Nomenclature of α-Amino Acids (Recommendations, 1974)[1]

IUPAC Commission on the Nomenclature of Organic Chemistry[2] and
IUPAC–IUB Commission on Biochemical Nomenclature[3]

CONTENTS

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 first codified by the former IUPAC Commission on the Nomenclature of Biological Chemistry in 1947 (3) and revised in 1960 (4). Further proposals for the revision of the rules for naming α-amino acids with two centers of chirality appeared in 1963 (5). Recommendations for symbols of amino-acid residues in peptide sequences, together with rules for their use, were made by the present IUPAC–IUB Commission on Biochemical Nomenclature (CBN)[3] (6, 7).

The present revision of the α-Amino-acid Rules brings the nomenclature of these substances into conformity with the IUPAC Rules for the Nomenclature of Organic Chemistry (8, 9). The revision was formulated by a committee composed of H. B. Vickery (convenor), K. Blaha, L. Fowden, W. Klyne, P. M. Scopes, and S. Veibel, with assistance from W. E. Cohn, J. S. Fruton, G. W. Kenner, P. O. Larsen, R. C. Sheppard, and G. T. Young. CNOC[2] and CBN[3] are grateful to these individuals for their efforts.

Recommendations

2AA-1. The Common α-Amino Acids

1.1. Trivial Names of the 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 (8), and formulas, are given in Table I.

1.2. Coining New Trivial Names

(a) The coining of trivial names for newly discovered α-amino acids should be avoided in the absence of compelling reasons.

(b) A new α-amino acid should be named, wherever possible, as a derivative of a well-known α-amino acid (see Recommendation 2AA-2). The configuration, when known, should be indicated in strict conformity with Recommendation 2AA-3.

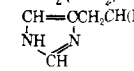
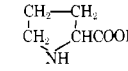
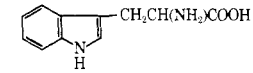
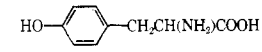
(c) Where names so constructed are unduly cumbersome and the substance is of sufficient importance, a new trivial name may be coined. This name should include either some element of its chemical structure or a reference to its biological origin. Attention should be paid to derivations from

[1] Approved by the authoring Commissions in 1974 and published by permission of IUPAC and IUB. Comments and suggestions for future revisions may be sent to any member of Commissions listed below. Reprints are available from W. E. Cohn, Director, NRC Office of Biochemical Nomenclature, Biology Division, Oak Ridge National Laboratory, Box Y, Oak Ridge, Tennessee, U.S., 37830.

[2] Composition: N. Lozac'h (Chairman), S. P. Klesney (Secretary), K. Bláha, L. C. Cross, H. Grünewald, W. Klyne, K. L. Loening, J. Rigaudy. Associate Members: K. Hirayama, W. H. Powell, S. Veibel, F. Vögtle.

[3] Composition: O. Hoffmann-Ostenhof (Chairman), W. E. Cohn (Secretary), A. E. Braunstein, B. L. Horecker, W. B. Jakoby, P. Karlson, B. Keil, W. Klyne, C. Liébecq, E. C. Webb.

Table I: α-Amino Acids under Direct Genetic Control

Trivial Name	Symbol	Systematic (and Semisystematic) Name	Formula
Alanine	Ala	2-Aminopropionic acid	$CH_3CH(NH_2)COOH$
Arginine	Arg	2-Amino-5-guanidinovaleric acid	$H_2NC(NH)NH(CH_2)_3CH(NH_2)COOH$
Asparagine	Asn	2-Aminosuccinamic acid	$H_2NCOCH_2CH(NH_2)COOH$
Aspartic acid	Asp	Aminosuccinic acid	$HOCOCH_2CH(NH_2)COOH$
Cysteine	Cys	2-Amino-3-mercaptopropionic acid (3-Mercaptoalanine)	$HSCH_2CH(NH_2)COOH$
Glutamine	Gln	2-Aminoglutaramic acid	$H_2NCO(CH_2)_2CH(NH_2)COOH$
Glutamic acid	Glu	2-Aminoglutaric acid	$HOCO(CH_2)_2CH(NH_2)COOH$
Glycine	Gly	Aminoacetic acid	$CH_2(NH_2)COOH$
Histidine	His	α-Amino-1*H*-imidazole-4-propionic acid (1*H*-Imidazole-4-alanine)	
Isoleucine	Ile	2-Amino-3-methylvaleric acid	$C_2H_5CH(CH_3)CH(NH_2)COOH$
Leucine	Leu	2-Amino-4-methylvaleric acid	$(CH_3)_2CHCH_2CH(NH_2)COOH$
Lysine	Lys	2,6-Diaminohexanoic acid	$H_2N(CH_2)_4CH(NH_2)COOH$
Methionine	Met	2-Amino-4-(methylthio)butyric acid	$CH_3S(CH_2)_2CH(NH_2)COOH$
Phenylalanine[a]	Phe	2-Amino-3-phenylpropionic acid	$C_6H_5CH_2CH(NH_2)COOH$
Proline[a]	Pro	2-Pyrrolidinecarboxylic acid	
Serine	Ser	2-Amino-3-hydroxypropionic acid	$HOCH_2CH(NH_2)COOH$
Threonine	Thr	2-Amino-3-hydroxybutyric acid	$CH_3CH(OH)CH(NH_2)COOH$
Tryptophan[a]	Trp	2-Amino-3-(3-indolyl)propionic acid (α-Amino-1*H*-indole-3-propionic acid)	
Tyrosine[a]	Tyr	2-Amino-3-(4-hydroxyphenyl)propionic acid [3-(4-Hydroxyphenyl)alanine]	
Valine	Val	2-Amino-3-methylbutyric acid	$(CH_3)_2CHCH(NH_2)COOH$

[a] See Sections 2.2–2.4 for numbering.

Greek or Latin roots (but not both in the same name) and ease of pronunciation. However, where a new trivial name is introduced or used, it is essential that a correctly constructed systematic or semisystematic name be stated at least once in each paper.

A number of existing trivial names are given in Appendixes A and B.

1.3. Ionic Forms

When it is desirable to mention or stress the ionic nature of an α-amino acid, the three kinds of ions possible may be indicated by adding the terms "anion," "cation," or "amphion" (alternatively, "dipolar ion") to the trivial name of the α-amino acid.

Examples:

glycine anion	$NH_2CH_2COO^-$
glycine cation	$NH_3^+CH_2COOH$
glycine amphion or glycine dipolar ion	$NH_3^+CH_2COO^-$

1.4. Names of Radicals and Residues (8)

1.4.1. The *acyl radical* of an α-amino acid is a structure that lacks the hydroxyl of the carboxyl group ($H_2NCHRCO-$). The names of such radicals are formed by replacing the ending "ine" (or "an" in tryptophan) by "yl."

Examples (8): alanyl, arginyl, leucyl, tryptophyl, tyrosyl, valyl, etc. "Cysteinyl" is used instead of "cysteyl," because of potential confusion with the radical of cysteic acid. "Cystyl" is the diacyl radical of cystine and "half-cystyl" is the acyl radical of cysteine lacking also the H of its SH group.

1.4.2. The monoacyl radicals of the dicarboxylic acids, $HOOC(CH_2)_nCHNH_2CO-$ and $-OC(CH_2)_nCHNH_2-COOH$, are designated α- and β-aspartyl (α- and γ-glutamyl), respectively (8). The diacyl radicals are aspartoyl and glutamoyl (8). The acyl radicals of asparagine and glutamine are termed asparaginyl and glutaminyl, respectively (8).

1.4.3. α-Amino-acid *residues* are structures that lack a hydrogen atom of the amino group (-NHCHRCOOH) or that lack both a hydrogen atom of the amino group and the hydroxyl moiety of the carboxyl group (-NHCHRCO-); all units of a polypeptide chain except for the amino-terminal *acyl radical* are therefore *amino-acid residues*. Residues are named from the trivial name of the α-amino acid (omitting the word "acid" from aspartic acid and glutamic acid). Examples: glycine residue, glutamic residue.

1.4.4. In many cases, biochemical names (see Appendixes) require that radicals formed at a locus other than the carboxyl group of an amino acid be named in terms of the parent amino acid and appear as prefixes rather than roots of the final names.[4] Such "radical-prefix" names representing radicals derived by loss of hydrogen from a nitrogen atom are formed by substituting o- for the terminal -e in those names ending in -e (by analogy with amine → amino, etc.); *e.g.*, alanino, valino, tyrosino. Tryptophan adds the -o directly, and the two dicarboxylic acids become asparto and glutamo. When there is more than one nitrogen atom in the

[4] In this document, a number of special procedures are used to form names based on the approved names of α-amino acids in an attempt to indicate important biochemical relationships. These procedures, yielding such names as alaninol (1.5), N^6-lysino (1.4.4), norleucin-6-yl (1.4.5), etc., should not be extended to other areas.

amino acid, a locant of the form N^x must precede the radical name. Examples: N^6-lysino (see Appendix B2); N^ω-arginino; N^5-glutamino; π-histidino (see 2.4).

1.4.5.[4] Radicals formed by loss of a hydrogen atom from carbon, sulfur, or oxygen atoms (excepting the carboxylic oxygen atoms) are named by substituting -*x*-yl for the terminal -e, *x* being the locant of the atom from which the hydrogen atom has been lost (examples: norleucin-6-yl, cystein-*S*-yl, threonin-O^3-yl; *re* norleucin-6-yl, see Appendix B2; exception: asparagine forms aspartamid-*x*-yl, glutamine forms glutamid-*x*-yl) or adding -*x*-yl to aspartic, glutamic, and tryptophan (examples: aspartic-2-yl, tryptophan-1-yl; see 2.3).

1.5. Aldehydes and Alcohols[4]

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, respectively, the endings "al" and "ol."

Examples ($RCHNH_2CHO$): alaninal, leucinal, lysinal, serinal, aspart-1-al, glutaminal, etc.
($RCHNH_2CH_2OH$): alaninol, leucinol, lysinol, serinol, aspart-1-ol, glutaminol, etc.

The aldehyde and alcohol derivatives of tryptophan take the names tryptophanal and tryptophanol.

1.6. Amides, Anilides, and Analogous Derivatives ($H_2NCHRCONHR'$)

Amides, anilides, and analogous derivatives of α-amino acids are structures in which 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 amino-acid name by the word "amide," "anilide," etc., or by adding these words to the name of the amino acid. Thus, glycinamide, argininamide, leucinanilide, etc.; glycine amide, leucine anilide, etc.

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

1.7. Esters and Salts

The names of esters of amino acids bearing trivial names are derived by replacing "ic acid" or the terminal "e" with "ate" (or adding "ate" to tryptophan) and prefixing these with the radical of the alcohol in the usual fashion.

Examples:

diethyl glutamate[5]	methyl cysteinate
α-ethyl glutamate[5]	methyl prolinate
ethyl glutaminate[5]	methyl tryptophanate

Salts of α-amino acids with bases are named in similar fashion (*e.g.*, sodium tryptophanate, monosodium glutamate). Salts of α-amino acids with acids are named by adding the name of the associated acid to the name of the amino acid (*e.g.*, glycine hydrochloride, histidine hydrochloride, lysine hydrochloride) or by using the " 'ium" form (*e.g.*, glycinium chloride, histidinium chloride, lysinium chloride).

It is permissible and customary to use the "ate" form without naming the cation when discussing the role of amino acids in biochemical reactions or where the co-ion is unimportant, unknown, or a mixture of ions. The "ate" form is preferred when the name of an amino acid is used adjectivally (*e.g.*, "glutamate-dependent reaction" instead of "glutamic-acid-dependent reaction").

1.8. Use of the Prefix "homo"

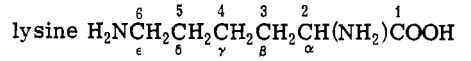
An α-amino acid that is otherwise similar to a common one (Table I) but that contains one more methylene group in the carbon chain may be named by prefixing "homo" to the name of that common α-amino acid. (Examples are included in Appendix A, alphabetized under "H").

2AA-2. Formation of Semisystematic Names of α-Amino Acids

Semisystematic names of substituted α-amino acids are formed according to the general principles of organic nomenclature (8), by attaching the name of the substituting radical to the trivial name of the amino acid. The position of substitution is preferably indicated by numerical locants, although Greek letters are frequently used, especially in the older literature.

2.1. Numbering

In the aliphatic α-amino acids, the carbon atom of the carboxyl group adjacent to the carbon atom carrying the amino group is numbered 1. Where Greek letters are used, the carbon atom adjacent to the carbon atom 1 and carrying the amino group is designated alpha (α). Example:



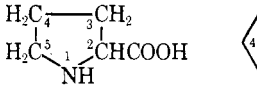
2.2. Proline

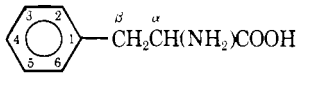
The carbon atoms in proline are numbered as in pyrrolidine, the nitrogen atom being numbered 1, and proceeding toward the carboxyl group.

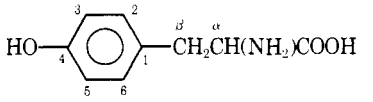
2.3. Aromatic Rings

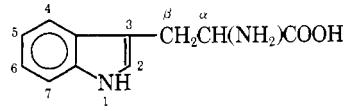
The carbon atoms in the aromatic rings of phenylalanine, tyrosine, and tryptophan are numbered as in systematic nomenclature, with 1 (or 3, for tryptophan) designating the carbon atom holding the aliphatic chain. The carbon atoms of the latter are designated α (for the carbon atom attached to the amino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

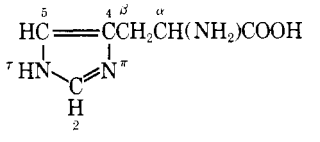
Note: This numbering system should also be used for the decarboxylated products. (*e.g.*, for tryptamine).











[5] Note that glutamate refers to glutamic acid whereas glutaminate refers to glutamine.

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 that of the side chain.[6] The carbon atom between the two imidazole nitrogen atoms is numbered 2 (as in imidazole), and the carbon atom adjacent to the τ nitrogen atom is numbered 5. The carbon atoms of the aliphatic chain are designated α and β, as in 2.1 and 2.3 above.

Note: this numbering should also be used for the decarboxylation product, histamine, and substituted histidines (*e.g.*, anserine, carnosine, homocarnosine, ophidin, thiolhistidine; see Appendix B) utilizing these trivial names.

2AA-3. Configuration at the α-Carbon Atom

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 letters D or L to indicate a formal relationship to D- or L-serine and thus to L- or D-glyceraldehyde. An additional symbol (*i.e.*, a plus or minus sign in parentheses) to denote the direction of rotation is not necessary although it may be inserted where it is essential or desirable to emphasize the direction of rotation under specified conditions. (See 2AA-6.)

Examples: L-leucine, D-valine, L-asparagine, L-canavanine, L(+)-alanine.

3.2. Position of Prefix and Exceptions

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 hyphens.

Examples: 4-methyl-L-glutamic acid, *S*-methyl-L-cysteine, 4-hydroxy-L-proline, 3,5-diiodo-L-tyrosine, 5-hydroxy-L-lysine.

Note: Admissible exceptions to this rule are L-phenylalanine, 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 (1.7), including peptides (2AA-7), the prefix is placed immediately before the trivial name of the parent acid or its radical.

Examples: L-histidine monohydrochloride monohydrate, cupric L-aspartate, D-ornithine dihydrochloride, *N*-acetyl-L-tryptophan, diethyl D-glutamate, 3-hydroxy-D_s-glutamic acid (see 3.4), 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 (D-cystathionine).

3.3. Omission of Prefix

The prefix may be omitted where the amino acid is stated to be or is obviously the enantiomer derived from a protein source and is thus assumed to be L. It may also be omitted where the amino acid is synthetic and not resolved and is therefore, save in rare exceptional cases, an equimolecular mixture of the enantiomers; or in a general statement that is true for either enantiomer or for any mixture of these; or in symbolic representations (6, 7).

3.4. Subscripts to D *and* L

Where confusion is possible between the use of the capital letter prefix for the configuration of the α-carbon atom in amino-acid nomenclature and for that of the highest numbered asymmetric carbon atom in carbohydrate nomenclature (13), a subscript is added to the small capital letter prefix. Where the prefix is used in the amino-acid sense, the subscript s is added; where the prefix is used in the carbohydrate sense, the subscript g is added. These subscripts (lower-case Roman letters) refer, respectively, to serine, the fundamental substance to which α-amino acids that bear structural resemblance to the carbohydrates can be formally related, and to glyceraldehyde, the fundamental substance to which the configuration of the carbohydrates is formally related.

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, they are highly desirable in naming α-amino acids that possess more than one center of chirality. (See 5.2.)

3.5. The Sequence Rule

A more general system of stereochemical designation, which is especially convenient for comparisons with other series of compounds, is the "Sequence Rule" of Cahn, Ingold, and Prelog (9–12). A full discussion of this system (the *RS* system) of naming chiral substances is to be found in the references cited.

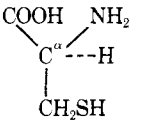
Note that almost all of the enantiomers of the α-amino acids derived from proteins and most of the α-amino acids and their derivatives found in plant or animal tissues have the α-L configuration, which corresponds in almost all cases to *S* in the Sequence Rule convention. The most important exceptions are L-cystine and L-cysteine and their derivatives, which are *R*.[7] A few α-amino acids of the α-D configuration, which generally corresponds to *R* in the Sequence Rule convention (except D-cystine, D-cysteine, and related compounds, which are *S*), are found among the products of hydrolysis of certain antibiotic polypeptides, although rarely elsewhere.

3.6. Amino Acids Derived from Amino Sugars

Amino acids derived from amino sugars containing five or more carbon atoms are named in conformity with the

[6] This recommendation (6) arose from the fact that two different systems of numbering the atoms in the imidazole ring of histidine have been used for so long a time (biochemists generally numbering as 1 the nitrogen atom adjacent to the side chain and organic chemists designating it as 3).

[7] In L-cysteine, the α-carbon atom bears the four groups NH_2, CH_2SH, COOH, and H. They fall in this order of precedence as the atomic number of sulfur is greater than that of oxygen and, if L-cysteine is written in "steering-wheel" fashion,

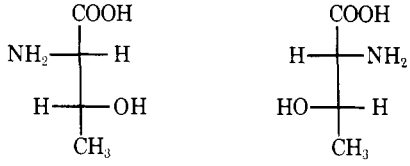


the groups are in the *rectus* (*R*) or clockwise order, thus (*2R*)-cysteine.

Rules of Carbohydrate Nomenclature (13) or with a recommended trivial name.

Examples: (1) D_g-glucosaminic acid for 2-amino-2-deoxy-D_g-gluconic acid, the α-carbon atom of which has the configuration of that in D-serine, and in which carbon atom 5, the highest numbered chiral center, also has the D configuration; (2) D_g-mannosaminic acid for 2-amino-2-deoxy-D_g-mannonic acid, the α-carbon atom of which has the configuration of that in L-serine, but in which carbon atom 5 has the D configuration. The subscript g may be omitted except where confusion with the use of the small capital letter prefixes in amino-acid nomenclature is possible.

Note: the structures of α-amino acids to show configurational relationships may be drawn in several ways. In Fischer-style formulas, the carbon chain is written vertically with the carboxyl group at the top. With L-amino acids, the amino group is shown at the left, with D-amino acids at the right.



L_s-threonine D_s-threonine

Alternatively, heavy and dotted lines may be used to represent bonds projecting respectively in front of and behind the plane of the paper.

2AA-4. Optically Inactive α-Amino Acids

4.1. Use of DL or ±

The optically inactive mixture or racemic compound of the enantiomers is designated by the prefix DL (no comma) or by the plus and minus sign (±) in parentheses.

Examples: DL-leucine, (±)-leucine.

4.2. Use of meso

The prefix *meso* or its abbreviation *ms* in lower case italic letters is used to denote those α-amino acids and derivatives of α-amino acids that are optically inactive because of internal compensation.

Examples: *meso*-lanthionine, *ms*-cystine.

2AA-5. Configuration at Chiral Centers Other than α-C

5.1. Use of Sequence Rule

In general, the Sequence Rule symbols are used to designate configuration, where known, at centers other than α-C. The configuration at α-C is customarily designated by D or L placed immediately before the trivial name.

Examples: (3*R*)-L_s-threonine, (3*S*)-L_s-isoleucine, (4*S*)-4-hydroxy-L_s-proline (see 5.3).

However, those who prefer not to use two different systems to indicate configuration in the same name may convey the same information as in the following examples: (2*S*,3*R*)-threonine, (2*S*,3*S*)-isoleucine, (2*S*,4*S*)-hydroxyproline.

5.2. Use of Carbohydrate Prefixes

Alternatively, names of α-amino acids having two or more chiral centers and in which a hydrogen atom is attached to each center may be formed with use of the prefixes of carbohydrate nomenclature[8] (13) to define the configurational relationships. The small capital letter prefixes L_s or D_s are placed immediately before the trivial name of the parent α-amino acid or its radical. The subscripts must invariably be inserted in order to show that the capital letter prefix refers to the configuration of the α-carbon atom (carbon atom 2). The carbohydrate-name prefix is italicized in print and is not capitalized even at the beginning of a sentence.

Examples:

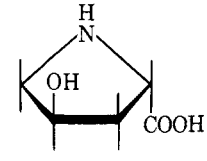
erythro-4-hydroxy-L_s-proline (from proteins)
threo-4-hydroxy-L_s-proline (otherwise allo-4-hydroxy-L_s-proline)
erythro-L_s-isoleucine (from proteins)
threo-L_s-isoleucine (otherwise allo-L_s-isoleucine)
erythro-5-hydroxy-L_s-lysine (from proteins)
threo-5-hydroxy-L_s-lysine (otherwise allohydroxy-L_s-lysine)
ribo-3,4-dihydroxy-L_s-glutamic acid
xylo-3,4-dihydroxy-L_s-glutamic acid
arabino-3,4-dihydroxy-L_s-glutamic acid
lyxo-3,4-dihydroxy-L_s-glutamic acid

Note that the *ribo* and *xylo* isomers belong to the L_g series, the *arabino* and *lyxo* isomers to the D_g series.

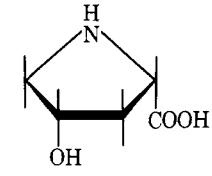
Note further that the relative configuration between the α center and the second center of chirality is already implied in the trivial name threonine.

5.3. Use of cis and trans

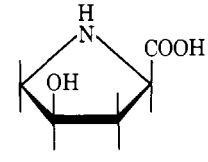
The amino acids 4-hydroxy-L-proline and 3-hydroxy-L-proline and analogous substituted prolines may also be named as follows (*cf.* 3.2, 5.1, 5.2)



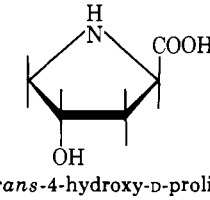
trans-4-hydroxy-L-proline
4-hydroxy-L-proline
erythro-4-hydroxy-L_s-proline



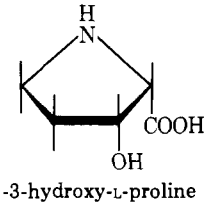
cis-4-hydroxy-L-proline
allo-4-hydroxy-L-proline
threo-4-hydroxy-L_s-proline



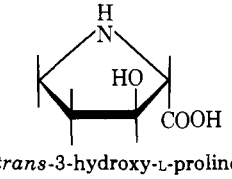
cis-4-hydroxy-D-proline
allo-4-hydroxy-D-proline
threo-4-hydroxy-D_s-proline



trans-4-hydroxy-D-proline
4-hydroxy-D-proline
erythro-4-hydroxy-D_s-proline



cis-3-hydroxy-L-proline
allo-3-hydroxy-L-proline
threo-3-hydroxy-L_s-proline



trans-3-hydroxy-L-proline
3-hydroxy-L-proline
erythro-3-hydroxy-L_s-proline

Cis and trans refer to the relative positions of the hydroxyl and carboxyl groups in each compound.

5.4. Use of allo

Where only the configuration of the α-chiral center of a naturally occurring α-amino acid has been ascertained and that of a second chiral center is unknown so that names

[8] It has been customary in amino-acid nomenclature to apply the carbohydrate prefixes according to Anglo-American usage (Rule Carb-8), in which the prefixes refer to a sequence of consecutive but not necessarily contiguous asymmetric groups.

constructed according to 5.1 or 5.2 cannot be assigned, the diastereoisomer is given the same trivial name with the prefix allo. Such diastereoisomers have the same configuration as the "parent" amino acid at α-C and the opposite configuration to that of the parent at the second chiral center. The examples are names that were assigned before the configuration of the second chiral center had been ascertained.

Examples: allo-L-isoleucine; allo-5-hydroxy-L-lysine; allo-β-hydroxy-L-glutamic acid; allo-L-threonine.

Note that the configuration at the α center is ascertained (a) by direct chemical correlation with substances of known configuration, or (b) by determination of absolute configuration by the methods of Bijvoet *et al.* (14, 15), or (c) tentatively, from the results of studies of biological properties, in particular response to either L-amino-acid oxidase or D-amino-acid oxidase, or (d) by examination of the change in optical rotation with change in the acidity of a solution of the substance.

Where a second chiral center is present, it has been customary in the past to designate the first diastereoisomer described as the L- (or D-) amino acid. The second diastereoisomer, when found or synthesized, is then assigned the same name but with the prefix allo-. Where choice of the names on these grounds is impossible or inappropriate, such designations as "isomer A" and "isomer B" are frequently employed until the full configurational relationships are known. Alternatively, the symbol ξ (Greek xi) for "unknown configuration" may be used if more precise specification seems desirable, for example, as in the name (2*S*)-2-amino-5ξ-hydroxyhexanoic acid, $CH_3CHOH(CH_2)_2$-$CH(NH_2)COOH$.

2AA-6. Optical Rotation

Where the configurational relationship of the α-carbon atom has not been established definitively, or where it is desired to emphasize the actual direction of rotation of an enantiomer of known configuration, the direction of rotation in a specified solvent is shown by a plus or a minus sign in parentheses or by the prefixes *dextro* or *levo* in lower case italic letters.

Examples: (+)-6-hydroxytryptophan, or *dextro*-6-hydroxytryptophan; (+)-glutamic acid, or *dextro*-glutamic acid (for dextrorotatory (in water) L-glutamic acid)

2AA-7. Peptides[9]

7.1. Construction of Names

Since peptides are, formally, aminoacylamino acids, the products of the condensation of two or more α-amino acids whereby the elements of water are eliminated between the α-carboxyl group of one amino acid and the α-amino group of the adjacent amino acid, they are named with use of the radical names ending in "yl." Thus, if amino acids "X-ine" and "Y-ine" (NH_2-X-COOH and NH_2-Y-COOH) condense in this order, the dipeptide (NH_2-X-CONH-Y-COOH) is named "X-yl-Y-ine." If they condense in the reverse order, the dipeptide (NH_2-Y-CONH-X-COOH) is named "Y-yl-X-ine." Tripeptides and higher oligopeptides are named similarly, *i.e.*, X-yl-Y-yl-Z-ine (or -an or -ic acid).

$NH_2CH(COOH)(CH_2)_2CONHCH(CH_2SH)CONHCH_2COOH$

γ-L-glutamyl-L-cysteinylglycine

[9] More comprehensive recommendations for naming peptides, including the use of symbols and prefixes, are to be found in ref 6 and 7.

7.2. Use of Prefixes in Peptide Names

The configurational prefixes are placed immediately before the trivial name of each unit, whether radical, residue, or terminal. The prefixes and radical names, with the exception of the achiral glycine or glycyl, are set off from each other by hyphens (6). (But see note following 7.4.)

Examples: L-alanyl-L-leucine, L-alanyl-D-leucine, DL-alanyl-DL-leucine (mixture of four isomers), glycyl-L-alanine, L-alanylglycine, L-leucyl-L-phenylalanyl-L-isoleucylglycine, glycyl-L-leucyl-L-alanine, L-alanylglycyl-L-leucine.

7.3. Order of Names of Residues[9]

The name of a polypeptide begins with the name of the radical that carries the free α-amino group (NH_2-terminal amino acid) followed in order by the name or names of the residue(s) of the internal α-amino acid(s). The final N-substituted amino acid carries the free carboxyl group (COOH-terminal amino acid). Formulas are usually written according to the same pattern, the NH_2-terminal at the left and the COOH-terminal at the right. The formula of reduced glutathione is written

7.4. Names of Simple Polymers of α-Amino Acids

Simple polymers of amino acids may, if preferred, be named with prefixes indicating the number of amino-acid units involved. More complex polymers are referred to as polyglycine, poly(L-lysine), etc. (7).

Examples: glycylglycylglycine may be named triglycine, but not (diglycyl)glycine; L-leucyl-L-leucyl-L-leucyl-L-leucine may be named tetra-L-leucine, or tri-L-leucyl-L-leucine.

Note: Although the units in a peptide chain—usually in abbreviated form—are usually separated by hyphens, it is necessary to use arrows when formulating a cyclic peptide, the arrow being pointed from the carboxyl of one unit to the amino group of the unit attached to that carboxyl group (6).

7.5. Conformation of Polypeptide Chains

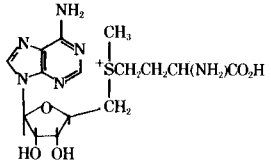
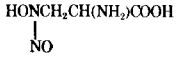
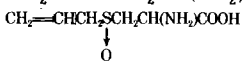
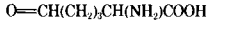
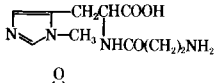
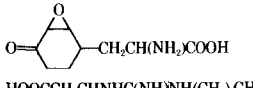
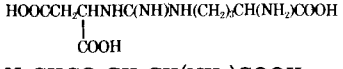
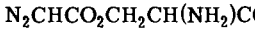
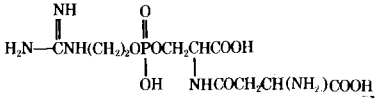
Abbreviations and symbols recommended by CBN[3] for the description of the conformation of polypeptide chains may be found in *J. Mol. Biol. 55,* 299 (1970), *Biochemistry 9,* 3471 (1970), and elsewhere.

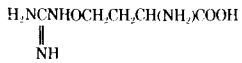
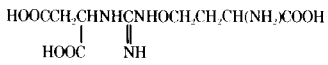
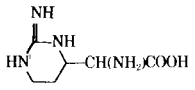
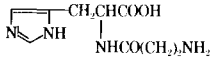
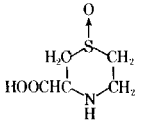
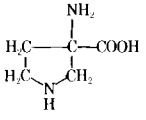
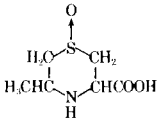
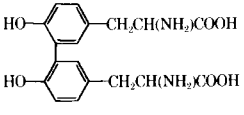
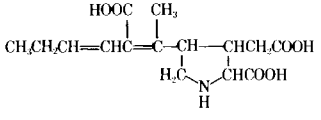
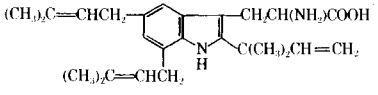
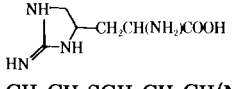
References

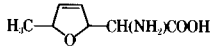
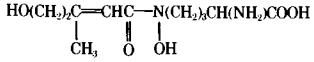
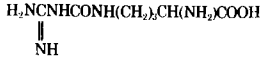
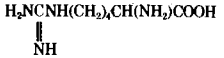
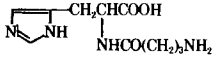
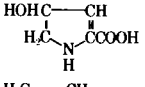
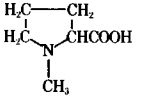
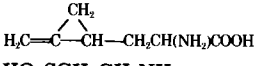
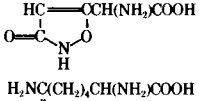
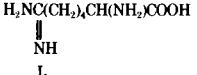
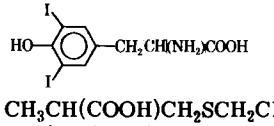
(1) Vickery, H. B., and Schmidt, C. L. A. (1931), *Chem. Rev. 9,* 169.
(2) Vickery, H. B. (1972), *Advan. Protein Chem. 26,* 82.
(3) Vickery, H. B. (1947), *J. Biol. Chem. 169,* 237.
(4) IUPAC, Definitive Rules for the Nomenclature of Amino Acids, (1960), *J. Amer. Chem. Soc. 82,* 5575.
(5) Addendum to ref 4 (H. B. Vickery) (1963), *J. Org. Chem. 28,* 291.
(6) IUPAC-IUB, Symbols for Amino-Acid Derivatives and Peptides (1972), *J. Biol. Chem. 247,* 977; (1972), *Biochim. Biophys. Acta 263,* 205; and elsewhere.
(7) IUPAC-IUB, Abbreviated Nomenclature of Synthetic Polypeptides (1972), *Biochem. J. 127,* 753; (1972) *Arch. Biochem. Biophys. 151,* 597; and elsewhere.
(8) IUPAC, Rules for the Nomenclature of Organic Chemistry, Sections A, B, and C (1971), Butterworths, London (Section C-421 in particular).
(9) IUPAC, Rules for the Nomenclature of Organic Chemistry, Section E, Fundamental Stereochemistry

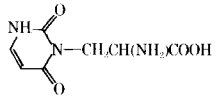
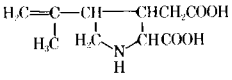
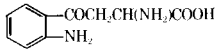
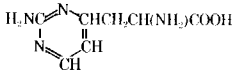
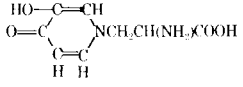
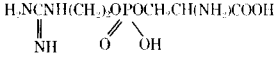
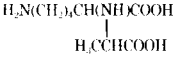
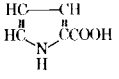
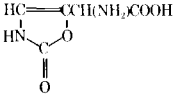
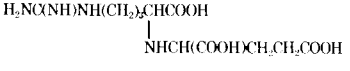
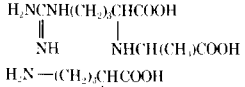
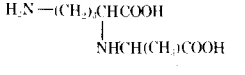
References continue following Appendix B.

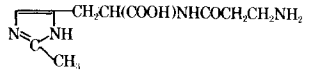
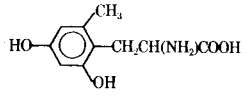
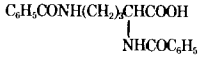
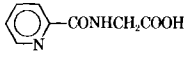
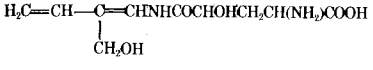
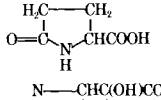
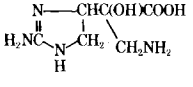
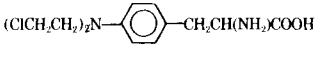
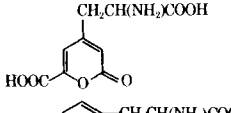
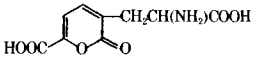
Appendix A: Naturally Occurring α-Amino Carboxylic Acids Bearing Unique Trivial Names[a] (excluding those listed in Table I)

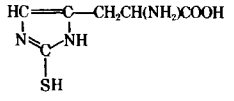
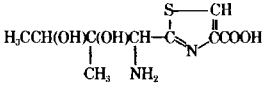
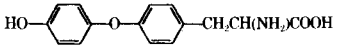
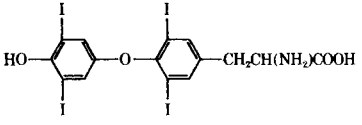
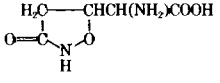
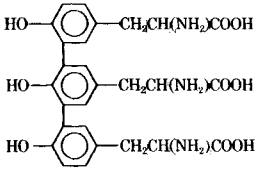
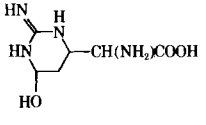
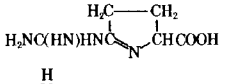
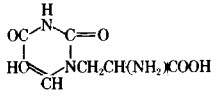
Structure	Trivial name	Biochemical Name(s)[c]	Systematic Name(s) or Other Names[d]
	Abrine[b]	N^{α}-Methyltryptophan	
	*S*-Adenosylmethionine	*S*-(5′-Deoxy-5′-adenosyl)methionine	*S*-5′-[(3-Amino-3-carboxypropyl)-methylsulfonio]-5′-deoxyadenosine
	Agaritine	N^{5}-[4-(Hydroxymethyl)anilino]glutamine	Glutamic 5-{[2-(4-hydroxymethyl)phenyl]-hydrazide}
	β-Alanine[a,b]		3-Aminopropionic acid
	Alanosine	3-(Hydroxynitrosamino)alanine	2-Amino-3-(hydroxynitrosamino)propionic acid
	Albizzine	3-Ureidoalanine	2-Amino-3-ureidopropionic acid
	Alliin	*S*-Allylcysteine *S*-oxide	3-(Allylsulfinyl)alanine
	Alloenduracididine[b]	(2*R*)-Enduracididine	
	Allohydroxylysine[b]	Allo-5-hydroxy-L$_s$-lysine	*threo*-5-Hydroxy-L$_s$-lysine; (5*S*)-5-hydroxy-L-lysine
	Allohydroxyproline[b]	Allo-4-hydroxy-L$_s$-proline	*threo*-4-Hydroxy-L$_s$-proline; *cis*-4-hydroxy-L-proline; (4*S*)-4-hydroxy-L-proline
	Alloisoleucine[b]	Allo-L$_s$-isoleucine	*threo*-L$_s$-Isoleucine
	Allokainic acid[b]	(See kainic acid)	
	Allothreonine[b]	Allo-L$_s$-threonine	(2*S*,3*S*)-Threonine
	Allysine	6-Oxonorleucine	2-Aminoadipaldehydic acid
	Anserine	π-Methylcarnosine; N^{α}-(β-Alanyl)-π-methylhistidine	N^{α}-(3-Aminopropionyl)-π-methylhistidine
	Anticapsin	3-(2,3-Epoxy-4-oxocyclohexyl)alanine	α-Amino-5-oxo-7-oxabicyclo[4.1.0]-heptane-2-propionic acid
	Argininosuccinic acid	N^{ω}-(1,2-Dicarboxyethyl)arginine	*N*-{[(4-Amino-4-carboxybutyl)amino]-iminomethyl}aspartic acid[a]
	Azaserine	O^{3}-(Diazoacetyl)serine	3-[(Diazoacetyl)oxy]alanine; serine diazoacetate ester
	Bonellidine	N^{2}-(γ-Aspartyl)lombricine	*N*-(1-Carboxy-2-hydroxyethyl)asparagine 2-guanidinoethyl hydrogen phosphate (ester)
	Butyrine[a,b]		2-Aminobutyric acid
	Canaline		2-Amino-4-(aminooxy)butyric acid

Structure	Trivial name	Biochemical Name(s)[c]	Systematic Name(s) or Other Names[d]
	Canavanine	N^{ω}-Amidinocanaline	2-Amino-4-(guanidinooxy)butyric acid
	Canavaninosuccinic acid	N^{ω}-(1,2-Dicarboxyethyl)canavanine	{*N*-[(3-Amino-3-carboxypropoxy)amino]-iminomethyl}aspartic acid
	Capreomycidine	2-(Hexahydro-2-iminopyrimidin-4-yl)-glycine; deoxytuberactidine	α-Aminohexahydro-2-imino-4-pyrimidineacetic acid
	Carnosine	N^{α}-(β-Alanyl)histidine	N^{α}-(3-Aminopropionyl)histidine
	Chondrine		3-Thiomorpholinecarboxylic acid 1-oxide; tetrahydro-2*H*-1,4-thiazine-3-carboxylic acid 1-oxide
$H_2NCH_2CH_2PO_3H_2$	Ciliatine[a]		(2-Aminoethyl)phosphonic acid
$H_2NCONH(CH_2)_3CH(NH_2)COOH$	Citrulline	N^5-Carbamoylornithine	2-Amino-5-ureidovaleric acid
	Cucurbitine		3-Amino-3-pyrrolidinecarboxylic acid
	Cycloalliin	5-Methylchondrine	5-Methyl-3-thiomorpholinecarboxylic acid 1-oxide
	Cycloserine[b] (oxamycin)		4-Amino-3-isoxazolidinone
$HO_3SCH_2CH(NH_2)COOH$	Cysteic acid	3-Sulfoalanine	
$HOOCCH(OH)CH_2SO_2CH_2SCH_2CH(NH_2)COOH$	Dichrostachinic acid	*S*-[(2-Carboxy-2-hydroxyethylsulfonyl)-methyl]cysteine	
	Dityrosine	3,3′-Bityrosine	
	Domoic acid	3-(Carboxymethyl)-4-(2-carboxy-1-methyl-1,3-hexadienyl)proline	
	Dopa[b]	(3,4-Dihydroxyphenyl)alanine	
	Echinine	5,7-Bis(3-methyl-2-butenyl)-2-(1,1-dimethylallyl)tryptophan	
	Enduracididine	2-Aminodihydrohistidine (tautomer)	α-Amino-2-imino-4-imidazolidine-propionic acid
$CH_3CH_2SCH_2CH_2CH(NH_2)COOH$	Ethionine	*S*-Ethylhomocysteine	

$HOCH_2CH_2C(CH_3)_2SCH_2CH(NH_2)COOH$	Felinine	*S*-(3-Hydroxy-1,1-dimethylpropyl)cysteine	
	Furanomycin	2-(2,5-Dihydro-5-methyl-2-furanyl)glycine	
	Fusarinine	N^5-Hydroxy-N^5-(5-hydroxy-3-methyl-2-pentenoyl)ornithine	2-Amino-5-[*N*-hydroxy-*N*-(5-hydroxy-3-methyl-2-pentenoyl)amino]valeric acid
	Gigartinine	N^ω-Amidinocitrulline; N^5-(amidinocarbamoyl)ornithine	5-(3-Amidinoureido)-2-aminovaleric acid
$HOOCCH(NH_2)(CH_2)_2CONHCH(CH_2SH)CONHCH_2COOH$	Glutathione	*N*-(*N*-γ-Glutamylcysteinyl)glycine	
$OCHN(OH)CH_2COOH$	Hadacidin[a]	*N*-Formyl-*N*-hydroxyglycine	
	Hippuric acid[b]	*N*-Benzoylglycine	
	Homoarginine	N^6-Amidinolysine	2-Amino-6-guanidinohexanoic acid
	Homocarnosine	N^α-(4-Aminobutyryl)histidine	
$H_2NCONH(CH_2)_4CH(NH_2)COOH$	Homocitrulline	N^6-Carbamoyllysine	2-Amino-6-ureidohexanoic acid
$HSCH_2CH_2CH(NH_2)COOH$	Homocysteine		2-Amino-4-mercaptobutyric acid
	Homoisoleucine[b]		2-Amino-4-methylhexanoic acid
$CH_3S(CH_2)_3CH(NH_2)COOH$	Homomethionine	5-(Methylthio)norvaline	2-Amino-5-(methylthio)valeric acid
$HOCH_2CH_2CH(NH_2)COOH$	Homoserine		2-Amino-4-hydroxybutyric acid
	Hydroxyminaline	2,3-Didehydro-4-hydroxyproline	4-Hydroxy-2-pyrroline-2-carboxylic acid
	Hygric acid	*N*-Methylproline	
	Hypoglycin A	3-(2-Methylenecyclopropyl)alanine	α-Amino-2-methylenecyclopropanepropionic acid
$HO_2SCH_2CH_2NH_2$	Hypotaurine[a]		2-Aminoethanesulfinic acid
	Ibotenic acid	2-(3-Oxo-4-isoxazolin-5-yl)glycine	α-Amino-3-oxo-4-isoxazoline-5-acetic acid
	Indospicine		6-Amidino-2-aminohexanoic acid
	Iodogorgoic acid	3,5-Diiodotyrosine	
$CH_3CH(COOH)CH_2SCH_2CH(NH_2)COOH$	Isobuteine	*S*-(2-Carboxypropyl)cysteine	3-[(2-Carboxypropyl)thio]alanine
$H_2N(CH_2)_3CH(NH_2)CH_2COOH$	Isolysine[a] (β-lysine)		3,6-Diaminohexanoic acid

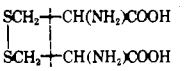
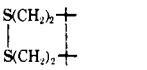
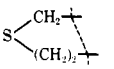
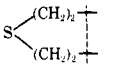
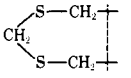
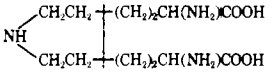
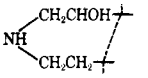
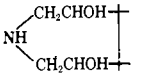
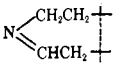
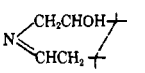
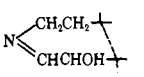
Structure	Trivial name	Biochemical Name(s)[c]	Systematic Name(s) or Other Names[d]
$(CH_3)_2CHCH(COOH)SCH_2CH(NH_2)COOH$	Isovalthine	*S*-(1-Carboxy-2-methylpropyl)cysteine	
	Isowillardine	3-(Uracil-3-yl)alanine	α-Amino-3,6-dihydro-2,6-dioxo-1(2*H*)-pyrimidinepropionic acid
	Kainic acid	3-(Carboxymethyl)-4-isopropenylproline	2-Carboxy-4-isopropenyl-3-pyrrolidine-acetic acid
	Kynurenine	3-Anthraniloylalanine	
$(CH_3)_3N^+(CH_2)_4CH(NH_2)COO^-$	Laminine	N^6-Trimethyllysine; N^6-trimethyllysine betaine	2-Amino-6-(trimethylammonio)hexanoate
$CN(CH_2)_2NHCO(CH_2)_2CH(NH_2)COOH$	Lathyrus factor	N^5-(2-Cyanoethyl)glutamine	
	Lathytine (tingitanine)	3-(2-Amino-4-pyrimidinyl)alanine	α,2-Diamino-4-pyrimidinepropionic acid
	Leucenol (mimosine)	3-[3-Hydroxy-4-oxo-1(4*H*)-pyridinyl]-alanine	
	Lombricine	O^3-(2-Guanidinoethoxy)phosphinicoserine	
	β-Lysine[b] (see isolysine)		
	Lysopine	N^2-(1-Carboxyethyl)lysine	
	Mimosine[b] (see leucenol)		
	Minaline	Tetradehydroproline	2-Pyrrolecarboxylic acid
	Muscazone	3-(2-Oxo-4-oxazolin-5-yl)glycine	α-Amino-2-oxo-4-oxazoline-5-acetic acid
	Nopaline	N^2-(1,3-Dicarboxypropyl)arginine; *N*-(1-carboxy-4-guanidinobutyl)glutamic acid	
	Norleucine[b]		2-Aminohexanoic acid
$HOOCCH(NH_2)(CH_2)_2CONHCH(CH_3)CONHCH_2COOH$	Norophthalmic acid	*N*-[*N*-γ-Glutamylalanyl)glycine	N^5-[1-((Carboxymethyl)carbamoyl)ethyl]-glutamine
	Norvaline[b]		2-Aminovaleric acid
	Octopine	N^2-(1-Carboxyethyl)arginine	
	Octopinic acid	N^2-(1-Carboxyethyl)ornithine	

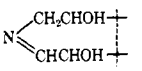
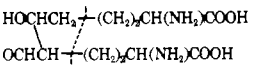
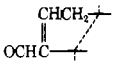
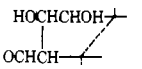
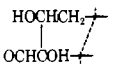
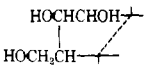
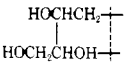
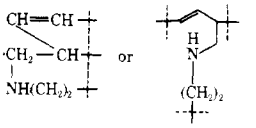
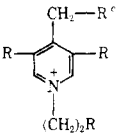
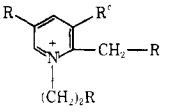
	Ophidin	2-Methylcarnosine; N^{α}-(β-alanyl)-2-methylhistidine	
	Orcylalanine	(2,4-Dihydroxy-6-methylphenyl)alanine	
$H_2N(CH_2)_3CH(NH_2)COOH$	Ornithine	5-Aminonorvaline	2,5-Diaminovaleric acid
	Ornithuric acid	N^2,N^5-Dibenzoylornithine	
	Oxamycin[b] (see cycloserine)		
	Oxypinnatanine	N^5-[2,5-Dihydro-3-(hydroxymethyl)-2-furanyl]-4-hydroxyglutamine	
$HOCH_2C(CH_3)_2CH(NH_2)COOH$	Pantonine	3,3-Dimethylhomoserine	2-Amino-4-hydroxy-3,3-dimethylbutyric acid
$HOCH_2C(CH_3)_2CHOHCONH(CH_2)_2COOH$	Pantothenic acid[a]	N-Pantoyl-(β-alanine); 3-pantoamidopropionic acid	N-(2,4-Dihydroxy-3,3-dimethylbutyryl)-(β-alanine)
$(CH_3)_2C(SH)CH(NH_2)COOH$	Penicillamine	3-Mercaptovaline; 3,3-dimethylcysteine	2-Amino-3-mercapto-3-methylbutyric acid
$C_6H_5CH_2CONHCH_2COOH$	Phenaceturic acid	N-(Phenylacetyl)glycine	
	Picolinuric acid	N-Picolinoylglycine; N-(2-pyridylcarbonyl)glycine	
	Pinnatanine	4-Hydroxy-N^5-[2-(hydroxymethyl)-1,3-butadienyl]glutamine	
	Piperidinic acid[a,b]		4-Aminobutyric acid
	Pyroglutamic acid	5-Oxoproline; 5-pyrrolidone-2-carboxylic acid	5-Oxopyrrolidine-2-carboxylic acid
	Roseanine[a]		3-Amino-2-(2-amino-2-imidazolin-4-yl)-2-hydroxypropionic acid; 2-amino-α-(aminomethyl)-α-hydroxy-2-imidazoline-4-acetic acid
$HOOC(CH_2)_2CH(COOH)NH(CH_2)_4CH(NH_2)COOH$	Saccharopine	N-(5-Amino-5-carboxypentyl)glutamic acid; N^6-(glutar-2-yl)lysine	2-Amino-6-[(1,3-dicarboxypropyl)amino]-hexanoic acid
	Sarcolysine	4-[Bis(2-chloroethyl)amino]phenylalanine	
	Sarcosine[b]	N-Methylglycine	
	Stizolobic acid	3-(6-Carboxy-2-pyron-4-yl)alanine	α-Amino-6-carboxy-2-oxo-2H-pyran-4-propionic acid
	Stizolobinic acid	3-(6-Carboxy-2-pyron-3-yl)alanine	α-Amino-6-carboxy-2-oxo-2H-pyran-3-propionic acid
	Surinamine[b]	N-Methyltyrosine	

Structure	Trivial name	Biochemical Name(s)[c]	Systematic Name(s) or Other Names[d]
$HOOCCH(NH_2)(CH_2)_2CH(OH)CH(NH_2)COOH$	Tabtoxinine	4-(Alanin-3-yl)threonine; 4-(2-amino-2-carboxyethyl)threonine	2,6-Diamino-3-hydroxyheptanedioic acid
$H_2NCH_2CH_2SO_3H$	Taurine[a]		2-Aminoethanesulfonic acid
$H_2NC(=NH)NH(CH_2)_2OP(O_2H)OCH_2CH(COOH)N(CH_3)_2$	Thallasemine	N^2,N^2-Dimethyllombrocine	*N,N*-Dimethylserine 2-guanidinoethyl hydrogen phosphate (ester); *N,N*-dimethyl-O^3-[(2-guanidinoethoxy)phosphinico]serine
$C_2H_5NHCOCH_2CH_2CH(NH_2)COOH$	Theanine	N^5-Ethylglutamine	
	Thiolhistidine	2-Mercaptohistidine	
	Thiostreptine		2-(1-Amino-2,3-dihydroxy-2-methylbutyl)-4-thiazolecarboxylic acid
	Thyronine	4-(4-Hydroxyphenoxy)phenylalanine	O^4-(4-Hydroxyphenyl)tyrosine
	Thyroxin	3,5,3′,5′-Tetraiodothyronine	O^4-(4-Hydroxy-3,5-diiodophenyl)-3,5-diiodotyrosine
	Tingitanine[b] (see lathytine)		
	Tricholomic acid	2-(3-Oxo-5-isoxazolidinyl)glycine	α-Amino-3-oxo-5-isoxazolidineacetic acid
	Trityrosine[e]	3,3′;5′,3″-Tertyrosine	
	Tuberactidine	2-(Hexahydro-6-hydroxy-2-iminopyrimidin-4-yl)glycine; 6-hydroxycapreomycidine	α-Aminohexahydro-6-hydroxy-2-imino-4-pyrimidineacetic acid
	m-Tyrosine[b]	(3-Hydroxyphenyl)alanine	
	Viomycidine	1,5-Didehydro-5-guanidinoproline	2-Guanidino-1-pyrroline-5-carboxylic acid
	Willardine	3-(Uracil-1-yl)alanine	α-Amino-3,4-dihydro-2,4-dioxo-1(2*H*)-pyrimidinepropionic acid

Footnotes to Appendix A: [a] Some amino acids that are not α-aminoalkanoic acids are included; their trivial names are so marked. No D-amino acids are included, hence L has been omitted from the names. No α-betaines are included. [b] Indicates no structure is given. [c] "Biochemical names" are based on the names of amino acids defined in this document, and attempt thereby to indicate biochemical relationships as far as possible. [d] Systematic or other names avoid the use of names not included among the most common α-amino acids (Table I) and avoid the radical-prefix names described in Recommendation 2AA-1.4.4. [e] As this name implies tyrosyltyrosyltyrosine, tertyrosine (column 3) should be used.

Appendix B: Trivially Named Di(α-amino acids) Other than Peptides

Name [Symbol]	Systematic Names	Formula
B1. Two α-Amino Acids Connected by S or S–S or S–CH_2–S		
Cystine [(Cys)$_2$;Cys(Cys)]	2,2′-Diamino-3,3′-dithiobis(propionic acid) 3,3′-Dithiobis(2-aminopropionic acid) 3,3′-Dithiodialanine	
Homocystine [(Hcy)$_2$;Hcy(Hcy)]	2,2′-Diamino-4,4′-dithiobis(butyric acid) 4,4′-Dithiobis(2-aminobutyric acid)	
Cystathionine [Ala;Hcy(Ala)]	*S*-(Alanin-3-yl)homocysteine *S*-(2-Amino-2-carboxyethyl)homocysteine 2-Amino-4-[(2-amino-2-carboxyethyl)thio]-butyric acid	
Lanthionine [Cys(Ala);Ala(Cys)]	*S*-(Alanin-3-yl)cysteine *S*-(2-Amino-2-carboxyethyl)cysteine 2,2′-Diamino-3,3′-thiobis(propionic acid)	
Methyllanthionine [Cys(3MeAla);3MeAla(Cys)]	*S*-(3-Methylalanin-3-yl)cysteine 2-Amino-3-[(2-amino-2-carboxyethyl)thio]-butyric acid	
Homolanthionine [Hcy(Abu)]	2,2′-Diamino-4,4′-thiobis(butyric acid) 4,4′-Thiobis(2-aminobutyric acid)	
Djenkolic acid [Cys$_2$(CH_2)]	3,3′-Methylenedithiobis(2-aminopropionic acid) 3,3′-(Methylenedithio)dialanine Methylenebiscysteine	
B2. Lysinonorleucines:[a] Two α-Amino Acids Connected by N		
Lysinonorleucine [Lys(ωNle)] [HN(ωNle)$_2$]	6-(N^6-Lysino)norleucine N^6-(Norleucin-6-yl)lysine N^6-(5-Amino-5-carboxypentyl)lysine Bis(6-norleucin-6-yl)amine	
, 5-Hydroxy- [o^5Lys(ωNle)]	6-[N^6-(5-Hydroxylysino)]norleucine N^6-(5-Hydroxynorleucin-6-yl)lysine 5-Hydroxy-N^6-(norleucin-6-yl)lysine N^6-(5-Amino-5-carboxypentyl)-5-hydroxylysine	
, 5,5′-Dihydroxy- [o^5Lys(o^5ωNle)] [NH(o^5ωNle)$_2$]	5-Hydroxy-6-[N^6-(5-hydroxylysino)]-norleucine 5-Hydroxy-N^6-(5-hydroxynorleucin-6-yl)-lysine N^6-(5-Amino-5-carboxy-2-hydroxypentyl)-5-hydroxylysine	
, N^6:6′-Dehydro- [Δ^6Lys(ωNle)]	6-(Δ^6-Lysino)norleucine N^6-(Norleucine-6,6-diyl)lysine N^6-(5-Amino-5-carboxypentylidene)-lysine	
, N^6:6′-Dehydro-5-hydroxy- [Δ^6Lys(o^5ωNle)]	5-Hydroxy-6-(Δ^6-lysino)norleucine N^6-(5-Hydroxynorleucin-6-yl)-Δ^6-lysine N^6-(5-Amino-5-carboxypentylidene)-5-hydroxylysine	
[Δ^6o^5Lys(ωNle)]	6-[Δ^6-(5-Hydroxylysino)]norleucine 5-Hydroxy-N^6-(norleucin-6-yl)-Δ^6-lysine N^6-(5-Amino-5-carboxy-2-hydroxypentylidene)lysine	

Name [Symbol]	Systematic Names	Formula
,N^6:6′-Dehydro-5,5′-dihydroxy-[Δ^6o^{5}Lys(o$^5\omega$Nle)]	5-Hydroxy-6-[Δ^6-(5-hydroxylysino)]-norleucine 5-Hydroxy-N^6-(5-hydroxynorleucin-6-yl)-Δ^6-lysine N^6-(5-Amino-5-carboxy-2-hydroxypentylidene)-5-hydroxylysine	
B3. Allysines:[b] Two (or More) Six-Carbon α-Amino Acids Connected by a Carbon–Carbon Bond		
Allysine aldol [(All)$_2$] Di(allysine)	5-(6-Hydroxynorleucin-6-yl)-6-oxonorleucine[a] 6-Hydroxy-6-(6-oxonorleucin-5-yl)norleucine 2,10-Diamino-5-formyl-6-hydroxyundecanedioic acid	
Dehydro(allysine aldol)[Δ(All)$_2$] Δ^5-(Allysine aldol)	2,10-Diamino-5-formyl-5-undecenedioic acid	
Syndesine [o(All)$_2$] Hydroxy(allysine aldol)	2,10-Diamino-5-formyl-6,7-dihydroxyundecanedioic acid	
Hydroxy(diallysine) Hydroxyallysine-allysine	2,10-Diamino-5-formyl-5,6-dihydroxyundecanedioic acid	
Syndesinol [o(All)$_2$H$_2$] Reduced syndesine Syndesine alcohol	2,10-Diamino-5,6-dihydroxy-7-(hydroxymethyl)undecanedioic acid	
Dihydroxydi(allysine)	2,10-Diamino-5,6-dihydroxy-5-(hydroxymethyl)undecanedioic acid	
Merodesmosine	2,10-Diamino-5-{[(5-amino-5-carboxypentyl)amino]methyl}-6-undecenedioic acid	
Desmosine	4-(4-Amino-4-carboxybutyl)-1-(5-amino-5-carboxypentyl)-3,5-bis(3-amino-3-carboxypropyl)pyridinium	
Isodesmosine	2-(4-Amino-4-carboxybutyl)-1-(5-amino-5-carboxypentyl)-3,5-bis(3-amino-3-carboxypropyl)pyridinium	

[a] For definition of lysino and norleucin-6-yl, see Section 1.4. [b] Allysine is difined in Appendix A as 6-oxonorleucine. [c] R = $(CH_2)_2CH(NH_2)COOH$.

(Tentative Rules) (1970), *J. Org. Chem. 35,* 2849; (1970) *Biochim. Biophys. Acta 208,* 1.
(10) Cahn, R. S., Ingold, C., and Prelog, V. (1966), *Angew. Chem., Int. Ed. Engl. 5,* 385, 511.
(11) Cahn, R. S., Ingold, C., and Prelog, V. (1966), *Angew. Chem. 78,* 413; (1956) *Experientia 12,* 81.
(12) Cahn, R. S. (1964), *J. Chem. Educ. 41,* 116.
(13) IUPAC–IUB, Tentative Rules of Carbohydrate Nomenclature (1972), *Biochemistry 10,* 3983; (1972) *Eur. J. Biochem. 21,* 455; and elsewhere.
(14) Bijvoet, J. M., Peerdeman, A. F., and van Bommel, A. J. (1951), *Nature (London) 168,* 271.
(15) Peerdeman, A. F., van Bommel, A. J., and Bijvoet, J. M. (1951), *Proc. Acad. Sci. Amsterdam B54,* 16.