Chemistry and Biochemistry of the Amino Acids

EDITED BY

G.C. Barrett

Chemistry and Biochemistry of the Amino Acids

EDITED BY

G.C. Barrett
Oxford Polytechnic, UK

LONDON NEW YORK
Chapman and Hall

First published 1985 *by*
Chapman and Hall Ltd
11 New Fetter Lane, London EC4P 4EE

Published in the USA by
Chapman and Hall
733 Third Avenue, New York NY 10017

Printed in Great Britain
by J.W. Arrowsmith Ltd, Bristol

ISBN 0 412 23410 6

British Library Cataloguing in Publication Data

Barrett, G.C.
Chemistry and biochemistry of the amino acids.
1. Amino acids
I. Title
547.7'5 QD431

ISBN 0–412–23410–6

Library of Congress Cataloging in Publication Data

Main entry under title:

Chemistry and biochemistry of the amino acids.

Bibliography: p.
Includes index.
1. Amino acids. I. Barrett, G.C., 1935–
QD431.C45155 1985 574.19'245 83–15109
ISBN 0–412–23410–6 (U.S.)

Contributors

J.L. Bada
Amino Acid Dating Laboratory (A-012B)
Scripps Institution of Oceanography
University of California, San Diego
La Jolla, California 92093, USA

G.C. Barrett
Oxford Polytechnic
Headington, Oxford OX3 0BP, UK

D.A. Bender
Courtauld Institute of Biochemistry
The Middlesex Hospital Medical School
Mortimer Street
London W1P 7PN, UK

V. Cody
Medical Foundation of Buffalo Inc.
73 High Street, Buffalo
New York 14203, USA

J.S. Davies
Department of Chemistry
University College of Swansea
Singleton Park
Swansea SA2 8PP, UK

C.N.C. Drey
Robert Gordon's Institute of Technology
School of Chemistry, St Andrew Street
Aberdeen AB1 1HG, UK

M.H. Engel
School of Geology and Geophysics
830 Van Vleet Oval
The University of Oklahoma
Norman, Oklahoma 73019, USA

P.M. Hardy
Department of Chemistry
University of Exeter
Stocker Road, Exeter EX4 4QD, UK

P.E. Hare
Geophysical Laboratory
Carnegie Institution of Washington
2801 Upton Street, N.W.
Washington DC 20008, USA

S. Hunt
Department of Biological Sciences
University of Lancaster
Lancaster LA1 4YQ, UK

R.A.W. Johnstone
Department of Organic Chemistry
University of Liverpool, PO Box 147
Liverpool L69 3BX, UK

M.J. Jung
Centre de Recherche Merrell International
16 rue d'Ankara
67084 Strasbourg Cedex, France

P.J. Lea
Department of Biochemistry
Rothamsted Experimental Station
Harpenden, Herts AL5 2JQ, UK

T.H. Lilley
Department of Chemistry
The University
Sheffield S3 7HF, UK

J. Meienhofer
Chemical Research Department
Hoffmann-La Roche Inc.
Nutley, New Jersey 07110, USA

B.J. Miflin
Department of Biochemistry
Rothamsted Experimental Station
Harpenden, Herts AL5 2JQ, UK

D. Perrett
Department of Medicine
St. Bartholomew's Hospital Medical College
West Smithfield
London EC1A 7BE, UK

M.E. Rose
Department of Chemistry
Sheffield City Polytechnic
Pond Street
Sheffield S1 1WB, UK

G.A. Rosenthal
Thomas Hunt Morgan School of Biological Sciences
University of Kentucky
Lexington, Kentucky 40506, USA

P.A. St. John
Geophysical Laboratory
Carnegie Institution of Washington
2801 Upton Street, N.W.
Washington DC 20008, USA

C. Toniolo
Centro di Studi sui Biopolimeri
Istituto di Chimica Organica
Universita' di Padova
35100 Padova, Via Marzolo 1, Italy

R.M. Wallsgrove
Department of Biochemistry
Rothamsted Experimental Station
Harpenden, Herts AL5 2JQ, UK

Preface

Amino acids are featured in course syllabuses and in project and research work over a wide spectrum of subject areas in chemistry and biology. Chemists and biochemists using amino acids have many common needs when they turn to the literature for comprehensive information. Among these common interests, analytical studies, in particular, have undergone rapid development in recent years. All other chemical and biochemical aspects of amino acids – synthesis, properties and reactions, preparation of derivatives for use in peptide synthesis, racemization and other fundamental mechanistic knowledge – have been the subject of vigorous progress.

This book offers a thorough treatment of all these developing areas, and is structured in the belief that biochemists, physiologists and others will profit from access to information on topics such as the physical chemistry of amino acid solutions, as well as from thorough coverage of amino acid metabolism, biosynthesis and enzyme inhibition; and that chemists will find relevant material in biological areas as well as in the analysis, synthesis and reactions of amino acids.

It has been an important objective that comprehensive coverage of the amino acids should be contained within one volume. This has been achieved partly because much of the early 'pre-spectroscopic' chemical literature is so thoroughly covered by J.P. Greenstein and M. Winitz in their classic *Chemistry of the Amino Acids* (published by John Wiley & Sons in 1961), and partly because excellent coverage of some topic areas (such as thin layer chromatography, amino acid routine analysis) is also easily accessible. Economy has also been achieved by the conflation of some topic areas.

Grateful thanks to the publishers and their staff, and to Mary Emerson in particular for her help in the planning stages and who seemed to know just when an injection of help and encouragement was needed.

Graham Barrett
Oxford, 1984

Contents

CHAPTER ONE

Nomenclature of Amino Acids

G.C. BARRETT

1.1 INTRODUCTION

Tables 1 and 2 of the following chapter list the names, structures, and IUPAC–IUB recommendations for abbreviated names of the protein amino acids. Tables 1–21 in Chapter 4 give the same information for the non-protein amino acids, although there are no IUPAC–IUB recommendations for compounds in this category; if a non-protein amino acid cannot be named as a derivative of a protein amino acid, then it is named according to the generally accepted conventions [1].

This chapter is intended to provide a working guide to current recommendations [2] covering the nomenclature of amino acids, since it is an area in which considerable interchangeability is permitted between trivial and systematic conventions. The IUPAC recommendations [1] and the recommendations of a joint IUPAC–IUB working party [2] have been reprinted in a number of sources, including several primary journals and in the secondary literature [2].

1.2 TRIVIAL NAMES OF AMINO ACIDS

Recommendations dating from 1960 and endorsed by the working party which reported in 1975 [2] support the continuing tradition of using long-established trivial names for the protein amino acids. However, the coining of trivial names for newly-discovered α-amino acids should be avoided except for compelling reasons (e.g. if the systematic name is unreasonably cumbersome and lengthy), and these new compounds should be named as far as possible as derivatives of protein amino acids.

Some trivial names incorporate features which contradict other IUPAC recommendations. For example, norvaline $CH_3CH_2CH_2CH(\overset{+}{N}H_3)CO_2^-$ uses the prefix 'nor' to indicate 'normal' (i.e. an unbranched carbon chain). The prefix 'nor' is used in terpene nomenclature to denote replacement of all CH_3 groups attached to a ring system by H atoms.

(a) Acyl radicals

The ending 'ine' of the trivial name of an amino acid becomes 'yl' (with some exceptions) when arriving at the name for the acyl radical $RCH(NH_2)CO-$. Thus, with the exceptions aspartyl, asparaginyl, cysteinyl, glutamyl, glutaminyl, and

tryptophyl, the names for the acyl radicals are in accordance with this rule: alanyl, arginyl, histidyl, cystyl, and so on. Half-cystyl is retained as the name for the acyl radical of cysteine in which the SH proton is substituted by some other grouping or the group is ionized. Di-acyl radicals of aspartic and glutamic acids are aspartoyl and glutamoyl respectively; mono-acyl radicals of these amino acid are α- or β-aspartyl or α- or γ-glutamyl, as appropriate.

(b) Esters

While 'the benzyl ester of methionine' or 'methionine benzyl ester' would be as likely to be as acceptable to most authors and editors as 'benzyl methioninate', the recommendation is for the latter style. Thus [2], 'ic acid' at the terminus of aspartic and glutamic acids, and the terminal 'e' of the other protein amino acids, is replaced by 'ate' in naming the ester of the amino acid (but 'ate' is added to tryptophan).

(c) Amides

Although recommendations corresponding to those for the naming of esters can be followed (thus, glycinamide, phenylalaninanilide, and so on), the 1974 report [2] also allows glycine amide, leucine anilide, etc. It is curious that the alternative ways of naming amino acid esters are not also recommended by the 1974 IUPAC–IUB working party.

(d) Radicals formed by links through other positions

The above recommendations refer to acylation through the α-carboxy-group. Specification of the linkages through other positions is covered in the 1974 recommendations; if the N–H proton is substituted then the letter 'o' replaces the terminal 'e' of the trivial name of the amino acid (and the terms asparto, glutamo, and tryptophano are generated for the trivial names that do not end in the letter 'e'). This would suggest 2-alanino-ethanol rather than the commonly used *N*-(2-hydroxyethyl)alanine for $HOCH_2CH_2NHCH(CH_3)CO_2H$, but the rule is useful in more complex cases. Thus, N^6-lysino- or N^ε-lysino- for the radical derived from lysine, and π-histidino- for N^π-substituted histidines, become convenient prefixes.

If positions other than the carboxy or amino groups are involved in links, i.e. C–H, S–H or S–R, or O–H, then the ending 'yl' replaces 'e' at the end of the trivial name, with the usual exceptions. Thus we have cystein-*S*-yl, threonin-O^3-yl, but aspartamid-x-yl, glutamid-x-yl, aspartic-2-yl, glutamic-x-yl, and tryptophan-1-yl.

1.3 SYSTEMATIC NOMENCLATURE FOR AMINO ACIDS

The naming of a substituted alkanoic acid recommended for a general case [1] has been used for several important amino acids. For example, γ-amino-butyric acid, GABA, is within these recommendations since the trivial name 'butyric acid'

remains acceptable. The fully systematic name would be 4-aminobutanoic acid, and most higher homologous amino acids of this type have been named systematically. The best-known exception is β-alanine, $H_3\overset{+}{N}CH_2CH_2CO_2^-$, but this system is more widely used (e.g. leucine and β-leucine).

1.4 MISCELLANEOUS NOMENCLATURE RECOMMENDATIONS

(a) Numbering of atoms

Although Greek letters are still frequently used, numbering is becoming more widely adopted, but often erroneously. The carbon atom adjacent to the carboxy group is C-2 (carboxyl carbon is C-1), so 2-cyanoalanine for $NCCH_2CH(\overset{+}{N}H_3)CO_2^-$ is incorrect because it mixes the trivial nomenclature and numbering, and also because the cyano-group is on C-3 in this molecule; it is either β-cyanoalanine or 2-amino-3-cyano-propanoic acid.

Proline carbon atoms are numbered as in pyrrolidine, the nitrogen atom being numbered '1' and numbering then proceeds towards the carboxy group; again numbering of a trivial name crosses the trivial–systematic divide, but 3-methylproline, for example indicates widely-used nomenclature which is probably more common than Greek lettering around the ring.

The carbon atoms in the aromatic rings of phenylalanine, tyrosine, and tryptophan are numbered systematically, with 1 (or 3 for tryptophan) designating the carbon atom carrying the aliphatic chain (in which carbon atoms are labelled α or β). The histidine imidazole nitrogen atoms are designated *pros-* and *tele-* (meaning close and distant) and shown in names by using the Greek letters π and τ respectively, to indicate the nitrogen atom closest to the aliphatic chain, and that furthest from it. Alternatively, numbers may be used, with uncertainty about tautomers reflected in the numbering.

(b) The prefix 'homo'

This indicates the existence of one more CH_2 group in the side chain of the amino acid relative to the structure implied by the trivial name (e.g. homoserine, 2-amino-4-hydroxybutanoic acid, $HOCH_2CH_2CH(\overset{+}{N}H_3)CO_2^-$).

(c) Designation of stereochemistry: the α-carbon atom in α-amino acids

The absolute configuration of the α-carbon atom in α-amino acids is designated with the prefix D- or L- to indicate chemical correlation with D- or L-serine and thus with L- or D-glyceraldehyde, respectively. Using the Sequence Rule (the '*RS* System'), the common protein L-amino acids are *S*-amino-alkanoic acids (but L-cysteine is a member of the *R*-series).

An equimolar mixture of the two enantiomers of an amino acid with one chiral centre is prefixed DL- or (*RS*)-. The prefix *meso-* or its abbreviation *ms-* (using lower case italic letters) is used to denote those diastereoisomers of amino acids with more than one chiral centre which are optically-inactive because of internal compensation (e.g. *meso*-lanthionine and *ms*-cystine).

The prefixes D- or L- are placed immediately before the trivial name in

derivatives of common amino acids. Examples are N^{ε}-methyl-L-lysine, 3,5-di-iodo-L-tyrosine (but note: L-2-phenylglycine, L-phenylalanine, L-hydroxyproline and L-hydroxylysine).

(d) Designation of stereochemistry: configuration at chiral centres other than the α-carbon atom

The sequence rule offers the best solution to this problem, leading to names (3*S*)-L_s-threonine, (3*R*)-L_s-isoleucine, and (4*S*)-4-hydroxy-L_s-proline; the first two of these examples are allo-L-threonine and allo-L-isoleucine in more colloquial usage. The subscript 's' refers to the Fischer convention as used for amino acids: s stands for serine, the amino acid which has been configurationally correlated with D-glyceraldehyde. All other α-amino acids have been correlated with L-serine and carbohydrates have been correlated with D-glyceraldehyde. The term L_s avoids any confusion with the Fischer convention applied to carbohydrates (for carbohydrates the D- or L- refers to the highest-numbered chiral centre). The obvious improvement accompanying the full use of the *RS*-system can be seen in the unambiguous names (2*S*,3*R*)-threonine, (2*S*,3*S*)-isoleucine, and (2*S*,4*S*)-hydroxyproline for stereoisomers of these protein amino acids. Ring structures, of which the last-named amino acid is an example, have traditionally used stereochemical terms of geometrical isomerism, and this compound is then also trans-4-hydroxy-L_s-proline or erythro-4-hydroxy-L_s-proline. Allo-L_s-isoleucine could also be named threo-L_s-isoleucine.

(e) Optical rotation

Although only rarely used to specify an enantiomer in the case of common amino acids, if indication *is* given of the sign of optical rotation, it should precede all other aspects of the name of the compound (e.g. (+)-6-hydroxytryptophan or *dextro*-6-hydroxytryptophan). Depending on the conditions of the measurement, the sign of optical rotation of an L-amino acid may be + or − (and vice versa for the D-enantiomer: cf. L-serine, $[\alpha]_D^{20} = -6.83°$ ($c = 0.01\ g\,cm^{-3}$, H_2O), or $[\alpha]_D^{25} = +14.45°$ ($c = 0.009\ g\,cm^{-3}$, 1 M hydrochloric acid).

Thus, all four parameters defining the magnitude and, in some cases (such as L-serine), also the sign, of optical rotation should be specified, though the sign of rotation used should be assumed to refer to solutions of amino acids in water.*

1.5 DI(α-AMINO ACIDS)

(a) Those formed by C−N linkage

Secondary amines such as $HN\{(CH_2)_4CH(\dot{N}H_3)CO_2^-\}_2$ can be considered to be derived from two amino acids and are named accordingly: this example is 6-(N^6-lysino)norleucine.

* Note that change of concentration can invert the sign of the rotation; thus, (+)-tartaric acid in water shows positive optical rotations for dilute solutions, but the rotation is negative through the wavelength range 365–589 nm for concentrated aqueous solutions [3].

(b) Those formed by C – S, S – S, and other linkages

S-(Alanin-β-yl) cysteine, (lanthionine), $S\{(CH_2)_2CH(\overset{+}{N}H_3)CO_2^-\}_2$ illustrates this structural type and the name to be used. Many examples of this class occur naturally, and trivial names can also be used. Djenkolic acid is $Cys_2(CH_2)$ or 3,3′-methylenethiobis(2-aminopropanoic acid), $\{^-O_2CCH(\overset{+}{N}H_3)CH_2S\}_2CH_2$.

(c) Those formed by C – C linkages

Di(allysine) is the extraordinary name in use for $^-O_2CCH(\overset{+}{N}H_3)CH_2$-$CH_2CH(CHO)CH(OH)(CH_2)_3CH(\overset{+}{N}H_3)CO_2^-$, a naturally occurring amino acid which can be considered to be the aldol formed between two molecules of the aldehyde derived from lysine. Its systematic name would be 5-(6-hydroxy-norleucin-6-yl)-6-oxonorleucine. Although a fully systematic name could be composed for this compound (norleucine ≡ 2-aminohexanoic acid), in terms of the recommended nomenclature for α-amino acids, norleucine is retained.

REFERENCES

1. *Nomenclature of Organic Chemistry* (1979) prepared for publication by J. Rigaudy and S.P. Klesney, 4th edn, Pergamon Press, Oxford, p. 193.
2. *IUPAC Information Bulletin* (1975) Appendix No. 46, September. Reproduced in *Biochemistry* (1975) **14**, 449.
3. Cf. Lowry, T.M. *Optical Rotatory Power*, Longmans, Green, and Co., London, 1935 and Dover Reprints, 1964.

CHAPTER TWO

The Protein Amino Acids

P.M. HARDY

2.1 INTRODUCTION

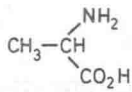
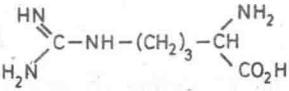
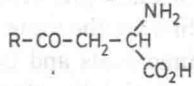
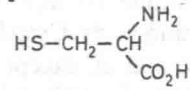
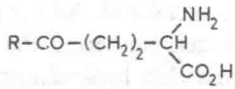
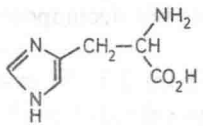
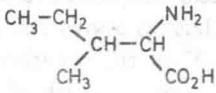
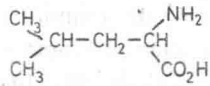
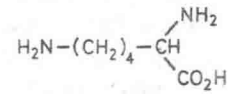
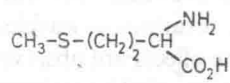
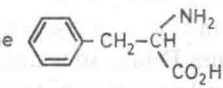
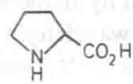
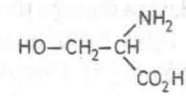
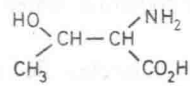
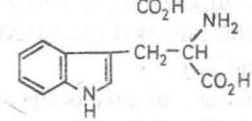
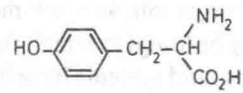
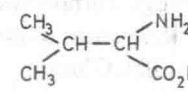
Amino acids are found in living organisms in both their free forms and bound by amide linkages in peptides and proteins. The diversity of structure observed in the free amino acids (several hundred are known) contrasts with the limited array of L-α-amino acids which are found in proteins and the peptides of higher organisms such as the mammals. The peptides of lower organisms, e.g. fungi and bacteria, are intermediate in the variety of their amino acids; cyclic structures, D-residues, and αβ-unsaturated residues are frequent components as well as α-hydroxy acids and non-protein amino acids.

The amino acid structures which occur in any given natural location reflect, of course, the local pathways of biosynthesis. Free amino acids arise solely through enzymic processes, but protein synthesis is a more complex procedure requiring both transfer RNAs and messenger RNA, and it takes place on a particular cellular particle, the ribosome. There is increasing evidence now that the peptides of higher organisms arise by cleavage of protein precursors, e.g. prohormones give rise to hormones, and therefore are constrained to the amino acid range found in proteins. The peptides of lower organisms are thought to originate in non-ribosomal enzymic processes, which are less discriminating: families of closely related compounds often occur.

It is possible to define protein amino acids in two ways. They may either be regarded as the components which are found in proteins as they emerge from the ribosome, i.e. the ones that are specifically coded for in the process of translation, or alternatively they may be considered as the components found in proteins which are of sufficient age for post-ribosomal modification of some residues to have taken place. For the purposes of this review, the former will be referred to as primary protein amino acids, and the latter as secondary. If modification involves the cross-linking of two amino acids, then the description tertiary protein amino acid is appropriate. In their definitive work on the amino acids in 1961, Greenstein and Winitz [1] listed twenty six protein amino acids, but six of these are now known to be secondary or tertiary amino acids. Of the twenty primary protein constituents (Table 2.1), nineteen are α-amino acids, but proline is a cyclic α-imino acid.

Although the genetic code is highly specific, non-coded amino acids can in certain circumstances be found in proteins. It has proved possible to incorporate structurally related analogues of the protein primary amino acids into bacterial

Table 2.1 The primary protein amino acids

Alanine		Arginine	
Aspartic acid (R = OH) Asparagine (R = NH_2)		Cysteine	
Glutamic acid (R = OH) Glutamine (R = NH_2)		Glycine	
Histidine		Isoleucine	
Leucine		Lysine	
Methionine		Phenylalanine	
Proline		Serine	
Threonine		Tryptophan	
Tyrosine		Valine	

protein in the absence of an adequate supply of the normal amino acid. Thus, for example, azetidine-2-carboxylic acid has been partially substituted into *E. coli* protein in place of proline, and norleucine in place of methionine [2]. However, such substitution occurs only under highly unnatural circumstances.

The structures of the primary protein amino acids were determined over the century following the characterization of glycine and leucine by Braconnet in 1820, threonine being the last of the twenty to be isolated in a pure form, in 1925. Some were prepared by chemical synthesis before being isolated from protein hydrolysates, e.g. alanine, and some initially obtained in a pure form from natural free amino acid sources, e.g. asparagine from asparagus juice. The fascinating history of the isolation and naming of the protein amino acids has been well reviewed[1,3,4], and will not be considered further here.

2.2 THE GENETICALLY CODED OR PRIMARY AMINO ACIDS

2.2.1 Physical properties

Table 2.2 outlines the more important physicochemical properties of the primary protein amino acids. Cysteine, lysine, and proline are the more water soluble of the twenty, and all except aspartic and glutamic acids and tyrosine exceed a solubility of 1 g per 100 cm^3 at 25°C. The aqueous solubility of tyrosine is too low to allow an accurate optical rotation to be measured. All the amino acids are more soluble in glacial acetic acid than water, rendering the former a useful component of solvent systems for paper and thin layer chromatography. The amino acids do not have sharp melting points but decompose over a range of several degrees above 200°C.

The pK of the α-carboxyl group varies from 1.71 in cysteine to 2.71 in threonine, while the pK of the α-amino groups extends from 8.18 in cysteine to 9.69 in alanine. In the case of fourteen of the primary amino acids, the specific optical rotations are higher in acidic solution than in simple aqueous solution. This is particularly useful in the case of alanine. The shapes of the optical rotatory dispersion curves are such as to give higher rotations at wavelengths lower than the sodium D line. In water, positive Cotton effects are observed with a peak at 216 nm or less; a change to 0.5 M HCl is accompanied by an increase in molecular rotation of 1000–1500° at the peak and a shift in the wavelength of the extremum to about 255 nm [5]. Circular dichroism curves of the primary amino acids show positive carboxyl-group Cotton effects at about 200 nm in water and 208–210 nm in acid [6].

The taste of most of the primary amino acids can be classified as either bitter or sweet, although authorities [4,7,8] vary in opinion, as assessment is essentially subjective. References to many of them can be found in which the tastes are cited as neutral or flat. Glutamic acid, methionine, and cysteine have flavours which do not fall into these categories, while lysine seems to be rather lacking in taste. On changing the configuration to the D-form, bitter amino acids become either neutral or sweet in taste, and sweet amino acids become sweeter. Protein amino acids are largely responsible for the taste of many types of foods, including soy sauce (soybean protein hydrolysed by microbial enzymes), seafood, and cheese. Many pleasant odours of cooked food are due to the interaction products of amino acids and sugars (the Maillard reaction); for example, proline and glucose give rise to the odour of newly backed bread [20].

The aliphatic members of the primary amino acids have no absorption in the ultraviolet region above 220 nm, but of course the aromatic amino acids histidine, phenylalanine, tryptophan, and tyrosine show characteristic maxima above 250 nm [9]. The zwitterionic character of the α-amino acids shows up clearly in their infrared spectra. No absorption due to the normal NH stretching frequency at 3300–3500 cm^{-1} is observed, indicating the lack of an NH_2 group. Instead, a peak near 3070 cm^{-1} due to the NH_3^+ is seen (except in proline, the NH_2^+