The authors' objective has been to concentrate on amino acids and peptides without detailed discussions of proteins, although the book gives all the essential background chemistry, including sequence determination, synthesis and spectroscopic methods, to allow the reader to appreciate protein behaviour at the molecular level. The approach is intended to encourage the reader to cross classical boundaries, such as in the later chapter on the biological roles of amino acids and the design of peptide-based drugs. For example, there is a section on enzyme-catalysed synthesis of peptides, an area often neglected in texts describing peptide synthesis.

This modern text will be of value to advanced undergraduates, graduate students and research workers in the amino acid, peptide and protein field.

Amino Acids and Peptides

Amino Acids and Peptides

G. C. BARRETT

AND

D. T. ELMORE

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, United Kingdom
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1998

Printed in the United Kingdom at the University Press, Cambridge

Typeset in Times NR 10/13pt [SE]

A catalogue record for this book is available from the British Library

ISBN 0 521 46292 4 hardback
ISBN 0 521 46827 2 paperback

Contents

Foreword

This is an undergraduate and introductory postgraduate textbook that gives information on amino acids and peptides, and is intended to be self-sufficient in all the organic and analytical chemistry fundamentals. It is aimed at students of chemistry, and allied areas. Suggestions for supplementary reading are provided, so that topic areas that are not covered in depth in this book may be followed up by readers with particular study interests.

A particular objective has been to concentrate on amino acids and peptides, as the title of the book implies; the exclusion of detailed discussion of proteins is deliberate, but the book gives all the essential background chemistry so that protein behaviour at the molecular level can be appreciated.

There is an emphasis on the uses of amino acids and peptides, and on their biological roles and, while Chapter 8 concentrates on this, a scattering of items of information of this type will be found throughout the book. Important pharmaceutical developments in recent years underline the continuing importance and potency of amino acids and peptides in medicine and the flavour of current research themes in this area can be gained from Chapter 9.

Supplementary reading
(see also lists at the end of each Chapter)

Standard Student Texts

Standard undergraduate Biochemistry textbooks relate the general field to the coverage of this book. Several such topic areas are covered in

Zubay, G. (1993) *Biochemistry*, Third Edition, Wm. C. Brown Communications Inc, Dubuque, IA

and

Voet, D. and Voet, J. G. (1995) *Biochemistry*, Second edition, Wiley, New York

Typically, these topic areas as covered by Zubay are

Chapter 3: 'The building blocks of proteins: amino acids, peptides and proteins'
Chapter 4: 'The three-dimensional structure of proteins'
Chapter 5: 'Functional diversity of proteins'

Removed more towards biochemical themes, are

Chapter 18: 'Biosynthesis of amino acids'
Chapter 19: 'The metabolic fate of amino acids'
Chapter 29: 'Protein synthesis, targeting, and turnover'

Voet and Voet give similar coverage in

Chapter 24: 'Amino acid metabolism'
Chapter 30: 'Translation' (i.e. protein biosynthesis)
Chapter 34: 'Molecular physiology' (of particular relevance to coverage in this book of blood clotting, peptide hormones and neurotransmitters)

Supplementary reading: suggestions for further reading

(a) Protein structure

Branden, C., and Tooze, J. (1991) *Introduction to Protein Structure*, Garland Publishing Inc., New York

(b) Protein chemistry

Hugli, T. E. (1989) *Techniques of Protein Chemistry*, Academic Press, San Diego, California
Cherry, J. P. and Barford, R. A. (1988) *Methods for Protein Analysis,* American Oil Chemists' Society, Champaign, Illinois

(c) Amino acids

Barrett, G. C., Ed. (1985) *Chemistry and Biochemistry of the Amino Acids*, Chapman and Hall, London
Barrett, G. C. (1993) in *Second Supplements to the 2nd Edition of Rodd's Chemistry of Carbon Compounds*, Volume 1, Part D: Dihydric alcohols, their oxidation products and derivatives, Ed. Sainsbury, M., Elsevier, Amsterdam, pp. 117–66
Barrett, G. C. (1995) in *Amino Acids, Peptides, and Proteins*, A Specialist Periodical Report of The Royal Society of Chemistry, Vol. 26, Ed. Davies, J. S., Royal Society of Chemistry, London (preceding volumes cover the literature on amino acids, back to 1969 (Volume 1))
Coppola, G. M. and Schuster, H. F. (1987) *Asymmetric Synthesis: Construction of Chiral Molecules using Amino Acids*, Wiley, New York
Dawson, R. M. C., Elliott, D. C., Elliott, W. H., and Jones, K. M. (1986) *Data for Biochemical Research*, Oxford University Press, Oxford

Greenstein, J. P., and Winitz, M. (1961) *Chemistry of the Amino Acid*s, Wiley, New York (a facsimile version (1986) of this three-volume set has been made available by Robert E. Krieger Publishing Inc., Malabar, Florida)
Williams, R. M. (1989) *Synthesis of Optically Active α-Amino Acids*, Pergamon Press, Oxford

(d) Peptides

Bailey, P. D. (1990) *An Introduction to Peptide Chemistry*, Wiley, Chichester
Bodanszky, M. (1988) *Peptide Chemistry: A Practical Handbook*. Springer-Verlag, Berlin
Bodanszky, M. (1993) *Principles of Peptide Synthesis*, Second Edition, Springer-Verlag, Heidelberg
Elmore, D. T. (1993) in *Second Supplements to the 2nd Edition of Rodd's Chemistry of Carbon Compounds*, Volume 1, Part D: Dihydric alcohols, their oxidation products and derivatives, Ed. Sainsbury, M., Elsevier, Amsterdam, pp. 167–211
Elmore, D. T. (1995) in *Amino Acids, Peptides, and Proteins*, A Specialist Periodical Report of The Royal Society of Chemistry, Vol. 26, Ed. Davies, J. S., Royal Society of Chemistry, London (preceding volumes cover the literature of peptide chemistry back to 1969 (Volume 1))
Jones, J. H. (1991) *The Chemical Synthesis of Peptides*, Clarendon Press, Oxford

1

Introduction

1.1 Sources and roles of amino acids and peptides

More than 700 amino acids have been discovered in Nature and most of them are α-amino acids. Bacteria, fungi and algae and other plants provide nearly all these, which exist either in the free form or bound up into larger molecules (as constituents of peptides and proteins and other types of amide, and of alkylated and esterified structures).

The twenty amino acids (actually, nineteen α-amino acids and one α-imino acid) that are utilised in living cells for protein synthesis under the control of genes are in a special category since they are fundamental to all life forms as building blocks for peptides and proteins. However, the reasons why all the other natural amino acids are located where they are, are rarely known, although this is an area of much speculation. For example, some unusual amino acids are present in many seeds and are not needed by the mature plant. They deter predators through their toxic or otherwise unpleasant characteristics and in this way are thought to provide a defence strategy to improve the chances of survival for the seed and therefore help to ensure the survival of the plant species.

Peptides and proteins play a wide variety of roles in living organisms and display a range of properties (from the potent hormonal activity of some small peptides to the structural support and protection for the organism shown by insoluble proteins). Some of these roles are illustrated in this book.

1.2 Definitions

The term '*amino acids*' is generally understood to refer to the *aminoalkanoic acids*, $H_3N^+—(CR^1R^2)_n—CO_2^-$ with $n = 1$ for the series of α-amino acids, $n = 2$ for β-amino acids, etc. The term '*dehydro-amino acids*' specifically describes *2,3-unsaturated (or 'αβ-unsaturated')-2-aminoalkanoic acids*, $H_3N^+—(C{=}CR^1R^2)—CO_2^-$.

However, the term '*amino acids*' would include all structures carrying amine and acid functional groups, including simple aromatic compounds, e.g. anthranilic acid,

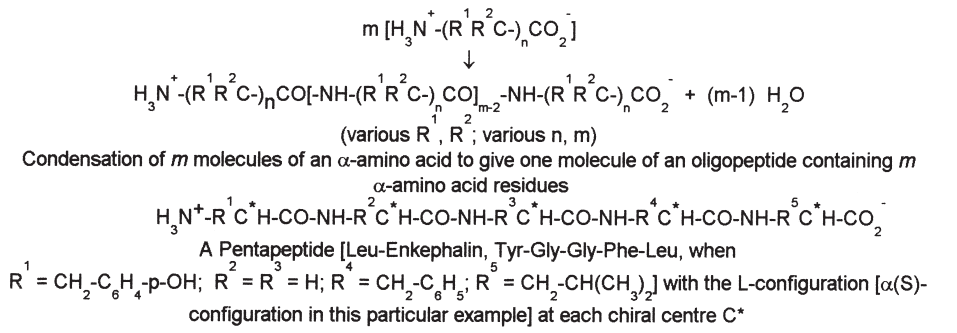


Figure 1.1. Peptides as condensation polymers of α-amino acids.

o-H_3N^+—C_6H_4—CO_2^-, and would also cover other types of acidic functional groups (such as phosphorus and sulphur oxy-acids, H_3N^+—$(R^1R^2C—)_nHPO_3^-$ and R_3N^+—$(R^1R^2C—)_nSO_3^-$, etc). The family of boron analogues $R_3N^{\bullet}BHR^1$—CO_2R^2 ($^{\bullet}$ denotes a dative bond) has recently been opened up through the synthesis of some examples (Sutton *et al*., 1993); it would take only the substitution of the carboxy group in these 'organoboron amino acids' ($R = R^1 = R^2 = H$) by phosphorus or sulphur equivalents to obtain an amino acid that contains no carbon! However, unlike the amino acids containing sulphonic and phosphonic acid groupings, naturally occurring examples of organoboron-based amino acids are not known.

The term '*peptides*' has a more restricted meaning and is therefore a less ambiguous term, since it covers polymers formed by the condensation of the respective amino and carboxy groups of α, β, γ . . . -amino acids. For the structure with $m = 2$ in Figure 1.1 (i.e., for a dipeptide) up to values of $m \simeq 20$ (an eicosapeptide), the term '*oligopeptide*' is used and a prefix *di-*, *tri-*, *tetra-*, *penta-* (see Leu-enkephalin, a linear pentapeptide, in Figure 1.1), . . . *undeca-* (see cyclosporin A, a cyclic undecapeptide, in Figure 1.4 later), *dodeca-*, . . . etc. is used to indicate the number of *amino-acid residues* contained in the compound. *Homodetic* and *heterodetic* peptides are illustrated in Chapter 7.

Isopeptides are isomers in which amide bonds are present that involve the *side-chain amino group* of an αω-di-amino acid (e.g. lysine) or of a poly-amino acid and/or *the side-chain carboxy-group of an* α*-amino-di- or -poly-acid* (e.g. aspartic acid or glutamic acid). Glutathione (Chapter 8) is a simple example. Longer polymers are termed '*polypeptides*' or '*proteins*' and the term '*polypeptides*' is becoming the most commonly used general family name (though proteins remains the preferred term for particular examples of large polypeptides located in precise biological contexts). Nonetheless, the relationship between these terms is a little more contentious, since the change-over from polypeptide to protein needs definition. The figure 'roughly fifty amino acid residues' is widely accepted for this. Insulin (a polymer of fifty-one α-amino acids but consisting of two crosslinked oligopeptide

chains; see Figure 1.4 later) is on the borderline and has been referred to both as a *small protein* and as a *large polypeptide*.

Poly(α-amino acid)s is a better term for peptides formed by the self-condensation of one amino acid; natural examples exist, such as poly(D-glutamic acid), the protein coat of the anthrax spore (Hanby and Rydon, 1946). In early research in the textile industry, poly(α-amino acid)s showed promise as synthetic fibres, but the synthesis methodology required for the polymerisation of amino acids was complex and uneconomic.

Polymers of controlled structures made from *N*-alkyl-α-amino acids (Figure 1.1; —NR^n instead of —NH—, $R^1 = R^2 = H$; $n = 1$), i.e. $H_2{}^+NR^n—CH_2CO—[NR^n—CH_2—CO—]_mNR^n—CH_2—CO_2^-$, which are poly(*N*-alkylglycine)s of defined sequence (various R^n at chosen points along the chain), have been synthesised as *peptide mimetics* (see Chapter 9) and have been given the name *peptoids*. These can be viewed as peptides with side-chains shifted from carbon to nitrogen; they will therefore have a very different conformational flexibility (see Chapter 2) from that of peptides and will also be incapable of hydrogen bonding. This is a simple enough way of providing all the correct side-chains on a flexible chain of atoms, in order to mimic a biologically active peptide, but the mimic can avoid enzymic breakdown before it reaches the site in the body where it is needed.

Using the language of polymer chemistry, polypeptides made from two or more different α-amino acids are *copolymers* or irregular poly(amide)s, whereas poly(α-amino acid)s, $H—[NH—CR^1R^2—CO—]_mOH$, are *homopolymers* that could be described as members of the nylon[2] family.

Depsipeptides are near-relatives of peptides, with one or more *amide bonds replaced by ester bonds*; in other words, they are formed by condensing α-amino acids with α-hydroxy-acids in various proportions. There are several important natural examples of these, of defined sequence; for example the antibiotic valinomycin and the family of enniatin antibiotics. Structures of other examples of depsipeptides are given in Section 4.8.

Nomenclature for conformational features of peptide structure is covered in Chapter 2.

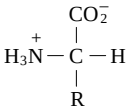
1.3 'Protein amino acids', alias 'the coded amino acids'

The twenty L-amino acids (actually, nineteen α-amino acids and one α-imino acid (Table 1.1)) which, in preparation for their role in protein synthesis, are joined *in vivo* through their carboxy group to tRNA to form α-aminoacyl-tRNAs, are organised by ribosomal action into specific sequences in accordance with the genetic code (Chapter 8).

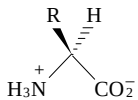
'Coded amino acids' is a better name for these twenty amino acids, rather than 'protein amino acids' or 'primary protein amino acids' (the term 'coded amino acids' is increasingly used), because changes can occur to amino-acid residues after they have been laid in place in a polypeptide by ribosomal synthesis. Greenstein and

Table 1.1. *The twenty 'coded' amino acids (nineteen 'coded'* L-α-*amino acids, and one 'coded'* L-α-*imino acid): structures and definitions*[a]

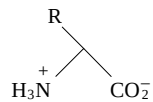
Structure conventions for the L-α-amino acids are



is equivalent to



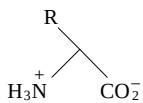
which is equivalent to

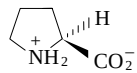


Fischer projection of an L-α-amino acid, requiring the carbon chain to be arranged vertically, with the carboxy group at the top

One of the commonly-used three-dimensional representations of an L-α-amino acid

Barrett representation of an L-α-amino acid

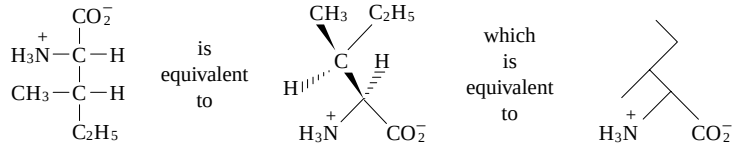
	Name of amino acid	Three-letter abbreviation	Single-letter abbreviation	Structures  *Amino acid side-chain, R =	Hydrophobicity High	Hydrophilicity High
One with no side-chain* (i.e. with a hydrogen atom)	Glycine	Gly	G	H		*
Four with saturated aliphatic side-chains* (hydrophobic side-chains)	Alanine	Ala	A	CH_3	*	
	Leucine	Leu	L	$CH_2CH(CH_3)_2$	*	
	Valine	Val	V	$CH(CH_3)_2$	*	
	Isoleucine	Ile	I	(S)-$CH(CH_3)C_2H_5$	*	

Ten with functionalised aliphatic side-chains* (mostly hydrophilic side-chains)	Arginine	Arg	R	$CH_2CH_2CH_2NHC(=NH)NH_2$					*
	Aspartic acid	Asp	D	CH_2CO_2H				*	
	Asparagine	Asn	N	CH_2CONH_2				*	
	Glutamic acid	Glu	E	$CH_2CH_2CO_2H$				*	
	Glutamine	Gln	Q	$CH_2CH_2CONH_2$				*	
	Lysine	Lys	K	$CH_2CH_2CH_2CH_2NH_2$					*
	Methionine	Met	M	$CH_2CH_2SCH_3$		*			
	Cysteine	Cys	C	CH_2SH		*			
	Serine	Ser	S	CH_2OH			*		
	Threonine	Thr	T	(R)-$CH(CH_3)OH$			*		
Four with aromatic or heteroaromatic side-chains* (most of these side-chains are hydrophobic)	Phenylalanine	Phe	F	$CH_2C_6H_5$	*				
	Tyrosine	Tyr	Y	CH_2-(p-OH-C_6H_4)				*	
	Histidine	His	H	CH_2-(imidazol-4-yl)					*
	Tryptophan	Trp	W	CH_2-(indol-3-yl)		*			
The 'coded' α-imino acid	Proline	Pro	P					*	

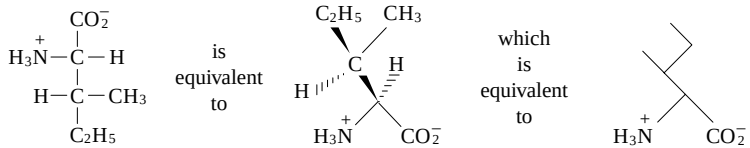
Notes:

1. The structure of each side-chain, R, is given for the 19 'coded α-amino acids', after each name. The full structure of the 'coded α-imino acid' proline is given. '*Three-letter*' and '*one-letter*' abbreviations are given for the 20. The *three-letter* abbreviation is the *first three letters of the name* for all twenty, *except* for asparagine (Asn), glutamine (Gln), isoleucine (Ile) and tryptophan (Trp). The *single-letter* abbreviated name is the first letter of their full name for *eleven* of them. Different letters are needed for the *other nine*, to avoid ambiguity: arginine (R), asparagine (N), aspartic acid (D), glutamic acid (E), glutamine (Q), lysine (K), phenylalanine (F), tryptophan (W) and tyrosine (Y).
2. All full names end in 'ine' except aspartic acid, glutamic acid and tryptophan. Adjectives are derived from the names by dropping the 'ine' or its equivalent ending and adding 'yl'; thus, alanyl, glutamyl, prolyl, tryptophyl, etc.
3. *Configurations.* The 'R/S' convention can easily be transferred to replace the Fischer 'D/L' system, while retaining the trivial names: L-enantiomers of all the coded amino acids are members of the S series except L-cysteine, which becomes R-cysteine through proper application of the R/S rules. Diastereoisomers (the isoleucine/allo-isoleucine and threonine/allothreonine pairs, 'allo' indicating inversion of the side-chain configuration of the coded amino acid) are less ambiguously named through the 'R/S' system, although the side-chain configuration can be indicated; for example, natural L-isoleucine is (2S,3S)-isoleucine:

Table 1.1. (*cont.*)



whereas L-alloisoleucine is (2S,3R)-isoleucine:



For the structures of natural L-*threonine* ((2S,3R)-threonine) and L-*allothreonine* ((2S,3S)-threonine), replace the side-chain ethyl group (C_2H_5) in isoleucine and alloisoleucine by OH.

4. *IUPAC–IUB nomenclature recommendations* (1983), reproduced in full in *Amino Acids, Peptides, and Proteins*, 1985, Vol. 16, The Royal Society of Chemistry, p. 387; and in *Eur.J.Biochem*., 1984, **138**, 9, encourage the retention of trivial names for the common α-amino acids, but systematic names are relatively straightforward; thus, L-alanine is 2S-aminopropanoic acid and L-histidine is 2S-amino-3-(imidazol-4-yl)-propanoic acid (the name for the predominant tautomer).

5. 'Hydrophilic' and 'hydrophobic' are terms used to denote the relative water-attracting and water-repelling property, respectively, of the side-chain when the amino acid is condensed into a polypeptide (see Chapter 5). The term 'hydropathy index' may be used to place the amino acids in order of their 'hydrophilicity' (Kyte and Doolittle, 1985), and their relative positions are shown here on an arbitrary scale.

[a] Selenocysteine (i.e. cysteine with the sulphur atom replaced by a selenium atom) has been found in certain proteins, e.g. formate dehydrogenase, an enzyme from *Escherichia coli*, and it has very recently been shown to be placed there through normal ribosomal synthesis (Stadtman, 1996). Thus selenocysteine can now be accepted as the 'twenty-first coded amino acid'.

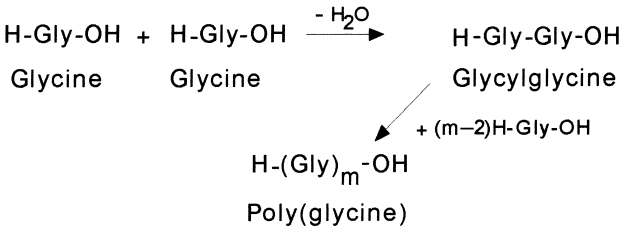


Figure 1.2. Polymerisation of glycine.

Winitz, in their 1961 book, listed 'the 26 protein amino acids', six of which were later found to be formed from among the other twenty 'protein amino acids' in the list of Greenstein and Winitz, after the protein had left the gene ('*post-translational* (sometimes called *post-ribosomal*) *modification*' or '*post-translational processing*'). Because of these changes made to the polypeptide after ribosomal synthesis, amino acids that are not capable of being incorporated into proteins by genes ('secondary protein amino acids', Table 1.2) can, nevertheless, be found in proteins.

1.4 Nomenclature for 'the protein amino acids', alias 'the coded amino acids'

The common amino acids are referred to through trivial names (for example, glycine would not be named either 2-aminoethanoic acid or amino-acetic acid in the amino acid and peptide literature). Table 1.1 summarises conventions and gives structures. The rarer natural amino acids are usually named as derivatives of the common amino acids, if they do not have their own trivial names related to their natural source (Table 1.2), but apart from these, there are occasional examples of the use of systematic names for natural amino acids.

1.5 Abbreviations for names of amino acids and the use of these abbreviations to give names to polypeptides

To keep names of amino acids and peptides to manageable proportions, there are agreed conventions for nomenclature (see the footnotes to Table 1.1). The simplest α-amino acid, glycine, would be depicted H—Gly—OH in the standard 'three-letter' system, the H— and —OH representing the 'H_2O' that is expelled when this amino acid undergoes condensation to form a peptide (Figure 1.2). The three-letter abbreviations therefore represent the 'amino-acid residues' that make up peptides and proteins.

So this '*three-letter system*' was introduced, more with the purpose of space-saving nomenclature for peptides than to simplify the names of the amino acids. A '*one-letter system*' (thus, glycine is G) is more widely used now for peptides (but is never used to refer to individual amino acids in other contexts) and is restricted to naming peptides synthesised from the coded amino acids (Figure 1.3).

Table 1.2. *Post-translational changes to proteins: the modified coded amino acids present in proteins, including crosslinking amino acids (secondary amino acids)*

Modifications to side-chain functional groups of coded amino acids

1. The aliphatic and aromatic coded amino acids may exist in αβ-dehydrogenated forms and the β-hydroxy-α-amino acids may undergo *post-translational dehydration*, so as to introduce αβ-dehydroamino acid residues, $-NH-(C=CR^1R^2)-CO-$, into polypeptides.

2. Side-chain OH, NH or NH_2 proton(s) may be substituted by *glycosyl*, *phosphate* or *sulphate*. These substituent groups are 'lost' during hydrolysis preceding analysis and during laboratory treatment of proteins by hydrolysis prior to chemical sequencing, which creates a problem that is usually solved through spectroscopic and other analytical techniques.

3. Side-chain NH_2 of lysine may be *methylated* or *acylated*: (N^{ε}-methylalanyl, N^{ε}-diaminopimelyl).

4. Side-chain NH_2 of glutamine may be *methylated*; giving N^5-methylglutamine, and the side-chain NH_2 of asparagine may be *glycosylated*.

5. Side-chain CH_2 may be *hydroxylated*, e.g. hydroxylysine, hydroxyprolines (trans-4-hydroxyproline in particular), or *carboxylated*, e.g. to give α-aminomalonic acid, β-carboxyaspartic acid, γ-carboxyglutamic acid, β-hydroxyaspartic acid, etc.

6. Side-chain aromatic or heteroaromatic moieties may be *hydroxylated*, *halogenated or* N-*methylated*.

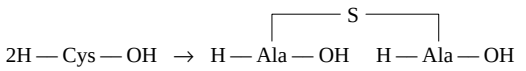
7. The side-chain of arginine may be modified (e.g. to give ornithine (Orn), $R=CH_2CH_2CH_2NH_2$, or citrulline (Cit), $R=CH_2CH_2CH_2NHCONH_2$).

8. The side-chain of cysteine may be modified, as in 1 above, also selenocysteine (CH_2SeH instead of CH_2SH; see footnote a to Table 1.1), lanthionine (see 10 below).

9. The side-chain of methionine may be S-alkylated (see Table 1.3) or oxidised at S to give methionine sulphoxide.

10. Crosslinks in proteins may be formed by condensation between nearby side-chains.
    (a) From lysine: e.g. lysinoalanine as if from [lysine+serine$-H_2O$]

    (b) From tyrosine: 3,3′-dityrosine, 3,3′,5′,3″-tertyrosine, etc.
    (c) From cysteine: oxidation of the thiol grouping ($HS-+-SH\rightarrow -S-S-$) to give the disulphide or to give cysteic acid (Cya): $-SH\rightarrow -SO_3H$ and alkylation leading to sulphide formation (e.g. alkylation as if by dehydroalanine to give lanthionine):



    (Further examples of crosslinking amino acids in peptides and proteins are given in Section 5.11.)

Nomenclature of post-translationally modified amino acids

Abbreviated names for close relatives of the 'coded amino acids' can be based on the 'three-letter' names when appropriate; thus, L-Pro after post-translational hydroxylation gives L-Hypro (trans-4-hydroxyproline, or (2S,4R)-hydroxyproline).

Current nomenclature recommendations (see footnote to Table 1.1) allow a number of abbreviations to be used for some non-coded amino acids possessing trivial names (some of which are used above and elsewhere in this book): Dopa, β-Ala, Glp, Sar, Cya, Hcy (homocysteine) and Hse (homoserine) are among the more common.

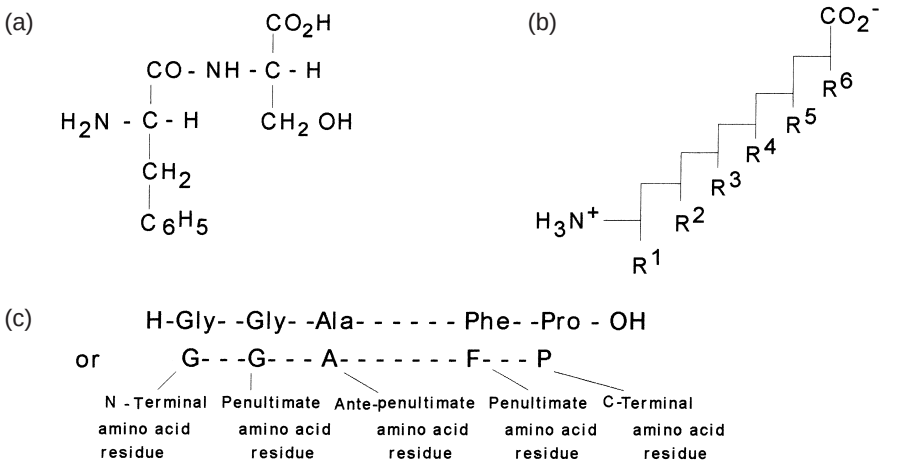


Figure 1.3. (a) The dipeptide L-phenylalanyl-L-serine in the Fischer depiction. (b) The schematic structure of a hexapeptide in the Fischer depiction, resulting in inefficient use of space. (c) The 'three-letter' and 'one-letter' conventions for a representative peptide, GGA---FP.

The 'three-letter system' has some advantages and has gradually been extended (Figure 1.4) to encompass several amino acids other than the coded amino acids. It is usually used to display schemes of laboratory peptide synthesis (Chapter 7) since it allows protecting groups and other structural details to be added, something that is very difficult and often confusing if attempted with the one-letter system.

The one-letter abbreviation (like its three-letter equivalent) represents 'an amino-acid residue' and the system allows the structure of a peptide or protein to be conveniently stated as a string of letters, written as a line of text, incorporating the long-used convention that the amino terminus (the '*N*-terminus') is to the LEFT and the carboxy terminus (the '*C*-terminus') is to the RIGHT. This convention originates in the Fischer projection formula for an L-α-amino acid or a peptide made up of L-α-amino acids; the L-configuration places the amino group to the left and the carboxy group to the right in a structural formula, as in Figure 1.3.

There are increasing numbers of violations of these rules; *N*-acetyl alanine, for example, being likely to be abbreviated Ac—Ala in the research literature or its correct abbreviation Ac—Ala—OH (but never Ac—A). This does not usually lead to ambiguity on the basis of the rule that peptide structures are written with the *N*-terminus to the left and the *C*-terminus to the right. Thus, Ac—Ala should still be correctly interpreted by a reader to mean CH_3—CO—NH—$CH(CH_3)$—CO_2H when this rule is kept in mind, since Ala—OAc (more correctly, H—Ala—OAc) would represent the 'mixed anhydride' NH_2—$CH(CH_3)$—CO—O—CO—CH_3 (there is a footnote about the term 'mixed anhydride' on p. 152).

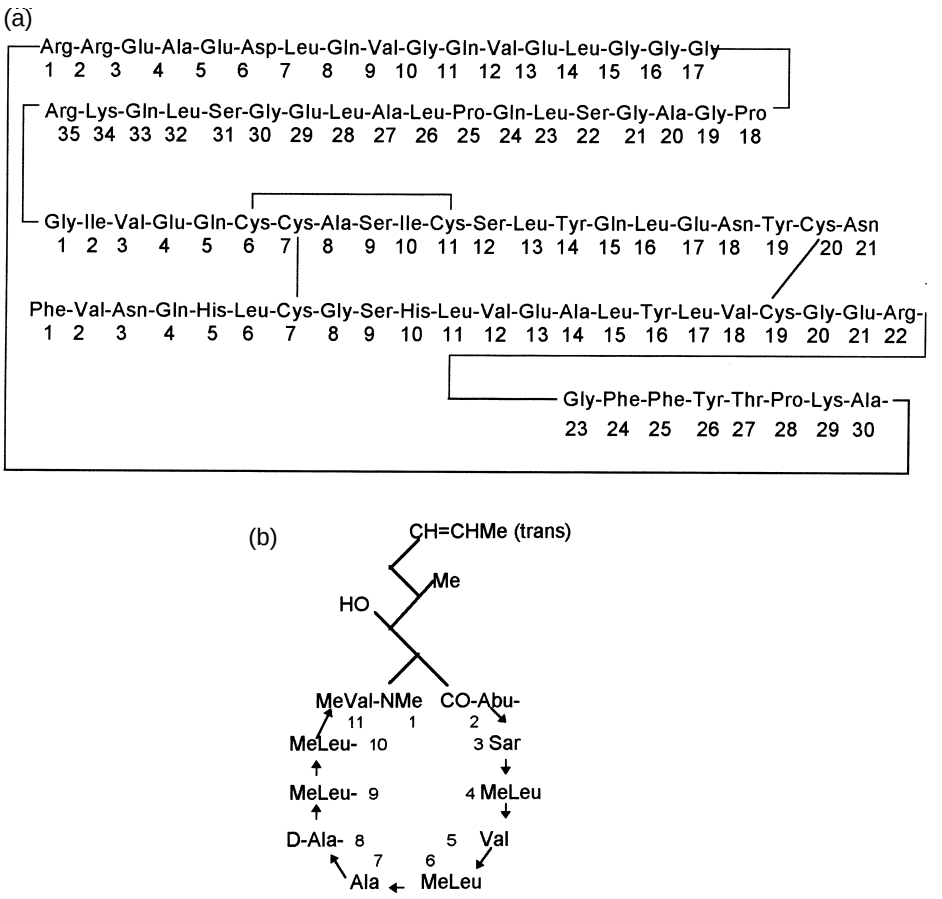


Figure 1.4. Post-translationally modified peptides: (a) Human proinsulin. (b) Cyclosporin A (Me is CH_3). As well as the post-translationally modified threonine derivative (residue 1, called 'MeBmt'), cyclosporin A contains one D-amino acid, four *N*-methyl-L-leucine residues, one 'non-natural' amino acid, Abu (butyrine, side-chain C_2H_5), Sar (sarcosine, *N*-methylglycine) and *N*-methyl-L-valine, but only two of the eleven residues are coded L-amino acids, valine and alanine.

Links through functional groups in side-chains of the amino-acid residues can be indicated in abbreviated structures of peptides (Figure 1.4). Cyclisation between the *C*- and *N*-termini to give a cyclic oligopeptide can also be shown in abbreviated structural formulae. Insulin (Figure 1.4) provides an example of the relatively common 'disulphide bridge' (there are three of these in the molecule), whereas cyclosporin A (a cyclic undecapeptide from *Trichoderma inflatum*, which is valuable for its immunosuppressant property that is exploited in organ-transplant surgery) is a product of post-translational cyclisation (Figure 1.4).

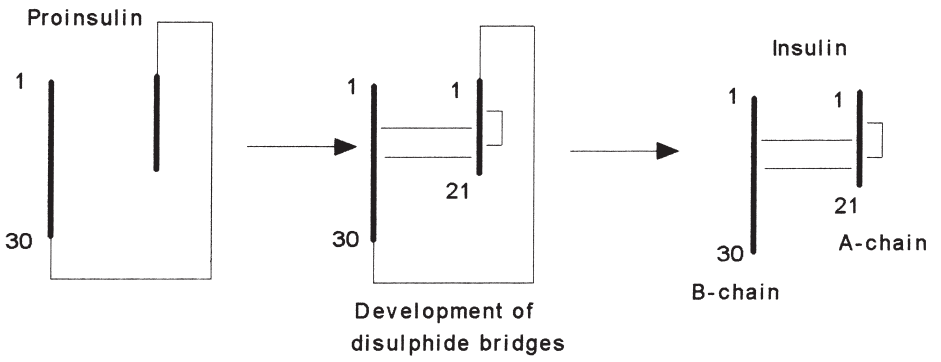


Figure 1.5. Generation of the active hormone, insulin, from the translated peptide, proinsulin (Chan and Steiner, 1977).

1.6 Post-translational processing: modifications of amino-acid residues within polypeptides

The major classes of *structurally altered amino-acid side-chains* within ribosomally synthesised polypeptides, which are achieved by intracellular reactions, are listed in Table 1.2 (see also Chapter 8).

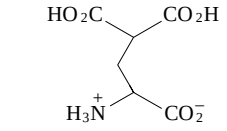
1.7 Post-translational processing: *in vivo* cleavages of the amide backbone of polypeptides

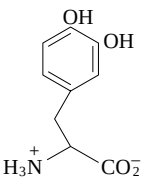
Changes to the amide backbone of the polypeptide through enzymatic cleavages transform the inactive polypeptide into its fully active shortened form. The polypeptide may be transported to the site of action after ribosomal synthesis and then processed there. Standard terminology has emerged for the extended polypeptides, *pre-*, *pro-* and *prepro-peptides* for the inactive *N*-terminal-extended, *C*-terminal-extended and *N*- and *C*-terminal extended forms, respectively, of the active compound. Figure 1.5 shows schematically the post-translational stages from human proinsulin (Figure 1.4) to insulin.

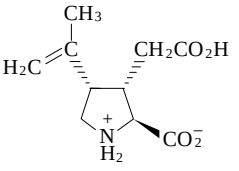
1.8 'Non-protein amino acids', alias 'non-proteinogenic amino acids' or 'non-coded amino acids'

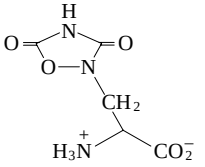
This further term is needed since there are several examples of higher organisms that utilise 'non-protein α-amino acids' that are available in cells in the free form (α-amino acids that are normally incapable of being used in ribosomal synthesis). Some of these free amino acids (Table 1.3) play important roles, one example being *S*-adenosyl-L-methionine, which is a 'supplier of cellular methyl groups'; for example, for the biosynthesis of neuroactive amines (and also for the biosynthesis of many

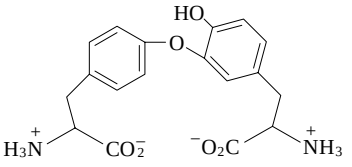
Table 1.3. *Some non-protein amino acids with biological roles, that are excluded from ribosomal protein synthesis*

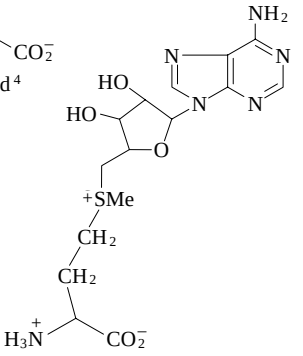

γ-Carboxy-L-glutamic acid[1]


L-Dopa[2]


Kainic acid[3]
(3S)-Carboxymethyl-(4S)-isopropenyl-(S)-proline


Quisqualic acid[4]


Di-L-tyrosine[5]


S-Adenosyl-L-methionine[6]

Notes:
1. γ-Carboxyglutamic acid, a constituent of calcium-modulating proteins (introduced through post-translational processing of glutamic acid residues).
2. Treatment for Parkinson's disease.
3. Potent excitatory effects, parent of a family of toxic natural kainoids present in fungi. Domoic acid, which has *trans*, *trans*-CHMe−CH=CH−CH=CH−CHMeCO_2H in place of the isopropenyl side-chain of kainic acid, is extraordinarily toxic, with fatalities ensuing through eating contaminated shellfish (Baldwin *et al*, 1990).
4. Exhibits potent NMDA receptor activity.
5. One of a number of protein crosslinks.
6. Widely distributed in cells.

other methylated species). Another physiologically important α-amino acid in this category is L-Dopa, the precursor of dopamine in the brain, which is used for treatment of afflictions such as Parkinson's disease and to bring about the return from certain comatose states (described in the book *Awakenings* by Oliver Sacks) that may be induced by L-Dopa.

Of course, most of the '700 or so natural amino acids' mentioned at the start of this chapter will be 'non-protein amino acids'. All these were, until recently, thought to be rigorously excluded from protein synthesis and other cellular events that are crucial to life processes, but a very few of these that are structurally related to the coded amino acids may be incorporated into proteins under laboratory conditions. This has been achieved by biosynthesising proteins in media that lack the required coded amino acid, but which contain a close analogue. For example, incorporation of the four-membered-ring α-imino acid azetidine-2-carboxylic acid instead of the five-membered-ring proline and incorporation of norleucine (side-chain $CH_2CH_2CH_2CH_3$) instead of methionine (side-chain $CH_2CH_2SCH_3$) have been demonstrated (Richmond, 1972) and β-(3-thienyl)-alanine has been assimilated into protein synthesis by *E. coli* (Kothakota *et al.*, 1995). Unusual amino acids that are not such close structural relatives of the coded amino acids have been coupled in the laboratory to tRNAs, then shown to be utilised for ribosomal peptide synthesis *in vivo* (Noren *et al.*, 1989).

Ways have been found, in the laboratory, of broadening the specificity of some enzymes (particularly the proteinases, but also certain lipases that can be used in laboratory peptide synthesis; see Chapter 7), for example by employing organic solvents, so that these enzymes catalyse some of the reactions of non-protein amino acid derivatives and some of the reactions of peptides that incorporate unusual amino acids. It has proved possible to involve D-enantiomers of the coded amino acids and D- and L-isomers of non-protein amino acids in peptide synthesis, to generate 'non-natural' peptides.

1.9 Coded amino acids, non-natural amino acids and peptides in nutrition and food science and in human physiology

The nutritional labels for some of the protein amino acids, such as 'essential amino acids', are an indication of their roles in this context. The meaning of the term '*essential*' differs from species to species and reflects the dependence of the organism on certain ingested amino acids that it cannot synthesise for itself, but which it needs in order to be able to generate its life-sustaining proteins. For the human species, the essential amino acids are the L-enantiomers of leucine, valine, isoleucine, lysine, methionine, threonine, phenylalanine, histidine and tryptophan. This implies that the other coded amino acids can be obtained from these essential amino acids, if not through other routes. There are some surprising pathways. For example, cysteine can be generated from methionine, the 'loss' of the side-chain carbon atoms being achieved through passage via cystathionine (Finkelstein, 1990); but homocysteine, the presence of which has been implicated as a causal factor in vascular disease, is also formed first in this route by demethylation of methionine. The D- and L-enantiomers of coded amino acids generally have different tastes and it has recently been appreciated that many fermented foods, such as yoghourt and shell-

fish (amongst many other food sources), contain substantial amounts of the D enantiomers of the coded amino acids.

The contribution of the D enantiomers to the characteristic taste of foods is currently being evaluated, but it is clear that the D enantiomers generally taste 'sweeter', or at least 'less bitter', than do their L isomers. Of course, kitchen preparation can involve many subtle chemical changes that enhance the attractiveness of natural foodstuffs, including racemisation (Man and Bada, 1987); therefore D enantiomers may be introduced in this way. Peptides are taste contributors, for example the bitter-tasting dipeptides Trp—Phe and Trp—Pro and the tripeptide Leu—Pro—Trp that are formed in beer yeast residues (Matsusita and Ozaki, 1993).

Some coded amino acids are acceptable as food additives and some are widely used in this way (e.g. L-glutamic acid and its monosodium salt). Addition of amino acids to the diet is unnecessary for people already eating an adequate and balanced food supply and the toxicity of even the essential amino acids (methionine is the most toxic of all the coded amino acids (Food and Drugs Administration, Washington USA, 1992)) should be better publicised, because some coded amino acids are easily available (for use in specialised diets by 'body-builders', for example) and are sometimes used unwisely. The use of L-tryptophan for its putative anti-depressant and other 'health' properties was responsible for the outbreak of eosinophilia myalgia syndrome that affected more than 1500 persons (with more than 30 fatalities) in the USA during 1989–90, though the problem was ascribed not to the amino acid itself but rather to an impurity introduced into the amino acid during manufacture (Smith *et al*., 1991). At the other end of the scale, some amino acids have more trivial uses, e.g. L-tyrosine in sun-tan lotion for cosmetic 'browning' of the skin.

Methionine is included in some proprietary paracetamol products (Pameton; Smith Kline Beecham), since it counteracts some serious side-effects that are encountered with paracetamol overdosing through helping to restore glutathione levels that are the body's natural defence against products of oxidised paracetamol. However, the recommended antidote (bearing in mind the toxicity of methionine) is intravenous *N*-acetyl-L-cysteine, which, in any case, reaches the liver of the overdosed patient faster.

Derivatives of aspartic acid have special importance in neurological research; the *N*-acetyl derivative is a putative marker of neurones and *N*-methyl-D-aspartic acid (NMDA) is creating interest for its possible links with Alzheimer's disease. NMDA is a potent excitant of spinal neurones; there are receptors in the brain for this α-imino acid, for which agonists/antagonists are being sought. A particular interaction being studied is that between ethanol and NMDA receptors (Collingridge and Watkins, 1994; see also Meldrum, 1991).

The industrial production base that has been developed to meet these demands (see Chapter 6) makes many amino acids cheaply available for other purposes such as laboratory use and has contributed in no small measure to the development of the biotechnological sector of the chemical industry.

1.10 The geological and extra-terrestrial distribution of amino acids

The development of sensitive analytical methods for amino acids became an essential support for the study of geological specimens (terrestrial ones and Lunar and Martian samples) from the 1970s. Some of the 'primary and secondary protein amino acids' (and some non-protein amino acids) were established to exist in meteorites (certainly in one of the largest known, the Murchison meteorite from Western Australia) though they have not been found in lunar samples. The scepticism that greeted an inference from this discovery – the inference that life as we know it exists, or once existed, on other planetary bodies – has also boosted interest in the chemistry of the amino acids to try to support alternative explanations for their presence in meteorites. The possibility that such relatively sensitive compounds could have survived the trauma experienced by meteorites penetrating the Earth's atmosphere was soon rejected. They must have been synthesised in the meteorites during the final traumatic stage of their journey. This conclusion was obtained bearing in mind the relevant amino-acid chemistry (Chapter 4); even the common, relatively much more gentle, laboratory practice of ultrasonic treatment of geological and biological samples prior to amino-acid analysis was hastily discouraged when it was found that this causes chemical structural changes to certain common amino acids (e.g. conversion of glutamic acid into glycine); and the injection of energy into mixtures of certain simple compounds also causes the formation of amino acids (Chapter 6).

The use of telescopic spectroscopy has revealed the existence of glycine in interstellar dust clouds. Since these clouds amount to huge masses of matter (greater than the total mass of condensed objects such as stars and planets), there must be universal availability of amino acids, even though they are dispersed thinly in the vast volume of space.

1.11 Amino acids in archaeology and in forensic science

Amino-acid analysis of relatively young fossils and of other archaeological samples has provided information on their age and on the average temperature profiles that characterised the Earth at the time of life of these samples. Samples from living organisms containing protein that has ceased turnover, i.e. proteins in metabolic *culs-de-sac* such as tooth and eye materials, can be analysed for their degree of racemisation of particular amino acids (Asp and Ser particularly; Leu for older specimens) in order to provide this sort of information. The D:L ratio for the aspartic acid present in these sources can be interpreted to assign an age to the organism, since racemisation of this amino acid is relatively rapid on the geological time scale and even in terms of life-span of a human being. The D:L ratio is easily measured through standard amino-acid analysis techniques (see Chapter 4; Bada, 1984).

D-Aspartic acid is introduced through racemisation into eye-lens protein in the

living organism at the rate of 0.14% per year, so that a 30-year-old person has accumulated 4.2% D-aspartic acid in this particular protein. It is in age determination of recently deceased corpses (and other 'scene-of-the-crime' artefacts, for which ^{14}C-dating is inaccurate), that forensic science interest in reliable amino-acid dating is centred. For older specimens, the method is wildly unreliable: thus, Otztal Ice Man – the corpse found at Hauslabjoch, high in the Austrian Tyrol, in 1991 – was dated to 4550 ± 27 BC by radiocarbon dating, but would have a grossly inaccurate assignment of birthday on the basis of amino-acid racemisation data (Bonani *et al.*, 1994). In unrelated areas, amino-acid racemisation has given useful information on the age of art specimens (e.g. dating of oil paintings through study of the egg-protein content).

Such inferences derive from data on the kinetics of racemisation, measured in the laboratory (described in Section 4.18.2) and there is a good deal of controversy surrounding the dating method since no account is taken of the catalytic influence on racemisation rates of molecular structures that surrounded the amino-acid residue for some or all the years. It is, for example, now known that the rate of racemisation of an amino acid, when it is a residue in a protein, is strongly dependent on the nature of the adjacent amino acids in the sequence; the particular amino acid on which measurement is made might have been located in a racemisation-promoting environment for many years after the death of the organism.

1.12 Roles for amino acids in chemistry and in the life sciences

1.12.1 Amino acids in chemistry

The physiological importance of α-amino acids ensures a sustained interest in their chemistry – particularly in pharmaceutical exploration for new drugs, and for their synthesis, reactions and physical properties. As is often the case when the chemistry of a biologically important class of compounds is being vigorously developed, an increasing range of uses has been identified for α-amino acids in the wider context of stereoselective laboratory synthesis (including studies of biomimetic synthetic routes).

1.12.2 Amino acids in the life sciences

Apart from their main roles, particularly their use as building blocks for condensation into peptides and proteins, α-amino acids are used by plants, fungi and bacteria as biosynthetic building blocks. Many *alkaloids* are derived from phenylalanine and tyrosine (e.g. Figure 1.6; and *penicillins* and *cephalosporins* are biosynthesized from tripeptides, Chapter 8).

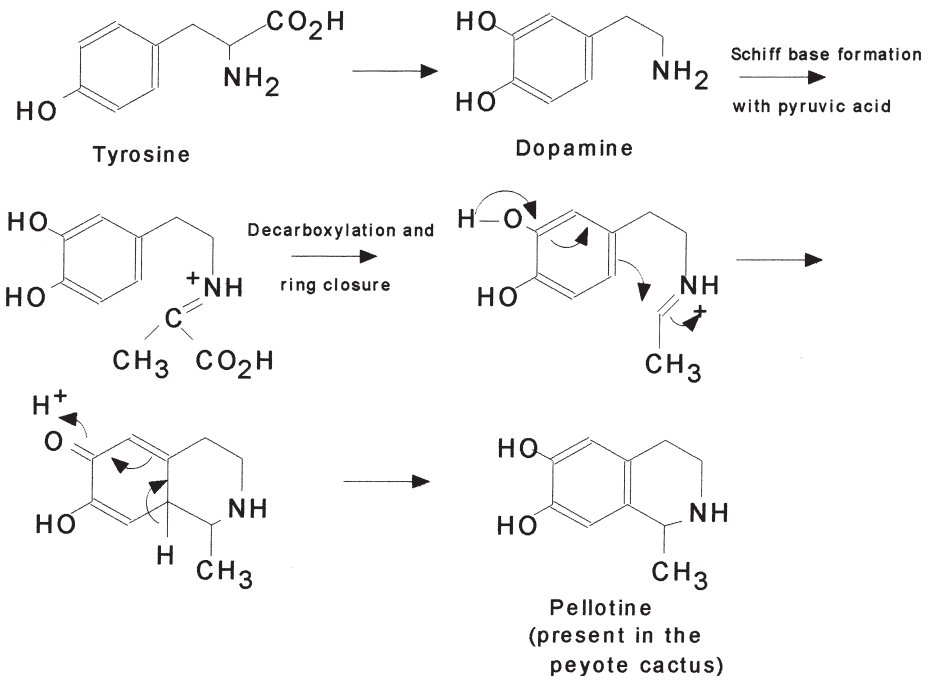


Figure 1.6. Routes from L-tyrosine to alkaloids. Alkaloid biosynthesis is often grouped into categories based on the initiating amino acid; i.e. the ornithine/cysteine route (e.g. nicotine); the phenylalanine/tyrosine/tryptophan route (e.g. the isoquinoline alkaloids, such as pellotine); etc.

1.13 β- and higher amino acids

There are relatively few examples; but there are increasing numbers of amino acids with greater separation of the amino and carboxy functions that have been found to play important biological roles (Drey, 1985; Smith, 1995). The coded amino acid, aspartic acid, could be classified either as an α- or as a β-amino acid. Glutamic acid (which can be classified either as an α-amino acid or as a γ-amino acid) is the biological source, through decarboxylation, of γ-aminobutyric acid (known as GABA; see Table 1.4), which functions as a neurotransmitter (as do glycine and L-glutamic acid and, probably, three other coded L-α-amino acids). The simple tripeptide glutathione (actually, an isopeptide; see Section 1.2 and Chapter 8) is constructed using the side-chain carboxy group rather than the α-carboxy group of glutamic acid and therefore could be said to be a peptide formed by the condensation of a γ-amino acid and two α-amino acids.

Numerous natural peptides with antibiotic activity and other intensely potent physiological actions incorporate α- and higher amino acids, as well as highly processed coded amino acids. The microcystins, which act as hepatotoxins, provide one

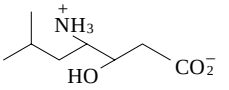
Table 1.4. *Some* β*-amino acids and higher amino acids found in biological sources*

Mentioned elsewhere in this chapter, as examples that are α-amino acids and also γ-, β- and δ-amino acids, respectively, are

Glutamic acid[b] $H_3N^+CH(CO_2^-)CH_2CH_2CO_2H$,
Aspartic acid[b] $H_3N^+CH(CO_2^-)CH_2CO_2H$ and
δ-Amino-adipic acid[c] $H_3N^+CH(CO_2^-)CH_2CH_2CH_2CO_2H$

β-Alanine[b] (β-Ala) $H_3N^+CH_2CH_2CO_2^-$
γ-Aminobutyric acid[a] (GABA) $H_3N^+CH_2CH_2CH_2CO_2^-$

Statine[c] (3S,4S)-3-hydroxy-4-amino-6-methylheptanoic acid



β-Phenylisoserine[c] [(2R,3S)-3-amino-2-hydroxy-3-phenylpropanoic acud; AHPA], $C_6H_5CH(^+NH_3)CH(OH)CO_2^-$, present in taxol (a potent anti-cancer agent) and present in bestatin, $^+NH_3CH(CH_2C_6H_5)CH(OH)CONHCH[CH_2CH(CH_3)_2]CO_2^-$ (an immunological response-modifying agent)

δ-Aminolaevulinic acid[a] $H_3N^+CH_2COCH_2CH_2CO_2^-$ (an analogue with a C=C grouping is the active constituent of light-activated ointments for the treatment of skin cancer)

Notes:
Some of these naturally occurring amino acids are:
[a] found only in the free state and not found in peptides;
[b] found in the free state and also found in peptides; and
[c] found only in peptides and other derivatised forms.

example. They are represented by the family structure cyclo[—D-Ala—X—D-MeAsp—Z—Adda—D-Glu—Mdha—), where X and Z are various coded L-amino acids and D-MeAsp is D-erythro-β-methylaspartic acid, found in the water bloom-forming cyanobacterium *Oscillatoria agardhii*. The structure of one of these, [D-Asp[3],DHb[7]]microcystin-RR (Sano and Kaya, 1995), is displayed in Chapter 3 (Figure 3.6).

Moving away from the simpler α-amino acids as constituents of peptides, the γ-amino acid (R)-carnitine $Me_3N^+CH_2CH(OH)CH_2CO_2^-$, is a rare example of a free amino-acid derivative with an important physiological role. This amino acid betaine is sometimes called 'vitamin B_T' and plays a part in the conversion of stored body fat into energy, through transport of fat molecules of high relative molecular mass to the sites of their conversion.

The (2R,3S)-phenylisoserine side-chain at position 13 of the taxane skeleton in the anti-cancer drug taxol (from the yew tree) is essential to its action.

1.14 References

Reviews providing information on all aspects of amino-acid science (Barrett, 1985; Greenstein and Winitz, 1961; Williams, 1989) and peptide chemistry (Jones, 1991) are listed at the end of the Foreword. References cited in the text of this chapter are the following.

Bada, J. L. (1984) *Methods Enzymol.*, **106**, 98.

Baldwin, J. E., Maloney, M. G. and Parsons, A. F. (1990) *Tetrahedron*, **46**, 7263.

Bonani, G., Ivy, S. D., Hajdas, I., Niklaus, T. R. and Suter, M. (1994) *Radiocarbon*, **36**, 247.

Chan, S. J. and Steiner, D. F. (1977) *Trends Biochem. Sci.*, **2**, 254.

Collingridge, G. L. and Watkins, J. C. (1994) *The NMDA Receptor*, Second Edition, Oxford University Press, Oxford.

Drey, C. N. C., in Barrett, G. C., Ed. (1985) *Chemistry and Biochemistry of the Amino Acids*, Chapman & Hall, London, p. 25.

Finkelstein, J. D. (1990) *J. Nutr. Biochem.*, **1**, 228.

Food and Drugs Administration, Washington, USA (1992) *Safety of Amino Acids used as Dietary Supplements*.

Hanby, W. S. and Rydon, H. N. (1946) *Biochem. J.*, **40**, 297.

Kothakota, S., Mason, T. L., Tirrell, D. A. and Fournier, M. J. (1995) *J. Am. Chem. Soc.*, **117**, 536.

Kyte, J. and Doolittle, R. F. (1985) *J. Mol. Biol.*, **157**, 105.

Man, E. H. and Bada, J. L. (1987) *Ann. Rev. Nutr.*, **7**, 209.

Matsusita, I., and Ozaki, S. (1993) *Peptide Chemistry; Proceedings of the 31st International Conference*, pp. 165–8 (*Chem. Abs.* **121**, 77 934).

Meldrum, B. S. (1991) *Excitatory Amino Acid Antagonists*, Blackwell, Oxford.

Noren, C. J., Anthony-Cahill, S. J., Griffiths, M. C. and Schultz, P. G. (1989) *Science*, **244**, 182.

Richmond, M. H. (1972) *Bacteriol. Rev.*, **26**, 398.

Sano, T. and Kaya, K. (1995) *Tetrahedron Lett.*, **36**, 8603.

Smith, B. (1995) *Methods of Non-α-Amino Acid Synthesis*, Dekker, New York.

Smith, M. J., Mazzola, E. P., Farrell, T. J., Sphon, J. A., Page, S. W., Ashley, D., Sirimanne, S. R., Hill, R. H. and Needham, L. L. (1991) *Tetrahedron Lett.*, **32**, 991.

Stadtman, T. C. (1996) *Ann. Rev. Biochem.*, **65**, 83.

Sutton, C. H., Baize, M. W. and Todd, L. J. (1993) *Inorg. Chem.,* **33**, 4221.