco cell assay. In tomato (left upper panel), peak #1 corresponds to HypSys 1 and there is a shoulder on the right of this peak that contains

from petunia (*P. hybrida*) and nightshade (*S. nigrum*), HypSys peptides were also purified from these species.

In addition to HypSys activity peaks, other peptide signals have been isolated using the alkalinization assay. Alkalinization profiles from HPLC fractions of plant extracts revealed an extremely large, late eluting peak with a very rapid response time (Figure 15.3). This activity peak was purified from tobacco and termed RALF, which was involved in root and root hair growth and development (see Section 15.3.2) [78].

The *Arabidopsis* extract shown in Figure 15.3 contained only the RALF peak when assayed using either tomato or tobacco suspension cells. When *Arabidopsis* suspension cells were used, an early eluting alkalinizing peak was found and was termed *At*Pep1 – a 23-amino-acid peptide involved in pathogen defense (see Section 15.2.3) [37].

The isolation of the defense-related peptides and RALF has been greatly facilitated by a bioassay that is quick and reliable, and by applicable HPLC methods based on the characteristics of the peptides. In all of these cases, the peptides have been basic in charge and strong cation-exchange HPLC has been a key step in isolation. For HypSys peptides, the carbohydrate moieties have added increased hydrophilicity that has facilitated separations from noncarbohydrate containing peptides. A change in elution solvent from acetonitrile to methanol caused the HypSys peptides to be retained longer than the remaining contaminating peptides. By utilizing the unique properties of the peptides, a separation scheme could be developed and the desired bioactive peptide purified. Also, new assay methods that exploit cellular processes other than the ATP-dependent proton pump inhibition that causes alkalinization could yield another set of peptides with completely different properties.

15.5.3
Isolation of Tyrosine-Sulfated Peptides

The tyrosine-sulfated peptides have been isolated by utilizing their unique characteristics (see Section 15.3.5). These peptides are found in suspension cell media, which is a rich source for secreted peptides and proteins. Sulfate ions have a higher charge and a smaller ion radius, and are highly retained on anion-exchange resin relative to nonsulfated peptides [151]. In addition, the sulfated tyrosines are usually surrounded by acidic amino acids and this characteristic would further increase the retention of the peptides. An enriched fraction of sulfated peptides was obtained by applying crude suspension cell media to an anion-exchange column at low salt concentrations and eluting the strongly retained fraction with high salt concentrations, followed by MS analysis. This technique was put to practical use in the isolation

HypSys 2 (#2). The next two peaks (#3 and #4) are isoforms of systemin peaks. Peak #3 is a methionine oxidized isoform. The next peak (#5) is HypSys 3 and the last larger peak (#6) is the RALF peak. HypSys 1, 2 and 3 were named by their retention times. After further purification by HPLC, peptide sequences and masses were determined by N-terminal sequencing and MS/MS [1, 11, 31, 32]. An *Arabidopsis* extract was added for comparison.

of PSY1 – an 18-amino-acid tyrosine-sulfated glycopeptide that promotes cellular proliferation and expansion (see Section 15.3.5.1) [151].

15.5.4
Use of Peptidomics

To search for unknown potentially bioactive peptides that may not have a unique structural feature poses a problem for the biochemist and requires genomic information. Although the genes give no information relating to the processing of the peptides into their mature forms or on their PTMs, a combination of *in silico* gene screening coupled with enrichment of secreted peptides and LC-MS analysis has led to the identification of a new peptide plant hormone – CEP1 (see Section 15.3.4) [5]. This was a good example of the use of peptidomics for the detection of novel plant bioactive peptides [4, 5, 145].

First, Ohyama *et al*. [5] grew *Arabidopsis* plants fully submerged in liquid medium. The authors predicted that small peptides would be secreted into the media due to the intercellular spaces being filled with water. To show that peptides were secreted, transgenic plants overexpressing the *CLE44* gene were sown into the liquid media. The peptide fraction was enriched by *o*-chlorophenol extraction followed by acetone precipitation. Comparison of the *CLE44* fraction with a wild-type fraction on HPLC produced two unique peaks that were subsequently analyzed by MS/MS, revealing the mature CLE peptide and a truncated form, both complete with hydroxyproline PTMs (see Section 15.3.1.2).

Having a method developed to search for secreted peptides, Ohyama *et al*. [5] looked for potential genes that had similar characteristics to other known gene families that code for secreted peptides. By narrowing their search to only genes containing signal peptides, small coding regions that were cysteine-poor, and by eliminating the precursors of the known peptide signals, a small group of 110 functionally unknown genes was produced. A novel five gene family with similarities at the C-terminal was identified. One of the genes was overexpressed and the HPLC profile peaks obtained from peptide extracts collected from the submerged culture of the transgenic plants was compared to wild-type plant cultures. Four unique peaks observed in the extracts of whole transgenic plant cultures were subsequently analyzed by MS/MS. One peak contained a 15-amino-acid peptide with two hydroxyproline modifications and the other peaks differed only by a single amino acid truncation at the C-terminal and a proline to hydroxyproline change at position 3 from the N-terminal. The new peptide was found to arrest root growth and was named CEP1 (see Section 15.3.4).

15.5.5
Fishing Ligands with Bait Receptors

PSK and its receptor are one of the few plant receptor–ligand pairs characterized to date. However, in the *Arabidopsis* genome there are as many as 600 RLKs and ligands have been found for only a few. A potential method for ligand fishing has been

reported in which the PSK LRR receptor-binding domain was fused to a HaloTag fusion protein and immobilized onto microbeads of HaloLink resin [152]. Experiments with [^{3}H]PSK demonstrated that the binding domain was functional. The receptor-fusion protein bead was incubated with conditioned medium from *Arabidopsis* cell culture, washed, and eluted with high salt buffer. The eluent was analyzed by LC-MS\MS and revealed that PSK had been purified from the medium. This method may be used in the future as an alternative to a genetic or a bioassay-based biochemical approach for the discovery of peptide signals.

In summary, extraction of peptide signals was a daunting task when systemin was first purified, but with the development of the alkalinization assay for rapid detection of bioactive peptides, the advances in HPLC purification techniques, and the development of MS analyses over the last decade, together with the use of bioinformatics, the discovery of new peptide signals should be forthcoming.

15.6 Conclusions

Peptide signals have been recognized as important regulatory molecules involved in plant growth and development, reproduction, and defense. Over a dozen families of plant peptide signals have been identified so far (Table 15.1). Like peptide hormones in other eukaryotes, plant regulatory peptides are extracellular signals derived by proteolytic processing, are active at low concentrations, and activate intracellular signal transduction events through their interaction with membrane-bound receptors. Some plant peptide families include numerous members, with little sequence homology between their protein precursors. However, the bioactive peptides show a high degree of conservation and often are functionally redundant. This suggests a common origin for large peptide signal families through gene duplication and elongation events. In some cases, the cell/tissue-specific and temporal regulation of peptide and receptor pairs may determine their functionality.

The regulation of synthesis, processing, storage, and secretion of peptide signals in plants is still poorly understood. The processing sites that release the active peptides are not known nor is the nature of the peptidases involved. Most regulatory peptides found in animals are synthesized as larger precursors and processed to form biologically active products, stored and secreted through the secretory pathway [153]. However, a number of extracellular peptides of significant biomedical relevance for human health are being found exported by nonclassical ER/Golgi-independent mechanisms, which are still not fully understood [154]. Several plant peptide signals are also synthesized from precursors lacking a signal peptide sequence (Tables 15.1 and 15.2), suggesting that they are also processed and transported through nonclassical mechanisms to the extracellular space.

Further studies on the biogenesis of peptide hormones, and the discovery of new peptide signals and their receptors in plants, will shed some light on the evolution of peptide signaling systems in eukaryotes. The application of bioinformatics tools together with new protocols for the extraction and purification of peptide signals

using modern LC-MS/MS techniques promise a great future for the identification of novel peptides, which may be the ligands for the many "orphan receptors" found in plants.

Acknowledgments

Dedicated to the memory of Clarence (Bud) A. Ryan, Charlotte Y. Martin Professor of Biochemistry and a Fellow of the Institute of Biological Chemistry, Washington State University (1931–2007).

References

1 Pearce, G., Strydom, D., Johnson, S. and Ryan, C.A. (1991) *Science*, **253**, 895–898.
2 Kende, H. and Zeevaart, J.A.D. (1997) *The Plant Cell*, **9**, 1197–1210.
3 Hedden, P. (2001) *Trends in Plant Science*, **6**, 498–501.
4 Oelkers, K., Goffard, N., Weiller, G.F., Gresshoff, P.M., Mathesius, U. and Frickey, T. (2008) *BMC Plant Biology*, **8**, 1–15.
5 Ohyama, K., Ogawa, M. and Matsubayashi, Y. (2008) *The Plant Journal*, **55**, 152–160.
6 Norman, G. (1997) *Hormones*, 2nd edn, Academic Press, San Diego, CA.
7 Narváez-Vásquez, J. and Orozco-Cárdenas, M.L. (2008) In *Induced Plant Resistance to Herbivory* (ed. A. Schaller), Springer, Berlin, pp. 313–328.
8 Wasternack, C., Stenzel, I., Hause, B., Hause, G., Kutter, C., Maucher, H., Neumerkel, J., Feussner, I. and Miersch, O. (2006) *Journal of Plant Physiology*, **163**, 297–306.
9 Bergey, D.R., Howe, G.A. and Ryan, C.A. (1996) *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 12053–12058.
10 Ryan, C.A., Huffaker, A. and Yamaguchi, Y. (2007) *Cellular Microbiology*, **9**, 1902–1908.
11 Pearce, G. and Ryan, C.A. (2003) *The Journal of Biological Chemistry*, **278**, 30044–30050.
12 Orozco-Cárdenas, M., McGurl, B. and Ryan, C.A. (1993) *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 8273–8276.
13 Corrado, G., Sasso, R., Pasquariello, M., Iodice, L., Carretta, A., Cascone, P., Ariati, L., Digilio, M.C., Guerrieri, E. and Rao, R. (2007) *Journal of Chemical Ecology*, **33**, 669–681.
14 Toumadje, A. and Johnson, W.C. (1995) *Journal of the American Chemical Society*, **117**, 7023–7024.
15 Vanhoof, G., Goossens, F., De Meester, I., Hendriks, D. and Scharpe, S. (1995) *The FASEB Journal*, **9**, 736–744.
16 Reiersen, H. and Rees, A.R. (2001) *Trends in Biochemical Sciences*, **26**, 679–684.
17 Pearce, G., Johnson, S. and Ryan, C.A. (1993) *The Journal of Biological Chemistry*, **268**, 212–216.
18 Meindl, T., Boller, T. and Felix, G. (1998) *The Plant Cell*, **10**, 1561–1570.
19 McGurl, B., Pearce, G., Orozco-Cárdenas, M. and Ryan, C.A. (1992) *Science*, **255**, 1570–1573.
20 Narváez-Vásquez, J. and Ryan, C.A. (2004) *Planta*, **218**, 360–369.
21 Seidah, N.G. and Chretien, M. (1997) *Current Opinion in Biotechnology*, **8**, 602–607.
22 Schaller, A. and Ryan, C.A. (1994) *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 11802–11806.

23 Schaller, A. (1998) *Phytochemistry*, **47**, 605–612.

24 McGurl, B., Orozco-Cárdenas, M., Pearce, G. and Ryan, C.A. (1994) *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 9799–9802.

25 Scheer, J.M. and Ryan, C.A. (1999) *The Plant Cell*, **11**, 1525–1535.

26 Scheer, J.M. and Ryan, A. (2002) *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 9585–9590.

27 Montoya, T., Nomura, T., Farrar, K., Kaneta, T., Yokota, T. and Bishop, G.J. (2002) *The Plant Cell*, **14**, 3163–3176.

28 Lanfermeijer, F.C., Staal, M., Malinowski, R., Stratmann, J.W. and Elzenga, J.T.M. (2008) *Plant Physiology*, **146**, 129–139.

29 Holton, N., Caño-Delgado, A., Harrison, K., Montoya, T. and Chory, J. (2007) *The Plant Cell*, **19**, 1709–1717.

30 Pearce, G., Moura, D.S., Stratmann, J. and Ryan, C.A. (2001) *Nature*, **411**, 817–820.

31 Pearce, G., Siems, W.F., Bhattacharya, R., Chen, Y.-C. and Ryan, C.A. (2007) *The Journal of Biological Chemistry*, **282**, 17777–17784.

32 Pearce, G., Bhattacharya, R., Chen, Y.-C. Barona, G., Yamaguchi, Y. and Ryan, C.A. (2009) *Plant Physiology*, Published online April 29, 2009; DOI:10.1104/pp.109.138669.

33 Chen, Y.-C., Siems, W.F., Pearce, G. and Ryan, C.A. (2008) *The Journal of Biological Chemistry*, **283**, **17**, 11469–11476.

34 Ren, F. and Lu, Y.-T. (2006) *Plant Science*, **171**, 286–292.

35 Narváez-Vásquez, J., Pearce, G. and Ryan, C.A. (2005) *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 12974–12977.

36 Narváez-Vásquez, J., Orozco-Cárdenas, M.L. and Ryan, C.A. (2007) *Plant Molecular Biology*, **65**, 711–718.

37 Huffaker, A., Pearce, G. and Ryan, C.A. (2006) *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 10098–10103.

38 Felix, G., Duran, J.D., Volko, S. and Boller, T. (1999) *The Plant Journal*, **18**, 265–276.

39 Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T. and Felix, G. (2004) *The Plant Cell*, **16**, 3496–3507.

40 Huffaker, A. and Ryan, C.A. (2007) *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 10732–10736.

41 Yamaguchi, Y., Pearce, G. and Ryan, C.A. (2006) *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 10104–10109.

42 Yamaguchi, Y. and Ryan, C.A. (2008) ASPB Annual Meeting, Poster 26010.

43 Fiers, M., Ku, K.L. and Liu, C.-M. (2007) *Current Opinion in Plant Biology*, **10**, 39–43.

44 Jun, J.H., Fiume, E. and Fletcher, J.C. (2008) *Cell and Molecular Science*, **65**, 743–755.

45 Ni, J. and Clark, S.E. (2006) *Plant Physiology*, **140**, 726–733.

46 Ito, Y., Nakanomyo, I., Motose, H., Iwamoto, K., Sawa, S., Dohmae, N. and Fukuda, H. (2006) *Science*, **313**, 842–845.

47 Clark, S.E., Running, M.P. and Meyerowitz, E.M. (1995) *Development*, **121**, 2057–2067.

48 Kondo, T., Sawa, S., Kinoshita, A., Mizuno, S., Kakimoto, T., Fukuda, H. and Sakagami, Y. (2006) *Science*, **313**, 845–848.

49 Fletcher, J.C., Brand, U., Running, M.P., Simon, R. and Meyerowitz, E.M. (1999) *Science*, **283**, 1911–1914.

50 Rojo, E., Sharma, V.K., Kovaleva, V., Raikhel, N.V. and Fletcher, J.C. (2002) *The Plant Cell*, **14**, 969–977.

51 Clark, S.E., Williams, R.W. and Meyerowitz, E.M. (1997) *Cell*, **89**, 575–585.

52 Ogawa, M., Shinohara, H., Sakagami, Y. and Matsubayashi, Y. (2008) *Science*, **319**, 294.

53 Kayes, J.M. and Clark, S.E. (1998) *Development*, **125**, 3843–3851.

54 Muller, R., Bleckmann, A. and Simon, R. (2008) *The Plant Cell*, **20**, 934–936.

55 Clark, S.E., Running, M.P. and Meyerowitz, E.M. (1993) *Development*, **119**, 397–418.

56 Fiers, M., Golemiec, E., Xu, J., van der Geest, L., Heidstra, R., Stiekema, W. and Liu, C.-M. (2005) *The Plant Cell*, **17**, 2542–2553.

57 Williams, R.W., Wilson, J.M. and Meyerowitz, E.M. (1997) *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 10467–10472.

58 Stone, J.M., Trotochaud, A.E., Walker, J.C. and Clark, S.E. (1998) *Plant Physiology*, **117**, 1217–1225.

59 Yu, L.P., Simon, E.J., Trotochaud, A.E. and Clark, S.E. (2000) *Development*, **127**, 1661–1670.

60 Yu, L.P., Miller, A.K. and Clark, S.E. (2003) *Current Biology*, **13**, 179–188.

61 Song, S.-K., Lee, M.M. and Clark, S.E. (2006) *Development*, **133**, 4691–4698.

62 Trotochaud, A.E., Hao, T., Wu, G., Yang, Z. and Clark, S.E. (1999) *The Plant Cell*, **11**, 393–405.

63 Ishiguro, S., Watanabe, Y., Ito, N., Nonaka, H., Takeda, N., Sakai, T., Kanaya, H. and Okada, K. (2002) *The EMBO Journal*, **21**, 898–908.

64 Casamitjana-Martinez, E., Hofhuis, H.F., Xu, J., Liu, C.-M., Heidstra, R. and Scheres, B. (2003) *Current Biology*, **13**, 1435–1441.

65 Mayer, K.F.X., Schoof, H., Haecker, A., Lenhard, M., Jurgens, G. and Laux, T. (1998) *Cell*, **95**, 805–815.

66 Brand, U., Fletcher, J.C., Hobe, M., Meyerowitz, E.M. and Simon, R. (2000) *Science*, **289**, 617–619.

67 Schoof, H., Lenhard, M., Haecker, A., Mayer, K.F.X., Jurgens, G. and Laux, T. (2000) *Cell*, **100**, 635–644.

68 Lenhard, M. and Laux, T. (2003) *Development*, **130**, 3163–3173.

69 Mitchum, M.G., Wang, X. and Davis, E.L. (2008) *Current Opinion in Plant Biology*, **11**, 75–81.

70 Chu, H., Qian, Q., Liang, W., Yin, C., Tan, H., Yao, X., Yuan, Z., Yang, J., Huang, H., Luo, D., Ma, H. and Zhang, D. (2006) *Plant Physiology*, **142**, 1039–1052.

71 Suzaki, T., Toriba, T., Fujimoto, M., Tsutsumi, N., Kitano, H. and Hirano, H.-Y. (2006) *Plant & Cell Physiology*, **47**, 1591–1602.

72 Opsahl-Ferstad, H.-G., Le Deunff, E., Dumas, C. and Rogowskiy, P.M. (1997) *The Plant Journal*, **12**, 235–246.

73 Strabala, T.J., O'Donnell, P.J., Smit, A.-M., Ampomah-Dwamena, C., Martin, E.J., Netzler, N., Nieuwenhuizen, N.J., Quinn, B.D., Foote, H.C.C. and Hudson, K.R. (2006) *Plant Physiology*, **140**, 1331–1344.

74 Kinoshita, A., Nakamura, Y., Sasaki, E., Kyozuka, J., Fukuda, H. and Sawa, S. (2007) *Plant & Cell Physiology*, **48**, 1821–1825.

75 Cock, J.M. and McCormick, S. (2001) *Plant Physiology*, **126**, 939–942.

76 Wang, X., Mitchum, M.G., Gao, B., Li, C., Diab, H., Baum, T.J., Hussey, R.S. and Davis, E.L. (2005) *Molecular Plant Pathology*, **6**, 187–191.

77 Olsen, A.N. and Skriver, K. (2003) *Trends in Plant Science*, **8**, 55–57.

78 Pearce, G., Moura, D.S., Stratmann, J. and Ryan, C.A. (2001) *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 12843–12847.

79 Haruta, M. and Constabel, C.P. (2003) *Plant Physiology*, **131**, 814–823.

80 Wu, J., Kurten, E.L., Monshausen, G., Hummel, G.M., Gilroy, S. and Baldwin, I.T. (2007) *The Plant Journal*, **52**, 877–890.

81 Scheer, J.M., Pearce, G. and Ryan, C.A. (2005) *Planta*, **221**, 667–674.

82 Olsen, A.N., Mundy, J. and Skriver, K. (2002) *In Silico Biology*, **2**, 441–451.

83 Escobar, N.M., Haupt, S., Thow, G., Boevink, P., Chapman, S. and Oparka, K. (2003) *The Plant Cell*, **15**, 1507–1523.

84 Narita, N.N., Moore, S., Horiguchi, G., Kubo, M., Demura, T., Fukuda, H., Goodrich, J. and Tsukaya, H. (2004) *The Plant Journal*, **38**, 699–713.

85 Wen, J., Lease, K.A. and Walker, J.C. (2004) *The Plant Journal*, **37**, 668–677.

86 Wen, J.Q. and Walker, J. (2006) In *Handbook of Biologically Active Peptides*, (ed. A.J. Kastin), Elsevier, New York, pp. 17–41.

87 Matsubayashi, Y. and Sakagami, Y. (1996) *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 7623–7627.

88 Matsubayashi, Y. and Sakagami, Y. (1998) *Plant Cell Reports*, **17**, 368–372.

89 Matsubayashi, Y. and Sakagami, Y. (1999) *European Journal of Biochemistry*, **262**, 666–671.

90 Kobayashi, T., Eun, C.H., Hanai, H., Matsubayashi, Y., Sakagami, Y. and Kamada, H. (1999) *Journal of Experimental Botany*, **50**, 1123–1128.

91 Yamakawa, S., Sakuta, C., Matsubayashi, Y., Sakagami, Y., Kamada, H. and Satoh, S. (1998) *Journal of Plant Research*, **111**, 453–458.

92 Chen, Y.-F., Matsubayashi, Y. and Sakagami, Y. (2000) *Planta*, **211**, 752–755.

93 Yang, H., Matsubayashi, Y., Nakamura, K. and Sakagami, Y. (2000) *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 6237.

94 Hanai, H., Nakayama, D., Yang, H., Matsubayashi, Y., Hirota, Y. and Sakagami, Y. (2000) *FEBS Letters*, **470**, 97–101.

95 Steiner, D.F. (1998) *Current Opinion in Chemical Biology*, **2**, 31–39.

96 Srivastava, R., Liu, J.-X. and Howell, S.H. (2008) *The Plant Journal*, **56**, 219–227.

97 Yang, H., Matsubayashi, Y., Nakamura, K. and Sakagami, Y. (2001) *Plant Physiology*, **127**, 842–851.

98 Yang, H., Matsubayashi, Y., Hanai, H. and Sakagami, Y. (2000) *Plant & Cell Physiology*, **41**, 825–830.

99 Matsubayashi, Y. and Sakagami, Y. (2006) *Annual Review of Plant Biology*, **57**, 649–674.

100 Matsubayashi, Y. and Sakagami, Y. (1999) *European Journal of Biochemistry*, **262**, 666–671.

101 Matsubayashi, Y. and Sakagami, Y. (2000) *The Journal of Biological Chemistry*, **275**, 15520–15525.

102 Matsubayashi, Y., Ogawa, M., Morita, A. and Sakagami, Y. (2002) *Science*, **296**, 1470–1472.

103 Matsubayashi, Y., Ogawa, M., Kihara, H., Niwa, M. and Sakagami, Y. (2006) *Plant Physiology*, **142**, 45–53.

104 Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M. and Matsubayashi, Y. (2007) *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 18333–18338.

105 Eun, C.-H., Ko, S.-M., Matsubayashi, Y., Sakagami, Y. and Kamada, H. (2003) *Plant Physiology and Biochemistry*, **41**, 447–452.

106 Casson, S.A., Chilley, P.M., Topping, J.F., Evans, I.M., Souter, M.A. and Lindsey, K. (2002) *The Plant Cell*, **14**, 1705–1721.

107 Topping, J.F., Wei, W. and Lindsey, K. (1991) *Development*, **112**, 1009–1019.

108 Lindsey, K., Casson, S.A. and Chilley, P.M. (2006) In *Handbook of Biologically Active Peptides*, (ed. A.J. Kastin), Elsevier, New York, pp. 23–27.

109 Chilley, P.M., Casson, S.A., Tarkowski, P., Hawkins, N., Wang, K.L.-C., Hussey, P.J., Beale, M., Ecker, J.R., Sandberg, G.K. and Lindsey, K. (2006) *The Plant Cell*, **18**, 3058–3072.

110 Patterson, S.E. (2001) *Plant Physiology*, **126**, 494–500.

111 Butenko, M.A., Patterson, S.E., Grini, P.E., Stenvik, G.-E., Amundsen, S.S., Mandal, A. and Aalen, R.B. (2003) *The Plant Cell*, **15**, 2296–2307.

112 Butenko, M.A., Stenvik, G.-E., Alm, V., Patterson, S.E., Saether, B. and Aalen, R.B. (2006) *Journal of Experimental Botany*, **57**, 3627–3637.

113 Stenvik, G.-E., Butenko, M.A., Urbanowicz, B.R., Rose, J.K.C. and Aalen, R.B. (2006) *The Plant Cell*, **18**, 1467–1476.

114 Jinn, T.-L., Stone, J.M. and Walker, J.C. (2000) *Genes and Development*, **14**, 108–117.

115 Watanabe, M., Takasaki, T., Toriyama, K., Yamakawa, S., Isogai, A., Suzuki, A. and Hinata, K. (1994) *Plant & Cell Physiology*, **35**, 1221–1229.

116 Komatsu, S. and Hirano, H. (1991) *FEBS Letters*, **294**, 210–212.

117 Hanada, K. and Hirano, H. (2004) *Biochemistry*, **43**, 12105–12112.

118 Komatsu, S., Koshio, O. and Hirano, H. (1994) *Bioscience, Biotechnology, and Biochemistry*, **58**, 1705–1706.

119 Yamazaki, T., Takaoka, M., Katoh, E., Hanada, K., Sakita, M., Sakata, K., Nishiuchi, Y. and Hirano, H. (2003) *European Journal of Biochemistry*, **270**, 1269–1276.

120 Nishizawa, N.K., Mori, S., Watanabe, Y. and Hirano, H. (1994) *Plant & Cell Physiology*, **35**, 1079–1085.

121 Gressent, F., Rahioui, I. and Rahbe, Y. (2003) *European Journal of Biochemistry*, **270**, 2429–2435.

122 Rea, A.C. and Nasrallah, J.B. (2008) *The International Journal of Developmental Biology*, **52**, 627–636.

123 Nou, S., Watanabe, M., Isogai, A. and Hinata, K. (1993) *Sexual Plant Reproduction*, **6**, 79–86.

124 Schopfer, C.R., Nasrallah, M.E. and Nasrallah, J.B. (1999) *Science*, **286**, 1697–1700.

125 Matsubayashi, Y., Yang, H. and Sakagami, Y. (2001) *Trends in Plant Science*, **6**, 573–577.

126 Stein, J.C., Howlett, B., Boyes, D.C., Nasrallah, M.E. and Nasrallah, J.B. (1991) *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 8816–8820.

127 Kachroo, A., Schopfer, C.R., Nasrallah, M.E. and Nasrallah, J.B. (2001) *Science*, **293**, 1824–1826.

128 Chookajorn, T., Kachroo, A., Ripoll, D.R., Clark, A.G. and Nasrallah, J.B. (2004) *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 911–917.

129 Naithani, S., Ripoll, D. and Nasrallah, J.B. (2006) In *Handbook of Biologically Active Peptides*, (ed. A.J. Kastin), Elsevier, New York, pp. 41–47.

130 Mishima, M., Takayama, S., Sasaki, K., Jee, J.G., Kojima, C., Isogai, A. and Shirakawa, M. (2003) *The Journal of Biological Chemistry*, **278**, 36389–36395.

131 Stein, J.C., Dixit, R., Nasrallah, J.B. and Nasrallah, M.E. (1996) *The Plant Cell*, **8**, 429–445.

132 Stone, S.L., Anderson, E.M., Mullen, R.T. and Goring, D.R. (2003) *The Plant Cell*, **15**, 885–898.

133 Kakita, M., Murase, K., Iwano, M., Matsumoto, T., Watanabe, M., Shiba, H., Isogai, A. and Takayama, S. (2007) *The Plant Cell*, **19**, 3961–3973.

134 Cabrillac, D., Cock, J.M., Dumas, C. and Gaude, T. (2001) *Nature*, **410**, 220–223.

135 Naithani, S., Chookajorn, T., Ripoll, D.R. and Nasrallah, J.B. (2007) *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 12211–12216.

136 Nasrallah, J.B., Kao, T.H., Chen, C.H., Goldberg, M.L. and Nasrallah, M.E. (1987) *Nature*, **326**, 617–619.

137 Takayama, S., Shimosato, H., Shiba, H., Funato, M., Che, F.-S., Watanabe, M., Iwano, M. and Isogai, A. (2001) *Nature*, **413**, 534–538.

138 Kandasamy, M.K., Paolillo, D.J., Faraday, C.D., Nasrallah, J.B. and Nasrallah, M.E. (1989) *Developmental Biology*, **134**, 462–472.

139 Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A. and Hinata, K. (2000) *Nature*, **403**, 913–916.

140 Nasrallah, M.E., Liu, P., Sherman-Broyles, S., Boggs, N.A. and Nasrallah, J.B. (2004) *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 16070–16074.

141 Sherman-Broyles, S., Boggs, N., Farkas, A., Liu, P., Vrebalov, J., Nasrallah, M.E. and Nasrallah, J.B. (2007) *The Plant Cell*, **19**, 94–106.

142 Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J.B. and

Nasrallah, M.E. (2001) *The Plant Cell*, **13**, 627–643.

143 Tang, W., Kelley, D., Ezcurra, I., Cotter, R. and McCormick, S. (2004) *The Plant Journal*, **39**, 343–353.

144 Ryan, C.A., Pearce, G., Moura, D.S. and Scheer, J. (2002) *The Plant Cell*, **14**, 251–264.

145 Farrokhi, N., Whitelegge, J.P. and Brusslan, J.A. (2008) *Plant Biotechnology Journal*, **6**, 105–134.

146 Green, T.R. and Ryan, C.A. (1972) *Science*, **175**, 776–777.

147 Ryan, C.A. (1967) *Analytical Biochemistry*, **19**, 434–440.

148 Bishop, P.D., Makus, D.J., Pearce, G. and Ryan, C.A. (1981) *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 3536–3640.

149 Baydoun, E.A.H. and Fry, S.C. (1985) *Planta*, **165**, 269–276.

150 Felix, G. and Boller, T. (1995) *The Plant Journal*, **7**, 381–389.

151 Amano, Y., Shinohara, H., Sakagami, Y. and Matsubayashi, Y. (2005) *Analytical Biochemistry*, **346**, 124–131.

152 Shinohara, H. and Matsubayashi, Y. (2007) *The Plant Journal*, **52**, 175–184.

153 Tanaka, S. (2003) *Zoological Science*, **20**, 1183–1198.

154 Nickel, W. (2005) *Traffic*, **6**, 607–614.

16
Nonribosomal Peptide Synthesis

Sean Doyle

16.1
Introduction

Nonribosomal peptide synthesis (NRPS) is a key mechanism used by microorganisms for the production of important low-molecular-mass (so-called "secondary") metabolites, often essential for organism growth in challenging environments. Despite the use of significant negative and somewhat inaccurate terminology, such as "*non*ribosomal" and "secondary", respectively, associated with the process of NRPS, it is now clear that it represents a fundamental biochemical mechanism that (i) provides a strategy for microorganisms to overcome environmental challenges (e.g., nutrient limitation and host immune response), (ii) is a source of current and new natural product medicines (e.g., antibiotics and immunosuppressants), and (iii) has the potential for antimicrobial drug development by targeting essential aspects of NRPS required for microbial survival in the host [1, 2]. The low-molecular-mass metabolites produced by NRPS are generally described as nonribosomal peptide (NRPs) and this terminology will be used throughout this chapter to describe nonribosomally synthesized peptides. These NRPs are synthesized by NRP synthetases, which are multifunctional, occasionally multisubunit, enzymes composed of distinct functional domains (Adenylation, Thiolation, and Condensation domains) that when grouped together are referred to as a "module" (Figure 16.1) [3]. Thiolation domains are also referred to as peptidyl carrier protein (PCP) domains. Additionally, other enzyme functions are occasionally present in NRP synthetases, including Epimerase and Thioesterase domains that are responsible for L- to D-amino acid conversion and release of the newly synthesized NRP from the NRP synthetase, respectively. An NRP synthetase can contain one or more modules and each module within an NRP synthetase is responsible for recognizing (via the Adenylation domain) and incorporating one amino acid (or nonproteogenic amino acid) into the NRP product [4]. The genes encoding NRP synthetases are either

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

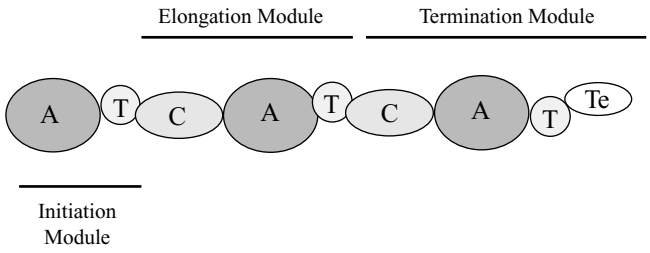


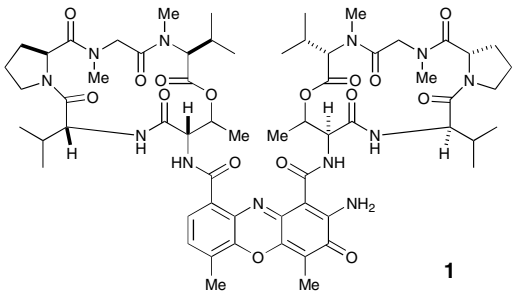
Figure 16.1 Schematic diagram of a hypothetical NRP synthetase showing the modular organization of the constituent functional domains [Adenylation (A), Condensation (C), Thiolation (T) (also referred to as PCP), and Thioesterase (Te)]. Each module is responsible for the incorporation of a single amino acid into the NRP. The first Condensation domain is responsible for peptide bond formation between the amino acid bound by the Initiation and the adjacent Elongation module. Domains conferring additional functionality may occasionally be present [e.g., Epimerase (E), Methylation (M), or Cyclization (Cy)]. Fungal NRP synthetases appear to lack Thioesterase domains and have a more variable domain organization than that found in bacteria.

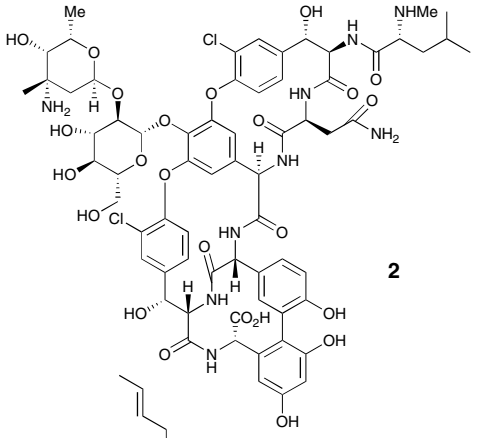
found within biosynthetic multigene clusters or are present individually within the genome of an organism. For example, monooxygenases, methyltransferases, dehydrogenases, metabolite transporters, and other enzyme functions are often encoded within these multigene clusters. Importantly, NRP synthetases are produced in the inactive (apoenzyme) form and require post-translational modification (PTM) to become functionally active [5]. PTM of NRP synthetases requires transfer of a 4′-phosphopantetheine group, derived from coenzyme A (CoA), to a specific serine residue within the Thiolation domain(s) of the NRP synthetase and this reaction is catalyzed by a 4′-phosphopantetheinyl transferase (4′-PPTases) [5].

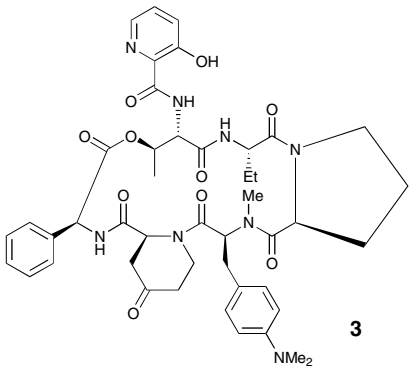
The purpose of this chapter is to describe the products and process of NRPS, and also to provide the reader with details of experimental strategies used to investigate NRP synthetase functionality.

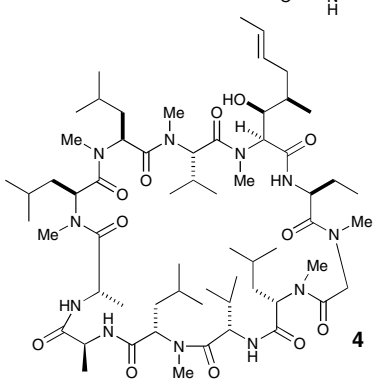
16.2 NRPs

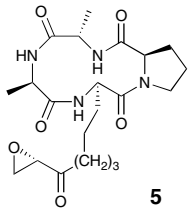
NRPS is responsible for the production of many secondary metabolites including important antibiotics and therapeutics, from both bacteria and filamentous fungi, and these include bacitracin, actinomycin C1 **1**, vancomycin **2**, pristinamycin 1_A**3**, cyclosporin **4**, HC-toxin **5**, enniatin A **6**, gramicidin S **7**, surfactin as well as precursors involved in the production of β-lactam antibiotics [3, 6]. Many antibiotics are known to incorporate D-amino acids, nonproteinogenic amino acids, and hydroxy acids, and require modifications such as *N*-methylation or cyclization. These antibiotics are synthesized nonribosomally via the thiotemplate mechanism, as described in Section 16.1 [7, 8].

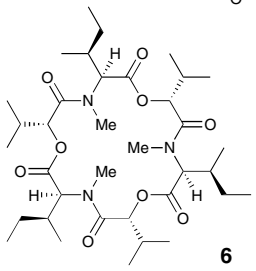


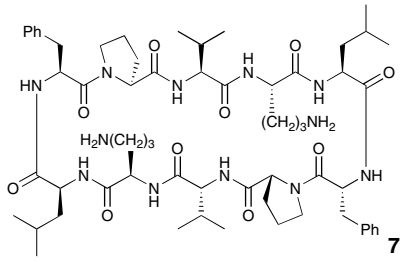




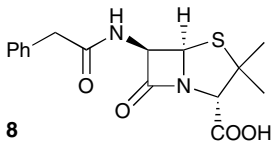




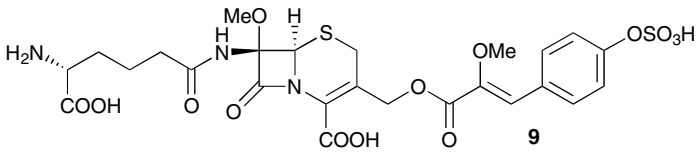




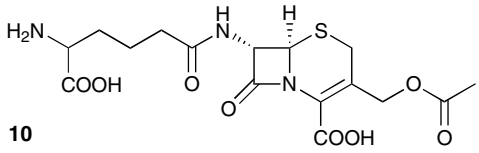
The tripeptide δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) required for the production of the β-lactam antibiotics penicillin (e.g., penicillin G **8)**, cephamycin A **9**, and cephalosporin C **10** is produced nonribosomally by ACV synthetase, encoded on the *pcbAB* gene (11 kb). This NRP synthetase has been isolated in *Acremonium chrysogenum*, *Acremonium nidulans*, *Streptomyces clavuligerus*, *Penicillium nalgiovense* and *Penicillium chrysogenum* [9], and is responsible for the condensation of the three L-amino acids, L-α-aminoadipic acid, L-cysteine, and L-valine, resulting in the formation of ACV. During this reaction epimerization of valine from the D- to L-configuration is required, and is carried out on an Epimerization domain present within the NRP synthetase located after a Thiolation domain.



8



9

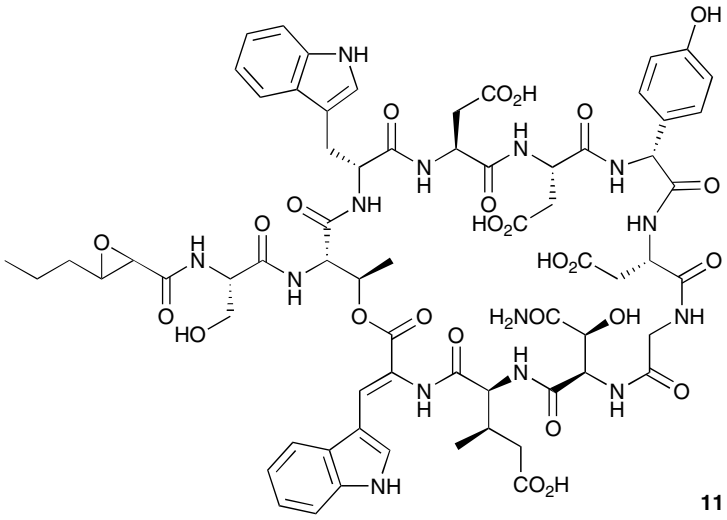


10

The enzyme isopenicillin N synthase (IPNS) is also involved in the production of β-lactam antibiotics, and catalyzes the formation of the ring structure. IPNS has been identified in *P. chrysogenum*, *Cephalosporium acremonium*, *A. nidulans*, *S. clavuligerus*, *Streptomyces lipmanii*, and *Streptomyces jumjinensis* [10]. Pristinamycin **3** produced by *Streptomyces pristinaespiralis* is a cyclohexadepsipeptide with antibiotic and bactericidal properties, and is also produced nonribosomally by a cluster of three pristinamycin synthetases [11]. Another NRP synthetase gene cluster is responsible for the production of balhimycin – a vancomycin-type antibiotic [12]. Vancomycin **2** is used to treat enterococcal infections including those caused by methicillin-resistant *Staphylococcus aureus*. This cluster involves four genes, *bpsA*, *bpsB*, *bpsC*, and *bpsD*, encoding eight modules, and is responsible for the activation of *N*-methyl-D-leucine, D-chloro-β-hydroxytyrosine, L-asparagine, D-4-hydroxyphenylglycine, L-chloro-β-hydroxytyrosine, and L-3,5-dihydrooxyphenylglycine [12].

Actinomycin **1** is an antibiotic produced by acyl peptide lactone synthetases in *Streptomyces* spp. The production of actinomycin involves a three-gene cluster of NRP synthetases [13]. Enniatins **6** are antibiotics produced by the plant pathogens *Fusarium* spp. These are produced by enniatin synthetases that have molecular masses in the order of 350 kDa and require *N*-methylation [13]. The *N*-methylation domain may be found between the Adenylation and Thiolation domains, and *N*-methylation occurs prior to peptide bond formation [14]. The cofactor *S*-adenosylmethionine (SAM) acts as the methyl donor in the transfer of the *N*-methyl group to the amino acid [15]. In the absence of SAM, small amounts of unmethylated depsipeptides are produced [7].

Calcium-dependent antibiotic (CDA) **11** is an acidic lipopeptide produced by *Streptomyces coelicolor* using the thiotemplate mechanism. It is a cyclic lactone undecapeptide produced by an NRP synthetase cluster involving either incorporation of D-4-hydroxyphenylglycine, D-3-phosphohydroxyasparagine, and L-glutamic acid/L-3-methylglutamic acid, into the peptide chain or subsequent modification of the respective amino acids, by tailoring enzymes, once assembled into the core NRP peptide structure [16]. Therefore, the NRP synthetase requires the presence of an Epimerization domain for putative conversion of L-asparagine to its D-configuration. Further studies of NRP synthesis systems may enable the design and engineering of new peptide antibiotics by swapping modules and domains [17–19].



11

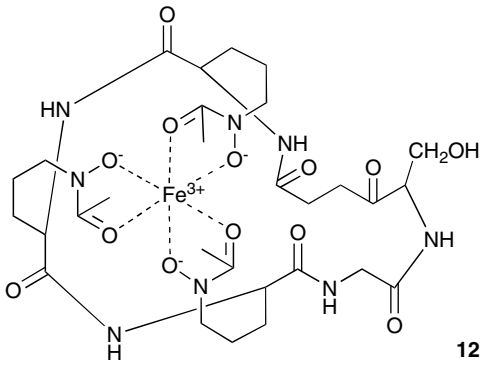
16.3
NRP Synthetase Domains

16.3.1
Adenylation Domains

Adenylation domains in NRP synthetases, which consist of approximately 500–600 amino acids, are responsible for (i) amino acid recognition from the cellular pool of proteogenic and nonproteogenic substrates, and (ii) the formation of the activated aminoacyl adenylate via concomitant ATP hydrolysis prior to transfer to the Thiolation domain. Substrate amino acid recognition and binding is facilitated by a nonlinear signature motif ("selectivity pocket") consisting of 10 amino acids that are distributed throughout the Adenylation domain. The first direct demonstration of Adenylation domain interaction with a substrate amino acid was by Conti *et al.* [20] who cocrystallized the L-phenylalanine-binding Adenylation domain of GrsA – an NRP synthetase responsible for the biosynthesis of gramicidin S. A number of substrate specificity prediction algorithms for Adenylation domain specificity of primarily bacterial NRP synthetases have been published [21–24]. In addition, domain specificity of both prokaryotic and eukaryotic NRP synthetases has been subject to analysis using a single algorithm whereby the substrate specificity of 95% of combined Adenylation–Thiolation and 78% of Adenylation domains, respectively, were correctly predicted [25]. Widespread application of this method will confirm its applicability to emerging NRP synthetase sequences.

To date, the vast majority of signature motifs have been predicted for bacterial NRP synthetases. However, as the number of fungal genome sequences continues to expand it is clear the Adenylation domain specificity code of eukaryotic NRP synthetases will become clearer.

The prototype and one of the most widely exploited NRPS pathways remains the industrial production of cephalosporins by the filamentous fungus *A. chrysogenum*, where, as noted above, the NRP synthetase, ACV sythetase, directs the biosynthesis of the penicillin precursor ACV from the cognate substrate amino acids [9, 10, 26, 27]. For a significant period this system provided the majority of information on fungal NRPS; however, Table 16.1 shows the Adenylation domain specificity of additional fungal NRP and polyketide/NRP synthetases that have recently become available.



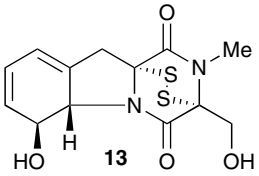
12

Table 16.1 Adenylation domain substrate specificity of fungal NRP or polyketide/NRP synthetases.

Enzyme type	Final product	Adenylation domains	Species	References
NRP synthetases				
SidD	TAFC	A1 (N^5-*cis*-anhydromevalonyl-N^5-hydroxy-L-ornithine	*Aspergillus fumigatus*	[28]
SidC	FC[a]	N^5-acetyl-N^5-hydroxy-L-ornithine, Ser, Gly[b]	*Aspergillus fumigatus*	[28]
GliP	gliotoxin	A1 (L-Phe) and A2 (L-Ser)	*Aspergillus fumigatus*	[33]
Hybrid polyketide/NRP synthetases				
FUSS	fusarin C	A1 (homoserine)	*Fusarium moniliforme*	[83]
TENS	2-pyridone tenellin	L-Tyr[c]	*Beauveria bassiana*	[84]
PsoA	pseurotin A	L-Phe	*Aspergillus fumigatus*	[85]
TdiA	terrequinone A	arylic acid[c]	*Aspergillus nidulans*	[86]
CheA	cytochalasan	L-Trp	*Penicillium expansum*	[67]

[a] FC can be converted to HFC by FC hydroxylase [28].
[b] Precise coding potential of each domain remains to be demonstrated.
[c] Proposed substrate.

Specifically, the high-affinity hydroxamate siderophores found in many fungal species, and which are responsible for environmental Fe^{3+} scavenging and storage in fungi, are NRPs [28–31]. *Aspergillus fumigatus* is an opportunistic human pathogen and the siderophore biosynthesis system in this organism has attracted much attention, especially because disruption of this pathway leads to ablated fungal virulence which means that components of the pathway may represent antifungal drug targets [28, 29]. The *sidC* gene in *A. fumigatus* encodes an NRP synthetase that contains three Adenylation domains, which directs the incorporation of N^5-acetyl-N^5-hydroxy-L-ornithine, serine, and glycine to form the intracellular, cyclic hexapeptide siderophore, ferricrocin (FC) **12** [28] (Table 16.1). Interestingly, FC can be enzymatically hydroxylated to form hydroxyferricrocin (HFC), which has been shown to be a major iron storage compound in *A. fumigatus* conidia [28] (Table 16.1). It has been conclusively shown the *A. fumigatus* genes *sidA* and *sidD*, which encode an L-ornithine-N^5-monooxygenase and NRP synthetase, respectively, are significant contributors to organismal virulence [2, 28, 32]. Bioinformatic analysis of SidD suggests that Adenylation domain 2 is not functional [31], yet it has been demonstrated that SidD is directly responsible for fusarinine C biosynthesis [28] so it is possible that Adenylation domain 1 of SidD is primarily responsible for N^5-*cis*-anhydromevalonyl-N^5-hydroxy-L-ornithine recognition and activation prior to fusarinine C biosynthesis, which, following acetylation via an N^2-transacetylase, is converted to triacetylfusarinine C (TAFC) [28] (Table 16.1). TAFC is secreted by *A. fumigatus* and is primarily located extracellularly.



13

Gliotoxin **13** biosynthesis, from precursor amino acids, L-phenylalanine and L-serine, is directed by a bimodular NRP synthetase, termed GliP, and the Adenylation domain specificity of the enzyme has been demonstrated [33] (Table 16.1). Although there has been considerable debate about the contribution which gliotoxin makes to the virulence of *A. fumigatus* [34, 35], recent studies [36, 37] have conclusively demonstrated that the type of immunosuppression regimen used during animal infection studies is important to demonstrate the potential of gliotoxin as an important virulence determinant. The Adenylation domain specificities of a number of fungal hybrid polyketide/NRP synthetases have been elucidated to date and it appears that aromatic amino acids (L-tyrosine, L-tryptophan, and L-phenylalanine), along with nonproteogenic amino acids, are prominent substrates (Table 16.1) [30].

Interestingly, Adenylation domains with apparently different signature motifs can, on occasion, activate the same amino acid; for example, the L-proline activating domain of tyrocidine synthetase B (TycB) possesses a somewhat different signature motif to Adenylation domains within the NRP synthetases RedM and PltF, which also activate L-proline. This observation, which possibly arises due to differential contact between signature motif residues and the substrate amino acid [38], implies that like

the genetic code, the nonribosomal code is also degenerate. Given that the number of Adenylation domains present in the UniProt database (www.uniprot.org) has been calculated to be 2824 by Minowa *et al.* [25], the challenge of deciphering the full extent of the nonribosomal code remains formidable.

16.3.2 Thiolation Domains

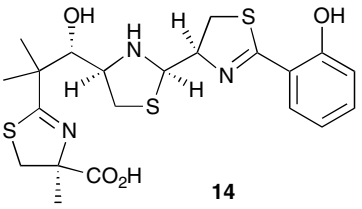
Thiolation/PCP domains within NRP synthetases generally consist of approximately 80 amino acids and have molecular masses in the region of 10 kDa. They serve as the site of attachment of the 4′-phosphopantetheine cofactor to a specific serine residue within a conserved motif, (L/I)GX(D/H)S(L/I) (X = Gly for PCP), which is present in both PCP and acyl carrier protein (ACP) domains [39]. In general, covalent attachment of substrate amino acids to the tethered 4′-phosphopantetheine cofactor contributes to the overall efficiency of NRPS by minimizing substrate diffusion and ensuring efficient movement of activated amino acids to the site of peptide bond formation on the Condensation domain.

16.3.3 Condensation Domains

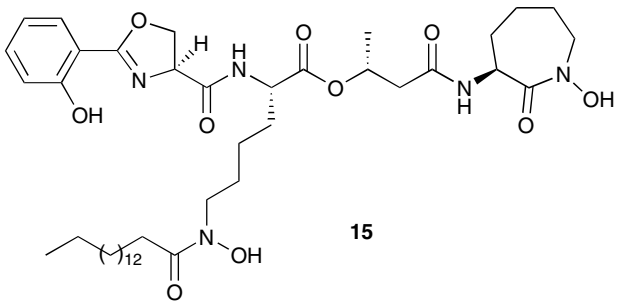
Condensation domains are composed of approximately 450 amino acid residues and contain the highly conserved consensus sequence, HHXXXDG. The conserved histidine residues play an important role in the activity of the Condensation domain [40]. Generally, the number of Condensation domains corresponds to the number of peptide bonds in the NRP peptide; however, in the case of iterative NRP synthetases, whereby domains are utilized more than once, this is not the case [41]. The Condensation domain is the site of peptide bond formation, in an N- to C-terminal direction, between two adjacent modules within NRP synthetases. Formation of the peptide bond requires either two adjacent aminoacyl intermediates for initiation of chain formation or one peptidyl donor with an acceptor aminoacyl intermediate during chain elongation [13, 42], and donor and acceptor sites within Condensation domains facilitate peptidyl and aminoacyl binding, respectively, during chain elongation [4, 41]. It has been noted that the specificity of the Condensation domain acceptor site is greater than that of the donor, the consequence of which is that greater selectivity is exhibited towards the Adenylation domain partner of the cognate Condensation domain than to Adenylation domain specificity elsewhere in the NRP synthetase [4]. Indeed, these authors speculate that this selectivity has evolved to ensure process integrity and may restrict combinatorial applications using Adenylation domains with attenuated substrate specificity.

It is now known from phlyogenetic analysis that Condensation domains cluster into three main functional groups, namely L,D-peptidyl and *N*-acyl donors, which is dependent upon the nature of the substrate donor molecule [42]. This finding has been elaborated upon by Rausch *et al.* [43] who use the terminology $^{L}C_{L}$ and $^{D}C_{L}$ to describe Condensation domains that catalyze peptide bond formation between two

L-amino acids and ones those that attach an L-amino acid to a peptide ending in a D-amino acid, respectively. In addition, these authors classify a third Condensation domain subtype, analogous to the *N*-acyl functional group [42], which they term a Starter domain, to be responsible for acylation of the initial amino acid in the NRP peptide with a β-hydroxy-carboxylic acid. A fourth type of Condensation domain can, in addition to peptide bond formation, also catalyze amino acid L- to D-epimerization and this is termed a dual Epimerase/Condensation domain. A fifth Condensation domain variant occurs infrequently in some NRP synthetases. This variant is the Cyclization domain that combines the Condensation domain function with additional heterocyclization or dehydration functions [44]. The Cyclization domain first identified in bacitracin synthetase, contains the conserved motif DxxxxDxxS, and is also present in yersiniabactin **14** and mycobactin (e.g., mycobactin T **15**) NRP synthetases [44]. Starter and $^{L}C_{L}$ domains are more closely related to each other than to the other Condensation domain subtypes and dual Epimerase/Condensation domains, and $^{D}C_{L}$ may share a common evolutionary ancestral protein [43]. Although these phylogenetic analyses have been performed using Condensation domains of bacterial origin only, it is clear that knowledge of NRP synthetase domain evolution and functionality, and future domain engineering, can be significantly enhanced by phylogenetic techniques exploiting hidden Markov models and other *in silico* strategies.



14



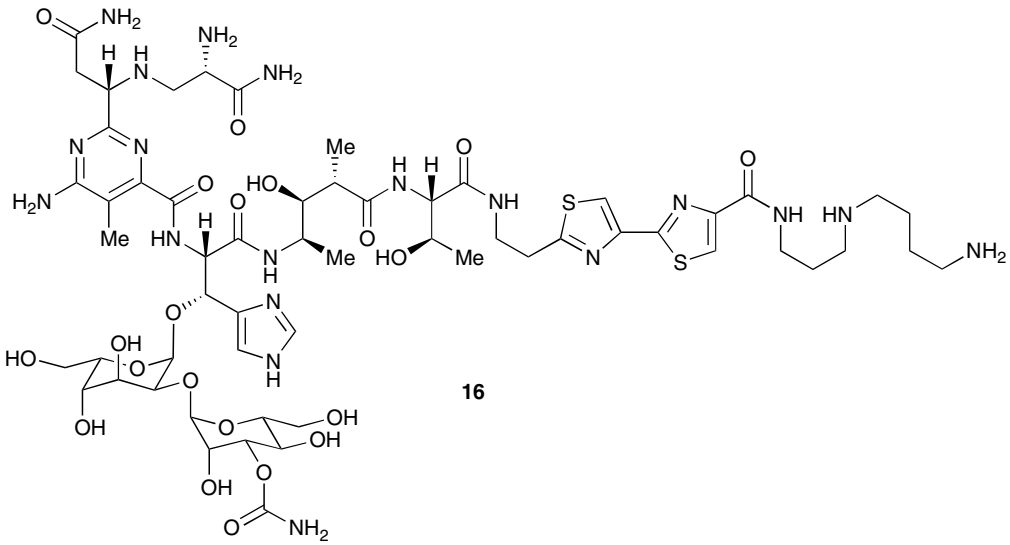
15

16.4 PPTases

CoA plays a pivotal role in NRPS because it provides a key component for NRP synthetase functionality. Moreover, CoA has also recently emerged as a key molecule in facilitating new avenues of NRPS research [45]. NRP synthetases are post-translationally modified from the inactive (apo) to active (holo) form via the transfer of the 4′-phosphopantetheine group (occasionally referred to as 4′Ppant, Ppant, or 4′PP) from CoA to the side-chain of the β-hydroxyl group of a conserved serine present in each Thiolation (PCP) domain. This modification is catalyzed by 4′-PPTases in a Mg^{2+}-dependent reaction [46]. 4′-PPTases are also responsible for the activation of ACPs (in polyketide and fatty acid biosynthesis) and α-aminoadipate semialdehyde reductases (in lysine biosynthesis), and so play a key role in polyketide, fatty acid, and lysine biosynthesis [47].

Classification of 4′-PPTases is based on structural characteristics and carrier protein specificity and, to date, three main groups: AcpS (approximately 120 amino acids in length), Sfp, and 4′-PPTases present in fatty acid synthase II have been identified [48]. Sfp-type 4′-PPTases are mainly responsible for activation of NRP synthetases and polyketide synthases [41]. A consensus sequence is evident in the Sfp-type enzymes (WxxKEAxxK), although overall, 4′-PPTases do not exhibit extensive homology to each other [49]. Many 4′-PPTases have been identified in bacteria; however, until recently, few have been found in filamentous fungi [5, 50].

In bacteria containing multiple 4′-phosphopantetheinylation-requiring pathways, each pathway had been thought to have or require a 4′-PPTase activity; however, it was demonstrated that a "promiscuous" 4′-PPTase was present, which could activate both type I and type II ACPs and PCPs from either *Streptomyces verticillus* or other *Streptomyces* spp. Indeed the biosynthesis of the hybrid peptide–polyketide drug bleomycin **16** in this organism was dependent on 4′-PPTase activity [48].



16

The role of 4′-PPTases in fungal NRPS has been reviewed elsewhere [30]; however, it has been suggested that a single 4′-PPTase (termed CfwA/NpgA) is responsible for total NRP synthetase and polyketide synthase activation in the filamentous fungus, *A. nidulans* [50]. Such dependence on a single enzyme for global post-translational activation of key enzymes is puzzling given the critical nature of NRPS to the organism. Fortunately, it strongly suggests that CfwA/NpgA may be a potential antifungal drug target.

16.4.1
4′PPTase Activity Determination

Historically, the determination of 4′-PPTase functionality or activity was not trivial and involved either gene deletion studies or protein mass spectrometry (MS) to

identify mass differences due to the presence of 4′-phosphopantetheine on carrier domains [5, 51, 52]. Additionally, no direct method was available to confirm *in vitro* 4′-phosphopantetheinylation of recombinant NRP synthetases or constituent domains for subsequent biochemical studies. Many of the difficulties associated with determination of 4′-PPTase activity have been circumvented by the findings of La Clair *et al*. [45], who developed a method to covalently label PCP domains of NRP synthetases, using biotinylated or fluorescently labeled CoA, via 4′-PPTase activity.

This chemoenzymatic approach, using a *Bacillus subtilis*-encoded 4′-PPTase, Sfp, normally responsible for activation of surfactin synthase, was used to modify recombinant carrier proteins with a single fluorescent or affinity reporter molecule (fluorophore or biotin), providing an attractive and sensitive means of protein visualization, Western blot identification, and affinity purification. This labeling technique also has utility in the identification of carrier protein domains from native cell lysates since a proportion of carrier proteins in a lysate may be present in apo- and inactive forms [45]. La Clair *et al*. [45] applied this technique to the 6-deoxyerythronolide B synthetase (DEBS) pathway from *Saccharapolyspora erythraea*, which lead to identification of a weak 150-kDa band proposed to be native DEBS. However, native 4′-PPTases in the producer strain modify the majority of carrier protein domains and exploitation of the full potential of this system to label native carrier proteins, in the future, will require the inhibition of native 4′-phosphopantetheinylation. Although the majority of reporter labeling of carrier proteins by 4′-PPTases has been limited to *in vitro* and cell surface protein labeling, a chemoenzymatic mechanism for *in vivo* protein modification using a fluorescent marker has been reported whereby the ability of native enzymes present in *Escherichia coli* (CoAA, CoAD, and CoAE) to sequentially modify a fluorescent pantetheine analog resulted in formation of a fluorescently labeled CoA analog [53]. Co-expression of the carrier protein VibB and the 4′-PPTase Sfp in the presence of the fluorescent pantetheine analog lead to the fluorescent modification of the VibB carrier protein *in vivo*. The ability of 4′-PPTases to post-translationally modify carrier proteins using low-molecular-mass chemical labels thus represents a novel and efficient tool for protein labeling, which has significant implications in NRPS and wider research areas.

Consequently, this approach has been adapted to imaging of cellular events involving transferrin-mediated iron uptake [54] – a system which represents a major mechanism employed by vertebrate cells to acquire iron from the environment. In this study, transferrin receptor 1 (TfR1) was fused to PCP and the resultant TfR1–PCP fusion was specifically labeled with a fluorophore via the action of Sfp. In a separate study, an 11-residue peptide, DSLEFIASKLA, predicted to be a suitable target for modification by the *B. subtilis* Sfp was fused to the C- and N-termini of a target protein and subsequently labeled with a biotin tag by the 4′-PPTase [55].

ACPs play a key role in fatty acid biosynthesis [56] and are analogous to Thiolation domains in that they can be 4′-phosphopantetheinylated by the action of AcpS. Novel anthranilic acid drugs have been designed to inhibit the 4′-phosphopantetheinylation activity of a *B. subtilis* AcpS [57]. Here, the availability of the AcpS

X-ray structure facilitated the modeling of numerous substrate analogs into the structure of AcpS and four analogs (anthranilic acid derivatives) were selected for use in *in vitro* AcpS inhibition studies. A novel 4′-phosphopantetheinylation assay, utilizing biotinylated CoA to assess AcpS activity via homogeneous time-resolved fluorescence (HTRF) was employed by these authors to determine the extent of enzyme inhibition by the putative inhibitors. This HTRF assay involved incubation of the AcpS (1.1 nM) in the presence of a glutathione-*S*-transferase–ACP carrier protein (35 nM) and a biotinylated CoA reporter analog (1.9 μM) in microtiter wells. Detection of the biotinylated carrier protein (i.e., AcpS activity) was achieved by an overnight incubation with a streptavidin–allophycocyanin conjugate followed by excitation and emission at 340/665 nm. All anthranilic acid drugs bound to the AcpS active site where the adenine, ribose, and 3′-phosphate of the CoA residues were located, and *in vitro* studies indicated up to 100% AcpS inhibition. Development of inhibitory agents of AcpS/4′-PPTase activity such as reported by Joseph-McCarthy *et al*. [57] provides a basis for future characterization of these enzymes and may potentially form the basis for the generation of novel antimicrobial agents.

16.5
Experimental Strategies for NRPS Investigations

Figure 16.2 shows the range of techniques that must be part of the technological arsenal to undertake comprehensive NRPS investigation. Bioinformatic techniques, such as BLAST analyses, are necessary to identify novel NRP synthetase genes in newly sequenced, or partially sequenced, genomes. In addition, commercially

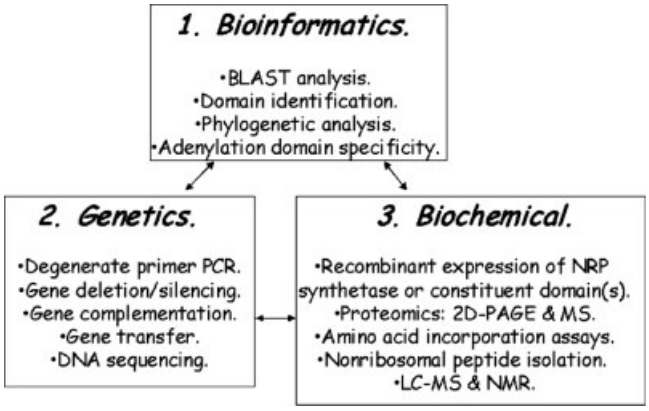


Figure 16.2 An integrated research strategy is required to determine the relationship between an NRP synthetase gene and the concomitant NRP. This "NRPS Research Pyramid" involves bioinformatics, molecular biology, and biochemical strategies to enable complete dissection of NRPS pathways in bacteria and fungi.

available software and freeware are available to assist the researcher in *in silico* gene annotation. Once hypothetical NRP synthetase genes have been identified, phylogenetic analysis of constituent domains (e.g., Adenylation or Condensation domains, respectively) can give clues as to the possible functionality and ultimate NRP encoded by the NRP synthetase and associated enzymes [58–60]. Algorithms are available [21–25] for the *in silico* prediction of bacterial Adenylation domain specificity which is a source of supporting, though not conclusive, data for the laboratory scientist to utilize towards the goal of solving the NRP code. As already mentioned, very limited data is in existence to aid delineation of the fungal code.

DNA sequence data can be readily used for primer design to amplify and confirm the presence of NRP synthetase genes, generate probes for Northern analysis of gene expression, and also to generate recombinant NRP synthetases in bacterial or eukaryotic expression systems [5, 61], or constituent domains/modules, for biochemical studies.

In the absence of DNA sequence data on the species of interest, a number of authors have developed degenerate primers for the amplification of NRP synthetase genes or gene fragments [62, 63] (Table 16.2 and Section 16.5.1), which in turn can be exploited for "genome walking" purposes or for cloning/expression analyses. It should be noted that cryptic intron–exon splice sites can be present in fungal NRP

Table 16.2 Selected examples of degenerate primers used to amplify NRP synthetase gene fragments from bacteria and fungi.

Primers	Sequence (5′–3′)	Species	References
A-F(A)	GCSTACSYSATSTACACSTCSGG	actinomycetes[a]	[71]
A-R(A)	SASGTCVCCSGTSCGGTAS		
P1-F	ATCTACAC(G/C)AGCGGGACGAC(G/C)GGC	*Streptomyces* sp. US24	[87]
P2-R	(G/C)AGGTCGCC(G/C)GTGCGGTACAT		
F1(C)	GCNGG(C/T)GG(C/T)GCNTA(C/T)GTNCC	cyanobacteria	[62]
R1(C)	CCNCG(AGT)AT(TC)TTNAC(T/C)TG		[72]
F1	GCNGGYGGYGCNTAYGTNCC	*Trichoderma harzianum*[b]	[63]
R1	CCNCGDATYTTNACYTG		
F2	TCCRCGGATYTDACCTGSKTATCC		
R2	TGCRGGYGGTGCITAYGTKCC		
A-F	TAYGGNCCNACNGA	*Neotyphodium*/*Epichloë*[c]	[88]
A-R	ARRTCNCCNGTYTTRTA		
M-F	AAYWSNGTNGYNCARTAYTTYCC		
M-R	YTTNGGNADNAYYTCNACRTG		

[a] *Streptomyces* spp., *Micromonospora* spp., *Actinoplanes* spp., *Pilimelia* spp., *Catenuloplanes* spp., *Couchioplanes* spp., *Nocardia* spp., *Rhodococcus* spp., *Kutzneria* spp., *Amycolatopsis* spp., *Saccharomonospora* spp., *Saccharopolyspora* spp., *Pseudonocardia* and so on.
[b] Primers (F1/R1) used with *T. harzianum* were based on from cyanobacterial sequence data [62], while F2/R2 were designed from *Trichoderma* sequence data. All primer sets amplified a region within an Adenylation domain of a fungal NRP synthetase.
[c] The *Neotyphodium*/*Epichloë* primers were based on conserved sequences within Adenylation (A-F/A-R) and Methyltransferase (M-F/M-R) domains, respectively.

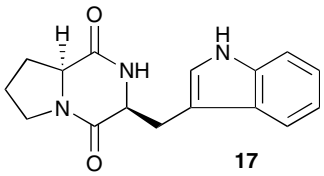
synthetases so care should be taken in primer and subsequent probe design. Moreover, a number of authors have underlined the requirement to confirm NRP synthetase gene activity with both RNA and protein expression studies to exclude the possibility of characterizing nonfunctional genes or gene fragments in the genome [63, 64].

The fundamental question asked by most NRPS investigations centers on the elucidation of orphan NRP synthetase functionality; in other words, what NRP does this new gene that I have identified ultimately encode? Three fundamental strategies exist to answer this question, and include (i) gene deletion (or disruption) and complementation studies [65, 66], (ii) gene silencing (both with subsequent wild-type versus mutant/silenced strain comparative studies) [67, 68], and (iii) interspecies/strain transfer of NRP synthetase coding regions followed by comparative metabolite analysis between original and recipient organisms [69]. Although gene deletion strategies are available for many organisms, this is not a trivial technique in some organisms and requires the input of expert knowledge. Likewise, complementation of mutant NRP synthetase genes is technically challenging, not least because of the large DNA fragment sizes associated with mutant gene replacement. To date, gene silencing has found limited application in NRPS research; however, using this technique it has been demonstrated that a fungal polyketide/NRP synthetase, CheA, directs biosynthesis of the cytochalasan, cheatoglobosin, in *Penicillium expansum* [67]. Interestingly, these authors noted that L-tryptophan is encoded by the NRP synthetase component. Moreover, they underlined the difficulty in gene deletion in *P. expansum*, along with the expression of *P. expansum* genes in heterologous hosts such as *Aspergillus* spp., which ultimately led to experimentation resulting in successful RNA silencing of *cheA*.

Identification and confirmation of NRP synthetase expression, and presence, in cell extracts is enabled by immunological techniques such as Western blotting using antisera raised against immunogenic, recombinant domain fragments, and also by matrix-assisted laser desorption/ionization (MALDI) or tandem MS analysis of high-molecular-mass proteins present after sodium dodecylsulfate (SDS)– polyacrylamide gel electrophoresis (PAGE) or two-dimensional PAGE fractionation of microbial cell lysates [31, 60]. Given the relatively low abundance of many NRP synthetases, prefractionation concentration steps such as trichloroacetic acid or ammonium sulfate precipitation are occasionally necessary prior to electrophoresis. Moreover, comparative Western analyses serves to complement Northern blotting or quantitative reverse transcription-polymerase chain reaction (PCR) results by confirming lack of target NRP synthetase gene expression in the mutant strain.

Once mutant, and complemented, strains have been generated, one can commence comparative metabolite analysis, generally by reverse-phase (RP)-high-performance liquid chromatography (HPLC) purification and ideally with photodiode array and/or fluorescence detection [70]. Organic extraction of culture supernatants, or cell, spore, or conidial lysates represents the first step in metabolite identification and generally precedes RP-HPLC analyses [69]. Metabolite(s) that are present in the wild-type, but absent in the mutant strain represent good targets for further structural and pharmacological studies [69]. Obviously,

reappearance of the relevant metabolite in a complemented strain serves as conclusive proof of gene function or coding potential. Combined liquid chromatography (LC)-MS and nuclear magnetic resonance (NMR) approaches can yield mass and structural information of putative NRPs. For instance, ^{1}H-NMR spectral analysis, in $CDCl_3$, of a purified NRP ($m/z = 283$) was compared to literature data to allow identification of the molecule as brevianamide F **17** which, in turn led to confirmation of the biosynthetic capacity of the cognate NRP synthetase [69].

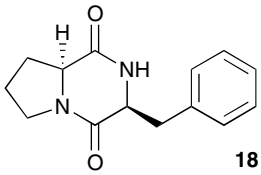


17

The following sections are focused on techniques that are somewhat specific to NRPS research and merit more consideration.

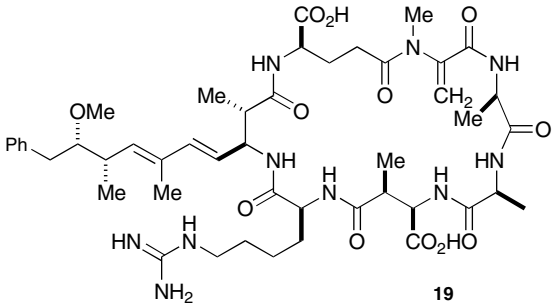
16.5.1 Degenerate PCR

The large number of bacterial and fungal genome sequences which have become available in recent years has greatly facilitated the search for, and contributed to the identification of, novel NRP synthetase genes. However, it is clear that a rich abundance of biosynthetic resource remains to be identified beyond the genomes which have already been sequenced. Fortuitously, the extent of sequence conservation within selected regions of NRP (and polyketide) synthetase genes has enabled the use of degenerate PCR to isolate and identify new coding potential.



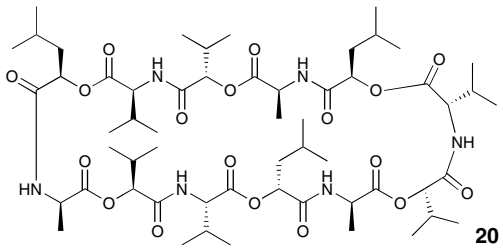
18

Sequence analysis of amplicons resulting from the use of degenerate PCR primers which amplify within Adenylation domains of actinomycete NRP synthetase genes (Table 16.2) has shown that these genes are widely distributed in both the major *Streptomyces* spp. and in related, though minor strains, where minimal data is currently available on the production of NRPs [71]. These authors tested 210 reference strains and found the majority contained NRP synthetase genes; moreover, simultaneous screening for polyketide sythetase genes using the degenerate PCR approach revealed a more limited distribution pattern. Furthermore, the identification of an NRP synthetase involved in cyclo (L-Phe–L-Pro) diketopiperazine **18** in *Streptomyces* spp. has been enabled using degenerate primers designed using conserved bacterial Adenylation domain sequences (Table 16.2).

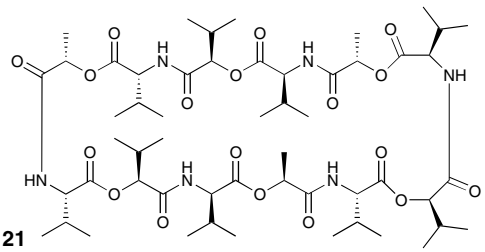


19

Cyanobacteria produce a range of metabolites, including alkaloids, NRPs, and polyketides, many of which exhibit potent toxicity. One of the best characterized cyanobacterial toxins is the nonribosomal heptapeptide, microcystin (e.g., microcystin AR **19)**, which is a known hepatotoxin. Using degenerate PCR primers (Table 16.2), Neilan *et al.* [62] detected putative NRP synthetase genes in a number of cyanobacterial strains from the genera *Anabaena*, *Aphanizomenon*, *Cylindrospermopsis*, *Microcystis*, *Nodularia*, *Nostoc*, *Oscillatoria*, *Plectonema*, and *Pseudanabaena*. Moreover, in what was a pioneering study, these authors were able to demonstrate the broad distribution of the microcystin gene cluster across a range of cyanobacteria using degenerate PCR to design probes, and characterized the phylogenetic interrelationships of microcystin gene clusters between different genera of cyanobacteria. Using the primers developed by Neilan *et al.* [62], others [72] further validated the use of degenerate PCR for the identification of NRP synthetase genes in, and establishment of evolutionary relationships between, cyanobacteria and proposed that although numerous in certain strains, NRP synthetase genes were absent from some cyanobacteria (e.g., *Prochlorococcus* spp. and some *Synechococcus* spp.). These authors also demonstrated that the filamentous and heterocystous cyanobacteria were the richest sources of NRP synthetase genes.



20



21

Degenerate PCR has been used to identify an NRP synthetase responsible for the biosynthesis of cereulide **20**, a valine-containing depsipeptide, related to valinomycin **21** [73]. Cereulide causes the emetic type of gastrointestinal disease following infection with *Bacillus cereus*. Following degenerate PCR, sequence analysis of one amplicon showed that it presented significant identity to bacterial valine activation NRP synthetase modules. Subsequent disruption of the NRP synthetase gene (*ces*) resulted in cereulide-deficient mutants, thereby confirming its role in cereulide biosynthesis. Moreover, because the valine-activating module was highly conserved

across emetic *B. cereus* strains from different geographical locations, a PCR assay detecting the *ces* gene was developed and showed high specificity of detection against cereulide-producing *B. cereus* strains. Thus, not only does degenerate PCR play a significant role in functional genomics of NRPS, but also contributes significantly to the identification of new diagnostic markers of bacterially induced disease.

16.5.2 Determination of Adenylation Domain Specificity

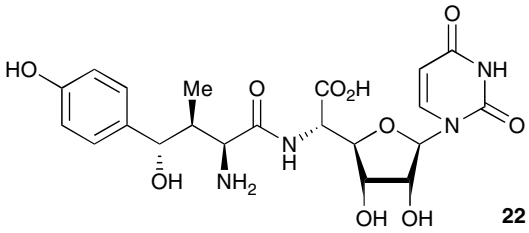
Even though improvements in *in silico* Adenylation domain specificity predictions are continually emerging [25], it is generally accepted that determination of the Adenylation domain specificity must be experimentally demonstrated to ensure that the computer predictions are accurate [74]. Historically, Adenylation domain specificity was evaluated by the detection of radioactive amino acid incorporation into either a native NRP synthetase or recombinant Adenylation domain, respectively. However, the expense associated with radioactive (proteogenic or nonproteogenic) amino acid use, and the experimental difficulty of ensuring optimal reaction and radioactive protein extraction, allied to the use of alternative strategies has meant that this approach has receded as the method of choice for Adenylation domain specificity determination.

Thus, the unambiguous assessment of Adenylation domain specificity requires direct biochemical analysis whereby the substrate amino acid binding specificity of individual recombinant adenylation domains is assessed by an amino acid-dependent ATP/PPi (pyrophosphate) exchange reaction [61, 75]. In addition, expression of functional apo-Adenylation–Thiolation domains, *in vitro* 4′-phosphopantetheinylation to the holo form, as has been demonstrated [5], followed by incubation with relevant amino acid substrates should facilitate the identification of the bound amino acid – either by whole-protein or peptide MS [74].

16.5.2.1 Protein MS

Kelleher *et al.* have pioneered the application of both whole-protein and large-fragment MS to the elucidation of the coding potential of NRP synthetases [74, 76] and the application of electrospray ionization-ion cyclotron resonance/Fourier transform MS (ESI-ICR/FTMS) to NRPS research has been extensively reviewed elsewhere [77]. ESI-ICR/FTMS allows accurate mass discrimination between 4′-phosphopantetheinylated and apo-Thiolation domains ($\Delta m/z = 340$ Da), and furthermore, enables detection and identification of covalently bound (substrate) amino acid to the holo-Thiolation domain. For example, the NRP synthetase module NikP1 was shown to encode the ligation of L-histidine into the NRP, nikkomycin **22**, by comparative FTMS analysis of apo-NikP1 and holo-NikP1 (approximately 75 kDa), in the presence of L-histidine and ATP [76]. The incorporation of L-phenylalanine into GrsA was also detectable by ESI-ICR/FTMS, which represented an even greater challenge as the size of the NRP synthetase module was approximately 130 kDa relative to the low molecular mass of L-phenylalanine (147 Da) [76]. In addition to using purified amino acid preparations in NRPS reactions, the

identification of substrate amino acids for NRP formation has also been performed using either commercially available algal hydrolyzates or representative *E. coli* metabolome extracts as the source of nonproteogenic amino acids by ESI-ICR/FTMS [74].



22

Although, whole-protein ESI-ICR/FTMS can avoid the use of enzymatic digestion to assist in identification of NRP synthetase functionality, analysis of tryptic fragments of recombinant modules primarily derived from TycB (molecular mass 237.5 kDa) provided important information on intermodular NRP synthetase interaction and the localization of Epimerase domains for combinatorial applications of NRPS [78].

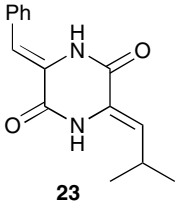
16.5.2.2 **Identification of NRP Synthetase Adenylation Domain Specificity (Strategy I)**

Optimization of recombinant Adenylation–Thiolation domain 4′-phosphopantetheinylation can be carried out as follows. Reaction mixtures for the *in vitro* 4′-phosphopantetheinylation assays should contain individual recombinant Adenylation–Thiolation domains (0–20 μM), 4′-PPTase (0–5 μM), biotinylated CoA (0–0.5 μM) [45, 57], 5 mM dithiothreitol, 10 mM $MgCl_2$, 75 mM Tris–HCl, pH 8.0, in final volumes of 100 μl and be incubated at 37 °C for 1 h. After incubation, reactions can be terminated by the addition of 10% (w/v) trichloroacetic acid (900 μl) and centrifuged at 10 000 × *g* for 10 min. Pellets should be resuspended in 1 M Tris–HCl, pH 8.0, and analyzed by SDS–PAGE and Western blot analysis, using streptavidin–horseradish peroxidase and chemiluminescent substrates (e.g., luminol), to visualize biotinylated Adenylation–Thiolation didomains. This approach confirms Thiolation domain functionality and will subsequently enable the confident performance of amino acid incorporation assays, using optimized 4′-PPTase assay conditions (with CoA instead of biotin-CoA), individual holo-Adenylation–Thiolation domains and proteogenic/nonproteogenic amino acids (e.g., D-amino acids). Two strategies can then be used to assess amino acid incorporation: (i) whole-protein MS by ICR/FTMS [74, 76–78] or (ii) MALDI-time-of-flight analysis of trypsin-digested recombinant holo-Adenylation–Thiolation didomains (pre- and postreaction) by accurate assessment of mass shifts, as we have previously used for assessing 4′- phosphopantetheinylation of *A. fumigatus* $Pes1_{TEA}$ [5]. Upon identification of any substrate amino acid using either of the aforementioned approaches, the incorporation assay should be repeated in the absence of that specific amino acid followed by MS analysis to confirm reaction specificity.

16.5.2.3 Identification of NRP Synthetase Adenylation Domain Specificity (Strategy II)

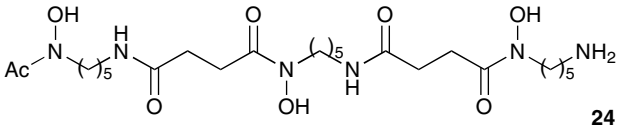
Adenylation domain specificity of NRP synthetases can also be assessed by expressing recombinant Adenylation or Adenylation-Thiolation didomains, respectively, and performing the well-established ATP/^{32}PPi ([^{32}P]sodium pyrophosphate) exchange assay [61, 75] to detect amino acid activation. The principle of this assay is that exchange of the radioactive label from ^{32}PPi into ATP is dependent on the ATPase activity of the Adenylation domain in the presence of the specific substrate amino acid. Postreaction, radioactive ATP is collected using activated charcoal and quantified by liquid scintillation counting [61]. Individual Adenylation, or Adenylation–Thiolation didomains, activity assays (each containing a different naturally occurring amino acid; $n = 20$) can be performed in the presence of [^{32}P]sodium pyrophosphate and ATP. Nonproteogenic and D-amino acids can also be used in the event of no evidence of proteogenic amino acid incorporation. This experimental strategy will yield biochemical identification of substrate amino acid incorporated by individual NRP synthetase Adenylation domains.

16.6 Non-NRPS



23

Although NRPS appears to be the primary biosynthetic mechanism for NRP formation, other mechanisms exist for the production of amino acid-containing metabolites in prokaryotes. One such mechanism is responsible for the production of the diketopiperazine, albonoursin **23**, in *Streptomyces noursei* [79, 80]. Albonoursin is an antibiotic with weak antitumor activity and is composed of two residues: α,β-dehydrophenylalanine and α,β-dehydroleucine. Like all other diketopiperazine derivatives it was thought to be produced via the thiotemplate mechanism of NRP biosynthesis. However, an enzyme, cyclic dipeptide oxidase, was identified that catalyzed the formation of albonoursin **23** [79]. Following this discovery, the biosynthetic gene cluster for albonoursin was identified by Lautru *et al.* [80] in 2002. The cluster was estimated to be approximately 2.7 kb in size, and consisted of four open reading frames: *albA*, *albB*, *albC*, and *albD*. *albA* and *albB* were found to be required for cyclic dipeptide oxidase activity, *albC* was responsible for cyclic dipeptide formation, and *albD* was thought to be a membrane protein, involved in the release of albonoursin. *albC* showed no similarity to NRP synthetases [80].



24

Desferrioxamine B **24** used to treat iron overload in humans (e.g., hemochromatosis) has been isolated from the actinomycete, *Streptomyces coelicolor* M145, and found to be the most prominent desferrioxamine siderophore produced by the organism [81]. A four-gene cluster (*desA–D*) encoding desferrioxamine B biosynthesis was identified and the proposed pathway for desferrioxamine B biosynthesis was found to be distinct from the classical NRPS thiotemplate mechanism [81]. It has subsequently been shown that the *des* cluster expression is controlled by the binding of a regulatory protein, DmdR1, to the *desA* promotor region and that *desA* encodes an enzyme with lysine decarboxylase activity confirming L-lysine as the substrate for desferrioxamine B biosynthesis in *S. coelicolor* [82].

It is unlikely that these findings represent the full extent of alternative nonribosomal strategies for metabolite biosynthesis, in either bacteria or fungi.

16.7 Conclusions

If there is one rule that pertains to NRPS, it is that the rules are there to be broken! This is clearly an overstatement, and the dogma of domain functionality and modular organization directing the sequence and structure of the resultant NRP is well established. However, the practically unlimited variety of possibilities with respect to Adenylation domain coding motifs, the effect of domain position on functionality, the presence of numerous ancillary enzyme functions, which are either part of NRP synthetase gene clusters or function to modify NRPs, and the presence of many orphan NRP synthetases (most of which have yet to be discovered) means that much work remains to be done to allow us to fully appreciate the complexity and potential of this exquisite biological process.

References

1 Chalut, C., Botella, L., de Sousa-D'Auria, C., Houssin, C. and Guilhot, C. (2006) The nonredundant roles of two 4′-phosphopantetheinyl transferases in vital processes of Mycobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 8511–8516.

2 Schrettl, M.E., Bignell, E., Kragl, C., Joechl, C., Rogers, T., Arst, H.N. Jr, Haynes, K. and Haas, H. (2004) Siderophore biosynthesis but not reductive iron assimilation is essential for *Aspergillus fumigatus* virulence. *The Journal of Experimental Medicine*, **200**, 1213–1219.

3 Grunewald, J. and Marahiel, M.A. (2006) Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides. *Microbiology and Molecular Biology Reviews*, **70**, 121–146.

4 Lautru, S. and Challis, G.L. (2004) Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology*, **150**, 1629–1636.

5 Neville, C.M., Murphy, A., Kavanagh, K. and Doyle, S. (2005) A 4′-phosphopantetheinyl transferase mediates non-ribosomal peptide synthetase activation in *Aspergillus fumigatus*. *ChemBioChem*, **6**, 679–685.

6 Keller, N.P.G., Turner, G. and Bennett, J.W. (2005) Fungal secondary metabolism – from biochemistry to genomics. *Nature Reviews. Microbiology*, **3**, 937–947.

7 Kleinkauf, H. and von Döhren, H. (1987) Biosynthesis of peptide antibiotics. *Annual Review of Microbiology*, **41**, 259–289.

8 Hubbard, B.K., Thomas, M.G. and Walsh, C.T. (2000) Biosynthesis of L-*p*-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics. *Chemistry & Biology*, **7**, 931–942.

9 Martin, J.F. (2000) Alpha-aminoadipyl-cysteinyl-valine synthetases in beta-lactam producing organisms. From Abraham's discoveries to novel concepts of non-ribosomal peptide synthesis. *The Journal of Antibiotics*, **53**, 1008–1021.

10 Kleinkauf, H. and von Döhren, H. (1990) Nonribosomal biosynthesis of peptide antibiotics. *European Journal of Biochemistry*, **192**, 1–15.

11 Thibaut, D., Bisch, D., Ratet, N., Maton, L., Couder, M., Debussche, L. and Blanche, F. (1997) Purification of peptide synthetases involved in pristinamycin I biosynthesis. *Journal of Bacteriology*, **179**, 697–704.

12 Recktenwald, J., Shawky, R., Puk, O., Pfennig, F., Keller, U., Wohlleben, W. and Pelzer, S. (2002) Nonribosomal biosynthesis of vancomycin-type antibiotics: a heptapeptide backbone and eight peptide synthetase modules. *Microbiology*, **148**, 1105–1118.

13 von Döhren, H., Keller, U., Vater, J. and Zocher, R. (1997) Multifunctional peptide synthetases. *Chemical Reviews*, **97**, 2675–2705.

14 Pieper, R., Haese, A., Schroder, W. and Zocher, R. (1995) Arrangement of catalytic sites in the multifunctional enzyme enniatin synthetase. *European Journal of Biochemistry*, **230**, 119–126.

15 Moffitt, M.C. and Neilan, B.A. (2000) The expansion of mechanistic and organismic diversity associated with non-ribosomal peptides. *FEMS Microbiology Letters*, **191**, 159–167.

16 Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., Wilkinson, B., Sidebottom, P.J., Rudd, B.A.M., Hayes, M.A., Smith, C.P. and Micklefield, J. (2002) Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chemistry & Biology*, **9**, 1175–1187.

17 Doekel, S. and Marahiel, M.A. (2000) Dipeptide formation on engineered hybrid peptide synthetases. *Chemistry & Biology*, **7**, 373–384.

18 Schauwecker, F., Pfennig, F., Schroder, W. and Keller, U. (1998) Molecular cloning of the actinomycin synthetase gene cluster from *Streptomyces chrysomallus* and functional heterologous expression of the gene encoding actinomycin synthetase II. *Journal of Bacteriology*, **180**, 2468–2474.

19 Stachelhaus, T., Schneider, A. and Marahiel, M.A. (1996) Engineered biosynthesis of peptide antibiotics. *Biochemical Pharmacology*, **52**, 177–186.

20 Conti, E., Stachelhaus, T., Marahiel, M.A. and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *The EMBO Journal*, **16**, 4174–4183.

21 Stachelhaus, T., Mootz, H.D. and Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology*, **6**, 493–505.

22 Challis, G.L., Ravel, J. and Townsend, C.A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & Biology*, **7**, 211–224.

23 Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. and Huson, D.H. (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Research*, **33**, 5799–5808.

24 Schwecke, T., Göttling, K., Durek, P., Dueñas, I., Käufer, N.F., Zock-

Emmenthal, S., Staub, E., Neuhof, T., Dieckmann, R. and von Döhren, H. (2006) Nonribosomal peptide synthesis in *Schizosaccharomyces pombe* and the architectures of ferrichrome-type siderophore synthetases in fungi. *ChemBioChem*, **7**, 612–622.

25 Minowa, Y., Araki, M. and Kanehisa, M. (2007) Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. *Journal of Molecular Biology*, **368**, 1500–1517.

26 Basch, J., Franceschini, T., Tonzi, S. and Chiang, S.-J.D. (2004) Expression of a cephalosporin C esterase gene in *Acremonium chrysogenum* for the direct production of deacetylcephalosporin C. *Journal of Industrial Microbiology & Biotechnology*, **31**, 531–539.

27 Rodríguez-Sáiz, M., Lembo, M., Bertetti, L., Muraca, R., Velasco, J., Malcangi, A., de la Fuente, J.L. and Barredo, J.L. (2004) Strain improvement for cephalosporin production by *Acremonium chrysogenum* using geneticin as a suitable transformation marker. *FEMS Microbiology Letters*, **235**, 43–49.

28 Schrettl, M., Bignell, E., Kragl, C., Sabiha, Y., Loss, O., Eisendle, M., Wallner, A., Arst, H.N. Jr, Haynes, K. and Haas, H. (2007) Distinct roles for intra- and extracellular siderophores during *Aspergillus fumigatus* infection. *PLoS Pathogens*, **3**, 1195–1207.

29 Hissen, A.H.T., Chow, J.M.T., Pinto, L.J. and Moore, M.M. (2004) Survival of *Aspergillus fumigatus* in serum involves removal of iron from transferrin: the role of siderophores. *Infection and Immunity*, **72**, 1402–1408.

30 Stack, D., Neville, C.M. and Doyle, S. (2007) Nonribosomal peptide synthesis in the human pathogen *Aspergillus fumigatus* and other fungi. *Microbiology*, **153**, 1297–1306.

31 Reiber, K., Reeves, E.P., Neville, C.M., Winkler, R., Gebhardt, P., Kavanagh, K. and Doyle, S. (2005) The expression of selected non-ribosomal peptide synthetases in *Aspergillus fumigatus* is controlled by the availability of free iron. *FEMS Microbiology Letters*, **248**, 83–91.

32 Hissen, A.H., Wan, A.N.C., Warwas, M.L., Pinto, L.J. and Moore, M.M. (2005) The *Aspergillus fumigatus* siderophore biosynthetic gene *sidA*, encoding L-ornithine N^5-oxygenase, is required for virulence. *Infection and Immunity*, **73**, 5493–5503.

33 Balibar, C.J. and Walsh, C.T. (2006) GliP, a multimodular nonribosomal peptide synthetase in *Aspergillus fumigatus*, makes the diketopiperazine scaffold of gliotoxin. *Biochemistry*, **45**, 15029–15038.

34 Cramer, R.A. Jr, Gamcsik, M.P., Brooking, R.M., Najvar, L.K., Kirkpatrick, W.R., Patterson, T.F., Balibar, C.J., Graybill, J.R., Perfect, J.R., Abraham, S.N. and Steinbach, W.J. (2006) Disruption of a nonribosomal peptide synthetase in *Aspergillus fumigatus* eliminates gliotoxin production. *Eukaryotic Cell*, **5**, 972–980.

35 Kupfahl, C., Heinekamp, T., Geginat, G., Ruppert, T., Hartl, A., Hof, H. and Brakhage, A.A. (2006) Deletion of the *gliP* gene of *Aspergillus fumigatus* results in loss of gliotoxin production but has no effect on virulence of the fungus in a low-dose mouse infection model. *Molecular Microbiology*, **62**, 292–302.

36 Sugui, J.A., Pardo, J., Chang, Y.C., Zarember, K.A., Nardone, G., Galvez, E.M., Müllbacher, A., Gallin, J.I., Simon, M.M. and Kwon-Chung, K.J. (2007) Gliotoxin is a virulence factor of *Aspergillus fumigatus*: *gliP* deletion attenuates virulence in mice immunosuppressed with hydrocortisone. *Eukaryotic Cell*, **6**, 1562–1569.

37 Spikes, S., Xu, R., Nguyen, C.K., Chamilos, G., Kontoyiannis, D.P., Jacobson, R.H., Ejzykowicz, D.E., Chiang, L.Y., Filler, S.G. and May, G.S. (2008) Gliotoxin production in *Aspergillus fumigatus* contributes to host-specific differences in virulence. *The Journal of Infectious Diseases*, **197**, 479–486.

38 Challis, G.L. and Naismith, J.H. (2004) Structural aspects of non-ribosomal peptide biosynthesis. *Current Opinion in Structural Biology*, **14**, 748–756.

39 Weber, T. and Marahiel, M.A. (2001) Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure*, **9**, R3–R9.

40 Stachelhaus, T., Mootz, H.D., Bergendahl, V. and Marahiel, M.A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis: catalytic role of the condensation domain. *The Journal of Biological Chemistry*, **273**, 22773–22781.

41 Mootz, H.D., Schwarzer, D. and Marahiel, M.A. (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem*, **3**, 490–504.

42 Roongsawang, N., Lim, S.P., Washio, K., Takano, K., Kanaya, S. and Morikawa, M. (2005) Phylogenetic analysis of condensation domains in the nonribosomal peptide synthetases. *FEMS Microbiology Letters*, **252**, 143–151.

43 Rausch, C., Hoof, I., Weber, T., Wohlleben, W. and Huson, D.H. (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evolutionary Biology*, **7**, 78.

44 Keating, T.A. and Walsh, C.T. (1999) Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Current Opinion in Chemical Biology*, **3**, 598–606.

45 La Clair, J.J., Foley, T.L., Schegg, T.R., Regan, C.M. and Burkart, M.D. (2004) Manipulation of carrier proteins in antibiotic biosynthesis. *Chemistry & Biology*, **11**, 195–201.

46 Walsh, C.T., Gehring, A.M., Weinreb, P.H., Quadri, L.E. and Flugel, R.S. (1997) Post-translational modification of polyketide and nonribosomal peptide synthases. *Current Opinion in Chemical Biology*, **1**, 309–315.

47 Guo, S. and Bhattacharjee, J.K. (2003) Molecular characterization of the *Candida albicans* LYS5 gene and site-directed mutational analysis of the PPTase (Lys5p) domains for lysine biosynthesis. *FEMS Microbiology Letters*, **224**, 261–267.

48 Sanchez, C., Du, L., Edwards, D.J., Toney, M.D. and Shen, B. (2001) Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide–polyketide antitumor drug bleomycin. *Chemistry & Biology*, **8**, 725–738.

49 Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C. and Walsh, C.T. (1996) A new enzyme superfamily – the phosphopantetheinyl transferases. *Chemistry & Biology*, **3**, 923–936.

50 Márquez-Fernández, O., Trigos, A., Ramos-Balderas, J.L., Viniegra-González, G., Deising, H.B. and Aguirre, J. (2007) Phosphopantetheinyl transferase CfwA/NpgA is required for *Aspergillus nidulans* secondary metabolism and asexual development. *Eukaryotic Cell*, **6**, 710–720.

51 Oberegger, H., Eisendle, M., Schrettl, M., Graessle, S. and Haas, H. (2003) 4′-Phosphopantetheinyl transferase-encoding *npgA* is essential for siderophore biosynthesis in *Aspergillus nidulans*. *Current Genetics*, **44**, 211–215.

52 Weissman, K.J., Hong, H., Oliynyk, M., Siskos, A.P. and Leadlay, P.F. (2004) Identification of a phosphopantetheinyl transferase for erythromycin biosynthesis in *Saccharopolyspora erythraea*. *ChemBioChem*, **5**, 116–125.

53 Clarke, K.M., Mercer, A.C., La Clair, J.J. and Burkart, M.D. (2005) *In vivo* reporter labeling of proteins via metabolic delivery of coenzyme A analogs. *Journal of the American Chemical Society*, **127**, 11234–11235.

54 Yin, J., Lin, A.J., Buckett, P.D., Wessling-Resnick, M., Golan, D.E. and Walsh, C.T. (2005) Single-cell FRET imaging of transferrin receptor trafficking dynamics by Sfp-catalyzed, site-specific protein labelling. *Chemistry & Biology*, **12**, 999–1006.

55 Yin, J., Straight, P.D., McLoughlin, S.M., Zhou, Z., Lin, A.J., Golan, D.E., Kelleher, N.L., Kolter, R. and Walsh, C.T. (2005)

Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 15815–15820.

56 Magnuson, K., Jackowski, S., Rock, C.O. and Cronan, J.E. (1993) Regulation of fatty acid biosynthesis in *Escherichia coli*. *Microbiological Reviews*, **57**, 522–542.

57 Joseph-McCarthy, D., Parris, K., Huang, A., Failli, A., Quagliato, D., Dushin, E.G., Novikova, E., Severina, E., Tuckman, M., Petersen, P.J., Dean, C., Fritz, C.C., Meshulam, T., DeCenzo, M., Dick, L., McFadyen, I.J., Somers, W.S., Lovering, F. and Gilbert, A.M. (2005) Use of structure-based drug design approaches to obtain novel anthranilic acid acyl carrier protein synthase inhibitors. *Journal of Medicinal Chemistry*, **48**, 7960–7969.

58 Cramer, R.A. Jr, Stajich, J.S., Yamanaka, Y., Dietrich, F.S., Steinbach, W.J. and Perfect, J.R. (2006) Phylogenomic analysis of non-ribosomal peptide synthetases in the genus *Aspergillus*. *Gene*, **383**, 24–32.

59 Lee, B.N., Kroken, S., Chou, D.Y.T., Robbertse, B., Yoder, O.C., Turgeon, B.G. and Gillian, B. (2005) Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, NPS6, involved in virulence and resistance to oxidative stress. *Eukaryotic Cell*, **4**, 545–555.

60 Reeves, E.P., Reiber, K., Neville, C., Scheibner, O., Kavanagh, K. and Doyle, S. (2006) A nonribosomal peptide synthetase (Pes1) confers protection against oxidative stress in *Aspergillus fumigatus*. *FEBS Journal*, **273**, 3038–3053.

61 Couch, R., O'Connor, S.E., Seidle, H., Walsh, C.T. and Parry, R. (2004) Characterization of CmaA, an adenylation–thiolation didomain enzyme involved in the biosynthesis of coronatine. *Journal of Bacteriology*, **186**, 35–42.

62 Neilan, B.A., Dittmann, E., Rouhiainen, L., Bass, R.A., Schaub, V., Sivonen, K. and Borner, T. (1999) Nonribosomal peptide synthesis and toxigenicity of cyanobacteria. *Journal of Bacteriology*, **181**, 4089–4097.

63 Vizcaino, J.A., Sanz, L., Cardoza, R.E., Monte, E. and Gutierrez, S. (2005) Detection of putative peptide synthetase genes in *Trichoderma* species: application of this method to the cloning of a gene from *T. harzianum* CECT 2413. *FEMS Microbiology Letters*, **244**, 139–148.

64 Harrison, P., Kumar, A., Lan, N., Echols, N., Snyder, M. and Gerstein, M. (2002) A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution. *Journal of Molecular Biology*, **316**, 409–419.

65 Gardiner, D.M., Cozijnsen, A.J., Wilson, L.M., Pedras, M.S.C. and Howlett, B.J. (2004) The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans*. *Molecular Microbiology*, **53**, 1307–1318.

66 Bok, J.W., Chung, D., Balajee, S.A., Marr, K.A., Andes, D., Nielsen, K.F., Frisvad, J.C., Kirby, K.A. and Keller, N.P. (2006) GliZ, a transcriptional regulator of gliotoxin biosynthesis, contributes to *Aspergillus fumigatus* virulence. *Infection and Immunity*, **74**, 6761–6768.

67 Schümann, J. and Hertweck, C. (2007) Molecular basis of cytochalasan biosynthesis in fungi: gene cluster analysis and evidence for the involvement of a PKS–NRPS hybrid synthase by RNA silencing. *Journal of the American Chemical Society*, **129**, 9564–9565.

68 Mouyna, I., Henry, C., Doering, T.L. and Latgé, J.P. (2004) Gene silencing with RNA interference in the human pathogenic fungus *Aspergillus fumigatus*. *FEMS Microbiology Letters*, **237**, 317–324.

69 Maiya, S., Grundmann, A., Li, S.M. and Turner, G. (2006) The fumitremorgin gene cluster of *Aspergillus fumigatus*: identification of a gene encoding brevianamide F synthetase. *ChemBioChem*, **7**, 1062–1069.

70 Larsen, T.O., Smedsgaard, J., Nielsen, K.F., Hansen, M.A.E., Samson, R.A. and Frisvad, J.C. (2007) Production of

mycotoxins by *Aspergillus lentulus* and other medically important and closely related species in section Fumigati. *Medical Mycology*, **45**, 225–232.

71 Ayuso-Sacido, A. and Genilloud, O. (2005) New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. *Microbial Ecology*, **49**, 10–24.

72 Ehrenreich, I.M., Waterbury, J.B. and Webb, E.A. (2005) Distribution and diversity of natural product genes in marine and freshwater cyanobacterial cultures and genomes. *Applied and Environmental Microbiology*, **71**, 7401–7413.

73 Ehling-Schulz, M., Vukov, N., Schulz, A., Shaheen, R., Andersson, M., Märtlbauer, E. and Scherer, S. (2005) Identification and partial characterization of the nonribosomal peptide synthetase gene responsible for cereulide production in emetic *Bacillus cereus*. *Applied and Environmental Microbiology*, **71**, 105–113.

74 Dorrestein, P.C., Blackhall, J., Straight, P.D., Fischbach, M.A., Garneau-Tsodikova, S., Edwards, D.J., McLaughlin, S., Lin, M., Gerwick, W.H., Kolter, R., Walsh, C.T. and Kelleher, N.L. (2006) Activity screening of carrier domains within nonribosomal peptide synthetases using complex substrate mixtures and large molecule mass spectrometry. *Biochemistry*, **45**, 1537–1546.

75 May, J.J., Finking, R., Wiegeshoff, F., Weber, T.T., Bandur, N., Koert, U. and Marahiel, M.A. (2005) Inhibition of the D-alanine:D-alanyl carrier protein ligase from *Bacillus subtilis* increases the bacterium's susceptibility to antibiotics that target the cell wall. *FEBS Journal*, **272**, 2993–3003.

76 Hicks, L.M., Mazur, M.T., Miller, L.M., Dorrestein, P.C., Schnarr, N.A., Khosla, C. and Kelleher, N.L. (2006) Investigating nonribosomal peptide and polyketide biosynthesis by direct detection of intermediates on >70 kDa polypeptides by using Fourier-transform mass spectrometry. *ChemBioChem*, **7**, 904–907.

77 Dorrestein, P.C. and Kelleher, N.L. (2006) Dissecting non-ribosomal and polyketide biosynthetic machineries using electrospray ionization Fourier-transform mass spectrometry. *Natural Product Reports*, **23**, 893–918.

78 Stein, D.B., Linne, U., Hahn, M. and Marahiel, M.A. (2006) Impact of epimerization domains on the intermodular transfer of enzyme-bound intermediates in nonribosomal peptide synthesis. *ChemBioChem*, **7**, 1807–1814.

79 Gondry, M., Lautru, S., Fusai, G., Meunier, G., Menez, A. and Genet, R. (2001) Cyclic dipeptdie oxidase from *Streptomyces noursei*. *European Journal of Biochemistry*, **268**, 1712–1721.

80 Lautru, S., Gondry, M., Genet, R. and Pernodet, J.-L. (2002) The albonoursin gene cluster of *S. noursei*. biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. *Chemistry & Biology*, **9**, 1355–1364.

81 Barona-Gómez, F., Wong, U., Giannakopulos, A.E., Derrick, P.J. and Challis, G.L. (2004) Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. *Journal of the American Chemical Society*, **126**, 16282–16283.

82 Tunca, S., Barreiro, C., Sola-Landa, A., Coque, J.J.R. and Martín, J.F. (2007) Transcriptional regulation of the desferrioxamine gene cluster of *Streptomyces coelicolor* is mediated by binding of DmdR1 to an iron box in the promoter of the *desA* gene. *FEBS Journal*, **274**, 1110–1122.

83 Rees, D.O., Bushby, N., Cox, R.J., Harding, J.R., Simpson, T.J. and Willis, C.L. (2007) Synthesis of [1,2-$^{13}C_2$, ^{15}N]-L-homoserine and its incorporation by the PKS–NRPS system of *Fusarium moniliforme* into the mycotoxin fusarin C. *ChemBioChem*, **8**, 46–50.

84 Eley, K.L., Halo, L.M., Song, Z., Powles, H., Cox, R.J., Bailey, A.M., Lazarus, C.M. and Simpson, T.J. (2007) Biosynthesis of the 2-pyridone tenellin in the insect pathogenic fungus *Beauveria bassiana*. *ChemBioChem*, **8**, 289–297.

85 Maiya, S., Grundmann, A., Li, X., Li, S.-M. and Turner, G. (2007) Identification of a hybrid PKS/NRPS required for pseurotin A biosynthesis in the human pathogen *Aspergillus fumigatus*. *ChemBioChem*, **8**, 1736–1743.

86 Bok, J.W., Hoffmeister, D., Maggio-Hall, L.A., Murillo, R., Glasner, J.D. and Keller, N.P. (2006) Genomic mining for *Aspergillus* natural products. *Chemistry & Biology*, **13**, 31–37.

87 Sioud, S., Karray-Rebai, I., Aouissaoui, H., Aigle, B., Bejar, S. and Mellouli, L. (2007) Targeted gene disruption of the cyclo (L-Phe, L-Pro) biosynthetic pathway in *Streptomyces* sp. US24 strain. *Journal of Biomedicine & Biotechnology*, DOI: 10.1155/2007/91409.

88 Johnson, R., Voisey, C., Johnson, L., Pratt, J., Fleetwood, D., Khan, A. and Bryan, G. (2007) Distribution of NRPS gene families within the *Neotyphodium/Epichloë* complex. *Fungal Genetics And Biology*, **44**, 1180–1190.

Index

Amino Acids, Peptides and Proteins in Organic Chemistry.
Vol.2 – Modified Amino Acids, Organocatalysis and Enzymes. Edited by Andrew B. Hughes

ISBN: 978-3-527-32098-1

f

i

j

k