

IUPAC-IUB Commission on Biochemical Nomenclature

Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino Acids)*

Revised Recommendations (1971)†

The numerous studies on the physical, chemical, and biological properties of synthetic polypeptides have brought with them different ways of describing, in abbreviated form, these products whose structures are often incompletely known. The use of a variety of nomenclatures complicates the literature; hence, a consistent and clearly defined system for naming such polypeptides is desirable. The proposals set forth here, which represent the consensus of many discussions and suggestions, should aid in systematizing the nomenclature of a wide variety of synthetic polypeptides.

These proposals are based in large part on the abbreviated nomenclature devised by Gill (2) and by Sela (3) and others. They utilize the symbols and conventions set forth in Section 2 of "Revised Tentative Rules for Abbreviations and Symbols of Chemical Names of Special Interest in Biological Chemistry" (4) and in "Abbreviated Designation of Amino Acid Derivatives and Peptides" (5), and they add only those terms or conventions needed for the specification of polymers but not encompassed by these schemes.

The symbols and conventions of the previous "Tentative Rules" (4, 5) used in this nomenclature system are summarized as follows. The symbols of the amino acid residues and their derivatives or modifications are those indicated in the "Tentative Rules" (4, 5) or formulated according to the principles set out in them. Hyphens or commas between the symbols for residues or groups of residues indicate known or unknown sequence, respectively, and involve only the α -NH₂ and α -COOH groups (the peptide link). Commas may be omitted when

other symbols (*e.g.* subscripts or superscripts) separate symbols in unknown sequences. Vertical strokes indicate covalent bonds involving functional groups or the remaining H-atom of the peptide bond, depending upon their placement (5). α -Amino acids and α -peptide links, read from left (NH₂ terminus) to right (COOH terminus), are assumed unless indicated otherwise (4, 5).

Definitions

- △ 1. **Linear Polymer:** All amino acid residues (constitutional units) are linked in an unbranched chain.
- 2. **Block:** A polymer that forms a distinct part of a larger polymer (*e.g.* a block or graft polymer may contain several blocks).
- 3. **Graft Polymer:** One or more blocks are linked to the functional groups of a linear polymer, thus creating a branch or branches. (Functional groups include ϵ -NH₂, β - or γ -COOH, etc., and the remaining H-atom of an α -peptide link.)
- 4. **Block Polymer:** Two or more species of block are linked to form a larger linear polymer.

Recommendations

1. **Designation of Blocks or Linear Polymers.** The prefix "poly" or the subscript n indicates "polymer of". It is attached to each main chain and is repeated in each block within a larger polymer unless there is sufficient indication of size and of structure to make this repetition unnecessary. For example, poly(Glu) or (Glu) _{n} represent poly(glutamic acid), and (Glu)₁₀, a decapeptide of glutamic acid. "Oligo" may replace "poly" for short chains.

Comments

- a) n replaces the p as originally, but no longer, used in the polymer nomenclature scheme devised by the IUPAC Subcommittee on the Nomenclature of Macromolecules (6). It is used in designating polynucleotides (7), and it is chosen in place of p in order to avoid confusion with the "p" used for a terminal phosphoric residue in the latter scheme. The n may be replaced by a definite number (*e.g.* 10 above), an average (*e.g.* $\overline{10}$), or a range (*e.g.* 8-12), as appropriate. However, two n 's should not appear in the same formula unless equal length is implied. When equal length is not the case, different letters
 - ▲ should be used, such as m, j, k.
 - b) If "poly" is used rather than the subscript n , the symbol(s) following "poly" should be enclosed in parentheses with no intervening space, *e.g.* poly(Lys). If "poly" is followed by a single, simple word, the whole is written as one word, *e.g.* polylysine. If what follows "poly" is complex, it should be enclosed in parentheses (again without following space), *e.g.* poly(amino acid), not polyamino acid or polyaminoacid; poly(glutamic acid) or polyglutamate, but not polyglutamic

* Document of the IUPAC-IUB Commission on Biochemical Nomenclature (CBN), originally approved and published by IUPAC and IUB in October 1967 (1), and again, in the present revised form, in 1971, approved and published by permission of the International Union of Pure and Applied Chemistry and the International Union of Biochemistry.

Comments on these Recommendations may be sent to any member of CBN: O. Hoffmann-Ostenhof (Chairman), W. E. Cohn (Secretary), A. E. Braunstein, P. Karlson, B. Keil, W. Klyne, C. Liébecq, E. C. Slater, E. C. Webb, W. J. Whelan.

Reprints of these Recommendations may be obtained from Waldo E. Cohn, Director, NAS-NRC Office of Biochemical Nomenclature, Biology Division, Oak Ridge National Laboratory, P. O. Box Y, Oak Ridge, Tenn., 37830, U. S. A.

† This revision differs from the original (1) essentially in the addition of comments after Recommendation 1 (marked with ▲ in the margin) and in the relevant "Examples." These revisions were made to conform with the practices in polymer chemistry and were recommended to CBN by the IUPAC Commission on Macromolecular Nomenclature (IUPAC Information Bulletin, Appendices on Nomenclature, etc., No. 13, February 1971) and the Nomenclature Committee of the Division of Polymer Chemistry of the American Chemical Society (see *Macromolecules*, **1**, 193, 1968). Minor revisions are indicated by △.

acid; poly(DL-alanine,L-lysine) for the substance shown in Example 2; and poly(DL-alanine-L-lysine) for the substance shown in Example 3. The format poly(L-lysine) is preferred to poly-L-lysine, *i.e.* L-lysine is regarded as a complex term. Similarly, poly(hydroxyproline), not polyhydroxyproline.

2. Designation of Branches and Branch Points. Branches (side chains) connected to the main chain can be designated in one of three ways: by a vertical line joining the main chain and the branch (side chain); by an extended bond joining the appropriate residues with the main chain written first; or by a horizontal double dash (not preferred).

The branch points are indicated by the origin and terminus of the vertical line. If the origin is unknown, the line originates at the "p" in "poly," if "poly" is used, or at the first parenthesis (bracket), if the subscript n is used (see Recommendation 1). If the origin is known, the line originates: (a) vertically at the initial letter of the appropriate symbol, if functional groups other than α -NH₂ or α -COOH residues are involved; (b) vertically at the position of the appropriate link, if substitution for the remaining H-atom of a peptide link is involved; or (c) horizontally to the left or right of the symbol, respectively, if α -NH₂ or α -COOH groups are involved. The same rules apply to the termination of the line. Thus, the linkage between a side chain functional group and an α -NH₂ or α -COOH group in the main chain is indicated by two perpendicular lines with the vertical line originating in the functional group and the horizontal line in the α -NH₂ or α -COOH group. A number in parentheses lying beside the line indicates the number of such links per 100 residues of polymer, if known.

Comment

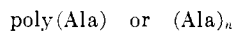
A limitation of the double dash as a connecting link lies in its inability to originate or to terminate definitively in a specific residue. Either the arrangement of the symbols must be such that connected ones are adjacent, or the information must be given independently.

3. Block size. A superscript outside the parentheses enclosing a block indicates the number of repeating sequences per 100 residues of polymer, and it is given to the first decimal place.

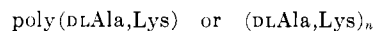
4. The Molar Percentage of a single type of amino acid residue within a copolymer, derived from the amino acid analysis and assuming copolymerization, is indicated by a superscript attached to the symbol of the residue. The molar percentages are given in whole numbers and should total 99 to 101%.

Examples

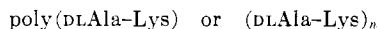
1. Simple homopolymer:



2. Linear copolymer, unknown sequence, composition not specified:



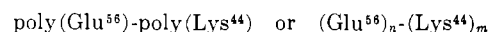
3. Linear copolymer, regular alternating sequence.



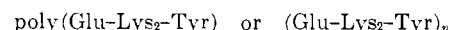
4. Linear sequence of unknown order [Composition: 56% Glu, 38% Lys, and 6% Tyr ($\Sigma = 100\%$)]:

- poly(Glu⁵⁶Lys³⁸Tyr⁶) or (Glu⁵⁶Lys³⁸Tyr⁶)_n (all L)
- poly(DGlu⁵⁶DLys³⁸Tyr⁶) (only Tyr is L)
- poly(DLGl⁵⁶Lys³⁸DTyr⁶) (Glu is DL, Tyr is D)

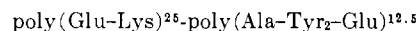
5. Block polymer of poly(Glu) combined through the α -COOH terminus to the α -NH₂ terminus of poly(Lys) [Composition: 56% Glu, 44% Lys ($\Sigma = 100\%$)]:



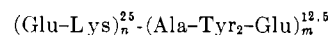
6. (a) Known, repeating sequence (a polymer of Glu-Lys-Tyr):



(b) Known, repeating sequences within each of two constituent blocks of a linear polymer [Composition: 37.5% Glu, 25% Lys, 25% Tyr, 12.5% Ala ($\Sigma = 100\%$)]:

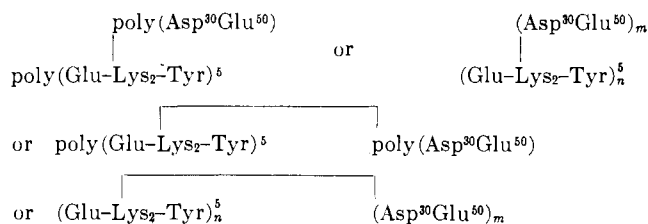


or

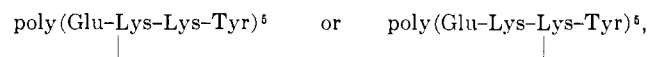


[The connection between the polymeric segments shown here is from the α -COOH of Lys to the α -NH₂ of Ala. Origin or termination in any other residue or functional group can be shown by rearranging the order of residues and by the orientation of the connecting line at its origin and terminus (see Examples 7, 8, and 9).]

(c) Known, repeating sequence in the main chain connected by the ϵ -NH₂ of a lysine (which of the two is not known) to an unknown point in an unknown sequence in the side chain [Composition: 30% Asp, 55% Glu, 10% Lys, 5% Tyr ($\Sigma = 100\%$)]:

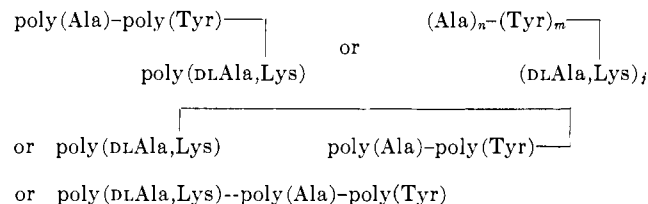


(Note: The double hyphen system is not applicable here.) If it is known which lysine residue is connected to the side chain, the main chain would be written in the form,



as appropriate.

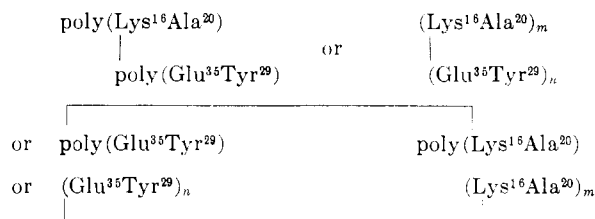
7. Graft polymer with the main chain of DL-alanine and L-lysine connected through the ϵ -NH₂ group of lysine to the α -COOH group of L-tyrosine in the side chain, which consists of a block polymer of L-tyrosine and L-alanine (no analytical data for the main chain):



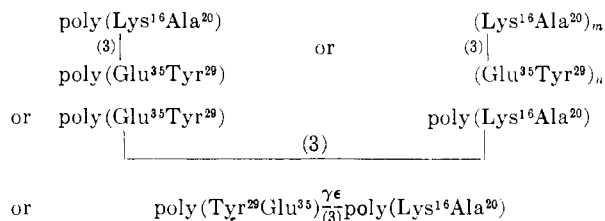
(Note: The points of attachment of Lys and Tyr cannot be specified in the last example.)

8. Graft polymer with an unknown sequence in the main chain and in the side chain [Composition: 16% Lys, 20% Ala, 35% Glu, 29% Tyr ($\Sigma = 100\%$)]:

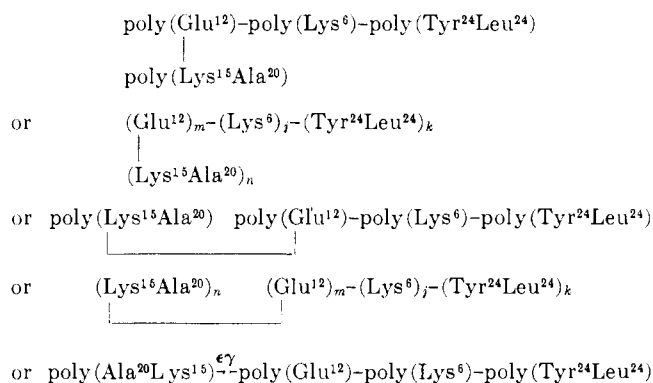
(a) Number and position of the points of attachment in the main chain unknown, but terminating in the lysine residues of the side chain:



(b) Same, but attachments are 3 in number and connect the ϵ -NH₂ groups of the lysine residues in the side chain and the γ -COOH groups of the glutamic acid residues in the main chain:



9. Graft polymer with a block polymer and an unknown sequence in the side chain (upper) attached to an unknown sequence in the main chain (lower); the points of attachment are between the γ -COOH groups of glutamic acid in the side chain and the ϵ -NH₂ groups of lysine in the main chain [Composition: 12% Glu, 21% Lys, 24% Tyr, 24% Leu, 20% Ala (Σ = 101%)].



REFERENCES

1. *Biopolymers*, **8**, 161 (1969); *Arch. Biochem. Biophys.*, **123**, 633 (1968); *Biochem. J.*, **106**, 577 (1968); *Biochemistry*, **7**, 483 (1968); *Biochim. Biophys. Acta*, **168**, 1 (1968); *Eur. J. Biochem.*, **3**, 129 (1968); *J. Biol. Chem.*, **243**, 2451 (1968); *Bull. Soc. Chim. Biol.*, **51**, 205 (1969); *Z. Physiol. Chem.*, **349**, 1013 (1969); *J. Mol. Biol.*, **5**, 492 (1971).
2. GILL, T. J., III, *Biopolymers*, **2**, 283 (1964); also *J. Biol. Chem.*, **240**, 3227 (1965); *Biochim. Biophys. Acta*, **124**, 374 (1966).
3. SELA, M., *Advan. Immunol.* **5**, 30 (1966).
4. *J. Biol. Chem.*, **241**, 527 (1966); *Biochemistry*, **5**, 1445 (1966); *Biochem. J.*, **101**, 1 (1966); *Virology*, **29**, 480 (1966); *Arch. Biochem. Biophys.*, **115**, 1 (1966); *Eur. J. Biochem.*, **1**, 259

(1967); *Z. Physiol. Chem.*, **348**, 245 (1967). See also Reference 7.

5. *J. Biol. Chem.*, **241**, 2491 (1966); *Biochemistry*, **5**, 2485 (1966); *Biochim. Biophys. Acta*, **121**, 1 (1966); *Biochem. J.*, **102**, 23 (1967); *Arch. Biochem. Biophys.*, **121**, 1 (1967); *Eur. J. Biochem.*, **1**, 375 (1967); *Z. Physiol. Chem.*, **348**, 256 (1967); *Bull. Soc. Chim. Biol.*, **49**, 121 (1967). **Revision in Preparation.**
6. *J. Polym. Sci.*, **8**, 257 (1952); revised in 1967 (unpublished; see references in footnote f.)
7. *Biochem. J.*, **120**, 449 (1970); *Biochemistry*, **9**, 4022 (1970); *Eur. J. Biochem.*, **15**, 203 (1970); *J. Biol. Chem.*, **245**, 5171 (1970); *Z. Physiol. Chem.*, **351**, 1055 (1970); *J. Mol. Biol.*, **55**, 299 (1971); and elsewhere.

Appendix

Reprints of all Recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature (CBN) are available from Waldo E. Cohn, Director, NAS-NRC Office of Biochemical Nomenclature, Oak Ridge National Laboratory, P. O. Box Y, Oak Ridge, Tennessee, 37830, U. S. A. These are as follows:

1. Abbreviations and Symbols for Chemical Names of Special Interest in Biological Chemistry (Reference 4 above).
 2. Abbreviated Designation of Amino Acid Derivatives and Peptides (Reference 5 above).
 3. Rules for Naming Synthetic Modifications of Natural Peptides (see *J. Biol. Chem.*, **243**, 2451 (1968) and elsewhere).
 4. Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino Acids) (Reference 1 above).
 5. A One-letter Notation for Amino Acid Sequences (*J. Biol. Chem.*, **243**, 3557 (1968) and elsewhere).
 6. Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains (*J. Biol. Chem.*, **245**, 6489 (1970) and elsewhere).
 7. Abbreviations and Symbols for Nucleic Acids, Polynucleotides and Their Constituents (Reference 7 above).
 8. 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 (*J. Biol. Chem.*, **241**, 2987 (1966) and elsewhere). (See also following paper, No. 9.)
 9. Nomenclature for Vitamins B6 and Related Compounds (*J. Biol. Chem.*, **245**, 4229 (1970) and elsewhere).
 10. The Nomenclature of Lipids (*J. Biol. Chem.*, **242**, 4845 (1967) and elsewhere).
 11. Nomenclature of Cyclitols (Jointly with IUPAC Commission on Nomenclature of Organic Chemistry) (*J. Biol. Chem.*, **243**, 5809 (1968) and elsewhere).
 12. Nomenclature of Steroids (Jointly with IUPAC Commission on Nomenclature of Organic Chemistry) (*Biochemistry*, **8**, 2227 (1969) and elsewhere).
- A document, OBN-5, describing the (American) NAS-NRC Office of Biochemical Nomenclature and listing other rules affecting biochemical nomenclature, is available from its Director, Dr. Waldo E. Cohn. (See also *J. Chem. Soc.*, **7**, 72 (1967)).

We are grateful to *The Journal of Biological Chemistry* for providing us with reproduction proofs of these rules.