PROGRESS IN

Nucleic Acid Research and Molecular Biology

Volume 14

PROGRESS IN
Nucleic Acid Research and Molecular Biology

edited by

WALDO E. COHN

Biology Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee

Volume 14

1974

ACADEMIC PRESS
New York and 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 0–12–540014–4

PRINTED IN THE UNITED STATES OF AMERICA

List of Contributors

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

WERNER ARBER (1), *Biozentrum, University of Basel, Basel, Switzerland*

MALCOLM L. GEFTER (101), *Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts*

GARY F. GERARD (187), *Institute for Molecular Virology, St. Louis University School of Medicine, St. Louis, Missouri*

MAURICE GREEN (187), *Institute for Molecular Virology, St. Louis University School of Medicine, St. Louis, Missouri*

KENNETH MURRAY (117), *Department of Molecular Biology, University of Edinburgh, Edinburgh, Scotland*

NIHAL K. NOTANI (39), *Biology and Agriculture Division, Bhabha Atomic Research Centre, Bombay, India*

ROBERT W. OLD (117), *Department of Molecular Biology, University of Edinburgh, Edinburgh, Scotland*

JANE K. SETLOW (39), *Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee*

Preface

The appearance of this volume marks the end of an era in this series of essays, for it is the first that has not had the benefit of the participation in its construction of one of its two founding Editors. The untimely death, in 1972, of Professor J. N. Davidson, whose many contributions to this area of science were set out so eloquently by his colleague, Professor R. M. S. Smellie, in Volume 13, is felt deeply by his coeditor, upon whom has fallen the responsibility of continuing the series in the manner developed by both Editors during the eleven years of its existence. Hopefully, the traditions established earlier and carried on by the contributors, who now number 167, will allow this responsibility to be discharged without noticeable changes in scope, content, or style.

It may be well, at this transition point, to repeat some of the ideas advanced in the Preface to Volume 1, ideas that were repeated in the succeeding volumes and that are still effective.

"We do not wish . . . an annual or fixed-date publication . . . or a bibliographic review or literature survey . . . [but] rather to encourage . . . 'essays in circumscribed areas' . . . by workers provided with an opportunity for more personal interpretation than is normally provided in review articles.

"While we expect each author to cover his particular field of interest and . . . closely related work of others as well, we encourage . . . discussion and speculation and the expression of points of view that may be controversial and certainly individualistic. It is to be expected that . . . authors will interpret this charge in different ways, some essaying a broad, philosophical vein, some developing or describing new theories or techniques, some assembling a number of fragmentary observations into a coherent pattern, and some reviewing a field in a more conventional manner. We do not attempt to define or restrict an author's approach . . . and confine our editing to ensuring maximal clarity to the reader . . . himself active in or concerned with the general field . . . ."

In connection with "editing to ensure maximum clarity to the reader," we have always been aware of the necessity for adhering to a common scientific language. Hence we have asked contributors to adhere to the spirit and, as far as possible, the letter of the nomenclatural recommendations of the relevant international commissions (Biochemical and Organic) and to the general policies of the leading journals with respect to the use of abbreviations. For the convenience of our readers, these are summarized in the following pages, together

with a set of contractions of the names of the most often quoted journals, introduced in the interests of conserving space and reducing costs.

I conclude with the older statement that "we seek to provide a forum for discussion . . . and we welcome suggestions from readers as to how this end may best be served," be it in the choice of subjects, of authors, or of style.

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.

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 parenthesis, 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 $(A_3,C_2)_n$, 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.:

$2[\text{poly(A)}{\cdot}\text{poly(U)}] \triangleq \text{poly(A)}{\cdot}2\ \text{poly(U)} + \text{poly(A)}$ (II-4a)

or $2[A_n{\cdot}U_m] \triangleq A_n{\cdot}2U_m + A_n$. (II-4b)

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

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 $tRNA^{fMet}$ or $tRNA_f^{Met}$]

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. 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

Some Articles Planned for Future Volumes

Mechanisms in Polypeptide Chain Elongation on Ribosomes
E. BERMEK AND H. MATTHAEI

Primary Structure of Ribosomal RNA
P. FELLNER

Bacterial Ribosomal Proteins
R. A. GARRETT, K. NIERHAUS, AND H. G. WITTMAN

Initiation of Protein Synthesis
M. GRUNBERG-MANAGO AND F. GROS

Immunogenic Polynucleotides
L. D. HAMILTON

X-Ray Diffraction Studies of Nucleic Acids and Their Components
R. LANGRIDGE, E. SUBRAMANIAN, AND P. J. BOND

Chemistry of Alkylation and Its Relationship to Mutagenesis and Carcinogenesis
B. SINGER

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

Contents

The Primary Structure of DNA

KENNETH MURRAY AND ROBERT W. OLD

RNA-Directed DNA Polymerase — Properties and Functions in Oncogenic RNA Viruses and Cells

MAURICE GREEN AND GARY F. GERARD

DNA Modification and Restriction

WERNER ARBER

Biozentrum, University of Basel
Basel, Switzerland

I. Introduction

The reader of scientific literature may have had his attention attracted recently to a growing number of highly interesting reports on research on DNA restriction endonucleases and DNA modification[1] methylases. A striking illustration is the November 1972 issue of the

[1] See article by Murray and Olds, this volume.

Proceedings of the National Academy of Sciences of the United States of America: it contains seven papers on restriction endonucleases, and none of the 16 authors signed more than one of these papers.

A. Definition of the Phenomena

Restriction endonucleases interact with double-stranded DNA molecules at specific sites leading to cleavage of the DNA into a number of fragments. The specificity of this interaction is thought to depend on the recognition by the enzyme of a particular sequence of base-pairs on the substrate DNA. Restriction endonucleases are found in many bacterial strains as products of genes carried either on the bacterial chromosome or on plasmid DNA. Among enzymes obtained from independent sources, each usually shows its own specificity of interaction.

As a rule, a bacterial strain can protect its own DNA from cleavage by its restriction endonucleases. This protection is brought about by site-specific methylation of the DNA, for which another activity, the DNA modification methylase, is responsible. Both endonuclease and methylase are thought to recognize the same base sequences on their substrate DNA. Each independent system of restriction and modification activities would then recognize its own particular target on the DNA. Therefore the modification given by a particular methylase protects the DNA only against restriction by the correlated endonuclease.

B. Sources of Restriction and Modification Enzymes

Mostly for reasons of a historical and practical nature, laboratory strains such as *Escherichia coli* K12 and B or *Haemophilus influenzae* are widely used in the experimental investigation of DNA restriction and modification, although many other bacteria are either known or are likely to have restriction and modification systems. Some bacterial strains carry only one restriction and modification system, a situation that facilitates analysis. This condition is fulfilled with *E. coli* K12, which has the additional advantage of being well suited to genetic investigations.

The genetic information for some of the extensively studied DNA restriction and modification systems is carried on plasmids, e.g., prophage P1 and some resistance transfer factors. Prophage and transfer factors can be incorporated at will into any convenient bacterial host, as can F′ factors carrying restriction and modification genes of chromosomal origin. Thus strains with a small number of known independent restriction activities can easily be constructed for par-

ticular experiments. It is still an open question whether the production of DNA restriction and modification enzymes is limited to bacterial cells or whether this capacity is found also in other microorganisms, or even in plant and animal cells.

C. Substrate DNA

As far as we know, any native double-stranded DNA molecule can serve as substrate in both restriction and modification reactions provided that it carries at least one specificity site producing the interaction of the DNA with the enzymes. Widely used substrates are bacteriophage DNA's. In fact the phenomenon of DNA host specificity was originally discovered in work with bacteriophage and described as host-controlled modification (*1–5*). This substrate lends itself well to experimentation both *in vivo* and *in vitro*. Some bacteriophage DNA molecules have the advantage of carrying only one specificity site, or a small, measurable number of specificity sites. If substrates with a large number of specificity sites are desired for particular experiments, bacterial DNA might be more useful. But it should be noted that DNA from sources other than bacteria is also susceptible to DNA restriction and modification. In particular, animal viral DNA has recently been used as a substrate for restriction (*6–9*).

D. Recent Reviews on This Topic

Several reviews on DNA restriction and modification have appeared in the past few years. The discovery of host-controlled modification and the general implications for cellular DNA have been outlined (*14a–17*). Some papers (*10, 14a, 17*) discuss in detail the genetic basis of restriction and modification enzymes and *in vivo* complementation experiments. Enzyme isolation and enzyme properties are described in others (*10, 17, 18*). Recognition sites have been characterized (*10, 17–19*), and the implication of methylation in the modification reaction has been outlined (*10, 17–19*). The special case of host-controlled modification of T-even phages has been comprehensively presented (*15, 20*). Other papers (*16, 17*) outline the wide distribution of R–M systems[1] among bacterial species and discuss evolutionary aspects.

[1] *Terminology:* Except for a few new or altered terms, the notations defined by Arber and Linn (*10*) are used in this paper, in which a relatively small number of well-studied restriction and modification systems (briefly called R–M systems) are discussed in detail. These R–M systems are:

K, determined by the genome of *E. coli* K12

B, determined by the genome of *E. coli* B

(footnote continued on page 4)

II. Facts from *in Vivo* Experiments

A. The Fate of Unmodified DNA

Teleologically, DNA restriction is considered to be a strain-specific defense mechanism against foreign genetic material that penetrates a particular cell. There are several mechanisms by which foreign genetic material can enter a cell: (a) uptake of free DNA in transformation; (b) transfer of cellular DNA in sexual conjugation; and (c) virus infection.

In each case, strains with an R–M system are able to ensure that only material carrying the modification type of the infected strain is accepted, while DNA not properly modified is subjected to rapid degradation (22). This is most readily shown in experiments with ^{32}P-labeled bacteriophage. Within a few minutes after penetration into a restricting cell, the label appears as acid-soluble material (23). However, since the restriction enzymes studied up to now are endonucleases producing very large DNA fragments, the acid solubilization observed *in vivo* must be caused by the subsequent action of exonucleases on the primary restriction cleavage products.

Both phage and bacterial DNA yield fragments that can be rescued from the restricted DNA by genetic recombination with superinfecting modified phage DNA and with the resident chromosome, respectively. In line with the idea that initial scission occurs at a certain, but not very large number of sites, closely linked markers have been reported to be jointly rescued at high probabilities. Simultaneous rescue of weakly linked markers, on the other hand, is much less frequent than in controls under nonrestrictive conditions

Footnote 1 (*continued*)

A, determined by the genome of *E. coli* 15 (*11*)

P1, determined by the genome of phage P1

P15, determined by the genome of a defective prophage carried by *E. coli* 15 (*11*)

RI, determined by the resistance transfer factors R124 (*12*) and RY-5 (*13*)

RII, determined by several resistance transfer factors, e.g., by R factor N-3 (*14*)

The symbols recently proposed for the R–M systems determined by strains of *Haemophilus* were not available when this article was written (*21*).

The symbol *hsd* (*h*ost *s*pecificity for *D*NA) is used for the genes determining DNA modification and restriction activities. Mutants affected in one or several *hsd* genes result in restriction-deficient (r^-) and/or modification-deficient (m^-) phenotypes. For example, a strain with $r_K^-m_K^+$ phenotype shows no K-specific restriction, but it gives K-specific modification. From determining the restriction and modification phenotype of a mutant, one usually cannot know precisely which of the *hsd* genes is affected by the mutation. Therefore it is operationally useful to indicate the phenotype rather than the genotype of a mutant as yet poorly characterized.