PROGRESS IN
Nucleic Acid Research and Molecular Biology

edited by

WALDO E. COHN

Volume 22

PROGRESS IN
Nucleic Acid Research and Molecular Biology

edited by

WALDO E. COHN
Biology Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee

Volume 22

1979

ACADEMIC PRESS
New York San Francisco London
A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 63–15847

ISBN 0–12–540022–5

PRINTED IN THE UNITED STATES OF AMERICA

79 80 81 82 9 8 7 6 5 4 3 2 1

List of Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

FRIEDRICH CRAMER (1), *Max-Planck-Institut für experimentelle Medizin, Abteilung Chemie, Hermann-Rein-Strasse 3, D-3400 Göttingen, Germany*

JAMES E. DARNELL, JR. (327), *The Rockefeller University, New York, New York 10021*

GEORGE JOHNSON (293), *Department of Biology, Washington University, St. Louis, Missouri 63130*

TOMAS LINDAHL (135), *Department of Medical Chemistry, University of Gothenburg, 400 33, Gothenburg, Sweden*

J. J. ROBERTS (71), *Institute of Cancer Research, Royal Cancer Hospital, Pollards Wood Research Station, Nightingales Lane, Chalfont Street, Giles, Bucks., United Kingdom*

MATHIAS SPRINZL (1), *Max-Planck-Institut für experimentelle Medizin, Abteilung Chemie, Hermann-Rein-Strasse 3, D-3400 Göttingen, Germany*

ROBERT J. SUHADOLNIK (193), *Department of Biochemistry, Temple University School of Medicine, Philadelphia, Pennsylvania 19140*

A. J. THOMSON (71), *School of Chemical Sciences, University of East Anglia, Norwich, Norfolk, United Kingdom*

Abbreviations and Symbols

All contributors to this Series are asked to use the terminology (abbreviations and symbols) recommended by the IUPAC–IUB Commission on Biochemical Nomenclature (CBN) and approved by IUPAC and IUB, and the Editor endeavors to assure conformity. These Recommendations have been published in many journals *(1, 2)* and compendia *(3)* in four languages and are available in reprint form from the Office of Biochemical Nomenclature (OBN), as stated in each publication, and are therefore considered to be generally known. Those used in nucleic acid work, originally set out in section 5 of the first Recommendations *(1)* and subsequently revised and expanded *(2, 3)*, are given in condensed form (I–V) below for the convenience of the reader. Authors may use them without definition, when necessary.

I. Bases, Nucleosides, Mononucleotides

1. *Bases* (in tables, figures, equations, or chromatograms) are symbolized by Ade, Gua, Hyp, Xan, Cyt, Thy, Oro, Ura; Pur = any purine, Pyr = any pyrimidine, Base = any base. The prefixes S–, H_2, F–, Br, Me, etc., may be used for modifications of these.

2. *Ribonucleosides* (in tables, figures, equations, or chromatograms) are symbolized, in the same order, by Ado, Guo, Ino, Xao, Cyd, Thd, Ord, Urd (Ψrd), Puo, Pyd, Nuc. Modifications may be expressed as indicated in (1) above. Sugar residues may be specified by the prefixes r (optional), d (=deoxyribo), a, x, l, etc., to these, or by two- or three-letter symbols, as in Ara-Cyt (for aCyd) or dRib-Ade (for dAdo).

3. *Mono-, di-, and triphosphates of nucleosides* (5′) are designated by NMP, NDP, NTP. The N (for "nucleoside") may be replaced by any one of the nucleoside symbols given in II-1 below. 2′-, 3′-, and 5′- are used as prefixes when necessary. The prefix d signifies "deoxy." [Alternatively, nucleotides may be expressed by attaching P to the symbols in (2) above. Thus: P-Ado = AMP; Ado-P = 3′-AMP.] cNMP′-cyclic 3′ : 5′-NMP; Bt_2cAMP = dibutyryl cAMP, etc.

II. Oligonucleotides and Polynucleotides

1. Ribonucleoside Residues

(a) Common: A, G, I, X, C, T, O, U, Ψ, R, Y, N (in the order of I-2 above).

(b) Base-modified: sI or M for thioinosine = 6-mercaptopurine ribonucleoside; sU or S for thiouridine; brU or B for 5-bromouridine; hU or D for 5,6-dihydrouridine; 1 for isopentenyl; f for formyl. Other modifications are similarly indicated by appropriate *lower-case* prefixes (in contrast to I-1 above) *(2, 3)*.

(c) Sugar-modified: prefixes are d, a, x, or l as in I-2 above, alternatively, by *italics* or boldface type (with definition) unless the entire chain is specified by an appropriate prefix. The 2′-*O*-methyl group is indicated by *suffix* m (e.g., -Am- for 2′-*O*-methyladenosine, but -mA- for 6-methyladenosine).

(d) Locants and multipliers, when necessary, are indicated by superscripts and subscripts, respectively, e.g., -m^6_2A- = 6-dimethyladenosine; -s^4U- or -^{4}S- = 4-thiouridine; -ac^{4}Cm- = 2′-*O*-methyl-4-acetylcytidine.

(e) When space is limited, as in two-dimensional arrays or in aligning homologous sequences, the prefixes may be placed *over the capital letter,* the suffixes *over the phosphodiester symbol.*

2. Phosphoric Acid Residues [left side = 5′, right side = 3′ (or 2′)]

(a) Terminal: p; e.g., pppN . . . is a polynucleotide with a 5′-triphosphate at one end; Ap is adenosine 3′-phosphate; C > p is cytidine 2′ : 3′-cyclic phosphate *(1, 2, 3)*; p < A is adenosine 3′ : 5′-cyclic phosphate.

(b) Internal: hyphen (for known sequence), comma (for unknown sequence); unknown sequences are enclosed in parentheses. E.g., pA-G-A-C(C_2,A,U)A-U-G-C > p is a sequence with a (5′) phosphate at one end, a 2′ : 3′-cyclic phosphate at the other, and a tetranucleotide of unknown sequence in the middle. **(Only codon triplets are written without some punctuation separating the residues.)**

3. Polarity, or Direction of Chain

The symbol for the phosphodiester group (whether hyphen or comma or parentheses, as in 2b) represents a 3′-5′ link (i.e., a 5′ . . . 3′ chain) unless otherwise indicated by appropriate numbers. "Reverse polarity" (a chain proceeding from a 3′ terminus at left to a 5′ terminus at right) may be shown by numerals or by right-to-left arrows. Polarity in any direction, as in a two-dimensional array, may be shown by appropriate rotation of the (capital) letters so that 5′ is at left, 3′ at right when the letter is viewed right-side-up.

4. Synthetic Polymers

The complete name or the appropriate group of symbols (see II-1 above) of the repeating unit, enclosed in parentheses if complex or a symbol, is either (a) preceded by "poly," or (b) followed by a subscript "n" or appropriate number. No space follows "poly" *(2, 5)*.

The conventions of II-2b are used to specify known or unknown (random) sequence, e.g.,

polyadenylate = poly(A) or A_n, a simple homopolymer;

poly(3 adenylate, 2 cytidylate) = poly(A_3C_2) or $(A_3,C_2)_n$, an *irregular* copolymer of A and C in 3 : 2 proportions;

poly(deoxyadenylate-deoxythymidylate) = poly[d(A-T)] or poly (dA-dT) or $(dA\text{-}dT)_n$ or $d(A\text{-}T)_n$, an *alternating* copolymer of dA and dT;

poly(adenylate,guanylate,cytidylate,uridylate) = poly(A,G,C,U) or $(A,G,C,U)_n$, a random assortment of A, G, C, and U residues, proportions unspecified.

The prefix copoly or oligo may replace poly, if desired. The subscript "n" may be replaced by numerals indicating actual size, e.g., $A_n \cdot dT_{12-18}$.

III. Association of Polynucleotide Chains

1. *Associated* (e.g., H-bonded) chains, or bases within chains, are indicated by a *center dot* (not a hyphen or a plus sign) separating the *complete* names or symbols, e.g.:

poly(A) · poly(U) or $A_n \cdot U_m$

poly(A) · 2 poly (U) or $A_n \cdot 2U_m$

△ poly(dA-dC) · poly(dG-dT) or $(dA\text{-}dC)_n \cdot (dG\text{-}dT)_m$.

2. *Nonassociated* chains are separated by the plus sign, e.g.:

$2[poly(A) \cdot poly(U)] \rightarrow poly(A) \cdot 2\ poly(U) + poly(A)$

or $2[A_n \cdot U_m] \rightarrow A_n \cdot 2U_m + A_n$.

3. Unspecified or unknown association is expressed by a comma (again meaning "unknown") between the completely specified chains.

Note: In all cases, each chain is completely specified in one or the other of the two systems described in II-4 above.

IV. Natural Nucleic Acids

RNA	ribonucleic acid or ribonucleate
DNA	deoxyribonucleic acid or deoxyribonucleate
mRNA; rRNA; nRNA	messenger RNA; ribosomal RNA; nuclear RNA
hnRNA	heterogeneous nuclear RNA
D-RNA; cRNA	"DNA-like" RNA; complementary RNA
mtDNA	mitochondrial DNA
tRNA	transfer (or acceptor or amino-acid-accepting) RNA; replaces sRNA, which is not to be used for any purpose
aminoacyl-tRNA	"charged" tRNA (i.e., tRNA's carrying aminoacyl residues); may be abbreviated to AA-tRNA
alanine tRNA or $tRNA^{Ala}$, etc.	tRNA normally capable of accepting alanine, to form alanyl-tRNA
alanyl-tRNA or alanyl-$tRNA^{Ala}$	The same, with alanyl residue covalently attached. [*Note:* fMet = formylmethionyl; hence $tRNA^{fMet}$, identical with $tRNA_f^{Met}$]

Isoacceptors are indicated by appropriate subscripts, i.e., $tRNA_1^{Ala}$, $tRNA_2^{Ala}$, etc.

V. Miscellaneous Abbreviations

P_i, PP_i	inorganic orthophosphate, pyrophosphate
RNase, DNase	ribonuclease, deoxyribonuclease
t_m (not T_m)	melting temperature (°C)

Others listed in Table II of Reference 1 may also be used without definition. No others, with or without definition, are used unless, in the opinion of the editor, they increase the ease of reading.

Enzymes

In naming enzymes, the 1972 recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature (CBN) *(4)*, are followed as far as possible. At first mention, each enzyme is described *either* by its systematic name *or* by the equation for the reaction catalyzed *or* by the recommended trivial name, followed by its EC number in parentheses. Thereafter, a trivial name may be used. Enzyme names are not to be abbreviated except when the substrate has an approved abbreviation (e.g., ATPase, but not LDH, is acceptable).

References*

1. *JBC* **241**, 527 (1966); *Bchem* **5**, 1445 (1966); *BJ* **101**, 1 (1966); *ABB* **115**, 1 (1966), **129**, 1 (1969); and elsewhere.†
2. *EJB* **15**, 203 (1970); *JBC* **245**, 5171 (1970); *JMB* **55**, 299 (1971); and elsewhere.†
3. "Handbook of Biochemistry" (G. Fasman, ed.), 3rd ed. Chemical Rubber Co., Cleveland, Ohio, 1970, 1975, Nucleic Acids, Vols. I and II, pp. 3–59.
4. "Enzyme Nomenclature," Elsevier Scientific Publ. Co., Amsterdam, 1973, and Supplement No. 1, *BBA* **429**, (1976).

* Contractions for names of journals follow.

† Reprints of all CBN Recommendations are available from the Office of Biochemical Nomenclature (W. E. Cohn, Director), Biology Division, Oak Ridge National Laboratory, Box Y, Oak Ridge, Tennessee 37830, USA.

5. "Nomenclature of Synthetic Polypeptides," *JBC* **247**, 323 (1972); *Biopolymers* **11**, 321 (1972); and elsewhere.*

Abbreviations of Journal Titles

Journals	*Abbreviations used*
Annu. Rev. Biochem.	ARB
Arch. Biochem. Biophys.	ABB
Biochem. Biophys. Res. Commun.	BBRC
Biochemistry	Bchem
Biochem. J.	Bj
Biochim. Biophys. Acta	BBA
Cold Spring Harbor Symp. Quant. Biol.	CSHSQB
Eur. J. Biochem.	EJB
Fed. Proc.	FP
J. Amer. Chem. Soc.	JACS
J. Bacteriol.	J. Bact.
J. Biol. Chem.	JBC
J. Chem. Soc.	JCS
J. Mol. Biol.	JMB
Nature, New Biology	Nature NB
Nucleic Acid Research	NARes
Proc. Nat. Acad. Sci. U.S.	PNAS
Proc. Soc. Exp. Biol. Med.	PSEBM
Progr. Nucl. Acid Res. Mol. Biol.	This Series

* Reprints of all CBN Recommendations are available from the Office of Biochemical Nomenclature (W. E. Cohn, Director), Biology Division, Oak Ridge National Laboratory, Box Y, Oak Ridge, Tennessee 37830, USA.

Some Articles Planned for Future Volumes

Chromatin Structure and Function
P. CHAMBON

Functional Aspects of the Interaction of Chemical Carcinogens with Nucleic Acids
D. GRUNBERGER AND I. B. WEINSTEIN

Ribonucleotide Reductase
F. D. HAMILTON

Mechanism of Interferon Action
P. LENGYEL AND G. SEN

Mitochondrial Nucleic Acids of Yeast
M. RABINOWITZ AND J. LOCKER

Ribosome Turnover in Eukaryotic Cells and Tissues
J. F. SCOTT

Patterns of Nucleic Acid Synthesis in *Physarum polycephalum*
G. TURNOCK

Structure and Functions of Ribosomal RNA
R. ZIMMERMANN

Contents

Naturally Occurring Nucleoside and Nucleotide Antibiotics

Robert J. Suhadolnik

Genetically Controlled Variation in the Shapes of Enzymes

George Johnson

Transcription Units for mRNA Production in Eukaryotic Cells and Their DNA Viruses

James E. Darnell, Jr.

The -C-C-A End of tRNA and Its Role in Protein Biosynthesis

MATHIAS SPRINZL AND FRIEDRICH CRAMER

Max-Planck-Institut für experimentelle Medizin Abteilung Chemie Göttingen, Germany

Progress in Nucleic Acid Research and Molecular Biology, Vol. 22

ISBN 0-12-540022-5

I. Introduction[1]

The adaptor role of tRNA requires that it have at least two functionally different sites—one for the specific interaction with the messenger, and one for accepting and transferring the particular amino acid to the growing polypeptide chain. The codon recognition mechanism of tRNA is explained in principle by the codon–anticodon interaction, including the wobble hypothesis (*1*), and has been confirmed by sequencing many tRNAs. The second function is much less clearly understood. Apart from the long-known fact that the -N-C-C-A end, common to all tRNAs, carries the amino acid in ester linkage (*2*), the problem of the specific attachment of an amino acid to its corresponding tRNA (the "recognition" problem) is still not solved (*3*). Clearly, the -N-C-C-A end is not responsible for the specificity of the attachment; there must be other features in the molecule or in the enzymic mechanism of aminoacylation responsible for this high degree of specificity. However, these characteristics—whatever they are—must be seen in relation to the -N-C-C-A end, since this part of the tRNA carries the amino acid. Thus the presence of an intact -N-C-C-A end is a general prerequisite for the aminoacylation and for amino-acid transfer. In this article, we discuss the structural and the chemical requirements for the functioning of the -N-C-C-A terminus. [In this connection, it should be mentioned that some viral RNAs can be aminoacylated. They also carry an -N-C-C-A end (*4*).[2]]

In the "cloverleaf" structure, the -C-C-A end always remains unpaired. The nature of the nucleotide prior to the -C-C-A terminus has been discussed in connection with a preselecting code that would

[1] *Abbreviations:*

$tRNA^{Phe}$ = $tRNA^{Phe}$-A-C-C-A = phenylalanine transfer RNA

Phe-$tRNA^{Phe}$-A-C-C-A = phenylalanyl-$tRNA^{Phe}$

$tRNA^{Phe}$-A-C-s^2C-A, $tRNA^{Phe}$-A-C-i^5C-A, $tRNA^{Phe}$-A-C-C-F, etc., are $tRNA^{Phe}$ species containing, respectively, 2-thiocytidine, 5-iodocytidine, formycin, etc.

$tRNA^{Phe}$-A-C-C-dA, $tRNA^{Phe}$-A-C-C-A(2′NH_2), etc., are $tRNA^{Phe}$ species containing, respectively, 2′-deoxyadenosine, 2′-amino-2′-deoxyadenosine, etc.

$tRNA^{Phe}$-A-C-C-A_{oxi} and $tRNA^{Phe}$-A-C-C-$A_{oxi-red}$ represent $tRNA^{Phe}$ after periodate oxidation, and after subsequent borohydride reduction, respectively

tRNA-N-C-C-A and corresponding formulas represent unfractionated mixtures of tRNAs

Aminoacyl-oligonucleotide nomenclature is as follows: C-A-Phe indicates C-A bearing a phenylalanine residue on the 2′ or 3′ hydroxyl group;

C-A(2′Phe)H and C-A(2′H) Phe indicate C-A bearing phenylalanine residues at the 2′- and 3′-position, with hydrogen in positions 3′ and 2′, respectively.

[2] See article by Waters and Mullin in Vol. 20 of this series [Ed.].

govern the lipophilic or hydrophilic character of the amino acid to be attached (5). However, this hypothesis no longer seems tenable in view of the many new sequences that have been obtained during recent years. Although the -N-C-C-A end does not seem to be involved in the secondary (or tertiary) base-pairing, it might nevertheless exist in a defined physical conformation.

The -N-C-C-A end does not merely serve as a spacer between the body of the tRNA molecule and the amino-acid residue, providing a sufficiently exposed position for the amino acid during the interaction with the aminoacyl-tRNA synthetase and on the ribosomal sites. Rather, it plays some highly specific roles: in the interaction with aminoacyl-tRNA synthetase, in the selection mechanism for the correct amino acid, and in the specific attachment to the ribosome.

An earlier review in this series by Deutscher (6) on the acceptor end of tRNA emphasized mainly the biosynthesis of the -N-C-C-A end and the function of the ATP(CTP) : tRNA nucleotidyltransferase. The present article summarizes recent results on the function of the -N-C-C-A end of tRNA as a reactive site during aminoacylation and ribosomal protein biosynthesis, with information about the structure of -N-C-C-A end and changes in its function after its modification.

II. Structure of the -N-C-C-A Terminus

All tRNAs sequenced to date can be presented in a "cloverleaf" structure in which the 3'-terminal -N-C-C-A is not base-paired. The general existence of such a structure is evident from a comparison of the sequences of different tRNAs (7) and has been proved directly by high-resolution nuclear magnetic resonance (NMR) studies (8). An important feature of this structure is the invariant number of nucleotides between the T-Ψ-C sequence and the terminal adenosine: the terminal adenosine is always separated from the first nucleoside in the T-Ψ-C loop (ribothymidine or uridine) by 21 nucleosides (7). This invariance underlines the functional importance of the single-stranded T-Ψ-C loop and the 3'-terminal sequence (9).

Several chemical and physicochemical studies seemed to indicate that the -N-C-C-A terminus of the tRNA in solution may be involved in tertiary interactions within the tRNA molecule (3). In view of the results that have accumulated in recent years, these interpretations are most probably not valid. However, the structure of the 3' terminus in free and aminoacylated tRNA with respect to stability, stacking interactions, and conformations still remains to be determined.

A. X-Ray Diffraction Studies

Successful interpretation of X-ray diffraction data of orthorhombic (*10*) and monoclinic (*11,12*) crystals of $tRNA^{Phe}$ from yeast up to a 2.5 Å resolution led to the elucidation of a three-dimensional crystal structure of this tRNA (for recent reviews, see *13–15*). Atomic coordinates and torsional angles have been published for the corresponding three-dimensional models (*16, 17*). The -A-C-C-A terminus in these models is attached as a single strand to a long double-stranded helical region formed by the base pairs of the T-Ψ-C and the aminoacyl stems. There are no additional tertiary interactions in which the -A-C-C-A end is involved. Unfortunately, the data concerning this part of the molecule are not sufficiently satisfactory to allow an unambiguous interpretation of the conformation of the -A-C-C-A end in the crystals (*13*). This could be due to a conformational perturbation of this single-stranded region or by an inhomogeneity of the terminal sequence with respect to the presence of the terminal adenosine. In the orthorhombic crystals, the 3′-terminal residues are helically stacked onto the aminoacyl stem except for the terminal adenosine-76, which is not stacked with respect to the penultimate cytidine-75 (*16*). In the monoclinic cell, the C_{75} and A_{76} residues are in a more extended form (*17*).

B. Chemical Modification

One of the earliest approaches to elucidation of the structure of tRNA in solution was its chemical modification by base-specific reagents (*18–37*), followed by sequence analysis of the product. Using this approach, the -N-C-C-A terminus and some other single-stranded regions of tRNA showed a high reactivity toward certain reagents. This was demonstrated by the reaction of tRNA with bisulfite (*18–26*), monoperphthalic acid (*27, 28*), methoxylamine (*29–32, 35*), sodium periodate (*33, 34*), and hydroxylamine (*36*), and by tritium exchange experiments (*37*).

Treatment of yeast $tRNA^{Phe}$ by monoperphthalic acid rapidly converts the terminal adenosine-76 to adenosine 1-oxide (*27*), but the unpaired adenosine-73 does not react. The same reaction has been used to modify yeast $tRNA^{Val}$ and $tRNA^{Phe}$ (*28*). From the rates of the reaction of particular nucleosides at the 3′ end of both tRNAs, it appears that after the modification of the 3′-terminal adenosine there is N-oxidation of the two penultimate cytidine residues. When the three terminal bases of the -A-C-C-A end are finally oxidized, the fourth unpaired base becomes reactive. This sequential step-by-step oxida-

tion does not occur in single-stranded oligonucleotides, such as C-A-C-C-A or A-A-A-U-C-A-C-C-A.

Similar observations were made during the investigation of the rate of conversion of cytidine to uridine in yeast $tRNA^{Phe}$ (38) and *E. coli* $tRNA^{Gly}$ (23) by $NaHSO_3$. Out of six cytidine residues in the unpaired regions of yeast $tRNA^{Phe}$, only the two at the -A-C-C-A end were reactive. The relative rates of modification of each individual reactive residue in $tRNA^{Phe}$, compared to those of single nucleotides and of the random copolymer poly(C_1,U_9), are shown in Fig. 1. This demonstrates that the cytidine residues in the -A-C-C-A end, not being base-paired, react at a lower rate than does CMP or the random copolymer. This implies that the tRNA sites are involved in some kind of ordered structure that is probably determined by stacking interactions. Varying degrees of modification of the two cytidine residues were also observed in some other cases using bisulfite (*18*) or hydroxylamine (*36*), although in these investigations the rates of modification of particular residues were not determined.

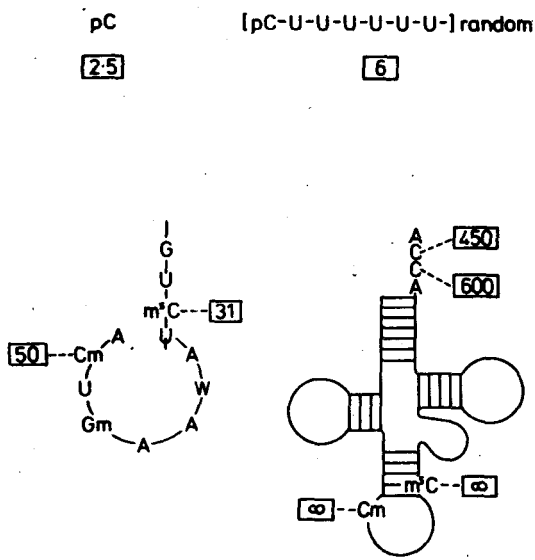

FIG. 1. Modification of cytidine with sodium bisulfite (38). The numbers in the boxes give the half-times (in hours) of the reaction converting cytidine to uridine for CMP, for the irregular copolymer poly(C_1,U_9), for the cytidine residues in a dodecanucleotide excised from the anticodon of $tRNA^{Phe}$ from yeast, and for the cytidine residues in intact yeast $tRNA^{Phe}$. Reaction conditions were as in Chambers *et al.* (*18*).