PROGRESS IN
Nucleic Acid Research and Molecular Biology

Volume 16

PROGRESS IN
Nucleic Acid Research and Molecular Biology

edited by

WALDO E. COHN

Biology Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee

Volume 16

（内部交流）

1976

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

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 63-15847

ISBN 0-12-540016-0

PRINTED IN THE UNITED STATES OF AMERICA

List of Contributors

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

E. I. BUDOWSKY (125), *M. M. Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, Moscow, USSR*

ERWIN CHARGAFF (1), *Cell Chemistry Laboratory, Department of Biochemistry, Columbia University, New York, New York*

L. EHRENBERG (189), *Wallenberg Laboratory, Stockholm University, Stockholm, Sweden*

I. FEDORCSÁK (189), *Department of Genetics, Eotvos University, Budapest, Hungary*

HIKOYA HAYATSU (75), *Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Tokyo, Japan*

JOHN D. SMITH (25), *Medical Research Council, Laboratory of Molecular Biology, Cambridge, England*

F. SOLYMOSY (189), *Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary*

Preface

In the best of all possible worlds, each volume of this serial publication would contain articles ranging over the whole family of fields and subfields encompassed by the title. Such wide-ranging volumes are indeed envisioned by the editor (as by every editor of a multi-authored book), but the perfect outcome planned is always subject to perturbation by human variabilities: articles that were to appear together are separated by unforeseen delays; subjects chosen at one time are altered by subsequent developments in the field; the emphasis in a given article may be other than what was planned; etc. Hence, the titles of the articles in a given volume, by themselves, may give an impression of a narrow, limited approach to the broad area blazoned on the cover.

Such might appear to be the case in this volume of *Progress in Nucleic Acid Research and Molecular Biology*. Three of the five articles appear, from their titles, to deal in rather limited fashion with organic chemistry rather than biology or biochemistry. But a closer inspection indicates that each of the three itself covers a wide range of molecular biology. Hayatsu (Tokyo) discusses not only the history and mechanism of the reaction of bisulfite with nucleic acids and their constituents, but also the many ways in which this reaction is employed to shed light on some of the mechanisms in the information-transfer chain. Budowsky (Moscow), who reviewed earlier work on hydroxylamine reactions with purine and pyrimidine materials in Volume 9 (as did Brown in Volume 7), deals with the important advances in our understanding of mutagenesis that have recently stemmed from this chemical approach; his discussion embraces such nonchemical subjects as point mutations and intracellular viruses. Likewise, in reviewing the manifold practical uses of diethyl pyrocarbonate in nucleic acid research, Ehrenberg (Sweden), Fedorcsák (Budapest), and Solymosy (Szeged) touch on most areas of contemporary molecular biology. They show how this reagent should be applied in various laboratory situations and indicate not only the pitfalls attending ill-considered use, but also the considerations necessary for maximum effectiveness. Thus all three "chemical" papers are essays of wide-ranging interest to molecular biologists.

Two articles explore the details of the biosynthesis of nucleic acids. Chargaff (New York) describes his experimental approach to

the problem of the initiation, by RNA primers, of the enzymic synthesis of DNA (his essay, "What Really Is DNA?," appeared in Volume 8; it might be considered a preface to the work described in the present article). Finally, John Smith (Cambridge) reviews a problem of considerable current interest: the transcription and subsequent maturation of transfer RNA. This involves consideration of such matters as the arrangement of the corresponding genetic material, the structure of the initial transcript, and the enzymic processing of the initial transcript or precursor.

As stated often in previous volumes, the ideas expressed are those of the authors. Editing is confined to matters of expression, to ensure maximum ease of comprehension and freedom from ambiguity. In this connection, readers should note the statement on Abbreviations and Symbols (pp. xi–xiv).

The comments and suggestions of readers are always welcome.

W.E.C.

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 Editors endeavor 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 NAS–NRC 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 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; i 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 *N*-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*).

(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 $(A_3C_2)_n$, a *random* 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.

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 $(A)_n\cdot 2(U)_m$

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.:

$2[\text{poly(A)}\cdot\text{poly(U)}] \xrightarrow{\Delta}$ poly(A)·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
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 editors, 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" (H. A. Sober, ed.), 2nd ed. Chemical Rubber Co., Cleveland, Ohio, 1970, Section A and pp. H130–133.
4. "Enzyme Nomenclature," Elsevier Scientific Publ. Co., Amsterdam, 1973.
5. "Nomenclature of Synthetic Polypeptides," *JBC* **247**, 323 (1972); *Biopolymers* **11**, 321 (1972); and elsewhere.†

° 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.

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
Proc. Nat. Acad. Sci. U.S.	PNAS
Proc. Soc. Exp. Biol. Med.	PSEBM
Progr. Nucl. Acid Res. Mol. Biol.	This Series

Some Articles Planned for Future Volumes

Mechanisms in Polypeptide Chain Elongation on Ribosomes
E. BERMEK

The Ribosome of *E. coli*
R. BRIMACOMBE, K. H. NIERHAUS, R. A. GARRETT, AND H. G. WITTMAN

Polyadenylate Polymerases
M. EDMONDS AND M. A. WINTERS

Structure and Function of 5 S and 5.8 S RNA
V. A. ERDMANN

Initiation of Protein Synthesis
M. GRUNBERG-MANAGO

Three-dimensional Structure of Transfer RNA
S. H. KIM

Interaction of Carcinogenic and Mutagenic *N*-Nitroso Compounds with Nucleic Acids
W. LIJINSKY

Ribosomal tRNA Binding Sites
H. MATTHAEI

Protein Synthesis
S. OCHOA

Antibiotics as Probes into Ribosome Function and Structure
S. PESTKA

Quantum Mechanical Investigation of the Electronic Structure of Nucleic Acids and Their Constituents
B. PULLMAN

The Biochemical and Microbiological Action of Platinum Compounds
A. J. THOMSON AND J. J. ROBERTS

Aliphatic Polyamines and the Regulation of Biosynthetic Reactions in Eukaryotes
H. G. WILLIAMS-ASHMAN AND A. CORTI

Structure and Functions of Ribosomal RNA
R. ZIMMERMANN

Contents

The Mechanism of the Mutagenic Action of Hydroxylamines

E. I. BUDOWSKY

Diethyl Pyrocarbonate in Nucleic Acid Research

L. EHRENBERG, I. FEDORCSÁK, AND F. SOLYMOSY

Initiation of Enzymic Synthesis of Deoxyribonucleic Acid by Ribonucleic Acid Primers

ERWIN CHARGAFF

Cell Chemistry Laboratory
Department of Biochemistry
Columbia University
New York, New York

I. Introduction

In 1954 the late J. N. Davidson and I wrote the Preface to Volume I of the treatise on nucleic acids (*1*) that we had edited together. It ended with the following words: "If this book helps to create an early need for a supplement to its present content, it will have fulfilled one of its purposes." Seldom has a wish been answered more promptly and more disastrously. It was almost as if the sorcerer's apprentice had invited the public to come and take a dip. Not that our book can be claimed to have contributed more than a minor share to the explosion of interest in the nucleic acids and their biological functions. But there can be little question that the publications in this field have multiplied at an unbelievable rate: what could be fitted comfortably into three volumes fifteen years ago would now require ten or more; and Procrustes would have to be hired as the editor.

A comprehensive and at the same time correct view of an entire discipline has, therefore, become almost impossible. It is necessary to restrict oneself to a tiny splinter, not without taking comfort from the fact that it is through an assiduous inspection of small parcels

that some understanding is often achieved in the study of that part of nature that is accessible to exploration by the natural sciences. This article will, hence, concern itself with only one facet of the great and largely unsolved problem of the synthesis of DNA in the cell, namely, the manner or the manners in which the initiation of this process, under conditions simulating those of biosynthesis, can be formulated.

That the mechanism of DNA replication in the living cell is a very important problem, both in biology and in biochemistry, requires little argument. But wherever in the first half of the preceding sentence a singular is used, we know very well that it ought to be a plural, and one comprising an enormous number of discrete varieties. If simplification has proved one of the sharpest and most useful tools in the exact sciences, it becomes a dangerous and destructive weapon when exercised on the awe-inspiring fabric of life. This is an old predicament, and I shall not discuss it here again. I merely wish to describe briefly a line of reasoning that led us to the suggestion that the synthesis of DNA requires the collaboration of two types of polymerases, namely, a DNA polymerase and an RNA polymerase.

II. The Basic Events

a. The DNA "Molecule." If a given DNA molecule really represents the genome of the cell, the task of the reproducing cell can be formulated succinctly. It is to replicate the gene complement faithfully and completely, so that each daughter cell is equipped with the entire and unchanged hereditary apparatus. It must, however, be pointed out that the apparent simplicity of these requirements is deceptive. First of all, if we consider the extremely elaborate ceremonial that attends the biosynthesis of a protein molecule, it is not unreasonable to assume that at least as many, and probably more, precautions accompany the synthesis of the genome itself. And, second, people brought up in chemistry may often not realize that in using such a term as "a DNA molecule" they are referring to a truly monstrous structure, for they are dealing here with chains or circles having masses ranging from 2×10^9 daltons in bacteria to 10^{11} daltons or bigger in eukaryotes. Only the much smaller DNA varieties occurring in bacteriophages are susceptible of chemical manipulation. I have discussed the conceptual problem of the DNA molecule some time ago (2); the cytological evidence has also been reviewed (3).

Considering that a DNA "molecule" of 10^{11} weight must contain

about 150 million nucleotides in each strand, we may conclude that we are confronted with a magnitude that transgresses chemical imagination. We are dealing here with an entire enormous landscape, as it were, in which many different events can take place simultaneously in different sites. The principal events can be listed. (1) Replication, i.e., the enzymic reproduction of the two original DNA strands, each acting as the *template*, giving rise to a new DNA strand complementary to the template strand by virtue of the base-pairing rules. (2) Transcription, i.e., the enzymic reproduction of a DNA template strand in the form of complementary RNA. (3) Denaturation, i.e., the rupture of all or of some of the hydrogen bonds holding the two halves of the DNA duplex together, leading to complete or to local strand separation. (4) Hydrolysis of the phosphodiester backbone by exo- and endonucleases some of which—especially the endonucleases—appear to possess truly remarkable specificities. (5) Chemical modification, e.g., a partial, and possibly nonrandom, depurination or depyrimidination; the enzymic introduction of methyl groups—again apparently a highly specific and directed process; the formation of dimers between adjacent pyrimidines, etc. (6) Repair: This can take several forms. There exist enzymic mechanisms that are able to recognize and to excise mismatched or otherwise damaged or incorrect tracts of a DNA strand. Other enzymes are capable of hydrolyzing ribonucleotide stretches that, via covalent links, terminate, or are interspersed in, one DNA strand and are hydrogen-bonded to the other. In all these cases, the resulting gaps must then be filled with the array of correct nucleotides, presumably with the aid of one of the replicating enzymes mentioned before. This must, finally, be followed by the action of an enzyme able to ligate two adjoining nucleotides by establishing a $5' \rightarrow 3'$ phosphodiester bridge between them, in order to heal a break in the DNA chain caused by the repair operations, to form a circle, once a linear DNA molecule has been completed, or to join shorter DNA precursor fragments. All these complex processes obviously must be hedged about with a multitude of checks and balances which we have not even begun to comprehend.

It is with a small segment of the first of the events listed above, namely, DNA replication, that this article deals; but, as will be seen, transcription also plays a role. The evidence to be considered will mostly be based on *in vitro* experiments; because, all things considered, we know really very little about the chemical mechanisms and the sequence of reactions comprising the synthesis of DNA under the conditions of life.

b. Primers and Templates. These designations are often used interchangeably in the literature. This is regrettable, since they stand for two entirely different concepts. To define the sense in which these terms are used here, I shall quote from a previous paper from this laboratory (*4*): "A *primer* is a poly- or, more frequently, an oligonucleotide that starts, or facilitates, the polymerization process by providing a free hydroxy group for the growing chain; it is incorporated into the product which in its composition does not have to reflect that of the primer. A *template* specifies the composition, and presumably the nucleotide sequence, of the product; it is an obligatory factor in the enzymic synthesis of sequentially specific polynucleotides."

All nucleic acid polymerases capable of effecting sequence-specific syntheses require a template; the DNA polymerases, in addition, require a primer. Under biological conditions, this priming is probably effected by ribo-oligonucleotides which are synthesized *in loco* by RNA polymerase. In the *in vitro* synthesis, either deoxyribo- or ribo-oligonucleotides can be used, with a chain length beginning at about 4 and reaching optimum efficiency at 8 to 12 (*5, 6*).

c. Nucleic Acid Polymerases. There exists an enormous number of publications on the enzymes capable of the sequence-specific synthesis of polynucleotides. I shall limit myself here to mentioning a few recent review articles on the enzymic synthesis of DNA (*7–11*) and of RNA (*12–15*). What all these enzymes have in common is that they require a polynucleotide, double- or single-stranded, as an obligatory template and that they function as nucleotidyltransferases, utilizing the nucleoside 5′-triphosphates as the precursors for the synthesis of a polynucleotide chain growing in the 5′ → 3′ direction. A cation, mostly magnesium, in some cases manganese, must also be supplied. The DNA polymerases are listed by the Enzyme Commission under the recommended name of DNA nucleotidyltransferases (EC 2.7.7.7), the RNA polymerases correspondingly as RNA nucleotidyltransferases (EC 2.7.7.6). Two other types of more recently discovered polymerases, which both utilize RNA as the template, will not concern us here; they are a DNA polymerase found in oncogenic viruses (*16*) and an RNA polymerase present in certain RNA-containing bacteriophages (*17*).

Whereas the *in vitro* mechanism of action of the several nucleic acid polymerases begins to be understood, this cannot be said at all of their functions *in vivo*, especially as regards the DNA polymerases. To begin with, the multitude of DNA polymerases encountered in the same cell is baffling: as primitive a system as *Escherichia*