Biochem. J. (1970) **120**, 449–454
Printed in Great Britain

IUPAC-IUB Commission on Biochemical Nomenclature (CBN)

Abbreviations and Symbols for Nucleic Acids, Polynucleotides and their Constituents

Recommendations 1970[1]

INTRODUCTION

The 1965 Revision of *Abbreviations and Symbols for Chemical Names of Special Interest in Biological Chemistry* was completed and published in 1965 and 1966 [1], almost coincident with the elucidation of the first complete nucleic acid sequence [2,3] and with the development of methods for the synthesis of specific polynucleotide sequences [4]. The latter developments and others (*e.g.*, modification of sugar components, synthesis of unnatural linkages) require a unified system for representing long sequences containing unusual or modified nucleoside residues. The system should facilitate comparisons between two or more such extended molecules, as in the search for homologies. At the same time, it must retain sufficient flexibility to accommodate the large variety of polymers synthesized by polymerases and be consistent, in this regard, with the rules governing the representation of polymerized amino acids [5].

The workers who first encountered these various needs invented a number of devices to achieve the representations required in their own papers, basing these for the most part upon the one-letter system presented in Section 5.4 of *Abbreviations and Symbols* [1]. Few of these devices have the capability of meeting all the situations that are now apparent. Hence the effort was undertaken to construct a system meeting as many of the latter as possible, preserving the previous, basic system and introducing additional conventions. This effort, as did the previous one, involved consultation with a large number of active workers in many countries over a period of some years. The conventions added here (indicated by ▲ for major additions, △ for minor revisions) are already in use by many of them, *e.g.* [3,6–8].

The present (1970) Recommendations are the result; they replace Section 5 of the previous Tentative Rules [1].

N-1. Abbreviations

N-1.1. SIMPLE NUCLEOTIDES[2]

The 5′-mono-, di-, and triphosphates of the common ribonucleosides may be represented by the customary abbreviations exemplified by AMP, ADP, ATP (AtetraP) in the adenosine series. The corresponding derivatives of other nucleosides are abbreviated similarly, using the

△ symbols in N-3.2, *i.e.*, A, C, G, I, T, U, Ψ, X for the known nucleosides; R and Y for unspecified purine and pyrimidine nucleosides, respectively; N for unspecified nucleoside (not X or Y); B, S, and D are reserved for 5-bromouridine, thiouridine, and 5,6-dihydrouridine, respectively. Orotidine may
△ be designated by O to give OMP for orotidine 5′-phosphate[2].

△ The di- and triphosphates may on occasion be better expressed in the alternate form ppN or pppN, as in the polymerization equation n ppN → $(pN)_n$ + n P_i, or when the outcome of specific labeling is to be indicated, *e.g.*, n p^*ppN → $(p^*N)_n$ + n PP.

Uridine diphosphate glucose may be represented as UDPG or UDP-Glc; the latter form is preferred if there is the possibility of confusing G for glucose with G for guanosine.

In the context of the chemistry of the nucleosides or nucleotides, the more systematic three-letter symbols (N-2) should be used, *e.g.*, Ado-5′*PPP* or Urd-5′*PP*-Glc (N-2.4.3).

N-1.2. NUCLEOTIDE COENZYMES AND RELATED SUBSTANCES

Riboflavin 5′-phosphate (flavin mononucleotide)	FMN
Flavin-adenine dinucleotide (oxidized and reduced)	FAD, $FADH_2$
Nicotinamide mononucleotide	NMN
Nicotinamide-adenine dinucleotide[3] (oxidized and reduced)	NAD^+, NADH
Nicotinamide-adenine dinucleotide phosphate[4]	$NADP^+$, NADPH

△ Analogues of NAD or NADP (the generic terms require neither the plus sign nor the H) may be named by substituting an appropriate defined symbol for the N or the A,

▲ [2] When abbreviations for single bases or nucleosides are required and permitted, the three-letter symbols listed in N-2.2 and N-2.3 should be used (see Comments in these sections), not single letters and not, *e.g.*, UR, TdR, etc. Examples:

	Proscribed	Proposed
Fluorouracil	FU	FUra
Fluorouridine	FUR	FUrd
Fluorodeoxyuridine	FUdR	FdUrd
Thymidine	TdR	dThd
Bromouracil	BU	BrUra
Bromodeoxyuridine	BUdR	BrdUrd

[3] Formerly diphosphopyridine nucleotide (DPN, DPN^+, DPNH) and coenzyme I.

[4] Formerly triphosphopyridine nucleotide (TPN, TPN^+, TPNH) and coenzyme II.

[1] These Recommendations were approved by the IUPAC-IUB Commission on Biochemical Nomenclature in 1969 and are published by permission of IUPAC and IUB.

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

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

e.g., AcPd (for acetyl-pyridine) in place of N; I (for inosine) in place of A, etc.

Semi-systematic names (see N-2) may often be used to advantage in discussing the chemistry of these dinucleotides, *e.g.*, NADP = Nir-5′-*PP*5′-Ado-2′*P*.

N-1.3. NUCLEIC ACIDS

N-1.3.1. The two main types of nucleic acids are designated by their customary abbreviations, RNA (ribonucleic acid or ribonucleate) and DNA (deoxyribonucleic acid or deoxyribonucleate). Ribonucleoprotein and deoxyribonucleoprotein should not be abbreviated.

N-1.3.2. *RNA Fractions*

Fractions of RNA or DNA, or functions exercised by preparations of RNA may be designated as follows:

messenger RNA	mRNA	transfer RNA[5]	tRNA ▲
ribosomal RNA	rRNA	complementary RNA	cRNA △
nuclear RNA	nRNA	mitochondrial DNA	mtDNA △

These are generic terms and apply to preparations as well as to specific molecules.

N-1.3.3. *Transfer RNA's*

Transfer RNA's that accept a specific amino acid are designated as follows (using alanine tRNA as an example):

a) Nonacylated: alanine tRNA or $tRNA^{Ala}$;

b) Aminoacylated: alanyl-tRNA or Ala-tRNA, or Ala-$tRNA^{Ala}$.

Comment. (i) The hyphen (in b) represents the aminoacyl bond and should not be used to connect a noun-adjective; (ii) the attached aminoacyl residue (in b) has the -yl ending, whereas the adjective describing the nonacylated form (a) does not; (iii) the superscript designator utilizes the conventional symbols for amino acid residues [1,9] exactly—one capital, two small letters.

Isoacceptors, i.e., two or more tRNA's accepting the same amino acid, are designated by subscripts, *e.g.*, $tRNA_1^{Ala}$, $tRNA_2^{Ala}$, etc.

Specification of source may be made in parentheses before or after the abbreviation, *e.g.*, (*E. coli*) $tRNA_1^{Ala}$, alanyl-$tRNA_2^{Ala}$ (*E. coli*).

The special problem of the particular methionine tRNA ($tRNA^{Met}$) that, once aminoacylated to give Met-tRNA, can be formylated to fMet-tRNA may be solved by the use of a subscript f (in the isoacceptor position) or by the use of $tRNA^{fMet}$. Thus $tRNA_f^{Met}$ (or $tRNA^{fMet}$) can be converted enzymically to Met-$tRNA_f^{Met}$ (or Met-$tRNA^{fMet}$) and then to fMet-$tRNA_f^{Met}$ (or fMet-$tRNA^{fMet}$); Met-$tRNA^{Met}$ cannot be formylated enzymically.

Symbols

General Concepts and Conventions

Two systems are recognized, designated the "three-letter" and the "one-letter" system, respectively. The first (N-2), patterned after the systems in use for amino acid and saccharide residues in polymers [1], is designed largely for descriptions of chemical work involving bases, nucleosides, nucleotides and very small oligonucleotides, or for abbreviating these in minimum space (as on chromatograms or figures or table headings). The "one-letter" system (N-3 and N-4) is designed for the representation of oligonucleotides or polynucleotides, or parts thereof, and for their noncovalent associations, not for mononucleotides or nucleotides. Neither system is intended to replace the names of the latter substances in the text of papers.

In both systems, it is assumed, in the absence of appropriate symbols, that (a) all nucleosides (except pseudouridine) are 1-(pyrimidine) or 9-(purine) glycosyls, (b) all nucleoside linkages are β, (c) all sugar configurations are D, (d) all sugar residues are ribosyls unless otherwise specified, (e) all deoxyribosyls are 2′-deoxyribosyls, and (f) only 3′ → 5′ linkages, read from left to right, are involved.

N-2. THREE-LETTER SYMBOLS[6]

N-2.1. *Phosphoric Acid Radical*

The phosphoric acid radical, whether monoesterified or diesterified, is designated by an italic capital *P*.

N-2.2. *Purines and Pyrimidines*

These are designated by the first three letters of their trivial names:

Ade	adenine	Thy	thymine
Gua	guanine	Cyt	cytosine
Xan	xanthine	Ura	uracil
Hyp	hypoxanthine	Oro	orotate
Pur	unknown purine	Pyr	unknown pyrimidine
	Base	unknown base	

Sur and Shy may be considered for thiouracil and thiohypoxanthine (6-mercaptopurine), respectively.

When abbreviations for single purines or pyrimidines are required and permitted, the above symbols should be used rather than A, C, G, T, U, etc.[2].

N-2.3. *Nucleosides*

N-2.3.1. *The ribonucleosides* are designated by the following symbols, chosen to avoid confusion with the corresponding bases:

Ado	adenosine	Thd	ribosylthymine (not thymidine)	
Guo	guanosine			
Ino	inosine	Cyd	cytidine	
△ Sno	thioinosine (mercaptopurine ribonucleoside)	Urd	uridine	
		Srd	thiouridine	△
Xao	xanthosine	Ψrd	pseudouridine	
△ Puo	"a purine nucleoside"	Ord	orotidine	△
△ Nuc	"a nucleoside"	Pyd	"a pyrimidine nucleoside"	△

Ribosylnicotinamide may be designated by Nir.

Comment. The prefix r (for ribo) may be used for emphasis or clarity. It may precede a single residue or, if applicable, a connected series.

N-2.3.2. *The 2′-deoxyribonucleosides* are designated by the above symbols (N-2.3.1) prefixed by d, *e.g.*, dAdo for 2′-deoxyribosyladenine (deoxyadenosine), dThd for 2′-deoxyribosylthymine (thymidine). The d may be used as a prefix to a connected series if all members of that series are 2′-deoxyribosyl derivatives. In mixed series, r and d should both be used before the appropriate residues, *e.g.*, *P*-dAdo-*P*-rThd-*P*.

△ Other sugar residues may be indicated by similar prefixes, *e.g.*, a for arabinose, x for xylose, l for lyxose.

[5] Replaces "soluble" RNA (sRNA), which should no longer be used for this purpose. RNA soluble in molar salt, or nonsedimentable at 100000 × *g*, or exhibiting a sedimentation coefficient of 4 S, should not be termed sRNA.

[6] The IUPAC Commission on Nomenclature in Organic Chemistry prefers these symbols to the one-letter ones (N-3), designed for polymer representation. The three-letter symbols should be used whenever chemical changes involving nucleosides or nucleotides are being discussed.

Comment. For special purposes, the base and the sugar may be designated separately, using the base abbreviations of N-2.2 and the standard sugar abbreviations [1], *i.e.*, Rib, Ara, Glc, etc. Thus, adenosine = Ado = Ade-Rib; thymidine = dThd = Thy-dRib. (The "de" used in section 3.5 of *Abbreviations and Symbols* [1] for deoxy may be shortened to "d" in this context).

When abbreviations for single nucleosides are required and permitted, the above symbols should be used, *e.g.*, Urd (not UR, Ur or U) and dThd (not TdR, Tdr, TDR, T or dT), for uridine and thymidine, respectively[2].

N-2.4. *Nucleotides*

N-2.4.1. *Mononucleotides.* In the three-letter symbols, mononucleotides are normally expressed as phosphoric esters, such as Ado-3′-*P* or *P*-3′-Ado for adenosine 3′-phosphate, *P*-2′-Guo or Guo-2′-*P* for guanosine 2′-phosphate, Cyd-5′-*P* or *P*-5′-Cyd for cytidine 5′-phosphate (see N-2.4.4).

N-2.4.2. *Cyclic phosphodiesters* are designated by two primed numerals, one for each point of attachment, as in Cyd-2′:3′-*P* (or *P*-2′:3′-Cyd) or in Ado-3′:5′-*P* (or *P*-3′:5′-Ado). (The corresponding bisphosphates would be Cyd-2′,3′-P_2 and Ado-3′,5′-P_2.)

△ N-2.4.3. *Nucleoside diphosphate sugars*, which center about a pyrophosphate group, are represented by, *e.g.*, Urd-5′*PP*-Glc for uridine diphosphate glucose, *i.e.*, uridine 5′-(α-D-glucopyranosyl diphosphate), often termed UDPG or UDP-Glc (see N-2.4.4 and N-1.1).

△ N-2.4.4. *Points of attachment* in oligo- or polynucleotides are designated by primed numerals, *e.g.*, 2′*P*5′, 5′*P*5′, etc. as in Ado-2′*P*5′-rThd-2′*P* or Ado-5′*PP*5′-Nir (for NAD; see N-2.4.3 and N-1.2). The positional numerals may precede a series, as in (2′-5′)Ado-*P*-Guo-*P*-Urd-*P* to specify Ado-2′*P*5′-Guo-2′*P*5′-Urd-2′*P*. They may be omitted when the series in the left-to-right direction is 3′*P*5′.

Comment. Phosphate groups at the ends of chains may appear without numerals. In this case it is understood that *P*- at the left end means a 5′-phosphate, -*P* at the right means a free 3′-phosphate. Thus AMP can be represented by Ado-5′-*P*, *P*-5′-Ado, or *P*-Ado, but not by Ado-*P* (which would represent the 3′-phosphate).

N-3. ONE-LETTER SYMBOLS

N-3.1. PHOSPHORIC ACID RESIDUES

A monosubstituted (terminal) phosphoric residue is represented by a small p. A phosphoric diester (internal) in 3′-5′ linkage is represented by a *hyphen* when the sequence is *known*, or by a *comma* when the sequence is *unknown*. Unknown sequences adjacent to known sequences are placed in parentheses; these replace, at the points where they occur, the need for other punctuation. All these symbols thus replace the classical 3′-5′ or 3′p5′ symbols (*cf.* N-3.3.1 and N-3.3.2). A 2′:3′-cyclic phosphate residue may be indicated by > or >p.

Comments

i) The terminal p's should be specified unless their presence is unknown, in doubt, or of no significance to the argument.

ii) "Polarity" (direction other than 3′→ 5′) is dealt with in N-3.3.2.

iii) Linkages other than 3′ and 5′ are specified by other means (see N-3.3.1).

iv) A codon triplet, in which definite left-to-right order and 3′-5′ linkages are assumed and in which the termini are not of importance, may be written without punctuation as, *e.g.*, AGC.

N-3.2. NUCLEOSIDES

N-3.2.1. *Ribonucleosides*[2]

The *common ribonucleoside residues* (radicals) are designated by single capital letters, as follows:

A adenosine	T ribosylthymine (not thymidine)
G guanosine	C cytidine
I inosine	U uridine
X xanthosine	Ψ pseudouridine[7]
△ R unspecified purine nucleoside	Y unspecified pyrimidine nucleoside

N unspecified or unknown nucleoside (do not use X, P, or any of the above).

Rare Nucleosides. It is often advantageous, *e.g.*, in comparing long sequences, to represent every nucleoside residue by a single letter rather than by a group of letters and numbers. In such cases, those capital letters not assigned to common nucleosides (above) may be arbitrarily defined and used. It is recommended that the following be reserved for the substances listed (*cf.* N-4.4):

△ D 5,6-dihydrouridine	B 5-bromouridine
△ S thiouridine (for locants, see N-4.4)	O orotidine (see N-1.1)

Other symbols for these and for other modifications are listed in N-4.

Comments

i) The prefix r for ribo should be used when there is need for the additional specification.

ii) Other sugars or modified sugars are considered in N-3.2.2, N-3.2.3 and N-4.2.

N-3.2.2. *Deoxyribonucleosides*

The common 2′-deoxyribonucleosides are designated by the above symbols, modified in one of the following ways:

a) When space is available and no other prefixes are required, the *prefix* d is used; thus (i) dA-dG-dC ... or d(A-G-C ...); (ii) poly[d(G-C)] or poly(dG-dC) (these are identical substances); d may precede each residue or a whole chain, as applicable.

▲ b) When space is available but other, possibly confusing, prefixes are involved, a *subscript* d is used; thus, mmtT_d-bzA_d-T_d-anC_d for a protected tetradeoxynucleotide [4]. (The prefixes are defined in N-4.1.)

▲ N-3.2.3. *Unusual Sugar Residues*

Sugar moieties other than ribosyl or 2′-deoxyribosyl may be indicated as described in N-3.2.2 above, depending on requirements for base-modifying prefixes (N-4.1) and space available, using a, x and l (see N-2.3.2) for the other pentosyls, *ad hoc* letters for others, each defined; thus -aC- or -C_a- for an arabinosylcytosine residue. Symbols for substituents on sugars are given in N-4.2 (see also N-4.4).

N-3.3. OLIGO- AND POLYNUCLEOTIDES

N-3.3.1. *Points of Attachment*

The diesterified phosphate residue, represented by hyphen or comma or parenthesis (*cf.* N-3.1) is considered to be attached to the oxygen atom of the 3′ carbon on its left and to that of the 5′ carbon on its right. For other types of linkage, the simple hyphen must be replaced by its numerical form, as in 2′-5′ (or 2′p5′), 5′-5′, etc. [6], *e.g.*, G3′p5′A2′p5′A or G3′-5′A2′-5′A. These locants may precede a chain or a polymer if the internucleotide linkage is identical throughout, *e.g.*, (2′-5′)A-U-G-C for the corresponding tetranucleotide.

[7] Q may replace Ψ for computer work.

▲ N-3.3.2. *Direction of the Phosphodiester Link*

The hyphen used in known sequences is a contraction of the arrow (→) that is understood to point to the 5′ terminus of the phosphodiester bond (unless other numerals are used, as in N-3.3.1). When left-to-right direction is *not* the case, this must be indicated by an appropriate locant preceding the chain, or by an arrow to indicate the 3′→5′ direction, as in the peptide rules [9]. Thus, associated hydrogen bonded segments (see N-3.4.2) may be represented by, *e.g.*,

(3′-5′)A-C-A-C-A-C etc.
. . . . . .
(5′-3′)U-G-U-G-U-G etc.

or by

A→ C→ A→ C→ A→ C etc.
. . . . . .
U←G←U←G←U←G etc.

Another device used to represent "reverse polarity" is rotation of the symbols [10,11]. Thus the above associated polymers may be shown as

A-C-A-C-A-C etc.
. . . . . .
U-G-U-G-U-G etc.

In such representation, the left-to-right 3′-5′ convention is assumed to hold when the letters appear right-side up.

Examples of Oligonucleotides

A-G-Up (for ApGpUp); 3′→5′ trinucleotide, terminal 3′ phosphate.

A-G-U>p; the same, with terminal 2′:3′-cyclic phosphate.

pA-G-U; the same, commencing with a 5′ phosphate, terminating in an uridine with unsubstituted 2′ and 3′ hydroxyls.

△ pppG-G . . . Ap; this nucleotide (of unspecified length and sequence) has a 5′-triphosphate residue on the G at one (the 5′) end and a 3′-phosphate on the A at the other (the 3′) end.

pG-A-Ψ(C_2U)T-C-C-A; a decanucleotide, commencing (5′ end) with a 5′ phosphate, including a trinucleotide of unknown sequence between the Ψ and the T, and terminating (3′ end) in an adenosine residue with unsubstituted 2′ and 3′ hydroxyl groups.

d(pG-A-C-T); tetranucleotide (all deoxy), with 5′ terminal phosphate on G.

d(T←C←A←Gp); the same (arrow indicates 5′←3′ direction).

pG_d-A_d-C-T; the same, but with two deoxy, two ribo residues (see N-3.2.2b).

(2′-5′)pG-A-C-T; the same, all ribo, all in 2′-5′ linkage.

pG2′-5′A-C-T; the same, with a single 2′-5′ linkage (between G and A).

AGC; a codon (Note: The symbols for phosphoric acid residues may be omitted in describing codons. This is an exception to N-3.1).

N-3.4. POLYMERIZED NUCLEOTIDES

N-3.4.1. *Single Chains*

Polynucleotides composed of repeating sequences or of unknown sequence may be represented by either of two systems essentially identical with those devised and recommended by the IUPAC Commission on Nomenclature of Macromolecules and by the American Chemical Society's Polymer Nomenclature Commission (see also *Synthetic Polypeptides* [5]).

a) The repeating unit is preceded by "poly", meaning "polymer of". Thus, polynucleotide or poly(N); polyadenylate or poly(A); poly(adenylate-cytidylate) or poly(A-C) (alternating); poly(adenylate, cytidylate) or poly(A,C) (random).

b) The repeating unit, enclosed in parentheses if complex, is followed by a subscript denoting length, *e.g.* a number $(A\text{-}C)_{50}$, an average number $(A\text{-}C)_{\overline{50}}$ or a range $(A\text{-}C)_{40-60}$, if desired. Where the number of residues has not been determined and this form is required by the context, the subscript "n" may be used (as in ref. [5]). However, two n's should not appear in the same formula unless equal length is implied. When equal length is not the case, additional letters should be used, such as m, k, j, etc.

In either case, the symbols may carry prefixes or subscripts as required for proper specification. Note that "poly" is not used in the second system.

Examples

poly(A-U), alternating copolymer of A and U [12];

poly(A,U), random copolymer of A and U; *not* poly AU or poly A + U;

poly(A_2,U), as above but 2:1 in average composition;

$(A_2,U)_{\overline{50}}$, as above, average length of chain, 150 residues;

poly[d(A-T)] or poly(dA-dT), for alternating dA and dT[8] (see N-3.1 and [12]).

Comment. Multiple parentheses or brackets may be used for blocks within polymers, and vertical lines for side chains, etc. [5,9]. "Oligo" may replace "poly" where applicable. Terminal phosphate residues need not be specified unless they are essential to the argument.

N-3.4.2. *Association between Chains*

Association (noncovalent) between two or more polynucleotide chains, such as that ascribed to hydrogen-bonding, is indicated by the *center dot* (not the hyphen, which indicates covalent linkage), *e.g.* (*cf.* [12,14]):

a) poly(A) · poly(U), not poly(A · U), nor poly AU, nor poly A + U[9]; poly(A · U) may be used when it is implied that each A is paired with a U, regardless of chain lengths.

b) poly(A) · 2 poly(U) not poly(A · 2U), nor poly(A · U_2); poly(A · 2U) indicates the same triple-stranded complex and that each A is matched by two U's, regardless of individual chain lengths.

c) poly[d(A-T)] · poly[d(A-T)] or poly[d(A-T) · d(A-T)].

d) A · poly(U) or A · $(U)_n$ for single adenosine residues associated with polyuridylate or poly(uridylic acid).

Absence of association between chains is indicated by the *plus* sign (traditional in chemistry for coexisting but nonassociated species) *e.g.*:

a) poly(dC) + poly(dT), not poly(dC + dT);

b) poly(dA,rT_2) + poly(dG);

c) 2 [(poly(A) · poly(U)] ⇌ poly(A) · 2 poly(U) + poly(A) [12].

The *absence of definite information on association* is indicated by the *comma* (as before, indicating "unknown"), *e.g.*:

a) poly(A), poly(A,U);

b) poly[d(G-C)], poly[d(A,T)].

---

[8] Poly[d(A-T)] or poly(dA-dT) was originally [13] termed poly dAT. While this has the advantage of brevity, it has proven ambiguous (see footnote [9]) in other situations and is inconsistent with the general principles of polymer symbolism (*e.g.* [5]). Hence, its use is not recommended.

[9] Poly AU and AU, etc., have been used for poly(A) · poly(U) [15]. The similarity of this system for associated homopolymers to that originally proposed for alternating copolymers (see foot-note [8]) can lead to confusion, in that it indicates one covalent chain rather than two. Its use is not recommended. Similar potential confusion attends the use of the other incorrect terms given in N-3.4.2.

Comments

i) Hyphens are *not* used for association (noncovalent); poly(A-U) specifies a single chain, not two chains.

▲ ii) The *center dot* should always be used to indicate base pairs involved in noncovalent associations (see N-3.3.2), *e.g.*, A · T base pair, or G · C hydrogen bonds (not A-T, or G-C which indicate covalent linkages). The center dot is located as shown, *above* the line.

▲ iii) In describing *base ratios*, the form (A + T)/(G + C) should be used, not AT/GC, nor A + T/G + C. Two capital letters should not be juxtaposed (except as in N-3.1, comment iv), to distinguish sequence G-C, from content G + C, from ratio G:C or G/C, from base pair G · C.

▲ N-4. MODIFIED BASES, SUGARS, OR PHOSPHATES IN POLYNUCLEOTIDES

N-4.1. *Designation of Substituents on Bases*

In long sequences, as in transfer RNA's, where it is preferable to have not more than one capital letter per nucleoside residue, the standard symbols for nucleosides [*i.e.*, A, U, G, C, etc. (see N-3.2.1)] may be modified by a symbol of lower case letter(s) placed immediately before the single capital letter. Those symbols recommended for more common modifications are listed below (for locants and multipliers, see N-4.4; for unusual sugar residues, see N-3.2.2 and N-3.2.3:

m, e, ac	methyl, ethyl, acetyl
n, o	amino (N replaces H), deamino (O replaces N)
z, c	aza (N replaces C), deaza (C replaces N)
h	dihydro (hU = dihydrouridine; see also N-3.2.1 and N-4.4)
hm, ho (or oh)	hydroxymethyl, hydroxy
aa	aminoacyl
f	formyl (as in the conventional fMet for formylmethionyl)
fa	formylaminoacyl
i	isopentenyl (= γ,γ-dimethylallyl)
s	thio or mercapto (sU = thiouridine; see also N-3.2.1 and N-4.4)
fl, cl, br, io	fluoro, chloro, bromo, iodo (not encountered in natural polynucleotides; see also N-3.2.1 and N-4.4).

Symbols for some N-protecting radicals used in synthetic work [4, 9] are:

bz, bzl, tos	benzoyl, benzyl, tosyl
tr, an, bh	trityl, anisoyl, benzhydryl (diphenylmethyl)
mmt	monomethoxytrityl(*p*-anisyldiphenylmethyl)
dmt	dimethoxytrityl (di-*p*-anisylphenylmethyl)
thp, dns	tetrahydropyranyl, dansyl
cmc	*N*-cyclohexyl-*N'*[β-(4-methylmorpholino) amidino] (reaction product from the corresponding carbodiimide) [16].

In simpler situations where the avoidance of multiple capital letters in a single residue symbol seems not to be necessary, the standard chemical symbols (Me, Br, etc.) may be used. In such cases, no punctuation should appear between modifier and nucleoside symbol, *e.g.*, $6Me_2A$, 5BrU. The prefix "di" should not be used; subscripts numerals suffice (*cf.* N-4.4).

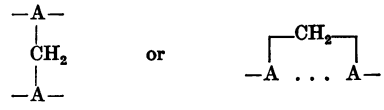
Comments

i) Symbols for other protecting groups may be constructed according to the principles indicated here and in Section 6 of *Amino Acids and Peptides* [9].

ii) When space is severely restricted, these symbols may appear above the nucleoside symbol (see N-4.4) [3, 4, 7, 8], *e.g.*,

$$\overset{ac}{C} \quad \text{for} \quad acC.$$

iii) Symbols for bifunctional adducts must lie above or below the chain (or chains) (see conventions for branched peptides in [5] and [9]) and hence may utilize any appropriate symbols. Thus a methylene bridge between two adenosines [17] could be represented as



for inter- or intra-chain linkage, respectively.

▲ N-4.2. *Designation of Substituents on Sugars*

N-4.2.1. *Internal Modifications.* The symbols are *lower case* when the modified sugar is internal; they are placed immediately *to the right* of the nucleoside symbol and indicate substitution at the (internal) 2′ position unless otherwise specified. Thus -Am- indicates a 2′-*O*-methyladenosine residue [7, 8] (see also N-4.4).

N-4.2.2. *Terminal Radicals.* The common, natural termini, phosphate and hydroxyl, are represented, if necessary, by p (N-3.1) and oh or ho (N-4.1); the latter is only required for emphasis as it is implied in the nucleotide symbol itself.

Other terminal radicals (hydroxyl-substituents) may utilize standard chemical symbols or abbreviations. These are placed in parentheses (following the appropriate nucleoside symbol, as noted above). Recommended abbreviations (aside from normal chemical symbols) are [4, 9]:

(EtOEt), (EtOMe)	1-ethoxyethyl, ethoxymethyl
(Ph_2CH), (Bzl), (Tr)	benzhydryl, benzyl, trityl
(MeOTr), [$(MeO)_2$Tr]	monomethoxytrityl, dimethoxytrityl
(Me), (Et), (Ac), (Tos)	methyl, ethyl, acetyl, tosyl
(Thp), F_3CCO-	tetrahydropyranyl, trifluoroacetyl
(AA), (Gly), (Leu), etc.	aminoacyl, glycyl, leucyl, etc.

Terminal glycol-protecting (bifunctional) radicals, bridging the 2′ and 3′ hydroxyls unless otherwise indicated, may require the following:

(>CMe_2)	isopropylidene; *e.g.*, -C-C-A(>CMe_2)
(>BOH), (>CO)	borate, carbonyl
>p or >	2′:3′-phosphate (cyclic) (*cf.* N-3.1)

▲ N-4.3. *Phosphoric Acid Protecting Groups*

Since these must be located at termini, standard chemical symbols should be used. These adjoin the appropriate hyphen (for phosphate; *cf.* N-3.1). Examples, in addition to any above [4]:

(CNEt)-; -(CNEt)	5′-cyanoethyl; 3′(or 2′)-cyanoethyl
(MeOPh), (Bzl), (Ph)	anisyl, benzyl, phenyl, with appropriate hyphen.

▲ N-4.4. *Locants and Multipliers*

Multipliers, when necessary, are indicated by the usual *sub*scripts [3, 8, 11]; thus -m_2A- signifies a dimethyladenosine residue, neither methyl being at the 2′-*O* position (see N-4.2.1). Locants are indicated by *super*scripts; thus -m^1_2A- indicates an N^6-dimethyladenosine residue [ribosyl-6-(dimethylamino) purine], -ac^4C indicates an N^4-acetylcytidine, -$m^{1,6}_2$A- or m^1m^6A a 1,N^6-dimethyladenosine, etc. [3, 8, 11]. Utilizing the convention of N-4.2.1, we can write -m^1_2Am- for the 2′-*O*-methyl-N^6-dimethyladenosine residue. Other examples are s^2U for 2-thiouridine and $h^{5,6}_2$U for 5,6-dihydrouridine (but see the alternates available in N-3.2.1 and N-4.1, namely ^{2}S, and hU or D, respectively; the locants and/or multipliers may be included in the definition). The prefix "di", which has no place in chemical symbolism, should not be used; subscript

numerals suffice. The prefix 2′-*O*-Me is best replaced by the suffix m (see N-4.2.1), especially when other substituents must be placed before the nucleoside symbol. Thus 2′*O*Me6Me$_2$A is better symbolized as m$^6_6$Am; similarly, 2MeS6iPeA becomes ms^{2}i^{6}A.

In presenting several homologous sequences, it is often desired to keep the capital letters representing nucleotides one below another. The presence of modifying symbols may interfere with such a presentation. One way of meeting this situation is to place the *prefixes* (including locants and multipliers) directly *over* the capital letter they modify, and to place the *suffix* (usually m for 2′-*O*-methyl) as a right-hand superscript (see also comment ii in N-4.1), *e.g.*, A^m_m; C^m.

Examples of this usage exist [3,7,8]. When so placed, smaller letters and/or numbers may be used to advantage [4,8]. Such positioning is consistent with the rules regarding designation of functional groups and their substituents in peptides [5,9].

REFERENCES

1. *Eur. J. Biochem.* 1 (1967) 259, and elsewhere. Section 5 appeared in *Biochim. Biophys. Acta*, 108 (1965) 1.
2. Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., and Zamir, A., *Science*, 147 (1965) 1462.
3. Holley, R. W., *Progr. Nucl. Acid. Res. Mol. Biol.* 8 (1968) 37.
4. Kössel, H., Büchi, H., and Khorana, H. G., *J. Amer. Chem. Soc.* 89 (1967) 2185.
5. *Eur. J. Biochem.* 3 (1967) 129, and elsewhere.
6. Richards, G. M., Tutas, D. J., Wechter, W. J., and Laskowski, M., Sr., *Biochemistry*, 6 (1967) 2908.
7. Woese, C. R., *Progr. Nucl. Acid. Res. Mol. Biol.* 7 (1967) 107.
8. *Handbook of Biochemistry* (edited by H. A. Sober), Chemical Rubber Co., Cleveland, Ohio, second edition 1970.
9. *Eur. J. Biochem.* 1 (1967) 375, and elsewhere. Revision in preparation.
10. Zachau, H. G., Dütting, D., and Feldmann, H., *Hoppe-Seyler's Z. Physiol. Chem.* 347 (1966) 212; *Angew. Chem.* 78 (1966) 392; *Angew. Chem. Int. Ed. Engl.* 5 (1966) 422.
11. Harada, F., Kimura, F., and Nishimura, S., *Biochim. Biophys. Acta*, 195 (1969) 590.
12. Michelson, A. M., Massoulié, J., and Guschlbauer, W., *Progr. Nucl. Acid Res. Mol. Biol.* 6 (1966) 83.
13. Inman, R. B., and Baldwin, R. L., *J. Mol. Biol.* 5 (1962) 172.
14. Ts'o, P. O. P., Rapoport, S. A., and Bollum, F. J., *Biochemistry*, 5 (1966) 4153.
15. Felsenfeld, G., and Miles, H. T., *Annu. Rev. Biochem.* 36 (1967) 407.
16. Ho, N. W. Y., and Gilham, P. T., *Biochemistry*, 6 (1967) 3632.
17. Feldman, M. Ya, *Biochim. Biophys. Acta*, 149 (1967) 20.

Acknowledgement. We are indebted to the Publishers and Editors of the *European Journal of Biochemistry* for permission to reproduce these Recommendations photographically.

All Tentative Rules of the IUPAC–IUB Commission on Biochemical Nomenclature (CBN) as well as a Document for Discussion are available from Waldo E. Cohn, Director, NAS–NRC Office of Biochemical Nomenclature, Oak Ridge National Laboratory, P.O. Box Y, Oak Ridge, Tenn. 37830, U.S.A.:

Abbreviations and Symbols for Chemical Names of Special Interest in Biological Chemistry. Revised Tentative Rules [see *Biochem. J.* (1966) **101**, 1].

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. Tentative Rules [see *Biochem. J.* (1967) **102**, 15].

Abbreviated Designation of Amino Acid Derivatives and Peptides. Tentative Rules [see *Biochem. J.* (1967) **102**, 23].

Rules of Naming Synthetic Modifications of Natural Peptides. Tentative Rules [see *Biochem. J.* (1967) **104**, 17].

The Nomenclature of Lipids. A Document for Discussion [see *Biochem. J.* (1967) **105**, 897].

Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino Acids) [see *Biochem. J.* (1968) **106**, 577].

The Nomenclature of Cyclitols. Tentative Rules [see *Biochem. J.* (1969) **112**, 17].

A One-Letter Notation for Amino Acid Sequences. Tentative Rules [see *Biochem. J.* (1969) **113**, 1].

Revised Tentative Rules for Nomenclature of Steroids [see *Biochem. J.* (1969) **113**, 5].

Nomenclature for vitamins B_6 and related compounds. Tentative Rules [see *Biochem. J.* (1970) **119**, 1].

Abbreviations and symbols for nuleic acids, polynucleotides and their constituents. Recommendations (this document).

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. Docum.* (1967) **7**, 72].