IUPAC-IUB Commission on Biochemical Nomenclature

Abbreviated Designation of Amino Acid Derivatives and Peptides

Tentative Rules

These Tentative Rules are an attempt to achieve a broad systematization of various types of abbreviated notation already in use (e.g., Brand and Edsall, *Ann. Rev. Biochem.*, 16, 224 (1947); *Report of the Committee on Abbreviations of the American Society of Biological Chemists*, December 18, 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,"* IUPAC Information Bulletin No. 20, July 1963, pages 16–18; 1965 revision of the latter9). They seek 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 of conveying more detailed chemical information in abbreviated form.4

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 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 formulas.

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 matter, 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\text{\textsubscript{6}} and S\text{\textsubscript{6}} of serine and cysteine, respectively, N\text{\textsubscript{6}} of lysine, N\text{\textsubscript{6}} 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 in a clearer manner, the hyphen between the configurational prefix and the symbol may be omitted (see 5.3.1.1 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 formulas of complicated features may be used along with the abbreviated notation wherever necessary for clarity.

2. ABBREVIATIONS FOR AMINO ACIDS

2.1 Common Amino Acids.

| Alanine | Ala | Lysine | Lys |
| Arginine | Arg | Methionine | Met |
| Asparagine | Asn | Ornithine | Orn |
| Aspartic acid | Asp | Phenylalanine | Phe |
| Cysteine | Cys | Proline | Pre |
| Glutamic acid | Glu | Serine | Ser |
| Glutamine | Gin | Threonine | Thr |
| Glycine | Gly | Tryptophan | Trp |
| Histidine | His | Tyrosine | Tyr |
| Isoleucine | Ile | Valine | Val |
| Leucine | Leu |

2.2 Less Common Amino Acids

2.2.1 Hydroxyamino Acids.

| Hydroxylysine | Hyl |
| 3-Hydroxyproline | 3Hyp |
| 4-Hydroxyproline | 4Hyp |

\[ \text{Asparagine and glutamine may also be denoted as Asp(NH}_2\text{)} \]

3. NH\text{\textsubscript{2}}

4. NH\text{\textsubscript{2}}
2.2.2 allo-Amino Acids.

- allo-Isoleucine  \( \text{aIle} \)
- allo-Hydroxylysine  \( \text{aHyl} \)

2.2.3 "Nor" 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.

- Norvaline  \( \text{Nva} \)
- Norleucine  \( \text{Nle} \)

2.2.4 Higher Unbranched Amino Acids—We suggest the following general rules for guidance in forming abbreviations: the functional prefix "amino" should be included in the symbol as the letter "A," diamino as "D."

The trivial name of the parent acid should be abbreviated to leave no more than two or three letters, as convenient and necessary for clarity. The word "acid" ("-sdiure," etc.) should be omitted from the symbol as carrying no significant information. Unless otherwise indicated (see paragraph below), single groups are in the \( \alpha \) position, two amino groups in the \( \alpha , \omega \) (monocarboxylic acids) or \( \alpha , \alpha ' \) positions (dicarboxylic acids). The location of amino acids in positions other than \( \alpha \) and \( \omega \) is shown by the appropriate Greek letter prefix.

Examples:
- \( \alpha \)-Aminobutyric acid  \( \text{Abu} \)
- \( \alpha \)-Aminoadipic acid  \( \text{Aad} \)
- \( \alpha \)-Aminopimelic acid  \( \text{Apm} \)
- \( \alpha , \gamma \)-Diaminobutyric acid  \( \text{Dbu} \)
- \( \alpha , \beta \)-Diaminopropionic acid  \( \text{Dpr} \)
- \( \alpha , \alpha ' \)-Diaminopimelic acid  \( \text{Dpm} \)
- \( \beta \)-Alanine  \( \beta \text{Ala} \)
- \( \epsilon \)-Aminocaproic acid  \( \epsilon \text{Acp} \)
- \( \beta \)-Aminoacrylate acid  \( \beta \text{Aad} \)

2.2.5 \( N^\omega \)-Alkylated Amino Acids—\( N^\omega \)-Methylamino acids are becoming more and more common (e.g. in the large group of depsipeptides). This justifies special symbols.

Examples:
- \( N \)-Methylglycine (sarcosine)  \( \text{MeGly or Sar} \)
- \( N \)-Methylisoleucine  \( \text{MeIle} \)
- \( N \)-Methylvaline, etc.  \( \text{MeVal, etc.} \)
- \( N \)-Ethylglycine, etc.  \( \text{EtGly, etc.} \)

3. Amino Acid Residues

3.1 Lack of Hydrogen on the \( \alpha \)-Amino Group—The \( \alpha \)-amino group is understood to be at the left-hand side of the symbol when using hyphens, and—in special cases—at the point of the arrow when using arrows to indicate the direction of the peptide bond (\( \text{--CO} \rightarrow \text{NH}--, \text{--NH} \leftarrow \text{CO--} \)).

Examples:
- \( \alpha \)-Gly:  \( \text{--HNCH}_{2}\text{COOH} \)
- \( \text{-Ala: --HNCH}_{2}\text{COOH} \)

3.2 Lack of Hydroxyl on the \( \alpha \)-Carboxyl Group—The \( \alpha \)-carboxyl group is always understood to be on the right-hand side of the symbol when using hyphens, and—in special cases—at the tail of the arrow when using arrows to indicate the direction of the peptide bond (\( \text{--CO} \rightarrow \text{NH}--, \text{--NH} \leftarrow \text{CO--} \)).

Example: \( \text{Gly: --HNCH}_{2}\text{COOH} \)

3.3 Lack of Hydrogen on \( \alpha \)-Amino, Imino, Guanidino, Hydroxyl, and Thiol Functions in the Side Chain.
3.4 Lack of Hydroxyl on Carboxyl Groups in the Side Chain.

Asp or Asp: \( \text{H}_2\text{NCHCOOH} \)

Glu or Glu: \( \text{H}_2\text{NCHCOOH} \)

4. SUBSTITUTED AMINO ACIDS

4.1 Substitution in the \( \alpha \)-Amino and \( \alpha \)-Carboxyl Groups—This follows logically from 3.1 and 3.2. The following examples will make the usage clear.

- N-Acetylglycine: Ac-Gly
- Glycine ethyl ester: Gly-OEt
- N\textsuperscript{*}-Acetylylsine: Ac-Lys
- Serine methyl ester: Ser-OMe
- \( \alpha \)-Ethyl-N-acetylglutamate: Ac-Glu-OEt
- Isoglutamine: Glu-NH\textsubscript{2}
- N-Ethyl-N-methylglycine: EtMeGly, GUY, \( \text{Me} > \text{Me} \)
- Et-GUY

4.2 Substitution in the Side Chain—Side chain substituents may be portrayed above or below the amino acid symbol, 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 a two-line abbreviation be used.

Aspartic acid \( \beta \)-methyl ester: Asp or Asp or Asp(OMe)

Acetyllysine: Ac-Lys

Acetylglycine: Ac-Gly

O-Methyltyrosine: Me

Tyr or Tyr or Tyr(Me)

Et

S-Ethyleysteine: Cys or Cys or Cys(Et)

Et

5. POLYPEPTIDES

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

Glycylglycine: Gly-Gly

Glu-Gly: Glu-Gly

\( \gamma \)-Glutamylglycine: Glu or Glu, Cys-Gly

\( \gamma \)-Glutamyllysine: Glu or Glu, Lys or Lys or Gly

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

N\textsuperscript{*}-\( \alpha \)-Glutamyllysine: Glu Lys or Glu Lys or Glu Lys

N\textsuperscript{*}-\( \gamma \)-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: H-Gly-Lys-Gly-OH

\( \text{H}_2\)-Gly-Lys-Gly-OH: 2Cl\textsuperscript{-}

\( \text{H}_2\)-Gly-Lys-Gly-OH

Its sodium salt: Gly-Lys-Gly-O\textsuperscript{-} Na\textsuperscript{+}

Its \( N\textsuperscript{*}\)-formyl derivative: Gly-Lys-Gly or Gly-Lys(CH\textsubscript{2})-Gly-CHO

Etc.
5.2 Peptides Substituted at N≥

Examples:

Glycynitrosoglycine
Gly — Gly  NO

Glycylsarcosine
Gly — Gly or Gly-MeGly or Gly-Sar  Me

Nα-Glycyl-Nα-
acetylglycine
Gly — Gly  Ac

Nα-Glycyl-Nα-
glycylglycine
Gly — Gly

Etc.

5.3 Cyclic Polypeptides.

5.3.1 Homodetic cyclic polypeptides (the ring consists of amino acid residues in peptide linkage only). Three representations are possible:

5.3.1.1 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-nPhe-Pro-Val-Orn-Leu-nPhe-Pro-)
or (see 1.5, sentence 2)
cyclo-(Val-Orn-Leu-nPhe-Pro-Val-Orn-Leu-nPhe-Pro-)

5.3.1.2 The terminal residues may be written on one line, as in 5.3.1.1, but joined by a lengthened bond. Using the same example in the two forms (see 1.5):

Val-Orn-Leu-nPhe-Pro-Val-Orn-Leu-nPhe-Pro—
or

Val-Orn-Leu-nPhe-Pro-Val-Orn-Leu-nPhe-Pro—

5.3.1.3 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):

Val — Orn — Leu — nPhe — Pro —

5.3.2 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  Cys-Tyr-Ile-Asn-Gln-Cys-Pro-Leu-Gly-NH₂

Cyclic ester of threonylglycylglycylglycine

Thr—Gly—Gly— or H—Thr—Gly—Gly—

6. ABBREVIATIONS FOR SUBSTITUENTS

Groups substituted for hydrogen or for hydroxyl may be indicated either by their structural formulas or by (accepted) abbreviations, e.g.

Denzoylglycine  \(\text{CO—Gly or C}_{11}\text{H}_{5}	ext{CO—Gly or Bz—Gly}\)

Glycine methyl ester  \(\text{Gly—OCH}_{3}\) or Gly—OMe

Suggestions for the abbreviations of protecting groups common in polypeptide chemistry follow. All such symbols (except those allowed by individual journals, e.g. Bz, Ac, Ph, Me, Et, etc.) should be defined in each paper. Although symbolization by the use of capital letters throughout would be useful for distinguishing these symbols from those of the amino acids, we propose the use of one capital letter followed by lower-case letters in order not to increase the flood of capital-letter abbreviations in biological chemistry.

6.1 N-Protecting Groups of the Urethane Type.

Benzoylcarbonyl  \(Z–\)

p-Nitrobenzoylcarbonyl  \(Z(\text{NO}_2)–\)

p-Bromobenzoylcarbonyl  \(Z(\text{Br})–\)

p-Methoxybenzoylcarbonyl  \(Z(\text{OMe})–\)

p-Methoxyphenylazobenzoylcarbonyl  \(\text{Mz}–\)

p-Phenylazobenzoylcarbonyl  \(\text{Pz}–\)

p-Butylxoycarbonyl  \(\text{Boc}–\)

Cyclohexylxoycarbonyl  \(\text{Poe}–\)

\(\text{Bz}–\) is the symbol generally used for benzoyl in organic chemistry. Its use for benzy1 (which has become rather common in polypeptide chemistry) should be discouraged. We propose \(\text{Bzl}–\) for benzy1.
### 6.2 Other N-Protecting Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetyl</td>
<td>Ac-</td>
</tr>
<tr>
<td>Benzoyl</td>
<td>Bz-</td>
</tr>
<tr>
<td>Tosyl</td>
<td>Toe-</td>
</tr>
<tr>
<td>Trifluoroacetyl</td>
<td>Tfa-</td>
</tr>
<tr>
<td>Phthalyl</td>
<td>Ph-</td>
</tr>
<tr>
<td>Benzyl</td>
<td>Bzl-</td>
</tr>
<tr>
<td>Trityl</td>
<td>Trt-</td>
</tr>
<tr>
<td>Tetrahydropyranyl</td>
<td>Thp-</td>
</tr>
<tr>
<td>Dinitrophenyl</td>
<td>Dnp-</td>
</tr>
<tr>
<td>Benzylthiomethyl</td>
<td>Btm-</td>
</tr>
<tr>
<td>o-Nitrophenylsulfenyl</td>
<td>Nps-</td>
</tr>
</tbody>
</table>

### 6.3 Carboxyl-protecting Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methoxy (methyl ester)</td>
<td>-OMe</td>
</tr>
<tr>
<td>Ethoxy (ethyl ester)</td>
<td>-OEt</td>
</tr>
<tr>
<td>Tertiary butoxy (tert-butyl ester)</td>
<td>-OBu'</td>
</tr>
<tr>
<td>Benzylxoy (benzyl ester)</td>
<td>-OBzl</td>
</tr>
<tr>
<td>Diphenylmethoxy (benzhydryl ester)</td>
<td>-OBzl</td>
</tr>
<tr>
<td>p-Nitrophenox (p-nitropheny l ester)</td>
<td>-ONp</td>
</tr>
<tr>
<td>Phenylthio (phenylthiolester)</td>
<td>-SPh</td>
</tr>
<tr>
<td>p-Nitrophenylthio</td>
<td>-SNp</td>
</tr>
<tr>
<td>Cyanomethoxy</td>
<td>-OCH$_2$CN</td>
</tr>
</tbody>
</table>

Note: Contrary to the symbols for amino acid residues, the position of the dashes in the symbols for substituents carries no significant information.