INTERNATIONAL UNION OF
PURE AND APPLIED CHEMISTRY
AND
INTERNATIONAL UNION OF BIOCHEMISTRY

SYMBOLS FOR AMINO-ACID DERIVATIVES AND PEPTIDES

RULES APPROVED 1974

Issued by the
IUPAC–IUB Commission on Biochemical Nomenclature

LONDON
BUTTERWORTHS

SYMBOLS FOR AMINO-ACID DERIVATIVES AND PEPTIDES†

IUPAC–IUB COMMISSION ON BIOCHEMICAL NOMENCLATURE (CBN)‡

Rules Approved 1974

The revised 'Tentative Rules' published by CBN in 1966[1] were an attempt to achieve a broad systematization of various types of abbreviated notation already in use [e.g. Brand and Edsall (1947) *Annu. Rev. Biochem.*, **16**, 224; Report of the Committee on Abbreviations of the American Society of Biological Chemists, 18 December 1959; Report of the Committee on Nomenclature of the European Peptide Symposium, Pergamon Press, 1963, pp. 261–269; 'Tentative Rules for Abbreviations and Symbols of Chemical Names of Special Interest in Biological Chemistry'[2]]. They sought to reconcile the needs of the protein chemist, i.e. indication of amino-acid sequences, with those of persons concerned more with the chemical reactions of proteins and the synthesis of polypeptides, i.e. the need for conveying more detailed chemical information in abbreviated form.

Recent progress in the field of peptide synthesis and in the chemical modification of proteins has made necessary a revision of these 'Tentative Rules'. This revision has been aided by the work of an expert group consisting of J. S. Fruton, B. S. Hartley, R. R. Porter, J. Rudinger, R. Schwyzer and G. T. Young. They are greatly indebted to many colleagues, notably W. H. Stein, for helpful suggestions.

1. GENERAL CONSIDERATIONS

1.1 The symbols chosen are derived from the trivial names or chemical names of the amino acids and of chemicals reacting with amino acids and polypeptides. For the sake of clarity, brevity and listing in tables, the symbols for amino-acid residues have been, wherever possible, restricted to three letters, usually the first letters of the trivial names.

1.2 The symbols represent not only the names of the compounds but also their structural formulae.

---

† This document is a revision of proposals published in provisional form as Tentative Nomenclature Appendix No. 23 (June 1972) to IUPAC *Information Bulletin* and in: *Arch. Biochem. Biophys.* **150**, 1 (1972); *Biochem. J.* **126**, 773 (1972); *Biochemistry*, **11**, 1726 (1972); *Biochim. Biophys. Acta*, **263**, 205 (1972); *Europ. J. Biochem.* **27**, 201 (1972); *J. Biol. Chem.* **247**, 977 (1972).

Comments on and suggestions for future revisions of these rules should be sent to Prof. O. Hoffmann-Ostenhof, Institut für Allgemeine Biochemie der Universität Wien, Währingerstrasse 38, A-1090 Wien, Austria.

‡ 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.

1.3 The amino-acid symbols by themselves represent the amino acids. The use of the symbols to represent the free amino acids is *not* recommended in textual material, but such use may occasionally be desirable in tables, diagrams or figures. Residues of amino acids are represented by addition of hyphens in specific positions as indicated in Section 3.

1.4 Heteroatoms of amino-acid residues (e.g. O^3 and S^3 of serine and cysteine, respectively, N^6 of lysine, N^2 of glycine, etc.) do not explicitly appear in the symbol; such features are understood to be encompassed by the abbreviation.

1.5 Amino-acid symbols denote the L configuration unless otherwise indicated by D or DL appearing before the symbol and separated from it by a hyphen. When it is desired to make the number of amino-acid residues appear more clearly, the hyphen between the configurational prefix and the symbol may be omitted (see 6.3.1.1. et seq.). (Note: The designation of an amino-acid residue as DL is inappropriate for compounds having another amino-acid residue with an asymmetrical center.)

1.6 Structural formulae of complicated features may be used along with their abbreviated notation wherever necessary for clarity.

1.7 All symbols listed below are to be printed or typed as one capital letter followed by two lower-case letters, e.g. Gln, not GLN or gln or GlN or glN, regardless of position in a sentence or structure. However, when used for purposes other than to represent an amino-acid residue (e.g. to designate a genetic factor), three lower-case italic letters (i.e. *gln*) should be used.

2. SYMBOLS FOR AMINO ACIDS

2.1 Common Amino Acids

Alanine	Ala	Leucine	Leu
Arginine	Arg	Lysine	Lys
Asparagine	Asn†	Methionine	Met
Aspartic acid	Asp	Phenylalanine	Phe
Cysteine	Cys	Proline	Pro
Glutamic acid	Glu	Serine	Ser
Glutamine	Gln†	Threonine	Thr
Glycine	Gly	Tryptophan	Trp (not Try)
Histidine	His	Tyrosine	Tyr
Isoleucine	Ile	Valine	Val

2.2 Less-common Amino Acids

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

---

† Asparagine and glutamine may also be denoted as $Asp(NH_2)$ or $\underset{\displaystyle NH_2}{\overset{|}{Asp}}$, and $Glu(NH_2)$ or $\underset{\displaystyle NH_2}{\overset{|}{Glu}}$, respectively, if necessary (as when the NH_2 is substituted, or its removal or modification is under discussion). See 4.2.

Glx may be used when the residue denoted could be 'glutamic acid or glutamine'; similarly, Asx for 'aspartic acid or asparagine'.

2.2.1 *Hydroxyamino acids*

		Preferred alternatives		
5-Hydroxylysine	5Hyl	Lys(5OH)	or	Lys 5| OH
3-Hydroxyproline	3Hyp	Pro(3OH)	or	OH 3| Pro
4-Hydroxyproline	4Hyp	Pro(4OH)	or	Pro 4| OH

2.2.2 *Allo-Amino Acids*

Alloisoleucine	aIle			OH 5|
Allohydroxylysine	aHyl	aLys(5OH)	or	aLys

2.2.3 *'Nor' and 'Homo' Amino Acids*

'Nor' (e.g. in norvaline) is not used in its accepted sense (denoting a lower homologue) but to change the trivial name of a branched-chain compound into that of a straight-chain compound (compare with 'iso', paragraph 2.1). 'Nor' should therefore be treated as part of the trivial name without special emphasis. 'Homo', used in the sense of a higher homologue, may also be incorporated into the trivial name.

Norvaline	Nva	Homoserine	Hse
Norleucine	Nle	Homocysteine	Hcy

2.2.4 *Higher Unbranched Amino Acids*

The functional prefix 'amino' is included in the symbol as the letter 'A', diamino as 'A_2'†. The trivial name of the parent acid is abbreviated to two letters. The word 'acid' ('-säure', etc.) is omitted from the symbol as carrying no significant information. Unless otherwise indicated, single groups are in the 2 position, two amino groups in the 2 and terminal positions (monocarboxylic acids) or 2 and 2′ positions (dicarboxylic acids). The location of amino groups in positions other than these is shown by appropriate prefixes.

Examples:

2-Aminobutyric acid	Abu
2-Aminoadipic acid	Aad
2-Aminopimelic acid	Apm
2,4-Diaminobutyric acid	A_2bu †
2,2′-Diaminopimelic acid	A_2pm †

† The symbols for diamino compounds previously[1] utilized the letter 'D' for 'diamino'. However, the overuse of D as the initial letter for many compounds beginning with 'di' (and of 'T' for 'tri' and 'tetra'), in addition to the fact that standard chemical symbolism utilizes subscript numerals for multipliers, leads to the proposal that diamino should be represented by A_2. This eliminates the ambiguity attached to 'D' and makes more clear the chemical relationship between the diamino and monoamino derivatives. It is in keeping with the increasing use of Me_2SO instead of DMSO and of Me_3Si- in place of TMS-, and with the earlier proposal of H_4 for tetrahydro[4].

2,3-Diaminopropionic acid	A_2pr[3]
$^{NH_2}_{3\vert}$ or $Ala(3NH_2)$ or Ala (see 4.3)	
β-Alanine	βAla
Orinithin (2,4-diaminovaleric acid)	Orn
6-Aminohexanoic	εAhx†
3-Aminoadipic acid	βAad

2.2.5 *N^2-Alkylated Amino Acids*

N^2-Alkylamino acids are becoming more and more common (e.g. in the large group of depsipeptides). This justifies special symbols.

Examples

N-Methylglycine (sarcosine) (see 6.2)	MeGly or Sar
N-Methylisoleucine	MeIle
N-Methylvaline, etc.	MeVal, etc.
N-Ethylglycine, etc.	EtGly, etc.

2.3 Nonamino-acid residues linked to peptides

For residues of muramic, sialic, neuraminic, etc., acids linked to amino-acid residues, as in bacterial cell-wall components, the symbols Mur, Sia, Neu, etc. (preceded by Ac if *N*-acetylated) are recommended. The symbols for sugar residues (Glc, Gal, etc.)[2] and nucleosides (Ado, Cyd, etc.)[3] may also be used.

3. AMINO-ACID RESIDUES

The links between residues have frequently been shown by peptide chemists as full points (periods, dots) and by carbohydrate chemists (generally) as short strokes (dashes, hyphens). At times, special symbols have been used (> or →) to show the direction of what is in all cases an unsymmetrical link (peptide or glycoside).

For consistency and ease of typing as well as economy in printing, the hyphen, representing the peptide bond, should be the standard connecting symbol[2].

The simple usage by which Gly-Gly-Gly stands for glycylglycylglycine appears to involve the employment of the same three letters (Gly) for three different residues or radicals (b), (c), (d) below. However, if the dashes or hyphens are considered as part of each symbol, we have four distinct forms, for the free amino acid and the three residues, viz.:

(a) Gly = $NH_2—CH_2—CO_2H$ the free amino acid
(b) Gly- = $NH_2—CH_2—CO—$ the left-hand unit
(c) -Gly- = $—NH—CH_2—CO—$ the middle unit
(d) -Gly = $—NH—CH_2—CO_2H$ the right-hand unit

† Recommended in place of the previous[1] εAcp, in which 'cp' for caproic may be confused with capric and caprylic.

For peptides, a distinction may be made between the *peptide*, e.g. Gly-Glu (shown *without* dashes at the ends of the symbols), and the *sequence*, e.g. -Gly-Glu- (shown *with* dashes at the ends of the symbols).

3.1 Lack of Hydrogen on the 2-Amino Group

The 2-amino group is understood to be at the left-hand side of the symbol when hyphens are used, and—in special cases—at the point of the arrow when arrows are used to indicate direction of peptide bond (—CO → NH—, —NH ← CO—). (For substitution for 2-amino hydrogen, see 4.1.)

Examples:

-Gly: —HNCH$_2$COOH

>Gly or $^{\perp}$Gly: >NCH$_2$COOH

-Ala: —HNCH(CH$_3$)COOH

>Ala or $^{\perp}$Ala: >NCH(CH$_3$)COOH

3.2 Lack of Hydroxyl on the 1-Carboxyl Group

The 1-carboxyl group is understood to be on the right-hand side of the symbol when hyphens are employed and—in such special cases as 6.3.1.3—at the tail of the arrow when arrows are used to indicate the direction of the peptide bond (—CO → NH—, —NH ← CO—).

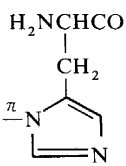
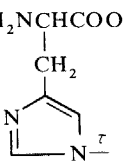
Example: Gly-: $H_2NCH_2CO—$

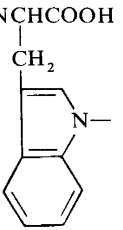
It is generally convenient to use the same abbreviated formula for a polypeptide no matter what its state of ionization. To show that a peptide is acting as a cation or anion the amino-terminal and carboxyl-terminal ends of the peptide are amplified with H and OH, respectively (I); these may be modified to show the appropriate state of ionization (II or III).

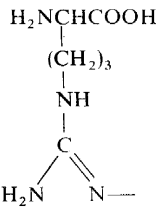
H-Gly-Val-Thr-OH	or	Gly-Val-Thr	(I)
$^{+}H_2$-Gly-Val-Thr-OH	or	$^{+}$HGly-Val-Thr	(II)
H-Gly-Val-Thr-O^{-}	or	Gly-Val-ThrO^{-}	(III)

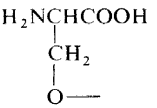
3.3 Lack of Hydrogen on Amino, Imino, Guanidino, Hydroxyl, and Thiol Functions in the Side Chain (for substitution in such positions, see 4.2)

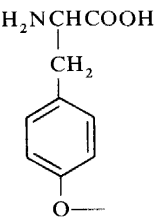
Lys or Lys: —NHCH(CH$_2$)$_4$NH—

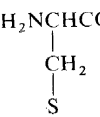
His or His:†  or 

Trp or Trp: 

Arg or Arg: 

Ser or Ser: 

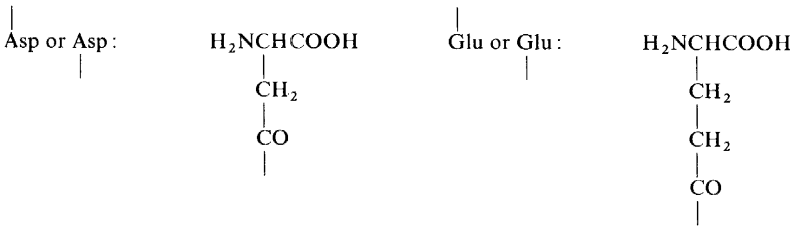
Tyr or Tyr: 

Cys or Cys:  ('half-cystine')

(Cystine would be: Cys–Cys, Cys Cys, or Cys Cys, *not* Cys–Cys).

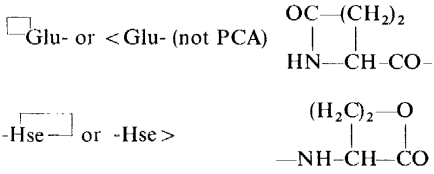
† See footnote † on page 325.

3.4 Lack of Hydroxyl on Carboxyl Groups in the Side Chain



3.5 Cyclic Derivatives of Amino Acid Residues

For the special cases of the residues derived from pyrrolid-2-one-5-carboxylic acid (also known as pyroglutamic acid) and from homoserine lactone, the following are recommended:

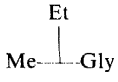


4. SUBSTITUTED AMINO ACIDS

4.1 Substitution in the 2-Amino and 1-Carboxyl Groups

This follows logically from 3.1 and 3.2. The following examples will make the usage clear. (See also 6.2.)

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
N-Ethyl-*N*-methylglycine	Et-MeGly, $\begin{matrix}\text{Et}\\\text{Me}\end{matrix}$ > Gly.

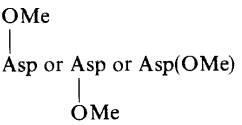
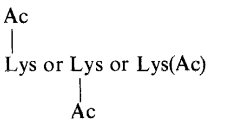
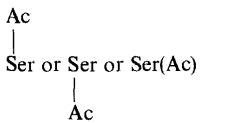
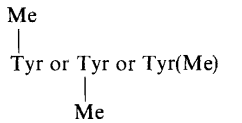
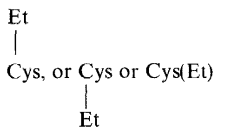
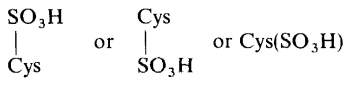
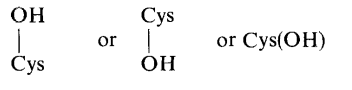
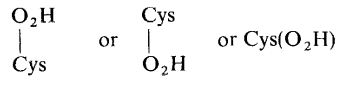
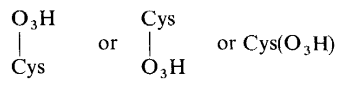
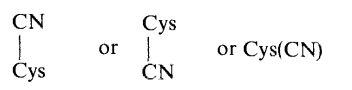


4.2 Substitution in the Side Chain

Side-chain substituents may be portrayed above or below the amino-acid symbol (see 3.3 and 3.4), or by placing the symbol for the substituent in parentheses immediately after the amino-acid symbol.

The use of parentheses should be reserved for a *single* symbol denoting a side-chain substituent. Where a more complex substituent is involved, it is

recommended that the vertical stroke and the two-line abbreviation be used[5]. In general, the one-line abbreviation should be used only when the structure of a substituted peptide is given in textual material.

O^4-Methyl hydrogen aspartate	
O^5-Ethyl hydrogen *N*-acetylglutamate	Ac-Glu(OEt)
N^6-Acetyllysine	
O^3-Acetylserine	
O^4-Methyltyrosine	
S-Ethylcysteine	
S-Sulfocysteine (*S*-Cysteinesulfonic acid)	
Cysteinesulfenic acid	
Cysteinesulfinic acid	
Cysteic acid (3-Sulfoalanine)	
S-Cyanocysteine	

Methionine sulfoxide	$\begin{matrix} & & O & \\ & & \vert & \\ \text{Met or Met or Met(O)} & & & \\ & & \vert & \\ & & O & \end{matrix}$
Methionine sulfone	$\begin{matrix} & & O_2 & \\ & & \vert & \\ \text{Met or Met or Met}(O_2) & & & \\ & & \vert & \\ & & O_2 & \end{matrix}$
O^3-Phosphonoserine (Phosphoserine)	$\begin{matrix} & & P & \\ & & \vert & \\ \text{Ser or Ser or Ser(P)} & & & \\ & & \vert & \\ & & P & \end{matrix}$
τ-Methylhistidine† (see 3.3) (*tele*methylhistidine)	$\begin{matrix} & & Me & \\ & & \tau\vert & \\ \text{His or His or His}(\tau\text{Me}) & & & \\ & & \tau\vert & \\ & & Me & \end{matrix}$

similarly for π substitution (*pros*methylhistine)†.

4.3 Substitution on Carbon Side Chain

This may use the same convention as in 4.2, with the addition of locant numerals where necessary, e.g.

3-Nitrotyrosine	$\begin{matrix} & & NO_2 & \\ & & 3\vert & \\ \text{Tyr or Tyr or Tyr}(3NO_2) & & & \\ & & 3\vert & \\ & & NO_2 & \end{matrix}$
2,3-Diaminopropionic acid (see 2.2.4) (3-aminoalanine)	$\begin{matrix} & & NH_2 & \\ & & 3\vert & \\ \text{Ala or Ala or Ala}(3NH_2) & & & \\ & & 3\vert & \\ & & NH_2 & \end{matrix}$
Diiodotyrosine	$\text{Tyr}(I_2)$

5. SYMBOLS FOR SUBSTITUENTS

Groups substituted for hydrogen or for hydroxyl may be indicated either by their structural formulae or by symbols or by combinations of both, e.g.

† The prolonged and well-entrenched ambiguity in the nomenclature of the *N*-methyl histidines (the chemist's *N*-1 being the biochemist's *N*-3 and *vice versa*) leads to the proposal that a new trivial system for designating these substances is necessary. It is therefore proposed that the imidazole N *near*er the alanine residue be designated *pros* (symbol π) and the one *far*ther *tele* (symbol τ), to give the following names and symbols:

*pros*methylhistidine or π-methylhistidine, His(πMe);

*tele*methylhistidine or τ-methylhistidine, His(τMe).

Benzoylglycine (hippuric acid)	Ph–CO-Gly or C_6H_5CO-Gly or †Bz-Gly or PhCO-Gly
Glycine methyl ester	Gly-OCH_3 or Gly-OMe
Trifluoroacetylglycine	CF_3CO-Gly

Suggestions for symbols designating substituent (or protecting) groups common in polypeptide and protein chemistry follow.

5.1 *N*-Substituents (Protecting Groups) of the Urethane Type

Benzyloxycarbonyl-	Z- or Cbz-
p-Nitrobenzyloxycarbonyl-	Z(NO_2)-
p-Bromobenzyloxycarbonyl-	Z(Br)-
p-Methoxybenzyloxycarbonyl-	Z(OMe)-
p-Methoxyphenylazobenzyloxycarbonyl-	Mz-
p-Phenylazobenzyloxycarbonyl-	Pz-
t-Butoxycarbonyl-	Boc- or Bu^tOCO-
Cyclopentyloxycarbonyl-	Poc- or *c*PeOCO-

5.2 Other *N*-Substituents

Acetyl-	Ac-
Benzoyl- (C_6H_5CO-)	PhCO- or Bz-
Benzyl- ($C_6H_5CH_2$-)	$PhCH_2$- or †Bzl
Benzylthiomethyl-	$PhSCH_2$- or Btm-
Carbamoyl-	NH_2CO- (preferred to Cbm)
1-Carboxy-2-nitrophenyl-5-thio-	‡Nbs-
3-Carboxypropionyl- (HOOC—CH_2—CH_2—CO—)§	Suc-
Dansyl- (5-dimethylaminonaphthalene-1-sulfonyl)	Dns-
Dinitrophenyl-	‖N_2ph- or Dnp
Formyl-	HCO- or CHO-
p-Iodophenylsulfonyl (pipsyl)	Ips
Maleoyl- (—OC—CH=CH—CO—)	—Mal- or Mal<
Maleyl- (HOOC—CH=CH—CO—)	Mal-
Methylthiocarbamoyl-¶	MeNHCS- or ¶Mtc-

† Bz- is the symbol generally used for *benzoyl* in organic chemistry. It should not be used for *benzyl* ($C_6H_5CH_2$– or $PhCH_2$–), for which the symbol is Bzl-. However, $PhCH_2$– is unambiguous.

‡ See Comment following 5.3.

§ Not succinyl, although it is the monovalent radical of succinic acid. See succinyl and Footnote †† on page 327.

‖ The use of D for 'di' and T for 'tri' or 'tetra' (and DH and TH for 'dihydro' and 'tetrahydro', respectively) is discouraged. Recognized symbols and subscripts are recommended. See also Footnote † on page 319.

¶ The symbol Pth has been used to denote a phenylthiohydantoin (e.g. Pth-Leu). Since this incorrectly implies the substitution of an amino acid by a 'phenylthiohydantoyl' group, it is suggested that the abbreviated symbol for such compounds be of the type CS-Leu-NPh or PhNCS-Leu (or Leu > PhNCS in textual material).

o-Nitrophenylthio-	Nps-
Phenylthiocarbamoyl-¶	PhNHCS- or ¶ Ptc-
Phthaloyl-	-Pht- or Pht<
Phthalyl-	Pht-
Succinyl-††(—OC—CH_2—CH_2—CO—)	-Suc- or Suc<
Tetrahydropyranyl-	H_4pyran- (preferred to Thp‖)
Tosyl- (*p*-tolylsulfonyl)	Tos-
Trifluoroacetyl-	‖CF_3CO- or F_3Ac-
Trityl- (triphenylmethyl)	Ph_3C- or Trt-

5.3 Substituents at Carboxyl Group

Benzyloxy- (benzyl ester)	—OCH_2Ph or —OBzl
Cyanomethoxy-	—OCH_2CN
Diphenylmethoxy- (benzhydryl ester)	—$OCHPh_2$ or —OBzh
Ethoxy- (ethyl ester)	—OEt
Methoxy- (methyl ester)	—OMe
p-Nitrophenoxy- (*p*-nitrophenyl ester)	—ONp
p-Nitrophenylthio-	—SNp
Phenylthio- (phenylthiolester)	—SPh
1-Piperidino-oxy-	—OPip
8-Quinolyloxy-	—OQu
Succinimido-oxy-	—ONSu
Tertiary butoxy- (*t*-butyl ester)	—OBu^t

Comment

Many reagents used in peptide and protein chemistry for the modification (protection) of amino, carboxyl and side-chain groups in amino-acid residues have been designated by a variety of acronymic abbreviations, too numerous to be listed here. Extensive and indiscriminate use of such abbreviations is discouraged, especially where the accepted trivial name of a reagent is short enough, e.g. tosyl chloride, bromosuccinimide, trityl chloride, dansyl chloride, etc., or may be formulated in terms of the group transferred, e.g. ‖N_2ph-F instead of FDNB for 1-fluoro-2,4-dinitrobenzene, Dns-Cl or dansyl-Cl in place of DNS. ‖Nbs_2 in place of DTNB for 5,5′-dithio-bis(2-nitrobenzoic acid) (Ellman's reagent), $(Pr^iO)_2PO$-F, Pr^i_2P-F, iPr_2P-F, or Dip-F‖ instead of DFP for diisopropylfluorophosphate. Other commonly-used substances that may be expressed more clearly in terms of symbols are MalNEt (instead of NEM) for *N*-ethyl-maleimide, Tos-PheCH_2Cl (instead of TPCK) for L-1-tosyl-amido-2-phenylethyl chloromethyl ketone, Tos-Arg-OMe (instead of TAME) for tosyl-L-arginine methyl ester, Me_3Si- (instead of TMS-) for trimethylsilyl, CF_3CO- (instead of TFA) for trifluoroacetyl (see 5.2), H_4furan (instead of THF), etc. See also Footnotes † on 319 and ‖ opposite).

Some additional symbolic terms for substituents (and reagents), as examples, are:

2-Aminoethyl-	-$(CH_2)_2NH_2$ (preferred to Aet)

†† Not succinoyl. For ¶ and ‖, see opposite.

Carbamoylmethyl-	$-CH_2CONH_2$ (preferred to Cam)
Carboxymethyl-	$-CH_2CO_2H$ (preferred to Cm)
Chloroethylamine	$Cl(CH_2)_2NH_2$
Ethyleneimine	$(CH_2)_2NH$
Chloroacetamide	$ClCH_2CONH_2$
Chloroacetic acid	$ClCH_2CO_2H$
p-Carboxyphenylmercuri-	—HgBzOH
p-Chloromercuribenzoate	*p*Cl-HgBzO$^-$
Diazoacetyl-	N_2CHCO—
Hydroxyethyl-	—$(CH_2)_2OH$
Ethylene oxide	$(CH_2)_2O$

6. POLYPEPTIDES

6.1 Polypeptide Chains[5]

Polypeptides may be dealt with in the same manner as substituted amino acids, e.g.

Glycylglycine — Gly-Gly

N-α-Glutamylglycine — Glu-Gly

N-γ-Glutamylglycine — Glu(Gly) or Glu(Gly) or Glu(Gly)

Glutathione — Glu(Cys-Gly) or Glu(Cys-Gly) or Glu(Cys-Gly) or Glu(Cys-Gly)

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

N^2-α-Glutamyllysine — Glu-Lys

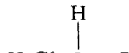
N^6-α-Glutamyllysine — Glu(Lys) or Glu(Lys)

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

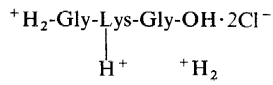
N^6-γ-Glutamyllysine — Glu(Lys) or Glu(Lys) or Glu(Lys)

The presence of free, substituted, or ionized functional groups can be represented (or stressed) as follows:

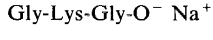
Glycyllysylglycine



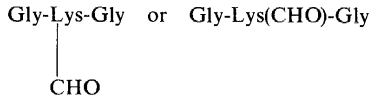
Its dihydrochloride



Its sodium salt



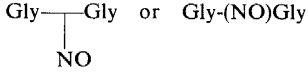
Its N^6-formyl derivative



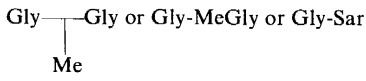
etc.

6.2 Peptides Substituted at N^2 (see 4.1).

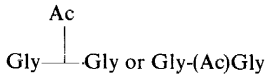
Glycylnitrosoglycine



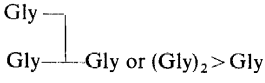
Glycylsarcosine (see 2.2.5)



N-Glycyl-*N*-acetylglycine



N,*N*-bisglycylglycine



etc.

6.3 Cyclic Polypeptides

6.3.1 *Homodetic cyclic polypeptides*

In homodetic cyclic polypeptides the ring consists of amino-acid residues in peptide linkage only. Three representations are possible:

6.3.1.1 The sequence is formulated in the usual manner but placed in parentheses and preceded by (an italic) *cyclo*.

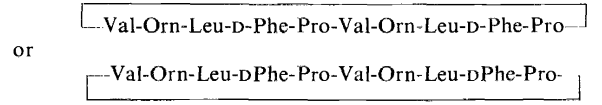
Example: Gramicidin S =

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

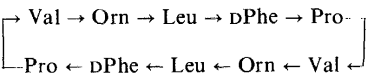
or (see 1.5, sentence 2).

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

6.3.1.2. The terminal residues may be written on one line, as in 6.3.1.1, but joined by a lengthened bond. Using the same example in the two forms (see 1.5):



6.3.1.3. The residues are written on more than one line, in which case the CO → NH direction must be indicated by arrows, thus (in the optional manner of 1.5):

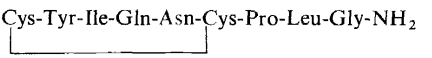


6.3.2. *Heterodetic cyclic polypeptides*

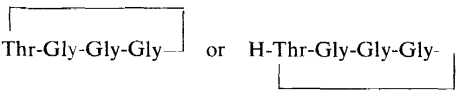
In heterodetic cyclic polypeptides the ring consists of other residues in addition to amino-acid residues in peptide linkage. These follow logically from the formulation of substituted amino acids.

Example:

Oxytocin



Cyclic ester of threonylglycylglycylglycine



REFERENCES

[1] Abbreviated Designation of Amino-Acid Derivatives and Peptides. (1966) *J. Biol. Chem.*, **241**, 2491; (1966) *Biochemistry*, **5**, 2485; (1966) *Biochim. Biophys. Acta*, **121**, 1; (1967) *Biochem. J.*, **102**, 23; (1967) *Arch. Biochem. Biophys.*, **121**, 1; (1967) *Europ. J. Biochem.*, **1**, 375; (1967) *Z. Physiol. Chem.*, **348**, 256; (1967) *Bull. Soc. Chim. Biol.*, **49**, 121; (1968) *Molek. Biol.*, **2**, 282; (1971) 5, 492.

[2] Abbreviations and Symbols for Chemical Names of Special Interest in Biological Chemistry. (Section 5 revised by 3 below). (1966) *J. Biol. Chem.*, **241**, 527; (1966) *Biochemistry*, **5**, 1445; (1966) *Biochem. J.*, **101**, 1; (1966) *Virology*, **29**, 480; (1966) *Arch. Biochem. Biophys.*, **115**, 1; (1967) *Europ. J. Biochem.*, **1**, 259; (1967) *Z. Physiol. Chem.*, **348**, 245; (1968) *Bull. Soc. Chim. Biol.*, **50**, 3; (1967) *Molek. Biol.*, **1**, 872.

[3] Abbreviations and Symbols for Nucleic Acids, Polynucleotides and Their Constituents. (1970) *Pure Appl. Chem.* **40**, 277.

[4] Nomenclature of Vitamins, Coenzymes and Related Compounds: Trivial Names of Miscellaneous Compounds of Importance in Biochemistry, Nomenclature of Quinones with Isoprenoid Side Chains, Nomenclature and Symbols for Folic Acid and Related Compounds, Nomenclature of Corrinoids. (1967) *Arch. Biochem. Biophys.*, **118**, 505; (1967) *Biochem. J.*, **102**, 15; (1965, 1966) *Biochim. Biophys. Acta*, **107**, 1; **117**, 285; (1967) *Europ. J. Biochem.*, **2**, 1;

(1966) *J. Biol. Chem.*, **241**, 2987; (1967) *Bull. Soc. Chim. Biol.*, **49**, 331; (1967) *Z. Physiol. Chem.*, **348**, 266.

[5] Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino Acids). (1973) *Pure Appl. Chem.*, **33**, 437.

[6] IUPAC: Nomenclature of Organic Chemistry (Definitive Rules for Sections A, B & C), Rule C-404 and Table VI. (1971) Butterworths: London.