Biochemistry

 Volume 5, Number 5 May 12, 1966

IUPAC–IUB Combined Commission on Biochemical Nomenclature. Abbreviations and Symbols for Chemical Names of Special Interest in Biological Chemistry.* Revised Tentative Rules (1965)†

Explanatory Memorandum[1]

The Commission on the Nomenclature of Biological Chemistry decided in 1958 that an attempt should be made to standardize the abbreviations and symbols used for chemical names of special interest in biological chemistry. A subcommittee, consisting of L. Hellerman, W. Klyne (Chairman), and E. C. Slater, was set up early in 1959 to deal with this problem.

The original draft proposals were based on the notes given at the beginning of each number of the *Journal of Biological Chemistry* (1958 *et seq.*) and of the *Suggestions to Authors* of the *Biochemical Journal* (*66*, 1 (1957)). These drafts were circulated to members of the Nomenclature Commission, editors of chemical and biochemical journals, and interested specialists in many fields.

The problems were discussed fully at the meeting of the Commission in Munich in September 1959—and also in joint sessions with the Organic Nomenclature Commission and the Enzyme Commission of the International Union of Biochemistry (IUB). A third draft, incorporating the results of the Munich discussions, was widely circulated in December 1959, and many useful comments on this were received.

A fourth draft, representing the "highest common factor" of all these comments and of many personal discussions, was prepared in August 1960, and circulated to members of the Commission on the Nomenclature of Biological Chemistry and to editors of some principal journals.

The meeting of editors of biochemical journals called together by the President of IUB in Cambridge in September 1960 invited W. Klyne to attend part of their meeting. He explained the history and purpose of the memorandum on abbreviations; he emphasized that this work lies on the borderline between the provinces of the two Unions, and that agreement of both Unions in principle was therefore very desirable.

After discussion, the Secretary General of IUB, R. H. S. Thompson, proposed the following statement, which was unanimously approved. "The contents of this memorandum (*i.e.* the fourth draft) were approved both by the Bureau of IUB and by the meeting of editors of biochemical journals called together by the President of IUB under the chairmanship of J. T. Edsall, at a meeting held in Cambridge on September 9, 1960."

The Tentative Rules were published in the IUPAC *Information Bulletin*, No. 13, in June 1961, and in

* The present document is being published at the request of the combined Commission on Biochemical Nomenclature of IUPAC–IUB. Publication in BIOCHEMISTRY does not necessarily constitute the endorsement or mandatory adoption of the rules by the editors.

† Changes made at the first revision (IUPAC *Information Bulletin* No. 20, July 1963, pp 13–26) are indicated by the symbol ▲. Those made in this second revision, which are more extensive, are indicated by ■. The latter changes were made by the IUPAC–IUB Combined Commission on Biochemical Nomenclature (CBN), which came into being in January 1964, with W. Klyne as Chairman and E. C. Slater as Secretary.

Comments on these re-revised proposals should be sent to the present Chairman (O. Hoffmann-Ostenhof) or Secretary (W. E. Cohn) or to any member of CBN (A. E. Braunstein, J. S. Fruton, B. Keil, W. Klyne, C. Liébecq, B. G. Malmström, R. Schwyzer, E. C. Slater), or corresponding member (N. Tamiya).

Published by permission of IUPAC, of IUB, and the official publishers to IUPAC, Messrs. Butterworths Scientific Publications.

Reprints of these Tentative Rules may be obtained from Waldo E. Cohn, Director, NAS–NRC Office of Biochemical Nomenclature, Oak Ridge National Laboratory, Box Y, Oak Ridge, Tennessee 37831.

[1] From the second version (1963).

April 1962 the Commission of Editors of IUB formally accepted the Tentative Rules thus published.

The IUPAC Commission has received a number of valuable suggestions for the amendment of the Tentative Rules and these were discussed at meetings in Amsterdam (April 1961), in Dobříš (near Prague) (September 1962), and in Zürich (April 1963). The *Revised Tentative Rules* (1963) were approved at the Zürich meeting.

It has been impossible to meet all the comments and criticisms that colleagues have kindly offered on the various drafts—since so many of the comments are mutually contradictory. However, these tentative proposals represent an honest attempt to give fair weight to all the diverse opinions that have been expressed.

Tentative Rules

1. Introduction

1.1 It is sometimes convenient to use abbreviations or symbols for the names of chemical substances, particularly in equations, tables, or figures, which would otherwise require the repeated use of unwieldy terms. The limited use of abbreviations and symbols of specified meaning is therefore accepted. However, clarity and unambiguity are more important than brevity.

1.2 Some chemists deprecate the use of *any* abbreviations or symbols for compounds. However, in the present state of biochemistry, increasing knowledge of the structure of large molecules such as proteins, polysaccharides, and polynucleotides makes it imperative to have some "shorthand" notation in which symbols are allotted to the monomeric units (monosaccharides, amino acids, and nucleosides), which are Nature's building bricks in these complex structures. Opponents of abbreviations should consider how unwieldy the formula of insulin would appear if the "three-letter" symbols for amino acids had not been used.

1.3 Titles and summaries of papers should be generally free of abbreviations. In the body of the paper, abbreviations and symbols should be used sparingly, and only if advantage to the reader results. Chemical equations, which traditionally consist of symbols, may use a shorthand expression for a term that appears in full in the neighboring text.

1.4 If, in exceptional circumstances, symbols or abbreviations are used in a summary, they should be defined in the summary, as well as in the body of the paper.

1.5 It is hoped that editors will adopt in their journals as many of the following rules as possible in the light of individual circumstances.

1.6 Even if a journal permits the use of these abbreviations without definition, nonstandard abbreviations should *always* be defined in each paper.

1.7 Nonstandard (*ad hoc*) abbreviations and symbols should not conflict with known ones, or with the general principles proposed in these rules (see also Section 8).

1.8 The symbols and abbreviations discussed here fall into two distinct classes.

(*a*) *Symbols* for monomeric units in macromolecules; these symbols are used to make up abbreviated structural formulas (sometimes called "shorthand" formulas), *e.g.*, Gly–Val–Thr for the tripeptide glycylvalylthreonine. These are generally used by structural organic chemists, and can be made fairly systematic.

(*b*) *Abbreviations* for semisystematic or trivial names, *e.g.*, ATP for adenosine triphosphate; FAD for flavin–adenine dinucleotide.

The abbreviations of the second kind are generally formed of three or four capital letters. They are required chiefly by biochemists and are generally introduced as required; the need is for brevity rather than for system. It is the indiscriminant coining of such abbreviations that has aroused objections to the use of abbreviations in general.

Symbols for Natural Macromolecules

1.9 There are three main series of symbols for monomeric units, *viz.*, those for amino acids, monosaccharides, and mononucleosides, of which the amino acid series is the oldest. An attempt has been made here to devise a standard treatment for all of the three great groups of macromolecules, which are built up from these units. The standardization of treatment will involve certain unimportant changes in the (as yet partly developed) systems for individual groups. This standardization is desirable for two reasons.

(*a*) The work of authors, editors, and readers is made simpler if the *same* principles apply to polypeptides, polysaccharides, and polynucleotides.

(*b*) Standard treatment is essential for dealing with "hybrid" compounds, built up of units of different kinds, *e.g.*, the nucleotide-peptides and glycopeptides.

1.10 It is much more difficult to be completely systematic in the planning of abbreviations and "shorthand" symbols for complex substances than in the construction of organic chemical formulas and physical symbols. Experience shows that it is not only difficult, but in some cases undesirable, to be rigidly consistent with these complex symbols.

The following example will illustrate these facts. For most purposes it is convenient to use the symbol Gly–Val–Thr to represent the tripeptide glycylvalylthreonine, as solid or in solution, whatever its state of ionization. We know that at certain defined pH values, the tripeptide will exist (mainly) as cation, or anion, or as dipolar ion, but it is usually unnecessary to make separate shorthand symbols to represent these different forms.

This deliberate lack of precision runs parallel with the convention by which biochemists talk about the "citric acid cycle" or "tricarboxylic acid cycle," in spite of the fact that the acids exist almost entirely as their anions at physiological pH.

In several cases it is thought desirable to recommend for the same substance two different forms of abbreviation or symbol, one or other of which is more convenient for specific purposes.

Alternative Abbreviations and Symbols

1.11 For some important compounds, it is in practice necessary to have *two* symbols or abbreviations. For example, most biological chemists will continue to speak of "adenosine diphosphate," or more often to abbreviate it as "ADP." Organic chemists interested in the structure and synthesis of this and related compounds will wish to call this compound "adenosine 5′-pyrophosphate," and to use a systematic symbol (Ado-5′-*P*-*P*). The abbreviation and the symbol should therefore coexist.

1.12 Abbreviations such as "ADP," which are to the organic chemist *trivial*, are used to form the systematic names of enzymes in the patterns proposed by the Enzyme Commission of the International Union of Biochemistry (Elsevier, 1965).

Language Differences

1.13 It is desirable that where trivial abbreviations (such as ACTH) are necessary, they should be identical in all languages—as are chemical symbols (*e.g.*, N standing for nitrogen, *azote*, and *Stickstoff*). It would be unfortunate if the substance called in English "ribonucleic acid," and abbreviated "RNA," were to retain two separate abbreviations, ARN (*acide ribonucléique*) and RNS (*Ribonukleinsäure*), in French and German, to say nothing of other languages. It is suggested that the international abbreviations should be taken from that language in which a given abbreviation first became common. Abbreviations introduced in the future may conveniently be based on Greek or Latin forms.

Structural Analogs

1.14 Structural analogs of a given compound should not generally be abbreviated as if they were derivatives of that compound.

2. Polypeptides and Proteins

This system is based on the original proposals of E. Brand and J. T. Edsall (*Ann. Rev. Biochem.*, *16*, 224 (1947)), as developed in the monograph of J. P. Greenstein and M. Winitz (The Chemistry of the Amino Acids, Vol. I–III, Wiley, New York, 1961). However, some modifications have been introduced so that it is possible to designate *all* amino acids found in proteins, including the acid amides and the hydroxylated compounds, by three-letter symbols—a capital followed by two lower case letters.

2.1 The following symbols denote the common amino acids and their residues (see also 2.5):

Alanine	Ala	Hydroxylysine	Hyl▲
Arginine	Arg	Hydroxyproline	Hyp▲
Aspartic acid	Asp	Isoleucine	Ile ▲
■ Asparagine	Asn	Leucine	Leu
		Lysine	Lys
	|	Methionine	Met
▲ Cystine (half)	Cys *or* Cys	Ornithine	Orn
	|	Phenylalanine	Phe
		Proline	Pro
▲ Cysteine	Cys	Serine	Ser
Glutamic acid	Glu	Threonine	Thr
■ Glutamine	Gln	Tryptophan	Trp■
Glycine	Gly	Tyrosine	Tyr
Histidine	His	Valine	Val

■ Modified amino acids, such as asparagine and glutamine, may also be represented as Asp(NH_2), Glu(NH_2) *or* Asp, Glu.

NH_2	NH_2

2.2 The abbreviations should *not* be used for the free amino acids in the *text* of papers, but only in tables, lists, and figures.

2.3 Where the sequence of residues in a peptide or protein is known, the symbols for the residues are written in order and joined by short lines (dashes, hyphens). Where the sequence is *not* known, the group of symbols, separated by *commas*, is enclosed in parentheses.[2]

In the formulation of linear polypeptides or proteins, the symbol written at the left-hand end of a known sequence is that of the amino acid carrying the free amino group, and the symbol written at the right-hand end is that of the residue of the amino acid carrying the free carboxyl group. Example: The condensed formula

Gly–Glu–Arg–Gly-Phe–(Phe, Tyr, Thr, Pro)–Lys–
Ala–Trp–Tyr–Val–Ile–Hyp–Cys–Ala

is that of a polypeptide in which the sequence of the first five amino acids has been established, the glycine at the left carrying the free amino group. The sequence of the next four amino acids is unknown, but the last nine amino acids are in known order with alanine carrying the free carboxyl group.

If the direction of the link must be specified, this may be done with an arrow thus (→), the point of the arrow indicating the nitrogen of the peptide bond ···CO→NH···

Example: Gly→Ala→Val.

The symbol → is desirable particularly in dealing with *cyclic* peptides.

Unless otherwise indicated, it is assumed that polyfunctional amino acids, such as glutamic acid, aspartic acid, and lysine, are joined by normal α-peptide bonds.

Abnormal links, *e.g.*, γ-peptide bonds or links formed through other functional groups, may be indicated by methods such as the following:

Glu	*or* Glu	
■ |	γ |	= glutathione (reduced)
└Cys–Gly	└Cys–Gly	

Comment

The links between residues have been shown previously by peptide chemists as full points (periods, dots; ·) and by carbohydrate chemists (generally) as

[2] It is preferable to display the polypeptide chain as a horizontal rather than as a vertical sequence.

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 printing, a short rule or dash (–), which is what we normally use for a chemical bond, should be the standard connecting symbol.

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 = $NH_2 \cdot CH_2 \cdot CO_2H$;[3] the free amino acid
(b) Gly– = $NH_2 \cdot CH_2 \cdot CO$—; the left-hand unit
(c) –Gly– = —$NH \cdot CH_2 \cdot CO$—; the middle unit
(d) –Gly = —$NH \cdot CH_2 \cdot CO_2H$; the right-hand unit

For peptides, a distinction may be made between the *peptide* itself, *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).

■ 2.4 The amino acid symbols represent the natural (L) form. Other forms are indicated by the appropriate symbols (D or DL) immediately preceding the amino acid symbol and separated from it by a hyphen. Example: Leu-D-Phe-Gly. When it is desired to make the number of residues appear in a more clear manner, the hyphen between the prefix and the symbol may be omitted. Example: Leu-D-Phe-Gly.

Rare Amino Acids

■ 2.5 The list in paragraph 2.1 is restricted to the more common amino acids. Symbols for the rare amino acids are included in *Abbreviated Designation of Amino Acid Derivatives and Polypeptides* (IUPAC *Information Bulletin*, in press; *J. Biol. Chem.*, in press).

State of Ionization

2.6 As stated in paragraph 1.10, it is generally convenient to use the same abbreviated formula for a polypeptide no matter what its state of ionization.

In some circumstances, however, an author will wish to show that a peptide is acting as a cation or anion; according to the convention of Greenstein and Winitz (The Chemistry of the Amino Acids), the amino-terminal and carboxyl-terminal ends of the peptide are marked 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 (I)
▲ $^+H_2$–Gly–Val–Thr–OH (II)
H–Gly–Val–Thr–O^- (III)

Derivatives

■ 2.7 Symbols for the functional groups of derivatives have been devised by specialists in the field (*cf.* reference in paragraph 2.5).

Comment

■ One-letter abbreviations for amino acid residues have been proposed. While recognizing the utility of such systems in computer analysis of sequences in proteins, the Commission does *not* recommend their use in printed material or teaching. The system proposed by Šorm *et al.* (*Collection Czech. Chem. Commun. 26*, 569 (1961)) has found relatively wide favor and may be recommended to those requiring a one-letter system for computer analysis.

3. Carbohydrates

A system of three-letter symbols for monosaccharide units and their residues, similar to that already in use for peptides, was introduced by the Carbohydrate Nomenclature Committees of the Chemical Society and the American Chemical Society (*cf. J. Chem. Soc.*, 5121 (1952); *Chem. Eng. News*, 1776 (1953); *J. Org. Chem. 28*, 281 (1963)). The following rules are based on this system.[4]

■ 3.1 The following symbols are used to indicate monosaccharide units and their residues in oligosaccharides and polysaccharides.

Glucose	Glc[5]	Fructose	Fru
Galactose	Gal	Ribose	Rib
Mannose	Man		

Other monosaccharides are represented similarly by the first three letters of their names, unless this would lead to confusion with an existing symbol (*e.g.*, Gly and Thr in the amino acid series).

3.2 Pyranose and furanose forms are designated where necessary by the suffixes *p* and *f*.

■ 3.3 Configurational symbols D and L (small Roman capital letters) and anomeric prefixes are shown where necessary as prefixes. Examples: (i) an α-D-glucopyranose unit, α-D-Glc*p or* Glc*p;* (ii) a β-D-fructofuranose unit, β-D-Fru*f or* Fru*f*.

3.4 Symbols thus formed are joined by short rules to indicate the links between units. The position and nature of the links are shown by numerals and the anomeric symbols α and β. Examples:

Maltose, Glc*p*α1–4Glc
Lactose, Gal*p*β1–4Glc
Stachyose, Gal*p*α1–6Gal*p*α1–6Gal*p*α1–2βFru

■ [3] Or corresponding ionized forms. Gly is ordinarily not to be used alone in text (see 2.2).

▲ [4] These proposals and those given in the Tentative Rules of Carbohydrate Nomenclature (separate to IUPAC *Information Bulletin*, April 1963) differ in some respects. Attempts are being made to resolve the differences by discussion between the Commissions involved.

▲ [5] Where no ambiguity can arise, the single-letter symbol G may be used.

A branched-chain tetrasaccharide:

Glcβ1–3Galβ1–4Glc
2
|
1αFuc

Arrows may be used to indicate the direction of the glycoside link, the arrow pointing away from the hemiacetal carbon of the link; *e.g.*, lactose may be represented as Gal*p*β1→4Glc.

3.5 A 2-deoxy sugar is designated by the symbol for its most common parent sugar with the prefix "de." Other deoxy sugars may be designated similarly with a positional numeral. Examples: 2-deoxyribose, deRib; 3-deoxyglucose, 3-deGlc.

Comment

It may sometimes be necessary to enclose such a symbol in parentheses to avoid confusion between the numeral indicating the "deoxy position" and numerals indicating the position of linkages.

3.6 Derived monosaccharide units—such as glyconic acids, glycuronic acids, 2-amino-2-deoxy saccharides, and their *N*-acetyl derivatives—may be designated by reasonable modified symbols, defined in each paper. Examples of symbols that have been used are as follows (all in the glucose, Glc, series):

▲

Gluconic acid	GlcA
Glucuronic acid	GlcUA
Glucosamine	GlcN
N-Acetylglucosamine	GlcNAc

3.7 Symbols should not be used for the monosaccharides themselves, except in tables, lists, and figures.

4. Phosphorylated Compounds: General

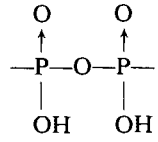
4.1 Phosphorylated compounds may be designated by the name (or abbreviation) of the parent compound with a capital italic *P* as a prefix or suffix.[6]

P is used as a prefix where it symbolizes "phospho-" at the beginning of a name. *P* is used as a suffix where it symbolizes "phosphoric acid" or "phosphate" at the end of a name.[7]

For compounds containing more than one position available for phosphorylation, the position of the phosphate group should always be indicated by number or Greek letter.

4.2 The capital *P* when linked to one radical indicates $-PO(OH)_2$ or any ion derived from it; when linked to two radicals it indicates $-PO(OH)-$, or the ion derived from it.

4.3 The pyrophosphate group (I) is represented by *-P-P-:*



(I)

Two separate phosphate groups, attached at different points to the same molecule, are represented by P_2.

Examples:

Glucose 6-phosphate	Glucose-6-*P* or Glc-6-*P*
Glycerol 3-phosphate *or* α-phosphoglycerol	Glycerol-3-*P* / α-*P*-Glycerol
3-Phosphoglyceric acid / Glycerate 3-phosphate	3-*P*-Glyceric acid / Glycerate-3-*P*
Phosphoenolpyruvate	*P*-Enolpyruvate
■ Fructose 1,6-bisphosphate[8]	Fructose-1,6-P_2 or Fru-1,6-P_2
Creatine phosphate / Phosphocreatine	Creatine-*P* / *P*-Creatine

■ 4.4 The term diphosphate (and the abbreviation DP, as in ADP) is correctly used only for the pyrophosphate group (IUPAC Inorganic Rules 7.5, 5.213, 2.251; also Organic Rules A2.5 and 67). Compounds with two or more orthophosphate residues are more properly termed bis-, tris-, tetrakis-, etc., phosphates. The older term, fructose 1,6-diphosphate, strictly interpreted, could indicate a pyrophosphate group connecting the 1 and 6 positions of fructose.

5. Nucleotides and Nucleic Acids ■

5.1 Two systems are recognized, one (5.3) using three-letter symbols for the more common nucleosides (like those used for amino acids and monosaccharides in Sections 2 and 3) and a capital italic *P* for the phosphate residue, the other (5.4) using single capital letters for the more common nucleosides and a lower-case p for the phosphate residue.

5.2 In either system, glycosyl linkages are assumed to be β and to involve only D-ribose or D-deoxyribose, and the phosphodiester linkage is assumed to be 3′–5′ from left to right unless otherwise specified by appropriate *ad hoc* symbols or numerals.

5.3 THREE-LETTER SYMBOLS. 5.3.1 The phosphate group is designated by an *italic* capital *P* (*cf.* Section 4), to distinguish it from *Roman* capital P for phosphorus.

5.3.2 The (ribo) nucleosides are designated by the following three-letter symbols, chosen to avoid confusion with the corresponding bases:

[6] *Comment:* This type of partial abbreviation (*e.g.*, glucose-6-*P*) is convenient in biochemical papers where there is much discussion of phosphorylated metabolites and intermediates. It is not commonly used in organic chemical papers.

[7] The *P* is italicized in order to avoid confusion with the accepted symbol for the phosphorus atom (Roman capital P).

■ [8] See comment on this change in 4.4.

Ado	adenosine	Thd	ribosylthymine[9] ■
Guo	guanosine	Cyd	cytidine
Ino	inosine	Urd	uridine
Xao	xanthosine	Ψrd	pseudouridine ■ (5-ribosyluracil)

Ribosylnicotinamide may be designated by Nir.

■ 5.3.3 The 2′-deoxyribonucleosides are designated by the symbols for the corresponding ribose derivatives (5.3.2) with the prefix d. Examples are dAdo for 2′-deoxyadenosine, dThd for 2′-(deoxy)thymidine[9] (2′-deoxyribosylthymine).

The letter d may also be used as a prefix to a series (an oligonucleotide) to indicate that *all* the sugars in the series are 2′-deoxyribosyl units.

5.3.4 The points of attachment of phosphate residues to a sugar, if other than 3′-*P*-5′, are designated by the appropriate primed numerals, separated by hyphens. Examples:

(i) Adenosine 2′-phosphate: Ado-2′-*P*
Adenosine 5′-phosphate: Ado-5′-*P* or *P*-5′-Ado

(ii) 5′-*O*-Phosphoryldeoxyadenylyl-(3′–5′)-thymidine: *P*-5′-dAdo-3′-*P*-5′-dThd[9] or *P*-dAdo-*P*-dThd[9]

The positional numerals may precede a series, as in 2′–5′-(Ado-*P*-Guo-*P*-Urd-*P*), which specifies Ado-2′-*P*-5′-Guo-2′-*P*-5′-Urd-2′-*P*. When the series in the left-to-right direction is 3′–5′, as in example ii above, they may be omitted (*cf.* 5.2).

5.3.5 A cyclic phosphate group is designated by the two positional numerals for the points of attachment of a single *P*, as in Cyd-2′:3′-*P*. (The corresponding bisphosphate would be Cyd-2′,3′-P_2 or *P*-2′-Cyd-3′-*P*.)

5.3.6 The so-called nucleoside diphosphate sugars, which are sometimes called pyrophosphates, may be represented as follows: Urd-5′-*P*-*P*-Glc for uridine diphosphate glucose [*i.e.*, uridine 5′-(α-D-glucopyranosyl diphosphate)]; Urd-5′-*P*-*P*-Gal.

5.4 ONE-LETTER SYMBOLS.[10] 5.4.1 The phosphate group is designated by a lower case p (to separate what would otherwise be a solid mass of capital letters).

■ 5.4.2 The common (ribo)nucleosides are designated by single capital letters, thus:

A	adenosine	T	ribosylthymine[9]
G	guanosine	C	cytidine
I	inosine	U	uridine
X	xanthosine	Ψ	pseudouridine (5-ribosyluracil)

The following general symbols are also useful: Pu, unspecified purine nucleoside; Py, unspecified pyrimidine nucleoside; N, unspecified nucleoside.

■ 5.4.3 The 2′-deoxyribonucleosides are designated by the same symbols (5.4.2) preceded by d (*cf.* 5.3.3). Thus, dA = 2′-deoxyribosyladenine, dT = 2′-deoxyribosylthymine (= thymidine).[9]

5.4.4 The points of attachment of phosphate residues, if other than 3′p5′, may be indicated as in 5.3.4. A regular 3′–5′ sequence (read left to right), as in the natural nucleic acids, need not be specified by positional numerals (5.2).

In this system, the substances in 5.3.4 become: A2′p; pA; d(pApT).

Other examples are:

ApGpUp (3′–5′ trinucleotide, ending at right in a 3′-phosphate).

ApGpU-cyclic p (the same, but ending in a 2′:3′-cyclic phosphate).[11]

pApApA (3′–5′ trinucleotide, starting with a 5′-phosphate at left, ending with unsubstituted 2′-and 3′-hydroxyl groups at right).

■ 5.5 MORE COMPLEX STRUCTURES. 5.5.1 *Sequence Designation.* For more complex structures (large oligonucleotides or polynucleotides), in which known and unknown sequences may be involved, the p for phosphate between two nucleosides may be replaced by a *hyphen* (for a *known* sequence) or a *comma* (*unknown* sequence) to give a system identical with that used for amino acid sequences (see Section 2.3 above). Regular 3′–5′ linking is assumed unless indicated otherwise. Thus GpApUp(CpCpUp)Gp—a 3′-ended heptanucleotide of partially known sequence—becomes G-A-U(C,C,U)Gp or G-A-U(C_2,U)Gp; d-pTpTpCpTpTpC becomes d(pT-T-C-T-T-C).

5.5.2 *Higher Polymers.* For sequences too long or repetitive or obscure for detailed exposition, the prefix "poly" may be used in conjunction with the comma and hyphen. For example, the alternating regular sequence dA-dT-dA-dT---- becomes poly d(A-T); the random copolymer of equal amounts of U and A becomes poly (U,A). The prefix "poly" may be omitted when the number (or proportions) of nucleoside residues is specified by subscript numerals, if known, or by the subscript *n* (in place of the *p* used in the system devised by the IUPAC Commission on Macromolecules (*J. Polymer Sci. 8*, 257 (1952)) for molecules of indefinite size. Thus, poly d(A-T) may be expressed as d(A-T)$_n$, and poly (U_2,A) as (U_2,A)$_n$. Composition and size may be shown by appropriate numerical subscripts, as in (U_2,A)$_{50}$, which contains 100 U's and 50 A's in random sequence, and in d(A-T)$_{50}$, which contains 50 dA's and 50 dT's in regular alternating sequence. Multiple parentheses or brackets may be used as in organic nomenclature for blocks within polymers, side chains, etc.

5.5.3 *Nucleoside Symbols.* The more common nucleoside residues are represented by single capital letters (see 5.4.2 above). All other nucleosides should be represented by single capital letters, insofar as possible, defined as introduced (*e.g.*, B for BrU). Where a nucleo-

■ [9] Because *thymidine* has traditionally been used for 2′-deoxyribosylthymine, arising at a time when the ribosyl analog was not known, it is recommended that the prefix r (for ribo) or d (for deoxyribo) be used with Thd or with T whenever there is a possibility of misunderstanding which substance is intended.

[10] Intended chiefly for oligonucleotides and polynucleotides. The IUPAC Commission for the Nomenclature of Organic Chemistry prefers the three-letter symbols (5.3).

■ [11] The symbol > for "cyclic" is useful in the one-letter system. Thus the example can be represented as ApGpU>p. Unless otherwise stated, this symbol indicates a 2′:3′-cyclic phosphate residue.

side symbol must include more than one character, it should contain neither hyphens nor commas (*e.g.*, -2MeG-, -6diMeA-, -BrU-). Linkages other than 3′–5′, or sugars other than ribose or deoxyribose, should be indicated by special *ad hoc* symbols. d and r may precede whole chains, groups within chains, or individual nucleoside residues, as appropriate.

5.5.4 *Association between Chains. Associations* between two or more nucleotide chains may be indicated by the *center dot*—as in $(A)_n \cdot (U)_n$ or $(dG)_n \cdot (dC)_n$ or $[(A)_n \cdot (U)_n \cdot (U)_n]$. The *absence of association* is indicated by the *plus* sign, as in poly (A) + poly (C). The *absence of definite information* on association is indicated by the comma (again meaning "unknown"), as in poly (A), poly (A-U) or $(A)_n$,$(A\text{-}U)_n$.

Special Abbreviations

5.6 The 5′-mono-, di-, and triphosphates of the common nucleosides may be designated by the customary special abbreviations, *e.g.*, AMP, ADP, and ATP for the derivatives of adenosine. The corresponding derivatives of cytidine, guanosine, inosine, uridine, and pseudouridine may be designated by similar abbreviations in which the initial letters are C, G, I, U, and Ψ, respectively. Thus, for example, IMP = inosine 5′-monophosphate; UDP = uridine 5′-diphosphate. Uridine diphosphate glucose may be designated by UDPG.

These compounds should, however, be designated in more "chemical" papers by systematic symbols as indicated in paragraph 5.3.4, *e.g.*, Ado-5′-*P*, Ado-5′-*P-P*, Ado-5′-*P-P-P*, when required for consistency with the other nucleotides.

5.7 Flavin mononucleotide (riboflavin 5′-phosphate) may be designated by the special abbreviation FMN.

5.8.1 The two types of nucleic acid are designated by their customary abbreviations:

RNA, ribonucleic acid or ribonucleate
DNA, deoxyribonucleic acid or deoxyribonucleate

■ 5.8.2 It is sometimes convenient to designate fractions or functions of RNAs by prefixes (*e.g.*, mRNA for "messenger" RNA, tRNA for "transfer" RNA, rRNA for "ribosomal" RNA, nRNA for "nuclear" RNA). Such terms should be defined in each paper unless defined by the journal in which the paper is published.

■ 5.8.3 Transfer RNAs that accept specific amino acids may be designated as, for example, $tRNA^{Ala}$ for the tRNA that accepts alanine (*i.e.*, "alanine tRNA"); in the case of more than one such species, they may be distinguished by subscripts, as $tRNA_1^{Ala}$, $tRNA_2^{Ala}$, etc. When such a tRNA species is bound to an amino acid, it may be designated as, for example, alanyl-$tRNA^{Ala}$. Specification of its source should be in parentheses before or after, as, for example, alanyl-$tRNA_2^{Ala}$ (*E. coli*) or (*E. coli*) alanyl-$tRNA_2^{Ala}$.

6. Coenzymes

There has been much controversy about names and symbols for the nucleotide coenzymes (DPN *vs.* CoI, etc.).

The Enzyme Commission of the International Union of Biochemistry decided in August 1959 to recommend the following names, for the reasons briefly stated in the comment below.

Nicotinamide–adenine dinucleotide for the compound hitherto commonly called diphosphopyridine nucleotide or Coenzyme I.

Nicotinamide–adenine dinucleotide phosphate for the compound hitherto commonly called triphosphopyridine nucleotide or Coenzyme II.

The IUPAC Commission on the Nomenclature of Biological Chemistry after discussion accepted these recommendations of the Enzyme Commission, which were formally adopted by the IUB Council at Moscow in August 1961.

Comment (*cf.* Dixon, M., *Nature 188*, 464 (1960))

The two main systems of nomenclature of the nicotinamide nucleotide coenzymes (the CoI and the DPN systems) are both unsatisfactory. The first gives no indication of the chemical structure at all; the second indicates a chemical structure that is incorrect.

Since no compromise between the two systems is possible, the only satisfactory solution is to abandon both and to adopt a name that indicates the correct chemical structure. The name adopted should be consistent with the existing names of three closely related compounds, namely, the corresponding mononucleotide (nicotinamide mononucleotide, NMN) and the two flavin nucleotides (flavin–adenine dinucleotide, FAD, and flavin mononucleotide, FMN).

The name the Enzyme Commission of IUB, after careful consideration of possible alternatives, has decided to recommend in place of CoI or DPN, namely nicotinamide–adenine dinucleotide (NAD), not only indicates the structure satisfactorily, but forms a logical system with the three that are already generally accepted.

CoII or TPN is a phosphorylated derivative of NAD, and may be called nicotinamide–adenine dinucleotide phosphate, conveniently abbreviated to NADP.

6.1 The dinucleotide coenzymes may be designated by the following abbreviations:

nicotinamide–adenine dinucleotide (formerly DPN, CoI)	NAD
nicotinamide–adenine dinucleotide phosphate (formerly TPN, CoII)	NADP

■ These abbreviations do not specify the state of oxidation of the compounds.

■ 6.2 The oxidized and reduced forms of the coenzymes may be designated by NAD^+ ($NADP^+$) and NADH (NADPH), respectively.[12] They may be used in an equation as follows:

$$NAD^+ + XH_2 \rightleftharpoons NADH + H^+ + X$$

■ [12] The IUB Standing Committee on Enzymes has used the descriptive terminology of "NAD (NADP)" and "reduced NAD (NADP)."

6.3 Other coenzymes may be designated as follows:

FAD, $FADH_2$	flavin-adenine dinucleotide, and its reduced form
FMN, $FMNH_2$	flavin mononucleotide, and its reduced form
GSH, GSSG	glutathione, and its oxidized form
CoA, acetyl-CoA *or* CoASH, CoASAc	Coenzyme A and its acetyl derivative (alternative forms)

■ 6.4 Systematic symbols may be built up for some of these coenzymes as shown in paragraphs 5.3.4 and 5.3.6. Examples:

NAD = Nir-5′-*P*-*P*-5′-Ado
NADP = Nir-5′-*P*-*P*-5′-Ado-2′-*P*

7. Miscellaneous Compounds

7.1 The following abbreviations are permitted; although they are fairly common, they should be defined in any paper if it is thought that readers might be unfamiliar with them. Some abbreviations are taken from the list published by *Annual Review of Biochemistry*.

ACTH	adrenocorticotropin, adrenocorticotropic hormone, or corticotropin
BAL	2,3-dimercaptopropanol
CM-cellulose	*O*-(carboxymethyl)cellulose
DDT	1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane
DEAE-cellulose	*O*-(diethylaminoethyl)cellulose
DFP	diisopropyl phosphorofluoridate
DNP-	2,4-dinitrophenyl-
DOC	11-deoxycorticosterone
DOCA	11-deoxycorticosterone acetate
DOPA	3,4-dihydroxyphenylalanine
DPT	diphosphothiamine (thiamine pyrophosphate, cocarboxylase) ▲
EDTA	ethylenediaminetetraacetic acid (or -acetate)
FDNB	1-fluoro-2,4-dinitrobenzene
Hb	hemoglobin (deoxygenated)
HbCO	"carboxy" hemoglobin—*i.e.* hemoglobin *plus* carbon monoxide
HbO_2	oxyhemoglobin
MetHb	methemoglobin
Mb	deoxygenated myoglobin (may be modified in the same way as Hb)
MSH	melanocyte-stimulating hormone
P_i	orthophosphate (inorganic) ▲
PP_i	pyrophosphate (inorganic) ▲
TEAE-cellulose	*O*-(triethylaminoethyl)cellulose
Tris	tris(hydroxymethyl)aminomethane; 2-amino-2-hydroxymethylpropane-1,3-diol

7.2 If one-letter symbols for steroids (Compound F, Substance S) are used, the systematic name of the compound should be given at least once in each paper. Derivatives, such as "tetrahydro-E" and "11-epi-F," should also be clearly defined by systematic names.

8. Standards for New Abbreviations

8.1 Abbreviations other than those listed or defined above should be constructed in accordance with the following principles.

8.2 The number should be limited; none should be introduced except where repeated use is required. Three-letter abbreviations are most convenient.

Use in another sense of an accepted abbreviation *must* be avoided.

Where a number of derivatives, salts, or addition compounds may be formed, the name of the common fundamental structure should be the one abbreviated, so that other symbols may be attached to it.

9. Alphabetical Lists

For convenience the symbols and abbreviations are collected in the alphabetical lists in Tables I and II.

TABLE I: Symbols for Monomeric Units in Macromolecules (or in Phosphorylated Compounds).

Symbol	Monomeric Unit in Macromolecule	Symbol	Monomeric Unit in Macromolecule
A, Ado	adenosine	Hyl	hydroxylysine
Ala	alanine	Hyp	hydroxyproline
Arg	arginine	I, Ino	inosine
Asp	aspartic acid	Ile	isoleucine
Asp(NH_2), Asn	asparagine	Leu	leucine
C, Cyd	cytidine	Lys	lysine
Cys[a]	cystine (half)	Man	mannose
Cys	cysteine	Met	methionine
■de, d	(indicates "deoxy" in carbohydrates and nucleotides)	Nir	ribosylnicotinamide
		Orn	ornithine
		P, p	phosphate
f	(suffix) furanose	*p*	(suffix) pyranose
Fru	fructose		
Gal	galactose	Phe	phenylalanine
G, Glc[b]	glucose	Pro	proline
G, Guo[b]	guanosine	Rib	ribose
GlcA	gluconic acid	Ser	serine
GlcN	glucosamine	Thr	threonine
GlcNAc	*N*-acetylglucosamine	Trp	tryptophan
		T, Thd	ribosylthymine[9]
GlcUA	glucuronic acid		
Glu	glutamic acid	Tyr	tyrosine
Glu(NH_2), Gln	glutamine	U, Urd	uridine
		Val	valine
Gly	glycine	Xao	xanthosine
His	histidine		

[a] With vertical bond above or below "s" (see Section 2). [b] The one-letter symbol G must *not* be used if confusion between its two meanings can arise.

TABLE II: Abbreviations for Semisystematic or Trivial Names.

ACTH	adrenocorticotropin, adrenocorticotropic hormone, or corticotropin	GSSG	oxidized glutathione
		GTP	guanosine 5′-triphosphate (pyro)
		Hb, HbCO, HbO_2	hemoglobin, carbon monoxide hemoglobin, oxyhemoglobin
ADP	adenosine 5′-diphosphate (pyro)		
AMP	adenosine 5′-phosphate	IDP	inosine 5′-diphosphate (pyro)
ATP	adenosine 5′-triphosphate (pyro)	IMP	inosine 5′-phosphate
BAL	2,3-dimercaptopropanol	ITP	inosine 5′-triphosphate (pyro)
CDP	cytidine 5′-diphosphate (pyro)	Mb, MbCO, MbO_2	myoglobin, carbon monoxide myoglobin, oxymyoglobin
CM-cellulose	*O*-(carboxymethyl)cellulose		
CMP	cytidine 5′-phosphate	MetHb, MetMb	methemoglobin, metmyoglobin
CoA (or CoASH)	coenzyme A	MSH	melanocyte-stimulating hormone
CoASAc	acetyl coenzyme A	NAD	nicotinamide–adenine dinucleotide (cozymase, Coenzyme I, diphosphopyridine nucleotide)
CTP	cytidine 5′-triphosphate (pyro)		
DEAE-cellulose	*O*-(diethylaminoethyl)cellulose		
DDT	1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane	NADP	nicotinamide–adenine dinucleotide phosphate (Coenzyme II, triphosphopyridine nucleotide)
DFP	diisopropyl phosphorofluoridate		
DNA	deoxyribonucleic acid	NMN	nicotinamide mononucleotide
DNP-	2,4-dinitrophenyl-	P_i	inorganic orthophosphate
DOPA	3,4-dihydroxyphenylalanine	PP_i	inorganic pyrophosphate
DPN[a]	diphosphopyridine nucleotide	RNA	ribonucleic acid
DPT	diphosphothiamine (thiamine pyrophosphate, cocarboxylase)	TEAE-cellulose	*O*-(triethylaminoethyl)cellulose
		TPN[b]	triphosphopyridine nucleotide
EDTA	ethylenediaminetetraacetate	Tris	tris(hydroxymethyl)aminomethane (2-amino-2-hydroxymethylpropane-1,3-diol)
FAD	flavin–adenine dinucleotide		
FDNB	1-fluoro-2,4-dinitrobenzene		
FMN	riboflavin 5′-phosphate	UDP	uridine diphosphate (pyro)
GDP	guanosine 5′-diphosphate (pyro)	UDPG	uridine diphosphate glucose
GMP	guanosine 5′-phosphate	UMP	uridine monophosphate
GSH	glutathione	UTP	uridine triphosphate (pyro)

[a] Replaced by NAD (see Section 6). [b] Replaced by NADP (see Section 6).