Pure & Appl. Chem., Vol. 56, No. 5, pp. 595-624, 1984.
Printed in Great Britain.

0033-4545/84 $3.00+0.00
Pergamon Press Ltd.

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

and

INTERNATIONAL UNION OF BIOCHEMISTRY

JOINT COMMISSION ON BIOCHEMICAL NOMENCLATURE*

NOMENCLATURE AND SYMBOLISM FOR AMINO ACIDS AND PEPTIDES

(Recommendations 1983)

*Membership of the Commission (JCBN) is as follows:

Chairman: H. B. F. DIXON (UK); *Secretary*: A. CORNISH-BOWDEN (UK); *Members*: C. LIÉBECQ (Belgium — as Chairman of IUB Committee of Editors of Biochemical Journals); K. L. LOENING (USA); G. P. MOSS (UK); J. REEDIJK (Netherlands); S. F. VELICK (USA); J. F. G. VLIEGENTHART (Netherlands).

Additional contributors to the formulation of these recommendations:

Nomenclature Committee of IUB (NC-IUB) (those additional to JCBN): H. BIELKA (GDR); N. SHARON (Israel); E. C. WEBB (Australia).

P. KARLSON (FRG, Past Chairman of JCBN); B. KEIL (France, a former Member of NC-IUB); W. E. COHN (USA); J. T. EDSALL (USA); J. S. MORLEY (UK); G. T. YOUNG (UK); and Members of IUPAC Commission on Nomenclature of Organic Chemistry.

IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN)

Nomenclature and Symbolism for Amino Acids and Peptides

Recommendations 1983

CONTENTS

These are recommendations of the IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), whose members are H.B.F.Dixon (chairman), A. Cornish-Bowden (secretary), C. Liébecq (as chairman of the IUB Committee of Editors of Biochemical Journals), K. L. Loening, G. P. Moss, J. Reedijk, S.F. Velick, and J. F. G. Vliegenthart. Comments may be sent to any member of the commission, or to its secretary: A. Cornish-Bowden, Department of Biochemistry, University of Birmingham, P. O. Box 363, Birmingham, England, B15 2TT.

JCBN thanks many who helped with drawing up the recommendations, especially P. Karlson, its former chairman, B. Keil, a former member of the Nomenclature Committee of IUB (NC–IUB), other members and former members of NC–IUB, namely H. Bielka, N. Sharon and E. C. Webb, and also W. E. Cohn, J. T. Edsall, J. S. Morley, G. T. Young, and members of the IUPAC Commission on Nomenclature of Organic Chemistry (CNOC).

Reproduced from *Eur. J. Biochem.*, Vol. 138, pp.9-37 (1984)
by courtesy of Springer-Verlag, Heidelberg, FRG.

INTRODUCTION

The traditional and well-known names of the common α-amino acids were, in general, given to them by their discoverers and bear no relationship to their chemical structures [1, 2]. The modification of these names to accommodate derivatives and to designate configuration was codified in 1947 [3] and revised in 1960 [4]. After proposals for the revision of the rules for naming α-amino acids with two centres of chirality had appeared in 1963 [5], a complete revision of the rules was made in 1974 [6] on the basis of a report by a committee convened by H. B. Vickery. Recommendations for symbols for amino-acid residues in peptide sequences made by Brand & Edsall (p. 224 in [8]) were revised in 1966 [9] and 1971 [10], and recommendations for a one-letter notation were approved in 1968 [11]. Recommendations for naming and symbolizing sequences derived from those of named peptides were made in 1966 [12].

The present revision combines all these documents. In Part 1 on *nomenclature*, the main changes are to propose names for particular ionic forms of residues (3AA-12.2) and to apply the stereochemical rules [13] more fully (3AA-3). Part 2 on *symbolism* introduces a few new symbols (3AA-15.2.7), simplifies the designation of ionized forms of peptides (3AA-19.3), explains the principles for giving symbols to reagents (3AA-18.2), presents a method for showing how parts of residues react (3AA-17.4), and describes the one-letter system for representing long sequences (3AA-20 and -21). Part 3, on the *modification of named peptides*, is extended to cover enantiomers and reversed sequences (3AA-22.7) and peptide analogues (3AA-22.8). Symbols for the twenty ribosomally incorporated (coded) amino acids are given in Table 1, and symbols used in these recommendations for other amino acids are mostly listed in the Appendix, although a few others are given in 3AA-15.2. Substantially new recommendations are marked by triangles in the margins.

Part 1. **Nomenclature**

Part 1, Section A: AMINO-ACID NOMENCLATURE

3AA-1. NAMES OF COMMON α-AMINO ACIDS

The trivial names of the α-amino acids that are commonly found in proteins and are represented in the genetic code, together with their symbols, systematic names [14] and formulas, are given in Table 1. Some other common amino acids are listed in the Appendix.

When the phrase 'amino acid' is a qualified noun it contains no hyphen; a hyphen is inserted when it becomes an adjective so as to join its components in qualifying another noun, e.g. amino-acid sequence.

3AA-2. FORMATION OF SEMISYSTEMATIC NAMES FOR AMINO ACIDS AND DERIVATIVES

3AA-2.1. *Principles of Forming Names*

Semisystematic names of substituted α-amino acids are formed according to the general principles of organic nomenclature [14], by attaching the name of the substituent group to the trivial name of the amino acid. The position of the substitution is indicated by locants (see 3AA-2.2). The configuration, if known, should be indicated (see 3AA-3, 3AA-4).

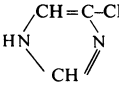
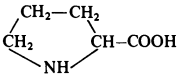
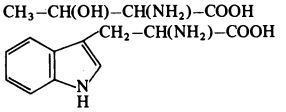
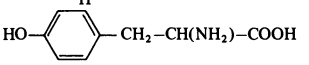
New trivial names should not be coined for newly discovered α-amino acids unless there are compelling reasons. When they are needed (e.g. because the substance is important and its semisystematic name is cumbersome), the name should be constructed according to the general principles for naming natural products [15], including either some element of its chemical structure or reference to its biological origin. It is important to use no elements in the trivial name that imply an incorrect structure; when a new trivial name is used, it is essential that it be defined by a correctly constructed systematic or semisystematic name.

A number of existing trivial names are given in the Appendix, and an extensive list has been published previously [6].

3AA-2.2 *Designation of Locants*

Note. The atom numbering given below is the normal chemical system for designating locants. A somewhat different system has been recommended for describing polypeptide conformations [16], in which Greek letters are used irrespective of the nature of the atom (unless it is hydrogen), so that in lysine N-6 becomes N^{ζ}, and in phenylalanine C-1, C-2 and C-6 become C^{γ}, $C^{\delta 1}$ and $C^{\delta 2}$ respectively.

Table 1. *α-Amino acids incorporated into protein under mRNA direction*
The systematic names and formulas given refer to hypothetical forms in which amino groups are unprotonated and carboxyl groups are undissociated. This convention is useful to avoid various nomenclatural problems but should not be taken to imply that these structures represent an appreciable fraction of the amino-acid molecules.

Trivial name[a]	Symbol	One-letter symbol[b]	Systematic name[c]	Formula
Alanine	Ala	A	2-Aminopropanoic acid	CH_3–$CH(NH_2)$–COOH
Arginine	Arg	R	2-Amino-5-guanidinopentanoic acid	H_2N–$C(=NH)$–NH–$[CH_2]_3$–$CH(NH_2)$–COOH
Asparagine	Asn[d]	N[d]	2-Amino-3-carbamoylpropanoic acid	H_2N–CO–CH_2–$CH(NH_2)$–COOH
Aspartic acid	Asp[d]	D[d]	2-Aminobutanedioic acid	HOOC–CH_2–$CH(NH_2)$–COOH
Cysteine	Cys	C	2-Amino-3-mercaptopropanoic acid	HS–CH_2–$CH(NH_2)$–COOH
Glutamine	Gln[d]	Q[d]	2-Amino-4-carbamoylbutanoic acid	H_2N–CO–$[CH_2]_2$–$CH(NH_2)$–COOH
Glutamic acid	Glu[d]	E[d]	2-Aminopentanedioic acid	HOOC–$[CH_2]_2$–$CH(NH_2)$–COOH
Glycine	Gly	G	Aminoethanoic acid	$CH_2(NH_2)$–COOH
Histidine	His	H	2-Amino-3-(1*H*-imidazol-4-yl)propanoic acid	
Isoleucine	Ile	I	2-Amino-3-methylpentanoic acid[e]	C_2H_5–$CH(CH_3)$–$CH(NH_2)$–COOH
Leucine	Leu	L	2-Amino-4-methylpentanoic acid	$(CH_3)_2CH$–CH_2–$CH(NH_2)$–COOH
Lysine	Lys	K	2,6-Diaminohexanoic acid	H_2N–$[CH_2]_4$–$CH(NH_2)$–COOH
Methionine	Met	M	2-Amino-4-(methylthio)butanoic acid	CH_3–S–$[CH_2]_2$–$CH(NH_2)$–COOH
Phenylalanine	Phe	F	2-Amino-3-phenylpropanoic acid	C_6H_5–CH_2–$CH(NH_2)$–COOH
Proline	Pro	P	Pyrrolidine-2-carboxylic acid	
Serine	Ser	S	2-Amino-3-hydroxypropanoic acid	HO–CH_2–$CH(NH_2)$–COOH
Threonine	Thr	T	2-Amino-3-hydroxybutanoic acid[e]	CH_3–$CH(OH)$–$CH(NH_2)$–COOH
Tryptophan	Trp	W	2-Amino-3-(1*H*-indol-3-yl)propanoic acid	
Tyrosine	Tyr	Y	2-Amino-3-(4-hydroxyphenyl)propanoic acid	
Valine	Val	V	2-Amino-3-methylbutanoic acid	$(CH_3)_2CH$–$CH(NH_2)$–COOH
▲ Unspecified amino acid	Xaa	X		

[a] The trivial name refers to the L or D or DL-amino acid; for those that are chiral only the L-amino acid is used for protein biosynthesis.
[b] Use of the one-letter symbols should be restricted to the comparison of long sequences (3AA-20).
[c] The fully systematic forms ethanoic, propanoic, butanoic and pentanoic may alternatively be called acetic, propionic, butyric and valeric, respectively. Similarly, butanedioic ≡ succinic, 3-carbamoylpropanoic ≡ succinamic, pentanedioic ≡ glutaric, and 4-carbamoylbutanoic ≡ glutaramic.
▲ [d] The symbol Asx denotes Asp or Asn; likewise B denotes N or D. Glx and Z likewise represent glutamic acid or glutamine or a substance, such as 4-carboxyglutamic acid, Gla (3AA-15.2.6), or 5-oxoproline, Glp (3AA-16.5), that yields glutamic acid on acid hydrolysis of peptides.
[e] See 3AA-3 and -4 for stereochemical designation.

2.2.1. Acyclic Amino Acids

In acyclic amino acids, the carbon atom of the carboxyl group next to the carbon atom carrying the amino group is numbered 1. Alternatively, Greek letters may be used, with C-2 being designated α. This practice is not encouraged for locants, although terms like 'α-amino acids' and 'α-carbon atom' are retained. Example:

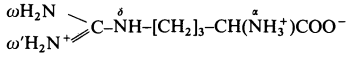
lysine H_3N^+–$\overset{6}{\underset{\varepsilon}{CH_2}}$–$\overset{5}{\underset{\delta}{CH_2}}$–$\overset{4}{\underset{\gamma}{CH_2}}$–$\overset{3}{\underset{\beta}{CH_2}}$–$\overset{2}{\underset{\alpha}{CH}}(NH_2)$–$\overset{1}{C}OO^-$

A heteroatom has the same number as the carbon atom to which it is attached, e.g. N-2 is on C-2. When such numerals are used as locants they may be written as N^6- or as 6-*N*, e.g. N^6-acetyllysine.

▲ The carbon atoms of the methyl groups of valine are numbered 4 and 4′; likewise those of leucine are 5 and 5′. Isoleucine is numbered as follows:

$$\begin{array}{l} \qquad\quad \overset{3^1}{CH_3} \\ \qquad\quad | \\ CH_3\text{–}CH_2\text{–}CH\text{–}CH(NH_3^+)\text{–}COO^- \\ 5 \quad 4 \quad 3 \quad 2 \quad\quad 1 \end{array}$$

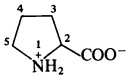
▲ The word 'methyl' can be italicized for use as a locant for substitution on (or isotopic modification {Section H in [14]} of) the methyl group of methionine, e.g. [*methyl*-^{14}C]methionine. The nitrogen atoms of arginine are designated as shown for the arginine (1+)cation:



It should be noted that the ω and ω' atoms of this cation are equivalent because of resonance. The carbon atom in the guanidino group may be called guanidino–C (it may be needed as a locant for isotopic replacement although it cannot carry a substituent).

2.2.2. Proline

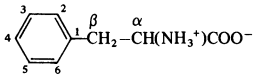
The carbon atoms in proline are numbered as in pyrrolidine, the nitrogen atom being numbered 1, and proceeding towards the carboxyl group.



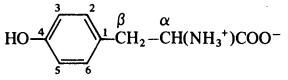
2.2.3. Aromatic Rings

The carbon atoms in the aromatic rings of phenylalanine, tyrosine and tryptophan are numbered as in systematic nomenclature, with 1 (or 3 for tryptophan) designating the carbon atom bearing the aliphatic chain. The carbon atoms of this chain are designated α (for the carbon atom attached to the amino and carboxyl groups) and β (for the atom attached to the ring system).

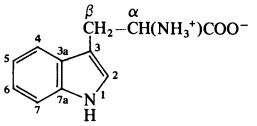
Note. This numbering should also be used for decarboxylated products (e.g. tryptamine).



phenylalanine



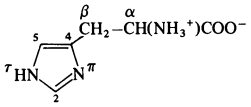
tyrosine



tryptophan

2.2.4 Histidine

The nitrogen atoms of the imidazole ring of histidine are denoted by *pros* ('near', abbreviated π) and *tele* ('far', abbreviated τ) to show their position relative to the side chain. This recommendation [6, 10] arose from the fact that two different systems of numbering the atoms in the imidazole ring of histidine had both been used for a considerable time (biochemists generally numbering as 1 the nitrogen atom adjacent to the side chain, and organic chemists designating it as 3). The carbon atom between the two ring nitrogen atoms is numbered 2 (as in imidazole), and the carbon atom next to the τ nitrogen is numbered 5. The carbon atoms of the aliphatic chain are designated α and β as in 2.2.1 and 2.2.3 above. This numbering should also be used for the decarboxylation product histamine and for substituted histidine.



histidine

2.2.5. Definition of Side Chain

▲ When amino acids are combined in proteins and peptides, C-1, C-2 and N-2 of each residue (the numbering being that of aliphatic amino acids) form the repeating unit of the main chain ('backbone') and the remainder forms a 'side chain'. Hence the words 'side chain' refer to C-3 and higher numbered carbon atoms and their substituents.

3AA-2.3. *Use of the Prefix 'homo'*

An α-amino acid that is otherwise similar to one of the common ones (Table 1), but that contains one more methylene group in the carbon chain, may be named by prefixing 'homo' to the name of that common amino acid. 'Homo' in the sense of a higher homologue (F-4.5 of [15]) is commonly used for homoserine (2-amino-4-hydroxybutanoic acid) and homocysteine (2-amino-4-mercaptobutanoic acid).

3AA-2.4. *Use of the Prefix 'nor'*

▲ The prefix 'nor' denotes removal of a methylene group (Sections F-4.2 and F-4.4 of [15]), but this is not the sense in which it has been used in the names 'norvaline' and 'norleucine'. Such names, although widely used, may therefore be misinterpreted, so

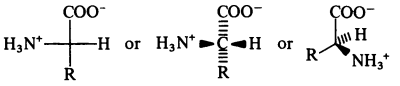
we cannot recommend them, especially since the systematic names for the compounds intended, 2-aminopentanoic acid and 2-aminohexanoic acid, are short.

3AA-3. CONFIGURATION AT THE α-CARBON ATOM

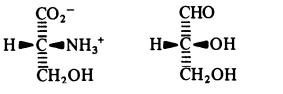
3AA-3.1. *Use of* D *and* L

The absolute configuration at the α-carbon atom of the α-amino acids is designated by the prefixed small capital letter D or L to indicate a formal relationship to D- or L-serine and thus to D- or L-glyceraldehyde. The prefix ξ (Greek xi) indicates unknown configuration.

The structures of amino acids may be drawn to show configuration in several ways [13]. In the Fischer-Rosanoff convention each chiral centre is projected onto the plane of the paper in the orientation such that the central atom appears as the point of intersection of two straight lines joining the attached groups in pairs, so that one straight line (which should be vertical) joins three atoms of the principal chain. The central atom is then considered to lie in the plane of the paper, the other atoms of the principal chain behind the plane from the viewer, and the remaining two groups in front of this plane. Thus an L-α-amino acid may be represented as



The relationship between serine and glyceraldehyde may therefore be represented as:



D-serine D-glyceraldehyde

3AA-3.2. *Position of Prefix*

In naming α-amino acids as derivatives of substances that have well-known trivial names, the prefix L or D is placed immediately before the trivial name of the parent amino acid and set off by a hyphen. Examples: *trans*-4-hydroxy-L-proline; 3,5-diiodo-L-tyrosine.

Note. Admissible exceptions to this rule are L-hydroxyproline and L-hydroxylysine, but only in general biochemical writing in a context such that the position of substitution is well understood. Note further that in the names of optically active derivatives of glycine, such as L-2-phenylglycine, the prefix must be placed before the name of the substituent as glycine itself is achiral. In the names of salts, esters and other derivatives, including peptides, the prefix is placed immediately before the trivial name of the parent acid or its radical. Examples: L-histidine monohydrochloride monohydrate; copper(II) L-aspartate; D-lysine dihydrochloride; *N*-acetyl-L-tryptophan; diethyl D-glutamate; N^6-methyl-L-lysine.

Other semisystematic names involving α-amino-acid configurations are treated similarly. Example: *S*-(D-2-amino-2-carboxyethyl)-D-homocysteine, or *S*-(D-alanin-3-yl)-D-homocysteine (3AA-8), i.e. D-cystathionine.

3AA-3.3. *Omission of Prefix*

The prefix may be omitted where the amino acid is stated to be or is obviously derived from a protein source and is therefore assumed to be L. It may also be omitted where the amino acid is synthetic and not resolved and is therefore, save in exceptional cases, an equimolecular mixture of the enantiomers. Likewise it may be omitted in a general statement that is true for either enantiomer or for any mixture of these.

3AA-3.4. *Subscripts to* D *and* L

Where confusion is possible between the use of the small capital letter prefix for the configuration of the α-carbon atom in amino-acid nomenclature and for that of the highest numbered chiral carbon atom in carbohydrate nomenclature [17], a subscript (lower case Roman letter) is added to the small capital letter prefix. If the prefix is used in the amino-acid sense, the subscript is s (for serine); if the prefix is used in the carbohydrate sense, the subscript is g (for glyceraldehyde).

Examples: L_s-threonine, for which the synonym in carbohydrate nomenclature is 2-amino-2,4-dideoxy-D_g-threonic acid; D_s-threonine, for which the synonym is 2-amino-2,4-dideoxy-L_g-threonic acid; L_s-allothreonine, for which the synonym is 2-amino-2,4-dideoxy-L_g-erythronic acid; D_s-allothreonine, for which the synonym is 2-amino-2,4- dideoxy-D_g-erythronic acid.

Note that the subscripts are essential only in discussions where both amino-acid names and those of carbohydrate derivatives occur. Nevertheless, these subscripts are highly desirable if D or L is used in naming α-amino acids that possess more than one centre of chirality (see 3AA-4).

3AA-3.5. *The* RS *System*

A more general system of stereochemical designation, which is especially convenient when there is no simple way of relating a compound to a defined standard, is the *RS* system of Cahn, Ingold & Prelog [13, 18]. In this system the ligands of a chiral atom are placed in an order of preference, based largely on atomic number. If the first three ligands appear clockwise in this order when viewed from the side remote from the least-preferred (fourth) ligand, the chiral centre is *R*; if anticlockwise, it is *S*.

The L-configuration, possessed by the chiral α-amino acids found in proteins, nearly always corresponds to *S* in the *RS* system. The most important exceptions are L-cysteine and L-cystine (see Appendix), which are *R* (in most amino acids the order of preference of the groups around C-2 is NH_3^+, COO^-, R, H, but in cysteine and cystine the group R takes precedence over carboxylate because the atomic number of sulfur attached to C-3 is higher than that of oxygen attached to C-1).

3AA-3.6. *Amino Acids Derived from Amino Sugars*

Amino acids that are derived from amino sugars and contain five or more carbon atoms are named in conformity with the system of carbohydrate nomenclature [17] or with a recommended trivial name.

Examples: (1) D_g-glucosaminic acid for 2-amino-2-deoxy-D_g-gluconic acid, the α-carbon of which has the configuration of that in D-serine, and in which C-5, the highest numbered chiral centre, also has the D-configuration; (2) D_g-mannosaminic acid for 2-amino-2-deoxy-D_g-mannonic acid, the α-carbon of which has the configuration of that in L-serine, but in which C-5 has the D configuration. The subscript g may be omitted unless confusion with the amino-acid use of the designations D and L is likely.

3AA-3.7. *Use of* meso

The prefix *meso-*, in lower case italic letters, is used to denote those amino acids or derivatives that, although they contain chiral groups, are achiral, usually because of a plane of symmetry, e.g. *meso*-lanthionine.

3AA-3.8. *Use of* DL

A mixture of equimolar amounts of D and L compounds is termed racemic and is designated by the prefix DL (no comma), e.g. DL-leucine. It may alternatively be designated by the prefix *rac-* (e.g. *rac*-leucine) or by the prefix (±)- (see 3AA-5).

3AA-4. CONFIGURATION AT CHIRAL CENTRES OTHER THAN THE α-CARBON

3AA-4.1. *The Sequence Rule*

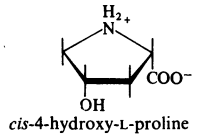
The *RS* system (3AA-3.5) is preferred for designating configuration at centres other than α-C, e.g. (2*S*,3*R*)-threonine. To avoid using two different systems of designation in the same name, (2*S*,4*S*)-4-hydroxyproline may be used instead of (4*S*)-4-hydroxy-L-proline.

3AA-4.2 *Carbohydrate Prefixes*

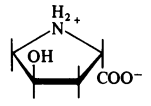
▲ The use of carbohydrate prefixes (e.g., D-*erythro*) cited in the 1974 version of these recommendations [6] as an alternative system for α-amino acids having two or more chiral centres is now discouraged.

3AA-4.3. *Use of* cis *and* trans

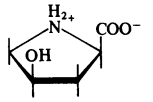
The amino acids 4-hydroxy-L-proline and 3-hydroxy-L-proline and analogous substituted prolines may also be named as follows (cf. 3AA-3.2).



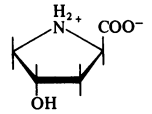
cis-4-hydroxy-L-proline



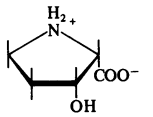
trans-4-hydroxy-L-proline



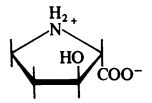
cis-4-hydroxy-D-proline



trans-4-hydroxy-D-proline



cis-3-hydroxy-L-proline



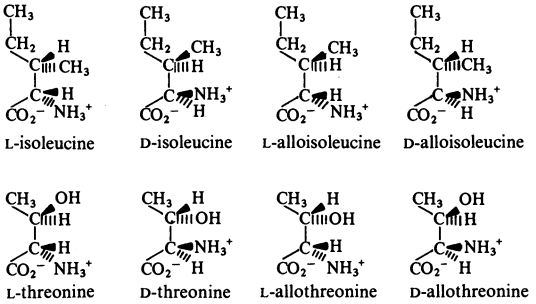
trans-3-hydroxy-L-proline

The prefixes *cis* and *trans* refer to the relative positions of the hydroxyl and carboxyl groups in each compound.

Comment. The hydroxyprolines found in collagen are *trans*-4-hydroxyproline (predominantly) and *trans*-3-hydroxyproline. The prefixes may be omitted when no ambiguity arises (cf. 3AA-3.3).

3AA-4.4. *Use of 'allo'*

Amino acids with two chiral centres were named in the past by allotting a name to the first diastereoisomer to be discovered. The second diastereoisomer, when found or synthesized, was then assigned the same name but with the prefix allo-. This method ▲ can be used only with trivial names (see 2.1) but not with semisystematic or systematic names. It is now recommended that allo should be used only for alloisoleucine and allothreonine, as follows:



3AA-4.5. *Designation of Centres with Unknown Configurations*

▲ When absolute or relative configurations at one or more centres are not known, such designations as 'isomer A' and 'isomer B' are frequently employed until the full configurational relationships are established.

▲ If the configuration is known at one centre but not at a second, the *RS* system is used for the known centre, with a Greek xi (ξ), meaning 'unknown configuration' for the other, e.g. (2*S*,5ξ)-2-amino-5-hydroxyhexanoic acid (a single stereoisomer). If the configuration at two centres is unknown, the ξ may be used as in the example (2ξ,5ξ)-2-amino-5-hydroxyhexanoic acid. If a racemate is to be designated, this is done by reference to its optical activity (3AA-5), e.g. ($\pm$)-(2ξ,5ξ)-2-amino-5-hydroxyhexanoic acid. If the relative configuration of two centres is known, but the absolute is unknown, '*R**' and '*S**' may be used, e.g. (2*R**,5*S**)-2-amino-5-hydroxyhexanoic acid.

3AA-4.6. *Other Stereochemical Features*

When other stereochemical elements are encountered, such as *E*/*Z* double-bond isomers, they are described according to the provisions of Section E of the IUPAC rules for organic nomenclature [13].

3AA-5. OPTICAL ROTATION

If it is desired to indicate the direction of rotation of plane polarized light of specified wavelength in a specified solvent, this can be done with a 'plus' or 'minus' sign in parenthesis (E-4.4 of reference [13]), e.g. (+)-6-hydroxytryptophan. This may be particularly useful if the configuration at C-2 is not known, but it may also be done for emphasis, with or without a configurational symbol D or L, when this configuration is known, e.g. (+)-glutamic acid, or (+)-L-glutamic acid. A racemic amino acid (3AA-3.8) may be indicated by ($\pm$), e.g. ($\pm$)-leucine.

Part 1, Section B: NOMENCLATURE OF NON-PEPTIDE DERIVATIVES OF AMINO ACIDS

If an amino acid is substituted on a saturated carbon, it remains an amino acid. Its naming is therefore described in 3AA-2. This section extends some of the procedures described there, and also covers modification of functional groups.

A number of special procedures are given below to allow names to be based on the trivial names of the α-amino acids, so that they may indicate biochemical relationships. These procedures, which yield names such as N^6-lysino (3AA-7), alanin-3-yl (3AA-8), leucinamide (3AA-9.2), phenylalanylchloromethane (3AA-10.2), alaninol (3AA-10.3), etc., should not be extended to other areas.

3AA-6. IONIZATION OF FUNCTIONAL GROUPS AND NAMING OF SALTS

The predominant form at pH 7 of a mono-amino-mono-carboxylic acid is R–CH(NH_3^+)–COO^- rather than R–CH(NH_2)–COOH. It is nevertheless often convenient to draw the latter conventional form (e.g. in Table 1) and to name alanine as 2-

aminopropanoic acid rather than as 2-ammoniopropanoate. This is particularly so for representing the isoelectric form of amino acids that contain other ionizing groups. A solution of lysine, for example, would contain appreciable amounts of both NH_3^+–$[CH_2]_4$–$CH(NH_2)$–COO^- and NH_2–$[CH_2]_4$–$CH(NH_3^+)$–COO^-.

When it is desirable to mention or stress the ionic nature of an amino acid, the three kinds of ions possible for a mono-amino-mono-carboxylic compound may be indicated as follows:

NH_3^+–CH_2–COO^-	glycine zwitterion (or dipolar ion, or amphion);
NH_3^+–CH_2–$COOH$	glycinium, or glycine cation;
NH_2–CH_2–COO^-	glycinate, or glycine anion.

In indicating an anion the ending 'ate' replaces 'ic acid' or the final 'e' of the trivial name, or is added to the name tryptophan.

Further forms are required for amino acids that contain ionizing side chains. The singly charged anions of aspartic and glutamic acids (strictly each has two negative and one positive charge, but this nomenclature refers to net charge) may be distinguished from the doubly charged anions by placing the charge after the name, or by stating the number of neutralizing ions. Thus the form of glutamate (glutamate refers to glutamic acid; glutaminate is the anion from glutamine) with a charge of minus one, ^-OOC–CH_2–CH_2–$CH(NH_3^+)$–COO^-, may be called glutamate(1-), glutamic acid monoanion, or hydrogen glutamate, and its sodium salt may be called sodium glutamate(1-), sodium hydrogen glutamate, or monosodium glutamate. The corresponding terms for the dianion, ^-OOC–CH_2–CH_2–$CH(NH_2)$–COO^-, include glutamate(2-), glutamic-acid dianion, and disodium glutamate. Unqualified, the word glutamate systematically means the dianion; hence the usage 'sodium hydrogen glutamate'; in normal use, however, it means the ion of net charge -1, since this is the form that predominates in neutral solution, and it is used in this way in, for example, 'a glutamate-dependent reaction' and 'glutamate dehydrogenase'.

Similarly, forms such as lysinium(1+) or lysine monocation may be used for the ion of unit net charge derived from lysine. Its salts may be indicated by adding the name of the anion to the lysinium form, e.g. lysinium(1+) chloride, or by naming it lysine monohydrochloride. The fully protonated form is the lysine dication or lysinium(2+).

3AA-7. AMINO ACIDS SUBSTITUTED ON NITROGEN

Since N-2 is the atom most easily modified in many amino acids, the locant can often be omitted without ambiguity, e.g. acetylglycine for *N*-acetylglycine.

It is sometimes convenient to use the name of a group derived by loss of hydrogen from a nitrogen atom of an amino acid as a prefix in forming another name. Such prefixes are formed by substituting 'o' for the terminal 'e' in those names that end in 'e' (by analogy with amine→amino); e.g. alanino, valino. Tryptophan adds the 'o' directly, and the two dicarboxylic acids become asparto and glutamo. Where there is more than one nitrogen atom in the amino acid, a locant of the form N^x must precede the group name. e.g. N^6-lysino, N^ω-arginino, N^5-glutamino, N^π-histidino.

3AA-8. SIDE-CHAIN MODIFICATIONS (excluding modifications of carboxyl or nitrogen)

Most modified amino acids can be named according to 3AA-2, e.g. *S*-(carboxymethyl)-L-cysteine. Groups formed by loss of hydrogen atoms from carbon, sulfur or oxygen atoms (excluding the carboxylic oxygen atoms, which are dealt with under 3AA-9) are named by substituting '-x-yl' for the terminal 'e' of the trivial name, where 'x' is the locant of the atom from which the hydrogen atom has been lost, e.g. cystein-*S*-yl, threonin-O^3-yl, alanin-3-yl, or by adding '-x-yl' to aspartic, glutamic and tryptophan, e.g. aspartic-2-yl, tryptophan-2-yl (see 3AA-2.2.3).

Comment. tryptophan-1-yl should be named 1-tryptophano according to 3AA-7.

A common side-chain modification is the oxidation of cysteine to yield cystine (formula in Appendix). Hydrogen atoms are removed from the –SH groups of two molecules, which are joined by an S–S bond. The term 'half cystine' refers to each half. It occurs seldom in naming compounds, since half a cystine molecule is a substituted cysteine and is named as such. In stating amount of substance, however, any specified entity may be used, so moles or numbers of residues of half cystine may usefully be compared with these quantities of other amino acids in stating protein composition.

3AA-9. ESTERS AND AMIDES OF THE CARBOXYL GROUP

3AA-9.1. *Esters*

Esters are named with the anion name (3AA-6), e.g. methyl prolinate, methyl cysteinate, or from the amino acid, e.g. proline methyl ester, cysteine methyl ester.

3AA-9.2. *Amides, Anilides and Analogous Derivatives* (H_2N–CHR–CO–NH–R′)

In amides, anilides and analogous derivatives of α-amino acids the hydroxyl group of the carboxyl has been replaced by an amino, anilino, or analogous group. They may be named by replacing the final 'e' of the trivial name of the amino acid by the word 'amide', 'anilide', etc., e.g. glycinamide, leucinamide, argininanilide. Alternatively, these compounds may be described as glycine amide, leucine amide, etc.

Note that the 4-amide of aspartic acid and the 5-amide of glutamic acid have specific trivial names, asparagine and glutamine. Their 1-amides are named aspartic 1-amide and glutamic 1-amide, or isoasparagine and isoglutamine.

3AA-9.3. *Acyl Groups*

The acyl group of an α-amino-mono-carboxylic acid is a structure that lacks the hydroxyl group of the carboxyl (H_2N–CHR–CO–). The names of such groups are formed by replacing the ending 'ine' (or 'an' in tryptophan) by 'yl' (C-421 of reference [14]), e.g. alanyl, arginyl, tryptophyl. 'Cysteinyl' is used instead of 'cysteyl', because of potential confusion with the group from cysteic acid. 'Cystyl' is the diacyl group of cystine, and 'half-cystyl' is the acyl group of cysteine lacking also the H of its SH group.

The monoacyl groups derived from aspartic acid, HOOC–CH_2–CH(NH_2)–CO– and –CO–CH_2–CH(NH_2)–COOH, are designated α-aspartyl (or aspart-1-yl) and β-aspartyl (or aspart-4-yl) respectively; the corresponding groups derived from glutamic acid are α-glutamyl (or glutam-1-yl) and γ-glutamyl (or glutam-5-yl) (C-421.3 of reference [14]). The diacyl groups formed from the dicarboxylic amino acids are aspartoyl and glutamoyl. The acyl groups derived from asparagine and glutamine are termed asparaginyl and glutaminyl respectively.

3AA-10. CARBOXYL GROUP MODIFICATIONS OTHER THAN ESTER AND AMIDE FORMATION

3AA-10.1. *Removal of the Carboxyl Group*

Several decarboxylated amino acids have trivial names terminated with 'amine': tyramine, histamine, cysteamine, tryptamine, methioninamine. Similarly cystine (see Appendix) forms cystamine.

3AA-10.2. *Ketones*

▲ If the hydroxyl group of the 1-carboxyl is replaced by an alkyl group, the name of the ketone formed can use the name of the amino acid by naming the compound as a substituted hydrocarbon, e.g. phenylalanylchloromethane for C_6H_5–CH_2–CH(NH_2)–CO–CH_2Cl, 3-amino-1-chloro-4-phenylbutan-2-one (see also 3AA-18.2). This type of name is based on the trivial names of amino acids (or peptides), so does not place the substituents of methane in alphabetical order (as systematic nomenclature does), but places 'chloromethane' at the end because this indicates C-terminal modification (see 3AA-13.1). The practice of using names such as 'phenylalanine chloromethyl ketone' is discouraged, because they erroneously specify the carbonyl group twice.

3AA-10.3. *Aldehydes and Alcohols*

Aldehydes and alcohols obtained by successive stages of reduction of the carboxyl group of α-amino acids are named by replacing the final 'e' of a trivial name ending in 'ine' (or the 'ic acid' of aspartic and glutamic acids) with the endings 'al' and 'ol' respectively.

Examples. R–CH(NH_2)–CHO: alaninal, leucinal, lysinal, serinal, aspart-1-al, glutaminal. R–CH(NH_2)–CH_2OH: alaninol, leucinol, lysinol, serinol, aspart-1-ol, glutaminol. The aldehyde and alcohol derivatives of tryptophan take the names tryptophanal and tryptophanol. The name glycinol is little used, because the systematic name 2-aminoethanol is short, and this already has the trivial name ethanolamine [19].

Note. The derivative of lysine in which the –CH_2–NH_2 group is replaced by –CHO has the trivial name allysine (see Appendix).

Part 1, Section C: PEPTIDE NOMENCLATURE

3AA-11. DEFINITION OF PEPTIDES

A peptide is any compound produced by amide formation between a carboxyl group of one amino acid and an amino group of another. The amide bonds in peptides may be called peptide bonds. The word peptide usually applies to compounds whose amide bonds are formed between C-1 of one amino acid and N-2 of another (sometimes called eupeptide bonds), but it includes compounds with residues linked by other amide bonds (sometimes called isopeptide bonds). Peptides with fewer than about 10–20 residues may also be called oligopeptides; those with more, polypeptides. Polypeptides of specific sequence of more than about 50 residues are usually known as proteins, but authors differ greatly on where they start using this term.

3AA-12. AMINO-ACID RESIDUES

3AA-12.1. *Definitions of Residues*

When two or more amino acids combine to form a peptide, the elements of water are removed, and what remains of each amino acid is called an amino-acid residue. α-Amino-acid residues are therefore structures that lack a hydrogen atom of the amino group (–NH–CHR–COOH), or the hydroxyl moiety of the carboxyl group (NH_2–CHR–CO–), or both (–NH–CHR–CO–); all units of a peptide chain are therefore amino-acid residues. (Residues of amino acids that contain two amino groups or two carboxyl groups may be joined by isopeptide bonds, and so may not have the formulas shown.)

The residue in a peptide that has an amino group that is free, or at least not acylated by another amino-acid residue (it may, for example, be acetylated or formylated), is called N-terminal; it is at the N-terminus. The residue that has a free carboxyl group, or at least does not acylate another amino-acid residue, (it may, for example, acylate ammonia to give –NH–CHR–CO–NH_2), is called C-terminal. If the amino group of the N-terminal residue is free, the residue may be named as an acyl group under 3AA-9.3; indeed any internal residue is an *N*-substituted amino-acyl group.

Residues are named from the trivial name of the amino acid, omitting the word 'acid' from aspartic acid and glutamic acid. Examples: glycine residue, lysine residue, glutamic residue.

3AA-12.2. *Ionized Forms of Residues*

When it is desirable to mention or emphasize the particular ionic form of a residue, this may be done as follows

Name of residue	*Protonated form*	*Deprotonated form*
arginine residue	argininium residue	arginine (base) residue
histidine residue	histidinium residue	histidine (base) residue
lysine residue	lysinium residue	lysine (base) residue
aspartic residue	aspartic (acid) residue	aspartate residue
cysteine residue	cysteine (acid) residue	cysteinate residue
glutamic residue	glutamic (acid) residue	glutamate residue
tyrosine residue	tyrosine (acid) residue	tyrosinate residue

This system cannot easily be applied to N- or C-terminal residues.

3AA-13. THE NAMING OF PEPTIDES

3AA-13.1. *Construction of Names*

To name peptides, the names of acyl groups ending in 'yl' (3AA-9.3) are used. Thus if the amino acids glycine, NH_3^+–CH_2–COO^-, and alanine, NH_3^+–$CH(CH_3)$–COO^-, condense so that glycine acylates alanine, the dipeptide formed, NH_3^+–CH_2–CO–NH–$CH(CH_3)$–COO^-, is named glycylalanine. If they condense in the reverse order, the product, NH_3^+–$CH(CH_3)$–CO–NH–CH_2–COO^-, is named alanylglycine. Higher peptides are named similarly, e.g. alanylleucyltryptophan. Thus the name of the peptide begins with the name of the acyl group representing the N-terminal residue, and this is followed in order by the names of the acyl groups representing the internal residues. Only the C-terminal residue is represented by the name of the amino acid, and this ends the name of the peptide. Formulas should normally be written in the same order, with the N-terminal residue on the left, and the C-terminal on the right, e.g.

$$NH_2–CH(COOH)–[CH_2]_2–CO–NH–CH(CH_2SH)–CO–NH–CH_2–COOH$$

glutathione, γ-glutamylcysteinylglycine

A multiplicative affix (p. 5 of reference [14]) placed before 'peptide' gives the total number of residues in the peptide, e.g. ▲ hexapeptide. Since the higher affixes are not well known, they may be replaced by numerals, e.g. a 22-peptide.

Higher oligopeptides and polypeptides of biological origin often have trivial names; their sequences are usually described more conveniently by symbols (3AA-14 to 3AA-19 below) than by constructing long names.

3AA-13.2. *Use of Prefixes in Peptide Names*

Configurational prefixes (3AA-3) are placed immediately before the trivial names of the residues they refer to. The prefixes are set off from the names before and after them with hyphens. Examples: L-alanyl-L-leucine; L-alanyl-D-leucine; glycyl-L-alanine; L-alanylglycine; L-leucyl-L-phenylalanyl-L-leucylglycine; L-alanylglycyl-L-leucine.

▲ The mixture of diastereoisomers formed by condensations between DL-amino acids will contain unspecified proportions of each pair of enantiomers. Names such as DL-alanyl-DL-leucine have been used in the past, but they are misleading because they contradict the accepted meaning of the prefix DL as signifying a racemate; here the racemate of L-alanyl-L-leucine and D-alanyl-D-leucine (which may be designated as *rac*-L-alanyl-L-leucine) is mixed in unspecified proportions with the racemate of L-alanyl-D-leucine and D-alanyl-L-leucine (which may similarly be designated as *rac*-L-alanyl-D-leucine). This is better indicated by the name *ambo*-alanyl-*ambo*-leucine (item 12c of reference [20]).

A mixture of L-alanyl-L-alanyl-L-alanine and L-alanyl-D-alanyl-L-alanine may likewise be called L-alanyl-*ambo*-alanyl-L-alanine.

3AA-13.3. *Name of Simple Polymers of α-Amino Acids*

Simple polymers of amino acids may, if preferred, be named with prefixes to indicate the number of amino-acid residues present, e.g. tetraglycine. Mixtures of polymers with varying numbers of residues may be given names like oligoglycine, polyglycine, poly(L-lysine), etc. [21].

3AA-13.4. *Numbering of Peptide Atoms*

The atoms of a peptide may need to be numbered as locants for substitution or isotopic replacement. Often no more numbering is required than that of atoms within a residue (see 3AA-2.2), e.g. alanyl-3-chloroalanylalanine. It may sometimes be convenient to indicate substitution of the peptide as a whole. This may be done by adding the residue number, obtained by ▲ numbering residues from the N-terminus, after the atom number, and separated from it by a point. The above compound may therefore be called 3.2-chloro(alanylalanylalanine). Thus the atom C-3.2 is C-3 of the second residue of the peptide. Example: Alanylthreonylglycylaspartylglycine 4.4-3.2-lactone for the compound that can be represented (3AA-16, -17 and -19 below) as Ala-Thr-Gly-Asp-Gly.

Such numbering is especially useful for peptides with trivial names (see 3AA-22.5), e.g. $N^{5,4}$-methyloxytocin would indicate a methyl substituent on N-5 of the glutamine residue at position 4 of oxytocin. If the peptide name that follows a substituent

indicated in this way is constructed residue by residue, it must be placed in parentheses to show that the numbering applies to the peptide as a whole, rather than to the first residue.

3AA-13.5. *Prefixes Formed from Peptide Names*

▲ When it is necessary to treat a peptide as a substituent, the point of attachment is specified by the suffix 'yl' (see 3AA-8) with the appropriate locant. If the group formed from the peptide is not the acyl group derived by removing hydroxyl from C-1 of the C-terminal residue, the position at which hydrogen (or hydroxyl from a side-chain carboxyl group) is removed should be indicated by a locant before the 'yl'; if the sequence of the peptide is given in full, it should be placed in parentheses to avoid implying that the group is formed by removing H or OH from the C-terminal residue. Examples:

(1) Leukotriene D, or (7*E*,9*E*,11*Z*,14*Z*)-(5*S*,6*R*)-6-[(cysteinylglycin)-*S*-yl]-hydroxyicosa-7,9,11,14-tetraenoic acid;

(2) Leukotriene C, or (7*E*,9*E*,11*Z*,14*Z*)-(5*S*,6*R*)-6-(glutathion-*S*-yl)-5-hydroxyicosa-7,9,11,14-tetraenoic acid, or (7*E*,9*E*,11*Z*,14*Z*)-(5*S*,6*R*)-6-[(γ-glutamylcysteinylglycin)-*S*-yl]-5-hydroxyicosa-7,9,11,14-tetraenoic acid;

(3) (2*S*)-2-*O*-[(serylalanylserin)-3.2-yl]lactic acid, or (2*S*)-2-[(serylserylserin)-$O^{3.2}$-yl]propanoic acid, or $O^{3.2}$-[(1*S*)-1-carboxyethyl](serylserylserine).

If the locant before 'yl' indicates the carbon of a carboxyl group, the prefix indicates the acyl group formed by removing hydroxyl from this atom. Example: 4-*O*-[(glutamylglutamylglutamic acid)-5.2-yl]-D-gluconic acid.

3AA-13.6. *Conformation of Polypeptide Chains*

Abbreviations and symbols for describing the conformation of peptide chains have been published separately [16].

Part 2. **Symbolism**

Part 2, Section A: THE THREE-LETTER SYSTEM (a revision and updating of [10])

3AA-14. GENERAL CONSIDERATIONS ON THREE-LETTER SYMBOLS

14.1. The symbol chosen for an amino acid (Table 1) is derived from its trivial name, and is usually the first three letters of this name. It is written as one capital letter followed by two lower-case letters, e.g. Gln (not GLN or gln), regardless of its position in a sentence or structure. If any other convention is used in representing residues, e.g. to emphasize homology, this should be stated clearly whenever it is used. When the symbol is used for a purpose other than representing an amino-acid residue, e.g. to designate a genetic factor, three lower-case italic letters may be used, e.g. *gln*.

14.2. The main use of the symbols is in representing amino-acid sequences. Inasmuch as the symbols by themselves represent the unsubstituted amino acids, they are modified (3AA-16) by hyphens to represent residues. We do not recommend use of the symbols to represent free amino acids in textual material, but such use may be desirable in tables, diagrams or figures. It may also be convenient to use them for indicating residue numbers, e.g. Tyr-110 for tyrosine residue 110. For substituents, supplementary symbols are used (3AA-17 and -18).

14.3. A symbol may represent either the name or the formula of a compound.

14.4. Heteroatoms of amino-acid residues (e.g. O-3 serine, N-6 of lysine) do not explicitly appear in the symbol, as it represents the whole molecule including them (but see 3AA-17.4).

14.5. Amino-acid symbols denote the L configuration of chiral amino acids unless otherwise indicated by the presence of D or DL before the symbol and separated from it with a hyphen (see also 3AA-19.2). L may similarly be inserted for emphasis.

14.6. Structural formulas may be used together with symbols to make complicated features or reactions clear (for examples see 3AA-17.4).

3AA-15. SYMBOLS FOR AMINO ACIDS

3AA-15.1. *Symbols for Common Amino Acids*

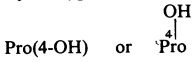
The symbols for the amino acids that are coded for by mRNA are listed in column 2 of Table 1.

3AA-15.2. *Symbols for Less Common Peptide Constituents*

Symbols for less common amino acids should be defined in each publication in which they appear. The following principles and notations are recommended.

15.2.1. Hydroxyamino Acids

The symbol 5Hyl is recommended for 5-hydroxylysine, and 4Hyp for 4-hydroxyproline (the numbers may be omitted, especially when limiting the symbols to three letters helps alignment of sequences, provided that the position of substitution is made clear in the text). Similarly 3Hyp would represent 3-hydroxyproline. Alternatively, symbols may be formed as shown in 3AA-17.3 below for substituted residues, so that 4-hydroxyproline may be written as:



15.2.2. Alloisoleucine and Allothreonine

Alloisoleucine and allothreonine (3AA-4.4) may be symbolized by aIle and aThr respectively.

15.2.3. 'Nor' Amino Acids

Since 'nor' in 'norvaline' and 'norleucine' is not used in its systematic sense of denoting a lower homologue, but to change the trivial name of a branched-chain compound to designate a straight-chain compound, its use for amino acids should be ▲ progressively abandoned (3AA-2.4), along with the earlier symbols Nva and Nle. Appropriate symbols for these compounds, 2-aminovaleric and 2-aminohexanoic acids, based on symbols proposed for the unsubstituted acids [19], are Avl and Ahx (see also 3AA-15.2.5).

15.2.4. 'Homo' Amino Acids

The prefix 'homo', used in the sense of a higher homologue, is commonly used for two amino acids (3AA-2.3). They are symbolized as follows:

Homoserine	Hse
Homocysteine	Hcy

15.2.5. Higher Unbranched Amino Acids

The functional prefix 'amino' is included in the symbol as the letter 'A' and 'diamino' as 'A_2'. The trivial name of the parent acid is abbreviated to two letters, based, when possible, on the symbols for lipid nomenclature [19]. Unless otherwise indicated, single groups are on C-2, two amino groups are in the 2 and terminal positions for monocarboxylic acids, and each is geminal with a carboxyl group for dicarboxylic acids. The location of amino groups other than these is shown by appropriate prefixes.

Examples	*Symbol*	*Note*
β-Alanine (3-aminopropanoic acid)	βAla	
2-Aminobutyric acid (2-aminobutanoic acid)	Abu	
▲ 2-Aminovaleric acid (2-aminopentanoic acid)	Ape	
2-Aminohexanoic acid	Ahx	
▲ 6-Aminohexanoic acid	εAhx	i
2-Aminoadipic acid (2-aminohexanedioic acid)	Aad	
3-Aminoadipic acid (3-aminohexanedioic acid)	βAad	
2-Aminopimelic acid (2-aminoheptanedioic acid)	Apm	
▲ 2,3-Diaminopropionic acid (2,3-diaminopropanoic acid)	A_2pr or Dpr	ii, iii, iv
▲ 2,4-Diaminobutyric acid (2,4-diaminobutanoic acid)	A_2bu or Dab	ii
Ornithine (2,5-diaminovaleric acid, 2,5-diaminopentanoic acid)	Orn	
▲ 2,6-Diaminopimelic acid (2,6-diaminoheptanedioic acid)	A_2pm or Dpm	ii, iii

Notes

i) This symbol is recommended in place of the previous εAcp, in which 'cp' stood for caproic, which may be confused with capric and caprylic.

ii) The previous edition of these recommendations (10) discouraged abbreviations starting 'D' for 'di' or 'T' for 'tri' or 'tetra'; because these letters were overused. We concur in preferring subscripts when these can be applied to well-known symbols, so that Me_2SO is preferable to DMSO, Me_3Si- to TMS-, and H_4 to TH. Nevertheless we are not convinced that 'A_2' easily suggests 'diamino', so alternative symbols are presented.

iii) 'Dap' should not be used as a symbol, since it could be construed to mean either diaminopropanoic acid or diaminopimelic acid.

iv) 2,3-Diaminopropanoic acid can be regarded as 3-aminoalanine, and so may be symbolized by 'side-chain substitution' (3AA-17.3 below) as Ala(NH_2) or Ala, but users should beware of the possibility that the former may be confused with Ala-NH_2 (3AA-17.1), the symbol for alaninamide. $\overset{|}{NH_2}$

15.2.6. Carboxylated and Oxidized Amino Acids

Symbols are recommended for two amino acids that have an additional acidic group and may occur in polypeptide sequences. They are:

4-Carboxyglutamic acid	Gla
Cysteic acid	Cya

15.2.7. Non-Amino-Acid Residues in Peptides

Symbols for sugar residues (e.g. Glc, Gal) have been proposed [22], as have ones for nucleoside residues (e.g. Ado, Cyd) [23], and these may be combined with amino-acid symbols to represent glycopeptides, etc. These symbols include [22] Neu for neuraminic acid, Neu5Ac for *N*-acetyl neuraminic acid, and Mur for muramic acid. Depsipeptides (3AA-19.6) contain hydroxy-acid residues; when symbols are used for these they should be defined.

3AA-16. SYMBOLISM OF AMINO-ACID RESIDUES

3AA-16.1. *General Principles for Symbolizing Residues*

The peptide glycylglycylglycine is symbolized as Gly-Gly-Gly. This involves modifying the symbol Gly for glycine, NH_2–CH_2–COOH, by adding hyphens to it, in three ways:

i) Gly-= NH_2–CH_2–CO– (normally as NH_3^+–CH_2–CO–)

ii) -Gly =–NH–CH_2–COOH (normally as –NH–CH_2–COO^-)

iii) -Gly-=–NH–CH_2–CO–

Thus the hyphen, which represents the peptide bond, removes OH from the 1-carboxyl group of the amino acid (written in the conventional un-ionized form) when it is placed on the right of the symbol (i), and removes H from the 2-amino group of the amino acid when it is placed on the left of the symbol (ii); both modifications can apply to one symbol (iii).

Thus the peptide Gly-Glu (without hyphens at its ends) is distinguished from the sequence -Gly-Glu- (with hyphens at its ends).

3AA-16.2. *Lack of Hydrogen on the 2-Amino Group*

A hyphen on the left of the symbol signifies removal of a hydrogen atom from the 2-amino group, as well as representing the bond formed by the group thus produced. If it should prove necessary to draw a bond to N-2 on the right of the symbol (e.g. in a cyclic peptide, 3AA-19.4 below), then the hyphen must be replaced by an arrow, which points from CO to NH within the peptide bond.

If both atoms on N-2 are replaced, two lines can be drawn on the left of the symbol, e.g.

$>$Ala or $\perp$Ala for $>$N–$CH(CH_3)$–COOH

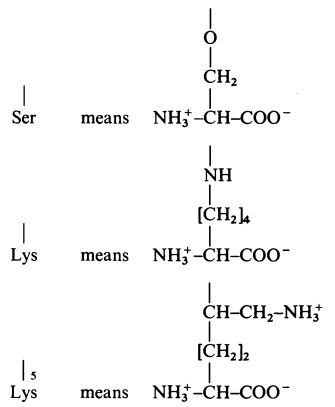
3AA-16.3. *Lack of Hydroxyl on the 1-Carboxyl Group*

A hyphen on the right of the symbol signifies removal of hydroxyl from the 1-carboxyl group as well as representing the bond formed by the group produced. If it is not possible to draw this bond on the right of the symbol, as in a cyclic peptide (3AA-19.4) then the hyphen must be replaced by an arrow, which has the same effect.

3AA-16.4. *Removal of Groups from Side Chains*

16.4.1. Monocarboxylic Acids

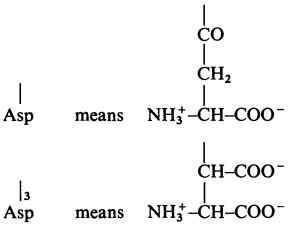
A vertical line drawn above or below the symbol for a monocarboxylic amino acid represents removal of hydrogen from the side chain so that a radical is formed. Replacement of this hydrogen by a substituent is treated in 3AA-17.2 below. Unless indicated by a locant placed beside the line, the hydrogen is assumed to be removed from a heteroatom in the residue. Examples:



Notes. (a) H is removed from N-ω rather than N-δ of arginine unless otherwise indicated; (b) a locant, π or τ (3AA-2.2.4), is always required for histidine.

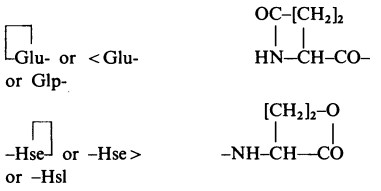
16.4.2. Dicarboxylic acids

A vertical line drawn above or below either of the symbols Asp and Glu represents removal of OH from the side-chain carboxyl group, as well as representing a bond to a substituent. If a hydrogen has to be removed from a saturated carbon of the side chain, then a vertical line may be used, but it must be accompanied by a locant. Examples:



3AA-16.5. *Cyclic Derivatives of Amino Acids*

Combination of horizontal lines, indicating removal of H from N-2 (3AA-16.1, 3AA-16.2) or OH from C-1 (3AA-16.1, 3AA-16.3), with the vertical lines that indicate removal of side-chain atoms (3AA-16.4) allows formation of symbols for 5-oxoproline (systematically 5-oxopyrrolidine-2-carboxylic acid, also known as pyroglutamic acid or pyrrolidonecarboxylic acid) and for homoserine lactone, as follows:



3AA-17. SUBSTITUTED AMINO ACIDS

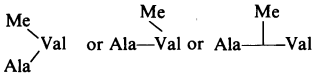
3AA-17.1. *Substitutions in the 2-Amino and 1-Carboxyl Groups*

This follows logically from 3AA-16.1, 3AA-16.2 and 3AA-16.3 by using symbols for atoms or groups to represent the substituents. Examples (see also 3AA-18.2):

N-Acetylglycine	Ac-Gly
Glycine ethyl ester	Gly-OEt
N^2-Acetyllysine	Ac-Lys
Serine methyl ester	Ser-OMe
O^1-Ethyl *N*-acetylglutamate	Ac-Glu-OEt
Isoglutamine	Glu-NH_2
O^1-Methyl hydrogen aspartate	Asp-OMe

A second substituent on N-2, when the first is shown with a line as above, may be represented with a second line to the left of the symbol for the substituted residue: $\geq$Xaa. It may be convenient to print this as a vertical line joining the first: $\perp$Xaa.

Example: alanyl-*N*-methylvaline may be represented



3AA-17.2. *Substitutions on Side-Chain Functional Groups*

Side-chain substituents may be portrayed above or below the amino-acid symbol (3AA-16.4) or by placing the symbol for the substituent in parentheses immediately after the amino-acid symbol. When the symbol for a substituent, such as an oligosaccharide, has a hyphen on its right-hand side to indicate the bond to the amino acid [22], then this symbol should be placed in parentheses before the amino-acid symbol rather than after it, e.g. -(Galβ1-4Xylβ1-)Ser- [22].

Symbols within parentheses written on one line should normally be used only in textual material and when the symbol for the substituent is short; otherwise the two-line symbols containing a vertical line will be clearer. Note that the substituents

represented replace hydrogen except when the amino acid is aspartic acid or glutamic acid, when they replace the OH of the carboxyl group unless otherwise specified (3AA-16.4.2).

If a locant is required, it is placed beside the vertical line that represents side-chain substitution, or is joined to the substituent symbol within the parentheses by a hyphen.

	Examples	*Symbols*	*Notes*
	O^4-Methyl hydrogen aspartate (β-methyl aspartate)	Asp or Asp(OMe)	i
	O^5-Ethyl hydrogen *N*-acetylglutamate	Ac–Glu or Ac–Glu(OEt)	i
	N^6-Acetyllysine	Lys or Lys(Ac)	
	O^3-Acetylserine	Ser or Ser(Ac)	
	O^4-Sulfotyrosine (tyrosine O^4-sulfate)	Tyr or Tyr(SO_3H)	
	S-Ethylcysteine	Cys or Cys(Et)	
▲	3-Sulfenoalanine	Cys or Cys(OH)	ii
▲	*S*-Sulfocysteine (*S*-cysteinesulfonic acid)	Cys or Cys(SO_3H)	ii, iii
	S-Cyanocysteine	Cys or Cys(CN)	
	Cystine	Cys or Cys Cys (not Cys-Cys)	
	D-Cystine	D-Cys D-Cys	
	meso-Cystine	L-Cys D-Cys or D-Cys L-Cys	
▲	Methionine *S*-oxide (methionine oxide)	Met or MetO	ii, iv
▲	Methionine *S*,*S*-dioxide (methionine dioxide)	Met or $MetO_2$	ii, iv
	Phosphoserine (O^3-phosphonoserine)	Ser or Ser(*P*)	v
	N^τ-Methylhistidine (*tele*methylhistidine, see 3AA-2.2.4)	His or His(τ-Me)	

Notes

i) Asp-OMe represents the O^1-methyl ester of aspartic acid (C-1 modification by 3AA-16.3), whereas Asp(OMe) represents the O^4-methyl ester (side-chain modification by 3AA-16.4.2).

ii) Names based on cysteine and symbols based on Cys already indicate sulfur in the molecule, and similarly with methionine and Met. Indication of modification of this sulfur should not suggest the addition of further sulfur. Hence calling 3-sulfenoalanine by the name cysteinesulfenic acid, 3-sulfinoalanine by the name cysteinesulfinic acid, methionine *S*-oxide by the name methionine sulfoxide, and methionine *S*,*S*-dioxide by the name methionine sulfone may be confusing and is not recommended.

iii) Care should be taken with this symbol because readers who fail to realize that the symbol Cys contains the sulfur may confuse it with cysteic acid, now symbolized Cya (3AA-15.2.6). The earlier [10] symbol Cys for cysteic acid has the disadvantage that the vertical line in it does not represent a single bond.

iv) The vertical lines or parentheses previously [10] in the symbols for MetO and $MetO_2$ are now omitted, because they wrongly implied removal of hydrogen.

v) $-P$ represents $-PO_3H_2$ [24].

3AA-17.3. *Substitution on Side-Chain Skeleton*

This may use the same convention as 3AA-17.2, with the addition of locant numerals where necessary (see 3AA-16.4), e.g.

4-Carboxyglutamic acid	$\overset{\overset{4}{\vert}}{\text{COOH}}$ Glu as an alternative to Gla (3AA-15.2.6) if it is desired to emphasize carboxylation
2,3-Diaminopropanoic acid (3-aminoalanine, see 3AA-15.2.5)	$\overset{\vert}{\text{NH}_2}$ Ala or Ala(NH_2)
3,5-Diiodotyrosine	Tyr(I_2) or Tyr(3,5-I_2) if the context does not imply the locants
3-Nitrotyrosine	$\overset{\overset{3}{\vert}}{\text{NO}_2}$ Tyr

3AA-17.4. *The Use of Symbols in Representing Reactions of Side-Chains*

▲ The symbols are designed primarily to indicate sequence, and care must be taken to avoid confusion when they are adapted to other uses.

Although the conversion of a cysteine residue in a protein into an *S*-carboxymethylcysteine can be adequately represented as

$$\text{–Cys–} \longrightarrow \text{–Cys(CH}_2\text{–COOH)–}$$

writers may wish to show the sulfur atom in order to indicate the chemistry of the reaction. Although it would be perfectly legitimate to write

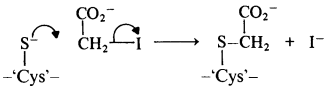
$$\text{–Ala(SH)–} \longrightarrow \text{–Ala(S–CH}_2\text{–COOH)–}$$

this would be confusing since the residue is thought of as cysteine rather than as modified alanine.

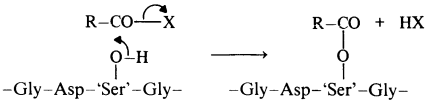
We therefore recommend putting the residue symbol into quotation marks if one of its groups is to be depicted separately (or, alternatively, using the symbol R_{Xaa}). Hence the thiol group of a cysteine residue may be shown as:

$$\overset{\overset{\text{SH}}{\vert}}{\text{–'Cys'–}} \text{ or } \overset{\overset{\text{SH}}{\vert}}{\text{–R}_{\text{Cys}}\text{–}}$$

Terminal amino and carboxyl groups can be shown similarly, e.g. H_2N–'Ala'– to show explicitly the amino group present in Ala- (in contrast with –Ala–NH_2 which shows the amide of C-terminal alanine). This convention allows mechanisms to be drawn out, e.g.



with the quotation marks to alert readers to the fact that the symbol here does not include the sulfur atom. Sequences may also be shown, so that the acylation of a serine proteinase could be drawn as:



When this convention is used it should be described.

3AA-17.5. *Modified Residue in Natural Peptides*

▲ If an unusual residue is to be symbolized within a particular context, it may be helpful to modify (e.g. with an asterisk) the symbol for the ribosomally incorporated residue, e.g. Ser* for 2-aminopropenoic acid (formed within a peptide chain by dehydration of a serine residue). Such an asterisk may be placed above the residue rather than after it to allow alignment with other 3-letter symbols. Symbols modified in this way should be defined when used.

3AA-17.6. *Lack of Substitution*

If it is desired to emphasize lack of substitution, H or OH may be added to the hyphen or vertical line that represents removal of one of these groups. Thus H-Ala may be contrasted with Ac-Ala, and Ala-OH with Ala-OMe.

3AA-18. SYMBOLS FOR SUBSTITUENTS

3AA-18.1. *Use of Symbols*

Groups substituted for hydrogen or hydroxyl may be indicated by their formulas or by symbols or by combination of both, e.g.

Benzoylglycine (hippuric acid) PhCO-Gly or C_6H_5CO-Gly

Note: the symbol Bz is often used for benzoyl in organic chemistry, and Bzl for benzyl, but because these symbols are so similar, the alternative PhCO and $PhCH_2$ are preferable.

Glycine methyl ester Gly–OCH_3 or Gly–OMe
Trifluoroacetylglycine CF_3CO–Gly (Table 3, Note ii)

Suggestions for symbols to designate substituent (or protecting) groups common in peptide and protein chemistry are given in Tables 2, 3 & 4.

Table 2. *Nitrogen substituents (protecting groups) of the urethane type*

Benzyloxycarbonyl-	Z- or Cbz-
2-(*p*-Biphenylyl)isopropyloxycarbonyl- [strictly 1-(biphenyl-4-yl)-1-methylethoxycarbonyl-]	Bpoc-
p-Bromobenzyloxycarbonyl-	Z(Br)-
t-Butoxycarbonyl-	Boc- or Bu^{t}OCO- or t-BuOCO- or Me_3C–OCO–
α,α-Dimethyl-3,5-dimethoxybenzyloxycarbonyl-	Ddz–
Fluoren-9-ylmethoxycarbonyl–	Fmoc–
p-Methoxybenzyloxycarbonyl–	Z(OMe)–
p-Nitrobenzyloxycarbonyl–	$Z(NO_2)$–
p-Phenylazobenzyloxycarbonyl–	Pz–

Table 3. *Non-urethane substituents for nitrogen, oxygen or sulfur*

Acetamidomethyl–	Acm–
Acetyl	Ac–
Benzoyl– (C_6H_5–CO–)	PhCO– (or Bz–; see note in 3AA-18.1)
Benzyl– (C_6H_5–CH_2–)	$PhCH_2$– (or Bzl; see note in 3AA-18.1)
Carbamoyl–	NH_2CO– (preferred to Cbm–)
(3-Carboxy-4-nitrophenyl)thio–	Nbs– (see 3AA-18.2)
3-Carboxypropanoyl– (HOOC–CH_2–CH_2–CO–)	Suc– (see Note i)
Dansyl–, 5-(dimethylamino)naphth-1-ylsulfonyl–	Dns–
2,4-Dinitrophenyl–	Dnp– or N_2ph (see Note ii)
Formyl–	HCO– or For– (see Note iii)
4-Iodophenylsulfonyl– (pipsyl–)	Ips–
Maleoyl– (–OC–CH=CH–CO–)	–Mal– or Mal< (C-404.1 of [14])
Maleyl– (HOOC–CH=CH–CO–)	Mal–
2-Nitrophenylthio–	NpS– (Nps– often used)
Phenyl(thiocarbamoyl)–	PhNHCS– or Ptc–
Phthaloyl–	–Pht– or Pht<
Phthalyl– (*o*-carboxybenzoyl–)	Pht–
Succinyl– (–OC–CH_2–CH_2–CO–)	–Suc– or Suc< (see Note i)
Tosyl–	Tos–
Trifluoroacetyl–	CF_3CO–
Trityl– (triphenylmethyl–)	Ph_3C– or Trt–

Notes

i) In organic nomenclature (C-404.1 of [14]), 'succinyl' signifies the bivalent group formed from succinic acid by removal of both hydroxyl groups, but in biochemical usage it usually signifies the 3-carboxypropanoyl group, e.g. succinyl-CoA.

ii) The use of D for 'di' and T for 'tri' and 'tetra' is discouraged if these apply to atoms or groups for which simple symbols exist, e.g. in CF_3CO–, Me_3Si– and H_4folate. We feel less strongly when their avoidance involves giving unusual meanings to symbols, e.g. N for nitro, so Dnp and N_2ph are offered as alternative symbols for dinitrophenyl. See also Note ii of 3AA-15.2.5.

iii) The symbol HCO– is preferred to CHO– for the formyl group, because CHO– has sometimes been used to indicate the attachment of carbohydrate.

3AA-18.2. *Principles of Symbolizing Substituent Groups and Reagents*

Many reagents used in peptide and protein chemistry for modifying (often protecting) amino, carboxyl and side-chain groups in amino-acid residues have been designated by a variety of acronymic abbreviations, too numerous to list here. Extensive and

▲ Table 4. *Substituents at the carboxyl group*

Group	Symbol	Name of glycine derivative (see note)
Benzotriazol-1-yloxy	–OBt	1-(Glycyloxy)benzotriazole
Benzyloxy	$-OCH_2Ph$ (or–OBzl, see note in 3AA-18.1)	Glycine benzyl ester
tert-Butoxy	$-OCMe_3$ or $-OBu^t$	Glycine *t*-butyl ester
Diphenylmethoxy	$-OCHPh_2$ or –OBzh	Glycine diphenylmethyl ester (or benzhydryl ester)
Ethoxy	–OEt	Glycine ethyl ester
Methoxy	–OMe	Glycine methyl ester
4-Nitrobenzyloxy	–ONb	Glycine 4-nitrobenzyl ester
4-Nitrophenoxy	–ONp	Glycine 4-nitrophenyl ester
4-Nitrophenylthio	–SNp	Thioglycine *S*-(4-nitrophenyl ester)
Pentachlorophenoxy	–OPcp	Glycine pentachlorophenyl ester
Phenylthio	–SPh	Thioglycine *S*-(phenyl ester)
Quinolin-8-yloxy	–OQu	Glycine quinolin-8-yl ester
Succinimido-oxy	–ONSu or –OSu	*N*-(Glycyloxy)succinimide
2,4,5-Trichlorophenyloxy	–OTcp	Glycine 2,4,5-trichlorophenyl ester

Note. Carboxyl substituents will not normally appear as prefixes in the names of derivatives of amino acids or peptides, so the name of the group, its prefix name, given in column 1, is little used in naming compounds. Column 3 is therefore given to show how derivatives containing the group are named (by one of the alternative methods of 3AA-9.1).

indiscriminate use of such abbreviations is discouraged, especially when the accepted trivial name of the reagent is short, e.g. tosyl chloride, trityl chloride, etc.

It can be useful to symbolize a reagent in such a way that the group transferred retains its identity in a reaction, e.g.

$$\text{Dns–Cl} + \text{Gly} \rightarrow \text{Dns–Gly} + H^+ + Cl^-$$
$$\text{Dnp–F} + NH_2\text{–R} \rightarrow \text{Dnp–NH–R} + H^+ + F^-$$

For this reason Dns–Cl is usually preferred to DNS for dansyl chloride (although the full name is short enough for most textual use), and Dnp–F to the original FDNB for 1-fluoro-2,4-dinitrobenzene, and similarly Nbs_2 in place of DTNB for 3,3′-dithiobis(6-nitrobenzoic acid) (Ellman's reagent) and $(Pr^iO)_2PO$–F or Dip–F for diisopropyl fluorophosphate.

Symbols constructed from known elements are more readily understood than arbitrary abbreviations, e.g. Tos-Arg-OMe rather than TAME for tosylarginine methyl ester, and Tos-Phe-CH_2Cl rather than TPCK for 'tosylphenylalanine chloromethyl ketone', a name incorrectly used for tosylphenylalanylchloromethane (3AA-10.2), but misleading because it erroneously specifies the carbonyl group twice.

3AA-19. PEPTIDE SYMBOLISM

3AA-19.1. *Peptide Chains*

The amino-acid symbols were developed for representing peptide sequences (3AA-16). Peptides containing bonds other than between C-1 and N-2 of adjacent residues are also easily represented (3AA-16 to -18). Examples:

Glycylglycine	Gly-Gly
N-α-Glutamylglycine	Glu-Gly
N-γ-Glutamylglycine	Glu(Gly) or Glu(Gly) or Glu(Gly) or Glu(-Gly)
Thyroliberin	Glp-His-Pro-NH_2
Angiotensin II	Asp-Arg-Val-Tyr-Ile-His-Pro-Phe
Glutathione	Glu(Cys-Gly) or Glu(Cys-Gly) or Glu(-Cys-Gly)

Note. Glu(Cys-Gly) would represent the corresponding thiol ester with a bond between the γ-carboxyl of glutamic acid and the thiol group of cysteine.

N^2-α-Glutamyllysine Glu-Lys

N^6-α-Glutamyllysine Glu(Lys)-Lys or Glu-Lys(Lys)

N^2-γ-Glutamyllysine Glu(Lys)-Lys or Glu-Lys(Lys)

N^6-γ-Glutamyllysine Glu(Lys)-Lys or Glu-Lys(Lys) or Glu(Lys)-Lys

Symbols for modified residues or names of compounds may be used in such formulas. Thus a peptide with a C-terminal aldehyde may be shown using either a name or a symbol constructed according to 3AA-16.3. Example:

▲ Ac-Leu-Leu-argininal or Ac-Leu-Leu-Arg-H

(If the second method is used, the symbol should be explained to avoid confusion.)

▲ If part of a sequence is unknown, but its composition can be specified, this may be indicated by parentheses, with commas between the residues listed as present, e.g. Ala-Lys-(Ala,Gly_3,Val_2)-Glu-Val.

▲ If a peptide must be written on more than one line, we advise placing a hyphen at the end of each line to be continued (where it has its usual meaning of a continuation symbol), and also at the start of the next line (where it represents the peptide bond), e.g.

Ala-Ser-Tyr-Phe-Ser-
-Gly-Pro-Gly-Trp-Arg

In diagrams the two lines can usually be joined, as in

Ala-Ser-Tyr-Phe-Ser
Gly-Pro-Gly-Trp-Arg,

but such a break may also be needed in textual material where this is not possible.

3AA-19.2. *Use of Configurational Prefixes*

Residue symbols written in a sequence denote the L configuration for chiral amino acids, unless otherwise indicated (3AA-14.5). A D residue is shown by inserting a D before the symbol, separated from it by a hyphen (which may be omitted to make the number of residues appear more clearly).

The symbol DL signifies a racemic mixture, so should not occur in the designation of peptides with more than one chiral residue; coupling of a DL-amino acid with a chiral peptide leads to a mixture of diastereoisomeric products whose ratio may ▲ depend on the conditions of the reaction and will not in general be unity. To indicate that both are present, *ambo* may be used (3AA-13.2), and thus the mixture of products formed by acylating L-leucine with DL-alanine may be represented as *ambo*-Ala-Leu, and a mixture of Phe-Ala-Leu and Phe-D-Ala-Leu may be represented as Phe-*ambo*-Ala-Leu.

▲ A residue of unknown configuration may be indicated by the prefix ξ (Greek xi), e.g. ξ-Ala.

3AA-19.3. *Representation of Charges on Peptides*

▲ It is usually convenient to use the same abbreviated formula for a peptide regardless of its state of ionization. To indicate or stress the charges on a peptide, plus and minus signs may be placed over residues with charged side chains and on either side of the formula to represent charged termini, e.g.

$^+$Gly-$\overset{-}{\text{Glu}}$-Ala-$\overset{+}{\text{Lys}}$-Cys-Val$^-$ and its form in alkali: Gly-$\overset{-}{\text{Glu}}$-Ala-Lys-$\overset{-}{\text{Cys}}$-Val$^-$

Such signs may be circled for clarity.

If, however, it is desired to indicate charge by formal modification of the symbols for residues, this may be done as follows.

i) *Protonation of the N-terminus.* The sign $^+$H is placed beside the symbol for the N-terminal residue without a hyphen between (since a hyphen would signify removal of H). This gives, for example, $^+$HGly-. We prefer this to the alternative recommendation [10] of adding $^+H_2$-, to give, for example, $^+H_2$-Gly-, because it seems artificial to remove one hydrogen before adding two, and because the hyphen here fails to represent a single bond.

ii) *Deprotonation of the C-terminus.* The symbol -*O*$^-$ is placed on the right of the C-terminal residue. Its hyphen signifies removal of -OH from the carboxyl group, so this is replaced by -O$^-$.

iii) *Protonation of Side-Chain basic groups.* 'H^+' is placed above the amino-acid symbol in the two-line representation, or after it, e.g. LysH^+, in the one-line system. No lines or parentheses are used, since they would imply removal of H. In earlier [10] recommendations 'H_2^+' was added with a vertical line or parentheses, but again (cf. i) the line represented no single bond.

iv) *Deprotonation of Side-Chain Acidic Groups.* The symbols Asp and Glu may have O^- placed at the end of a vertical line above or below them, or in parentheses after them (cf. ii), since O^- replaces the OH removed. Other acidic residues, e.g. Cys, have the charge alone at the end of the vertical line or in parentheses, since the group removed here is H.

Hence the two ionic forms shown above for a peptide could be drawn as

$$^{+}HGly\text{-}Glu(O^-)\text{-}Ala\text{-}Lys(H^+)\text{-}Cys\text{-}Val\text{-}O^- \quad \text{and} \quad Gly\text{-}Glu(O^-)\text{-}Ala\text{-}Lys\text{-}Cys(^-)\text{-}Val\text{-}O^-$$

An isoelectric form of Gly-Lys-Gly could be drawn as

$$Gly\text{-}Lys(H^+)\text{-}Gly\text{-}O^-$$

whereas its dihydrochloride could be drawn as

$$^{+}HGly\text{-}Lys(H^+)\text{-}Gly\text{-}OH \cdot 2\,Cl$$

3AA-19.4. *Peptides Substituted at N-2* (see 3AA-16.2 and -17.1)

Glycylnitrosoglycine	Gly(NO)–Gly or Gly–Gly(NO)
Glycylsarcosine (see Appendix)	Gly-Sar or Gly(Me)–Gly or Gly–Gly(Me)
Glycyl-*N*-acetylglycine	Gly(Ac)–Gly or Gly–Gly(Ac)
N,*N*-diglycylglycine	Gly(Gly)–Gly or Gly–Gly(Gly)

3AA-19.5. *Cyclic Peptides*

3AA-19.5.1. Homodetic Cyclic Peptides

Cyclic peptides in which the ring consists solely of amino-acid residues in eupeptide linkage may be called homodetic cyclic peptides. Three representations are possible:

i) The sequence is formulated in the usual manner but placed in parentheses and preceded by 'cyclo'. Example: gramicidin S

cyclo(-Val-Orn-Leu-D-Phe-Pro-Val-Orn-Leu-D-Phe-Pro-)

or (see 3AA-19.2, sentence 2)

cyclo(-Val-Orn-Leu-DPhe-Pro-Val-Orn-Leu-DPhe-Pro-)

ii) The sequence is again written in one line, but the residues at each end of the line are joined by a lengthened bond, e.g.

$$\text{Val-Orn-Leu-D-Phe-Pro-Val-Orn-Leu-D-Phe-Pro}$$

or (3AA-19.2, sentence 2)

$$\text{Val-Orn-Leu-DPhe-Pro-Val-Orn-Leu-DPhe-Pro}$$

iii) The residues are written on two lines, so that the sequence is reversed on one of them. Hence the CO to NH direction within the peptide bond must be indicated by arrows (3AA-16.2 and -16.3). Hence gramicidin S may by written (using the option of 3AA-19.2, sentence 2):

$$\begin{array}{l} \text{Val}\rightarrow\text{Orn}\rightarrow\text{Leu}\rightarrow\text{DPhe}\rightarrow\text{Pro} \\ \text{Pro}\leftarrow\text{DPhe}\leftarrow\text{Leu}\leftarrow\text{Orn}\leftarrow\text{Val} \end{array}$$

3AA-19.5.2. Heterodetic Cyclic Peptides

Heterodetic cyclic peptides are peptides consisting only of amino-acid residues, but the linkages forming the ring are not solely eupeptide bonds; one or more is an isopeptide, disulfide, ester, or other bond.

Their symbolic representation follows logically from that of substituted amino acids (3AA-16.4). Examples:

Oxytocin $\overline{\text{Cys-Tyr-Ile-Gln-Asn-Cys}}\text{-Pro-Leu-Gly-NH}_2$

Cyclic ester of threonylglycylglycylglycine $\overline{\text{Thr-Gly-Gly-Gly}}$ or (3AA-17.6) $\text{H-}\overline{\text{Thr-Gly-Gly-Gly}}$

3AA-19.6. *Depsipeptides*

Depsipeptides are oligomers formed from amino acids and other bifunctional acids, usually hydroxy acids. They are often cyclic. In symbolic representation, any special symbols used for the hydroxy acids should be defined.

3AA-19.7. *Peptide Analogues*

▲ Analogues of peptides in which the –CO–NH– group that joins residues is replaced by another grouping may be indicated [25] by placing a Greek psi, followed by the replacing group in parenthesis, between the residue symbols where the change occurs. Examples:

Ala-ψ(NH – CO) – Ala for $NH_3^+ - CHMe - NH - CO - CHMe - COO^-$

Ala-ψ(CH = CH, *trans*)-Ala for $NH_3^+ - CHMe - CH = CH - CHMe - COO^-$

3AA-19.8. *Alignment of Peptide and Nucleic-Acid Sequences*

▲ Although hyphens between residues are important in representing peptide sequences (3AA-16), they may be omitted (I) if it is necessary to align sequences with those of nucleic acids; this is an alternative to separating triplets (II):

```
                MetSerIleGlnHis                          Met-Ser-Ile-Gln-His
(I)   AGTATGAGTATTCAACAT        (II)    AGT ATG AGT ATT CAA CAT
      TCATACTCATAAGTTGTA                TCA TAC TCA TAA GTT GTA
```

Part 2, Section B: THE ONE-LETTER SYSTEM (revision and updating of [11])

3AA-20. THE NEED FOR A CONCISE REPRESENTATION OF SEQUENCE

3AA-20.1. *General Considerations Regarding the One-Letter System*

There are difficulties in using the three-letter system (3AA-14 to 3AA-19) in presenting long protein sequences. A one-letter code is much more concise, and is helpful in summarizing large amounts of data, in aligning and comparing homologous sequences, and in computer techniques for these processes. It may also be used to label residues in three-dimensional pictures of protein molecules.

The possibility of using one-letter symbols was mentioned by Gamow & Yčas [26] in 1958. Šorm *et al.* [27] systematized the idea in 1961 (see, for example, [28]), and Dayhoff and Eck used one-letter symbols derived partly from the code of Šorm *et al.* in their compilations of protein sequences ([29], latest edition [30]). IUB-IUPAC recommendations [11] were approved in 1968 on the basis of proposals of a subcommittee of W. E. Cohn, M. O. Dayhoff, R. V. Eck, and B. Keil, and these recommendations are given here with no substantial change.

3AA-20.2. *Limits of Application of the One-Letter System*

The one-letter system is less easily understood than the three-letter system by those not familiar with it, so it should not be used in simple text or in reporting experimental details of sequence determination. It is therefore recommended for comparisons of long sequences in tables and lists, and in other special uses where brevity is important. If both it and the single-letter system for nucleotide sequences [23] are used in the same paper, particular care should be taken to avoid confusion.

3AA-21. DESCRIPTION OF THE ONE-LETTER SYSTEM

3AA-21.1. *Use of the Code*

The letter written at the left-hand end is that of the amino-acid residue carrying the free amino group, and the letter written at the right-hand end is that of the residue carrying the free carboxyl group. The absence of punctuation beyond either end of a sequence implies that the residue indicated at that end is known to be terminal. A fragmentary sequence is preceded or followed by a slash (/) if its end is not known to be the end of the complete molecule.

3AA-21.2. *The Code Symbols*

The symbols are listed, in alphabetical order of amino-acid names, in Table 1. Table 5 gives them in alphabetical order of symbols.

Note on the Choice of Symbols

Initial letters of the names of the amino acids were chosen where there was no ambiguity. There are six such cases: cysteine, histidine, isoleucine, methionine, serine and valine. All the other amino acids share the initial letters A, G, L, P or T, so arbitrary assignments were made. These letters were assigned to the most frequently occurring and structurally most simple of the amino acids with these initials, alanine (A), glycine (G), leucine (L), proline (P) and threonine (T).

Other assignments were made on the basis of associations that might be helpful in remembering the code, e.g. the phonetic associations of F for *ph*enylalanine and R for *ar*ginine. For tryptophan the double ring of the molecule is associated with the bulky letter W. The letters N and Q

Table 5. *The One-Letter Symbols*

One-letter symbol	Three-letter symbol	Amino acid
A	Ala	alanine
B	Asx	aspartic acid or asparagine
C	Cys	cysteine
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
X	Xaa	unknown or 'other' amino acid
Y	Tyr	tyrosine
▲ Z	Glx	glutamic acid or glutamine (or substances such as 4-carboxyglutamic acid and 5-oxoproline that yield glutamic acid on acid hydrolysis of peptides)

were assigned to asparagine and glutamine respectively; D and E to aspartic and glutamic acids respectively. K and Y were chosen for the two remaining amino acids, lysine and tyrosine, because, of the few remaining letters, they were close alphabetically to the initial letters of the names. U and O were avoided because U is easily confused with V in handwritten material, and O with G, Q, C and D in imperfect computer print-outs, and also with zero. J was avoided because it is absent from several languages.

Two other symbols are often necessary in partly determined sequences, so B was assigned to aspartic acid or asparagine when these have not been distinguished; Z was similarly assigned to glutamic acid or glutamine. X means that the identity of an amino acid is undetermined, or that the amino acid is atypical.

3AA-21.3. *Spacing*

An important use of the one-letter notation is in presenting alignment of homologous sequences. It is therefore vital not to destroy alignment by variable punctuation or variable width of letters. A single space is therefore left between symbols, as a blank if not occupied by punctuation (3AA-21.4 and 21.5), so that such punctuation can be inserted without destroying alignment. Exactly the same spacing is given to each letter, each blank and each punctuation mark, as in typewritten material or, if printed, as in 'typewriter type font'.

3AA-21.4. *Known Sequences*

A blank between letters indicates that the sequence was determined experimentally. For example, A C D E F G H I K L M N P Q means Ala-Cys-Asp-Glu-Phe-Gly-His-Ile-Lys-Leu-Met-Asn-Pro-Gln

3AA-21.5. *Punctuation in Partly Known Sequences*

Parentheses are used to indicate regions of a sequence in which the composition is known, but the sequence undetermined; they are also placed round the symbol for a single residue to show that its identification is tentative. The one-space symbol '=' can be used for ')(' to indicate the end of one unknown sequence and the beginning of another.

If the residue inside parentheses can be positioned with confidence by homology with related proteins, the letters are separated by dots. If their position is arbitrary for lack of even indirect evidence the letters are separated by commas. A slash (/) may be used to separate the symbols for residues that have not been shown experimentally to be corrected, because they are derived from different peptides. A slash before or after a sequence shows that termination has not been demonstrated (3AA-21.1).

This punctuation is illustrated in the comparison of three sequences, where two partly known (a,c) are aligned with a known one (b):

```
a)   (A,C,D)E F G(H.I.K.L=M,N)P Q
b)    R S T E F G H I K L A D P Q
c)    A C D E F/G H I K L(M,N)P Q
```

Thus the sequence of one of the fragments (H.I.K.L) can be inferred with confidence for (a) whereas that of fragments (A,C,D) and (M,N) cannot. Two fragments were sequenced independently in (c). Their positioning is made only by analog with (b).

If more elaborate punctuation is required for special circumstances, it is essential that only one character (or a blank of similar size) should appear between the letters of the code.

Part 3. **Modification of Named Peptides** (a revision and updating of [12])

3AA-22. NAMES AND SYMBOLS FOR DERIVATIVES OF NAMED PEPTIDES

It is often convenient to specify the structure of a peptide by reference to a named sequence of which it is a variant. The recommendations that follow allow this, but they apply only to modifications of the sequence involving normal amide links between residues.

Note. To exemplify any named peptide, the imaginary peptide 'iupaciubin' (to symbolize the harmonious co-operation of IUPAC and IUB), Ala-Lys-Glu-Tyr-Leu, is used in formulating the recommendations below.

3AA-22.1. *Replacement of Residues*

In a peptide of trivial name iupaciubin, if the q^{th} amino-acid residue, starting from the N-terminal end of the chain, is replaced by the amino acid Xaa, the semitrivial name of the modified peptide is [*q*-amino acid]iupaciubin, and the abbreviated form is [Xaa^{q}]iupaciubin. A designation of the chain may be placed before the residue number, e.g. [Ala^{B12}]insulin(cattle) (see Comment f). Examples:

[8-Citrulline]vasopressin, [Cit^{8}]vasopressin;

[5-Isoleucine,7-alanine]angiotensin II, [Ile^{5},Ala^{7}]angiotensin II.

Notes

a) In the full name, the replacement amino acid is designated by its residue name, not the name of its acyl group (e.g. glycine, not glycyl). This name, and the position of replacement, are given in square brackets.

b) In the abbreviated form, the amino-acid residues are designated by standard 3-letter symbols (Table 1), the first letter only being a capital (3AA-14), in square brackets.

c) In the abbreviated form the position of substitution is indicated in a special fashion, i.e. by a superior numeral, to indicate that it is a residue, not an individual atom, that is being replaced.

d) The residue replaced is not designated in these semitrivial names in order to keep the names short, and because this form of nomenclature of 3AA-22.1 clearly differs from ordinary substitution nomenclature.

e) The replacement of an amino-acid residue by its enantiomer may be shown by application of this rule as follows: the ▲ replacement in iupaciubin of L-tyrosine at position 4 by D-tyrosine results in [4-D-tyrosine]iupaciubin with the abbreviations [D-Tyr^{4}]iupaciubin. A mixture of this with iupaciubin gives [4-*ambo*-tyrosine]iupaciubin or [*ambo*-Tyr^{4}]iupaciubin (3AA-13.2 and 3AA-19.2). Examples: [D-Ser^{1}]corticotropin; [D-Asp^{1}]angiotensin II.

▲ f) Specification of a sequence may require the species as well as the peptide to be named. If so, the name of the species should be attached, in parenthesis, to the name of the peptide whenever a modifying prefix is present. Thus a substitution in cattle insulin could give [Ala^{B12}]insulin(cattle). (We prefer 'cattle' as the adjective for this species, since 'bovine' is not used in common speech to designate the species, but to compare human attributes with those of the cow, and 'ox' can be misleading.)

▲ g) It may be convenient to represent replacement of Gln by Glu or of Asn by Asp with the prefix 'desamido'. Thus [Glu^{30}]corticotropin(pig) could be called desamido^{30}-corticotropin(pig). Similarly replacement of Gla by Glu can be designated with the prefix 'decarboxy'.

3AA-22.2. *Extension of the Peptide Chain*

The compounds obtained by extension of a peptide at either the N-terminus on the C-terminus are designated by the kinds of names and abbreviations shown below; these are in accordance with general principles of peptide nomenclature (3AA-13.1). Examples:

a) Extension at the N-terminus

Aminoacyliupaciubin	Xaa-iupaciubin
Valyliupaciubin	Val-iupaciubin
Valylglycyliupaciubin (for extension by two residues)	Val-Gly-iupaciubin

b) Extension at the C-terminus

Iupaciubinylamino acid	Iupaciubinyl-Xaa
Iupaciubinylleucine	Iupaciubinyl-Leu

▲ This rule is not directly applicable to extension at the C-terminus of natural peptides that possess a terminal amide group, such as oxytocin and α-melanotropin. For these, a new name should be given to the corresponding peptide with a free carboxyl group by adding 'oic acid' to the trivial name, e.g. oxytocinoic acid from oxytocin, so that extension can then be denoted as above, e.g. oxytocinoyl-Xaa.

Note

The enkephalins are the two peptides Tyr-Gly-Gly-Phe-Leu and Tyr-Gly-Gly-Phe-Met. Designations such as Leu-, leucyl- and leucine-enkephalin have been given to the former, with corresponding terms for the latter. These all wrongly imply N-terminal extension, and could not be used together with any indication of such extension. Morley [25] has advocated [Leu]enkephalin, if necessary [Leu^5]enkephalin, in accordance with 3AA-22.1, implying that enkephalin means Tyr-Gly-Gly-Phe-Xaa. We believe that [Leu^5]enkephalin is the best designation.

3AA-22.3. *Insertion of Residues*

▲ The compound obtained by insertion of an additional amino-acid residue Xaa in the position between the q^{th} and the $(q+1)^{th}$ residue of the peptide iupaciubin is named endo-*q*a-amino acid-iupaciubin (abbreviated form: endo-Xaa^{qa}-iupaciubin). Example:

endo-4a-tyrosine-angiotensin II, endo-Tyr^{4a}-angiotensin II.

Notes

a) This form has analogies in other fields where endo implies the insertion of something into a structure (e.g. endo-methylene). The prefix or index *q*a is based on analogies with the steroids where the atoms inserted into a ring after atom no. *q* are designated *q*a, *q*b, etc.

b) The prefix 'homo' is not suitable for designating the insertion of a whole residue, especially since it is commonly used to modify the names of individual amino acids, e.g. homoserine (3AA-2.3).

c) Multiple insertions, and insertion of two or more residues together in the same place in the chain, are shown by a logical extension of this recommendation. Thus the insertion into the peptide iupaciubin of threonine between residues 1 and 2, and of valine and glycine (in that order) between residues 4 and 5, is shown by the name 'endo-1a-threonine,4a-valine,4b-glycine-iupaciubin' and the abbreviation 'endo-Thr^{1a},Val^{4a},Gly^{4b}-iupaciubin'.

3AA-22.4. *Removal of Residues*

The compound obtained by the formal removal of an amino-acid residue from the peptide iupaciubin in position *q* is designated by the name des-*q*-amino acid-iupaciubin, abbreviated des-Xaa^q-iupaciubin. Example:

des-7-proline-oxytocin, des-Pro^7-oxytocin

Notes

a) Removal of a whole residue is indicated in a way similar to that for removal of a ring in steroids, e.g. des-*A*-androstane.

b) The form 'de' is not suitable as a prefix because it is easily confused, in speaking, with D (for configuration).

▲ c) Multiple deletions are designated similarly, e.g. des-Ile^3,Asn^5-oxytocin. If a complete sequence is to be removed, the first and last loci of this sequence are all that need be specified, and they should be put in parentheses with a hyphen between them, e.g. des-(B24-B28)-insulin(mouse).

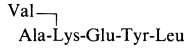
3AA-22.5. *Substitution of Side Chains of Residues*

The compound formed by introducing an additional amino-acid residue as a substituent of the side chain of a residue in a peptide is named by applying the rules of peptide nomenclature (3AA-7, -9 and -13) to the trivial name, as follows.

22.5.1. Acylation of a Side-Chain Amino Group

If the additional residue acylates an amino group of a peptide, the name of the additional amino acid is written, in its acyl group form (3AA-9.3), and prefixed by symbols indicating the position of substitution (atom and residue number, see 3AA-13.4).

Example: the imaginary compound



derived by acylation of the ε-amino group of the lysine residue at position 2 of iupaciubin (Ala-Lys-Glu-Tyr-Leu) with a valyl group is named $N^{\varepsilon 2}$-valyl-iupaciubin (abbreviated $N^{\varepsilon 2}$-Val-iupaciubin).

22.5.2. Other Substituents Named as Prefixes

Other substituents that can be named as prefixes are treated similarly to amino acyl groups. Examples:

Me
|
$N^{6.2}$-Methyliupaciubin for Ala-Lys-Glu-Tyr-Leu

COOH
4|
$C^{4.3}$-Carboxyiupaciubin for Ala-Lys-Glu-Tyr-Leu

(alternative to replacement with Gla under 3AA-22.1)

SO_3H
|
$O^{4.4}$-Sulfoiupaciubin or iupaciubin $O^{4.4}$-sulfate for Ala-Lys-Glu-Tyr-Leu

$O^{4.12}$-Sulfogastrin or gastrin $O^{4.12}$-sulfate for gastrin II.

$N^{\alpha B1}$,$N^{\varepsilon B29}$-bis(Boc)insulin or $N^{2.B1}$,$N^{6.B29}$-bis(Boc)insulin

22.5.3. Acylation by a Side-Chain Carboxyl Group

If an additional amino acid is joined by amide formation between its 2-amino group and the side-chain carboxyl of the peptide iupaciubin, so that the additional amino acid has a free 1-carboxyl group, the derivative is named by specifying the position of substitution (atom and residue number, 3AA-13.4) and it is given the designation 'iupaciubinyl-amino acid'.

Example: the imaginary compound

Val
|
Ala-Lys-Glu-Tyr-Leu

in which valine is acylated by the γ-carboxyl group of a glutamic residue in position 3 (Glu-3) of iupaciubin (Ala-Lys-Glu-Tyr-Leu), is named *N*-(iupaciubin-$C^{\delta 3}$-yl)valine, or *N*-(iupaciubin-$C^{5.3}$-yl)valine, abbreviated to iupaciubin-$C^{5.3}$-yl-Val.

Prefixes may also need to be formed from peptide names for use as substituents in other types of compound, as described in 3AA-13.5.

3AA-22.6. *Partial Sequences* (fragments)

▲ A peptide derived from a named peptide iupaciubin by removal of all residues before the p^{th} and all after the q^{th} is named as iupaciubin-(*p*-*q*)-peptide. Examples:

From α-melanotropin

Ac-Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val-NH_2
1 2 3 4 5 6 7 8 9 10 11 12 13

we may have

Met-Glu-His-Phe-Arg-Trp-Gly or α-melanotropin-(4-10)-peptide
4 10

or, to illustrate the naming of a peptide that contains two fragments, and also a C-terminal amide group:

His-Phe-Arg-Lys-Pro-Val-NH_2 may be called α-melanotropin-(6-8)-(11-13)-peptide amide.

For oligopeptides it may be convenient to state the length of peptide, e.g. iupaciubin-(2-4)-tripeptide, but this is not normally useful for peptides of over about twelve residues, because the larger multiplying affixes are not widely known. Such a check on the number of residues is particularly useful when two or more sequences are joined, e.g. α-melanotropin-(6-8)-(11-13)-hexapeptide amide.

3AA-22.7. *Peptides with Reversed Sequence and Enantiomers*

The peptide whose sequence is the reverse of a named peptide may itself be named with the prefix '*retro*-', giving *retro*-iupaciubin from iupaciubin. The enantiomer of a named peptide may be specified with the prefix '*ent*-' (a contracted form of *enantio*-, F-6.4 of [15]), giving *ent*-iupaciubin from iupaciubin.

3AA-22.8. *Peptide Analogues*

▲ Analogues of peptides in which the –CO–NH– group that connects two residues is replaced by another group (3AA-19.7) may be indicated by placing ψ in square brackets before the name [25]. The ψ is placed between superscripts indicating the residues modified, and is followed by a comma and the replacing group, e.g. [$^{3}\psi^{4}$,CH_2-S]iupaciubin.

3AA-22.9. *Summary of Modification Nomenclature*

The application of the principles of 3AA-22 is illustrated in Table 6.

Table 6. *Application of the methods of modifying named peptides*

Section	Operation	Short name	Structure
	—	Iupaciubin	^{1}Ala-^{2}Lys-^{3}Glu-^{4}Tyr-^{5}Leu
22.1	Replacement	[Phe^{4}]iupaciubin*	Ala-Lys-Glu-^{4}Phe-Leu
22.2a	Extension (N-terminal)	Arg-iupaciubin	Arg-^{1}Ala-Lys-Glu-Tyr-^{5}Leu
22.2b	Extension (C-terminal)	Iupaciubinyl-Met	Ala-Lys-Glu-Tyr-Leu-Met
22.3	Insertion	Endo-Thr^{2a}-iupaciubin	Ala-^{2}Lys-^{2a}Thr-^{3}Glu-Tyr-Leu
22.4	Removal	Des-Glu^{3}-iupaciubin	Ala-^{2}Lys-^{4}Tyr-Leu
22.5.1	Side-chain substitution on amino group	$N^{\varepsilon 2}$-Val-iupaciubin	Val┐ Ala-Lys-Glu-Tyr-Leu
22.5.2	Side-chain substitution on carboxyl group	$C^{\delta 3}$-Iupaciubinyl-Val	Ala-Lys-Glu-Tyr-Leu ┌Val
22.6	Partial sequence	Iupaciubin-(2-4)-peptide	Lys-Glu-Tyr
22.7	Reversal of sequence	*retro*-Iupaciubin	Leu-Tyr-Glu-Lys-Ala
22.7	Enantiomer	*ent*-Iupaciubin	D-Ala-D-Lys-D-Glu-D-Tyr-D-Leu
22.8	Replacement of CO–NH of peptide bond	[$^{3}\psi^{4}$,CH_2–S]iupaciubin	Ala-Lys-Glu-$\psi(CH_2$–S)-Tyr-Leu

* Square brackets are required to indicate replacement, but are not used for most other modifications.

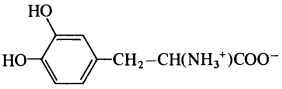
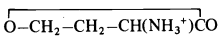
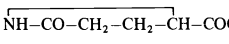
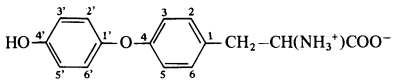
REFERENCES

1. Vickery, H. B. & Schmidt, C. L. A. (1931) *Chem. Rev. 9*, 169–318.
2. Vickery, H. B. (1972) *Adv. Protein Chem. 26*, 81–171.
3. Vickery, H. B. (1947) *J. Biol. Chem. 169*, 237–245.
4. International Union of Pure and Applied Chemistry (IUPAC), Definitive Rules for the Nomenclature of Amino Acids (1960) *J. Amer. Chem. Soc. 82*, 5575–5577.
5. Vickery, H. B. (1963) *J. Org. Chem. 28*, 291–293.
6. IUPAC Commission on the Nomenclature of Organic Chemistry (CNOC) and IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Nomenclature of α-Amino Acids, Recommendations 1974, *Biochem. J. 149*, 1–16 (1975); *Biochemistry, 14*, 449–462 (1975); *Eur. J. Biochem. 53*, 1–14 (1975); also pp. *64–77* in [7].
7. International Union of Biochemistry (1978) *Biochemical Nomenclature and Related Documents*, The Biochemical Society, London.
8. Brand, E. & Edsall, J. T. (1947) *Annu. Rev. Biochem. 16*, 223–272.
9. IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Abbreviated Designation of Amino-Acid Derivatives and Peptides, Recommendations 1966, *Arch. Biochem. Biophys. 121*, 1–5 (1967); *Biochem. J. 102*, 23–27 (1967); *Biochemistry, 5*, 2485–2489 (1966); *Biochim. Biophys. Acta, 121*, 1–7 (1967); *Bull. Soc. Chim. Biol. 49*, 121–129 (1968) (in French); *Eur. J. Biochem. 1*, 375–378 (1967); *Hoppe-Seyler's Z. Physiol. Chem. 348*, 256–261 (1967) (in German); *J. Biol. Chem. 241*, 2491–2495 (1966); *Mol. Biol. 2*, 282–288 (1968) (in Russian).
10. IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Symbols for Amino-Acid Derivatives and Peptides, Recommendations 1971, *Arch. Biochem. Biophys. 150*, 1–8 (1972); *Biochem. J. 126*, 773–780 (1972), corrected *135*, 9 (1973); *Biochemistry, 11*, 1726–1732 (1972); *Biochim. Biophys. Acta, 263*, 205–212 (1972); *Eur. J. Biochem. 27*, 201–207 (1972), corrected *45*, 2 (1974); *J. Biol. Chem. 247*, 977–983 (1972); *Pure Appl. Chem. 40*, 315–331 (1974); also pp. *78–84* in [7].
11. IUPAC-IUB Commission on Biochemical Nomenclature (CBN), A One-Letter Notation for Amino Acid Sequences, 1968, *Arch. Biochem. Biophys. 125* (3), i–v (1968); *Biochem. J. 113*, 1–4 (1969); *Biochemistry, 7*, 2703–2705 (1968); *Biochim. Biophys. Acta, 168*, 6–10 (1968); *Bull. Soc. Chim. Biol. 50*, 1577–1582 (1968) (in French); *Eur. J. Biochem. 5*, 151–153 (1968); *Hoppe-Seyler's Z. Physiol. Chem. 350*, 793–797 (1969) (in German); *J. Biol. Chem. 243*, 3557–3559 (1968); *Mol. Biol. 3*, 473–477 (1969) (in Russian); *Pure Appl. Chem. 31*, 641–645 (1972), also pp. *91–93* in [7].
12. IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Rules for Naming Synthetic Modifications of Natural Peptides, 1966, *Arch. Biochem. Biophys. 121*, 6–8 (1967); *Biochem. J. 104*, 17–19 (1967), corrected *135*, 9 (1973); *Biochemistry, 6*, 362–364 (1967); *Biochim. Biophys. Acta, 133*, 1–5 (1967); *Bull. Soc. Chim. Biol. 49*, 325–330 (1967) (in French); *Eur. J. Biochem. 1*, 379–381 (1967), corrected *45*, 3 (1974); *Hoppe-Seyler's Z. Physiol. Chem. 348*, 262–265 (1967) (in German); *J. Biol. Chem. 242*, 555–557 (1967); *Mol. Biol. 2*, 466–469 (1968) (in Russian); *Pure Appl. Chem. 31*, 647–653 (1972); also pp. *85–87* in [7].
13. IUPAC Commission on Nomenclature of Organic Chemistry (CNOC), Nomenclature of Organic Chemistry, Section E: Stereochemistry, Recommendations 1974, *Pure Appl. Chem. 45*, 11–30 (1976); also pp. *1–18* in [7] and pp. *473–490* in [14].

14. International Union of Pure and Applied Chemistry (1979) *Nomenclature of Organic Chemistry, Sections A, B, C, D, E, F and H*, Pergamon Press, Oxford.
15. IUPAC Commission on the Nomenclature of Organic Chemistry (CNOC), Nomenclature of Organic Chemistry, Section F: Natural Products and Related Compounds, Recommendations 1976, *Eur. J. Biochem. 86*, 1 – 8 (1978); also pp. *19 – 26* in [7] and pp. *491 – 511* in [14].
16. IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains, 1969, *Arch. Biochem. Biophys. 145*, 405 – 421 (1971); *Biochem. J. 121*, 577 – 585 (1971); *Biochemistry, 9*, 3471 – 3479 (1970); *Biochim. Biophys. Acta, 229*, 1 – 17 (1971); *Eur. J. Biochem. 17*, 193 – 201 (1970); *J. Biol. Chem. 245*, 6489 – 6497 (1970); *Mol. Biol. 7*, 289 – 303 (1973) (in Russian); *Pure Appl. Chem. 40*, 291 – 308 (1974); also pp. *94 – 102* in [7].
17. IUPAC Commission on the Nomenclature of Organic Chemistry (CNOC) and IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Tentative Rules for Carbohydrate Nomenclature, Part 1, 1969, *Biochem. J. 125*, 673 – 695 (1971); *Biochemistry, 10*, 3985 – 4004 & 4995 (1971); *Biochim. Biophys. Acta, 244*, 223 – 302 (1971); *Eur. J. Biochem. 21*, 455 – 477 (1971), corrected *25*, 4 (1972); *J. Biol. Chem. 247*, 613 – 635 (1972); also pp. *174 – 195* in [7].
18. Cahn, R. S., Ingold, C. K. & Prelog, V. (1966) *Angew. Chem. 78*, 413 – 447 (in German); *Angew. Chem. Int. Ed. Engl. 5*, 385 – 415, errata 511.
19. IUPAC-IUB Commission on Biochemical Nomenclature (CBN), The Nomenclature of Lipids, Recommendations 1976, *Biochem. J. 171*, 21 – 35 (1978); *Eur. J. Biochem. 79*, 11 – 21 (1977); *Hoppe-Seyler's Z. Physiol. Chem. 358*, 617 – 631 (1977); *Lipids, 12*, 455 – 468 (1977); also pp. *122 – 132* in [7].
20. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), Nomenclature of Tocopherols and Related Compounds, Recommendations 1981, *Arch. Biochem. Biophys. 218*, 347 – 348 (1982); *Eur. J. Biochem. 123*, 473 – 475 (1982); *Pure Appl. Chem. 54*, 1507 – 1510 (1982).
21. IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino Acids), Recommendations 1971, *Arch. Biochem. Biophys. 151*, 597 – 602 (1972); *Biochem. J. 127*, 753 – 756 (1972); *Biochemistry, 11*, 942 – 944 (1972); *Biochim. Biophys. Acta, 278*, 211 – 217 (1972); *Eur. J. Biochem. 26*, 301 – 304 (1972); *J. Biol. Chem. 247*, 323 – 325 (1972); *Mol. Biol. 5*, 492 – 496 (1971) (in Russian); *Pure Appl. Chem. 33*, 439 – 444 (1973); also pp. *88 – 90* in [7].
22. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), Abbreviated Terminology of Oligosaccharide Chains, Recommendations 1980, *Eur. J. Biochem. 126*, 433 – 437 (1982); *J. Biol. Chem. 257*, 3347 – 3351 (1982); *Pure Appl. Chem. 54*, 1517 – 1522 (1982).
23. IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Abbreviations for and Symbols for Nucleic Acids, Polynucleotides and their Constituents, Recommendations 1970, *Arch. Biochem. Biophys. 145*, 425 – 436 (1971); *Biochem. J. 120*, 449 – 454 (1970); *Biochemistry, 9*, 4022 – 4027 (1970); *Biochim. Biophys. Acta, 247*, 1 – 12 (1971); *Eur. J. Biochem. 15*, 203 – 208 (1970), corrected *25*, 1 (1972); *Hoppe-Seyler's Z. Physiol. Chem. 351*, 1055 – 1063 (1970) (in German); *J. Biol. Chem. 245*, 5171 – 5176 (1970); *Mol. Biol. 6*, 166 – 174 (1972) (in Russian); *Pure Appl. Chem. 40*, 277 – 290 (1974); also pp. *116 – 121* in [7].
24. IUPAC-IUB Commission on Biochemical Nomenclature (CBN), Nomenclature of Phosphorus-Containing Compounds of Biochemical Importance, Recommendations 1976, *Biochem. J. 171*, 1 – 19 (1978); *Eur. J. Biochem. 79*, 1 – 9 (1977); *Hoppe-Seyler's Z. Physiol. Chem. 358*, 599 – 616 (1977); *Proc. Natl Acad. Sci. USA, 74*, 2222 – 2230 (1970); also pp. *203 – 311* in [7].
25. Morley, J. S. (1981) *Neuropeptides, 1*, 231 – 235.
26. Gamow, G. & Yčas, M. (1958) *Symposium on Information Theory in Biology*. Pergamon Press, New York.
27. Šorm, F., Keil, B., Vaněček, J., Tomášek, V., Mikeš, O., Meloun, B., Kostka, V. & Holeyšovský, V. (1961). *Collect. Czech. Chem. Commun. 26*, 531 – 578.
28. Keil, B., Prusík, Z. & Šorm, F. (1963) *Biochim. Biophys. Acta, 78*, 559 – 578.
29. Dayhoff, M. O., Eck, R. V., Chang, M. A. & Sochard, M. R. (1965) *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Silver Spring, Maryland.
30. Dayhoff, M. O., in *Atlas of Protein Sequence and Structure*, vol. 5 (1972), suppl. 1 (1973), suppl. 2 (1976) and suppl. 3 (1979). National Biomedical Research Foundation, Washington, DC.

Appendix. **Amino Acids with Trivial Names** (excluding those listed in Table 1)

It is often helpful to use trivial names in order to avoid cumbersome systematic or semisystematic names, particularly if the substance has to be named frequently. Coining of trivial names is treated in 3AA-2.1, and a number of existing trivial names are listed in the appendices to the previous edition of recommendations on amino-acid nomenclature [6]; only the commoner are listed here.

Trivial name	Symbol	Structure of substance or of derived ion in the form predominating at neutral pH
β-Alanine	βAla	$NH_3^+-CH_2-CH_2-COO^-$
Allysine	—	$HCO-[CH_2]_3-CH(NH_3^+)COO^-$
Citrulline	Cit	$NH_2-CO-NH-[CH_2]_3-CH(NH_3^+)COO^-$
Cystathionine	Ala | Hcy	$CH_2-CH(NH_3^+)COO^-$ | $S-[CH_2]_2-CH(NH_3^+)COO^-$
Cysteic acid	Cya	$^-O_3S-CH_2-CH(NH_3^+)COO^-$
Cystine	Cys | Cys	$S-CH_2-CH(NH_3^+)COO^-$ | $S-CH_2-CH(NH_3^+)COO^-$
Dopa	—	
Homocysteine	Hcy	$HS-CH_2-CH_2-CH(NH_3^+)COO^-$
Homoserine	Hse	$HO-CH_2-CH_2-CH(NH_3^+COO^-$
Homoserine lactone	Hsl	
Lanthionine	Ala | Cys	$CH_2-CH(NH_3^+)COO^-$ | $S-CH_2CH(NH_3^+)COO^-$
Ornithine	Orn	$NH_3^+-[CH_2]_3-CH(NH_3^+)COO^-$
5-Oxoproline	Glp	
Sarcosine	Sar	$CH_3-NH_2^+-CH_2-COO^-$
Thyronine	—	
Thyroxine	Thx