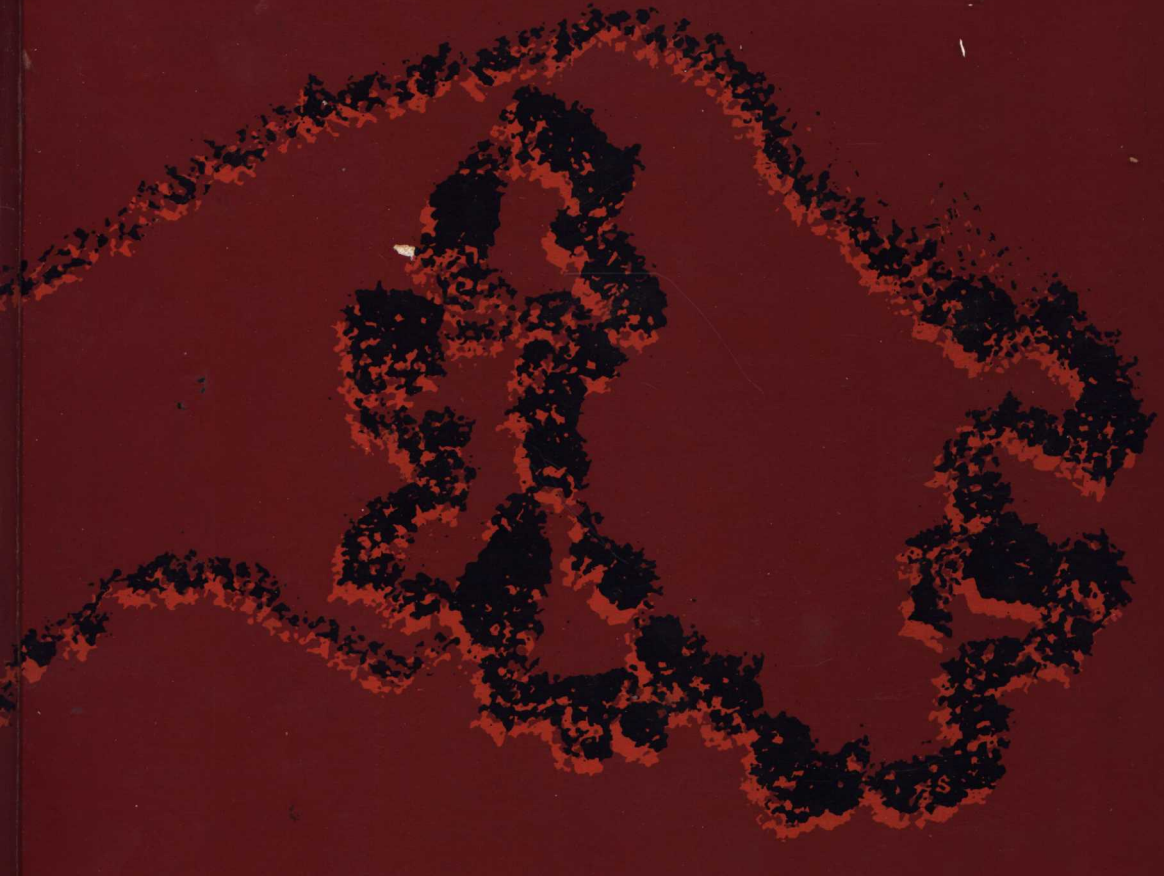
DNA Replication



Arthur Kornberg

DNA Replication

Arthur Kornberg
STANFORD UNIVERSITY

W. H. FREEMAN AND COMPANY
San Francisco

Cover: A looped rolling circle of the duplex replicative form of phage ϕX174 (based on an electron micrograph provided by Dr. Jack Griffith). See Figure 11-25 for more detail. (Design by Marjorie Spiegelman.)

Sponsoring Editor: Arthur C. Bartlett; Project Editor: Patricia Brewer; Copyeditor: Robert McNally; Production Coordinator: Linda Jupiter; Artist: Charlene Levering; Compositor: Advanced Typesetting Services; Printer and Binder: R. R. Donnelley & Sons Company.

Library of Congress Cataloging in Publication Data

Kornberg, Arthur, 1918–
DNA replication.

Includes bibliographical references and indexes.
1. Deoxyribonucleic acid synthesis.
I. Title.
QP624.K66 574.8'732 79–19543
ISBN 0-7167-1102-8

Printed in the United States of America

9 8 7 6 5 4 3 2

Preface to *DNA Replication*

I wrote *DNA Synthesis* in 1974 to review and record some of the early discoveries, which were receding from fashionable attention, and to discuss the advances and their wider ramifications. I anticipated that the pace of progress would require a major revision of the book in a few years. I did not anticipate how little I would retain five years later and how much would have to be added. But writing this virtually new and more massive book has been a pleasure rather than a chore. Besides extensive new information, recent advances in DNA synthesis have produced a qualitative change and justify a larger scope and a new title, *DNA Replication*.

The term DNA replication embraces biochemical, genetic, and physiological aspects, and also the numerous DNA transactions that determine the structure and function of genetic material. My goal is an up-to-date account of DNA replication and metabolism with a strong biochemical emphasis that can be used for orientation and as an information resource. I hope the book will serve students with a beginning interest in DNA synthesis as well as those working directly on the subject.

In this reorganization, I have devoted the first half of the book to the greatly expanded field of enzymology of DNA. I have chosen again to treat *E. coli* DNA polymerase I in detail as a prototypical polymerase. I have added discussions of supercoiling, binding and twisting proteins, and expanded the accounts of ligases, nucleases, and inhibitors of replication. For a grasp of the variety of replication mechanisms, I have surveyed the replicative life cycles of bacterial and animal viruses and of plasmids and organelles.

Proper acknowledgments for help in preparing this book could easily fill a chapter. I want to express my gratitude to the Rockefeller Foundation for their hospitality at the Bellagio Study Center, where I was able to start this book in October 1977. Specialists in many areas generously gave me their points of view and the most recent information, often unpublished. These people have read and influenced the contents and style of one or more of the chapters. I hope they will feel rewarded with a useful book and to them my gratitude is unbounded: Bruce Alberts, Paul Berg, Maurice Bessman, Douglas Brutlag, Michael Chamberlin, Nicholas Cozzarelli, David Dressler, Adayapalam Ganesan, Mehran Goulian, Philip Hanawalt, Nicholas Hoogenraad, Dale Kaiser, David Korn, Roger Kornberg, Sylvy Kornberg, Gordon Lark, Robert Lehman, Stuart Linn, Robert Low, Mark Pearson, Peter Reichard, Charles Richardson, Joseph Shlomai, George Stark, Jean Thomas, and Olke Uhlenbeck.

I owe major debts to Charlene Levering who did the illustrations with artistic skill and unstinting devotion, to Patricia Brewer whose superb styling and editorial judgment smoothed a rough manuscript, and to Leroy Bertsch, whose knowledgeable and meticulous final review reduced the errors to what I hope will be a forgivable level.

June 1979 *Arthur Kornberg*

Preface to *DNA Synthesis*

The rapid flow of facts and ideas in biochemistry makes it difficult to write an article, let alone a book. But such turbulence submerges useful facts and ideas, which become too specialized for general textbooks and are lost sight of even in detailed annual reviews. That this was true of the biochemistry of DNA synthesis became clear during preparation of the Robbins Lectures, given at Pomona College in April 1972, and it prompted me to undertake this effort.

This book emphasizes biochemical rather than physiological aspects of DNA synthesis. The scope has been broadened beyond that of an earlier book (*Enzymatic Synthesis of DNA*, 1962) to include topics clearly pertinent to DNA synthesis: precursors, repair, recombination, restriction, and transcription. It is hoped that with an enlarged scope and simplified language the book will serve students with a beginning interest in DNA synthesis as well as those working directly on the subject. Citations of the literature favor reviews and recent papers, which will in turn give the interested reader more complete bibliographies.

At the conclusion of writing this book, I am surprised and embarrassed at the large number of people whom I have enlisted in the preparation of this relatively modest effort. I am most indebted to my wife, Sylvy, who helped me write this book and do the early work on DNA synthesis. I am also grateful to Fred Robbins for the initial stimulus, to Charlene Levering for the illustrations, to Inge Loper for the typing, to Stephanie Lee Rowen for a careful early reading of the text, to I. Robert Lehman, David S. Hogness, A. Dale

Kaiser, and R. David Cole for a critical reading of the entire text, and to colleagues too numerous to mention who gave me information and advice in preparing many sections of the book. The best requital for all these contributions is a useful book, and this has been my primary goal.

February 1974 *Arthur Kornberg*

Contents

1

Structure and Functions of DNA

1-1 DNA: Past and Present[1–4]

1869–1943. The Discovery

This period opened with the discovery of a new organic phosphate compound in cells rich in nuclear substance.[3] At first called nuclein and later chromatin, this compound was subsequently shown to consist of deoxyribonucleic acid (DNA) and protein. Other discoveries followed. Analysis of DNA showed it to contain four kinds of building blocks, called nucleotides. DNA was distinguished from ribonucleic acid (RNA) in having a different sugar, deoxyribose, in place of ribose, and a distinctive base, thymine, in place of uracil.

1. Fruton, J. S. (1972) *Molecules and Life*, John Wiley, New York.
2. McElroy, W. D., and Glass, B. (eds.) (1957) *The Chemical Basis of Heredity*, Johns Hopkins Press, Baltimore.
3. Mirsky, A. E. (1968) *Sci. Amer. 218*, no. 6, 78.
4. Olby, R. (1974) The Path to the Double Helix, University of Washington Press, Seattle.

Although there was some reason to believe that DNA might be the genetic material, there was even more reason to assign this role to proteins. Each molecule of DNA was thought to be a repeating polymer of one kind of tetranucleotide unit. Proteins, since they are larger and are composed of twenty different amino acids, were thought more suitable for a genetic role. One must also recall that in the 1930s DNA was still called thymus nucleic acid, and it was widely believed to occur only in animal cells. RNA had been isolated only from plant cells and was called yeast nucleic acid. In fact, plant and animal cells were sometimes distinguished on the basis of this chemical feature.

1944–1960. The Genetic Substance

This "golden age"[5] began with the first important evidence that DNA is the genetic substance: the discovery, reported in 1944, that DNA prepared from one strain of pneumococcus could "transform" another strain.[6] The purified DNA carried a genetic message that could be assimilated and expressed by cells of another strain. DNA was further recognized to be a molecule far larger and more complex than a repeating tetranucleotide, varying in composition from organism to organism.[7] Yet, since traces of protein could not be ruled out as contaminants in the transforming DNA, some doubt remained that DNA was the genetic material, and these findings had relatively little impact on genetics.

Two persuasive discoveries were eventually made. The first was the demonstration in 1952 that infection of *Escherichia coli* by T2 bacteriophages involved injection of the DNA of the virus into the host cell.[8] The viral protein structures appeared to serve merely to inject the DNA into the bacterium and then to be largely discarded outside the cell. The DNA from the virus thus directed the bacterial cell to produce many identical copies of the infecting virus. This experiment dramatized the role of DNA as the carrier of information for producing the unique proteins of the virus and for duplicating its DNA many times over.

A second, remarkable event was the discovery, in 1953, of the complementary, double-stranded (duplex) structure of DNA and with it the recognition of how the molecule can be replicated.[9] Complementary pairing of the nucleotide constituents of one strand to

---

5. Stent, G. (1969) *The Coming of the Golden Age*, Natural History Press, Garden City, NY; (1978) *Paradoxes of Progress*, W. H. Freeman and Company, San Francisco.
6. Avery, O. T., MacLeod, C. M., and McCarty, M. (1944) *J. Exp. Med. 79*, 137; Hotchkiss, R. D. (1957) in *The Chemical Basis of Heredity*, (McElroy, W. D., and Glass, B., eds.), Johns Hopkins Press, Baltimore, p. 321.
7. Chargaff, E. (1950) *Experientia 6*, 201.
8. Hershey, A. D. (1953) *CSHS 18*, 135.
9. Watson, J. D., and Crick, F. H. C. (1953) *Nat. 171*, 737.

those of the second strand was postulated to explain in a simple way how one DNA duplex can direct the assembly of two molecules identical to itself. In this model, each strand of the duplex serves as a template upon which the complementary strand is made.

These discoveries and other important ones that followed led to the realization that DNA has two major and discrete functions. One is to carry the genetic information that brings about the specific phenotype of the cell. DNA is transcribed into RNA, and the RNA is then translated into the amino acid language of the proteins. In the "central dogma" of molecular biology, information is transferred from nucleic acids to protein and never in the reverse direction.[10] The other major function of DNA is its own replication. For duplicating the genotype of the cell, DNA serves as a template for converting one chromosome into two identical chromosomes.

1960–1973. Consolidation

The beginning of this age was not marked by a specific event. It was an age in which the generally held conceptions of both the structure and the dual functions of DNA were expanded. Without epochal discoveries, this age nevertheless brought a radical change in viewpoint toward DNA. Genetics and DNA became a branch of chemistry. Despite its chemical complexity, DNA was modified, dissected, analyzed, and synthesized in the test tube. There were insights into a metabolic dynamism of DNA that had not been anticipated. DNA suffers lesions and is repaired. DNA molecules exchange parts with one another. DNA molecules are specifically modified and degraded, twisted and relaxed, transcribed in reverse from RNA as well as directly into RNA. DNA functions not only in the nucleus but also in mitochondria and chloroplasts. These insights served as a stimulus to determine the total base sequence of DNA and to resynthesize it. There was a confidence that the metabolic gyrations of DNA in the cell could be understood in as explicit detail as those of, say, glucose or glutamate.

1973–The Present. Another Golden Age

Refinements in analytic methods uncovered unanticipated complexities and subtleties in the organization, replication, and expression of DNA. Revision of earlier concepts of chromosome organization, of expression of genes, and of replication and recombination has made the present era one of continuing drama and promise.

The concept that one stretch of DNA represents a single gene, expressed colinearly as one protein, has been shaken. One stretch

10. Crick, F. H. C. (1970) *Nat. 227*, 561.

of bacteriophage ϕX174 DNA is used in coding five different proteins. In animal cells, one gene can code for a multifunctional protein in which several enzymatic activities are contained within a single polypeptide. The genetic sequence coding for a polypeptide in higher organisms is not necessarily continuous, as previously believed. Rather, in the cases so far examined, it is often interspersed among regions that are functionally silent or have yet to be heard.

These discoveries were made with novel methods for dissecting, cloning, and amplifying genomes. A 5000-nucleotide-long genome can be sequenced in a few weeks. DNA insertion elements are cropping up everywhere; they facilitate a dynamic transposition of DNA among plasmids, viruses, and chromosomes, merging their identities. The intricacy of enzymatic machinery operating on DNA is proving to be awesome: Already twenty polypeptides are known to be engaged in the task of copying a small, single-stranded DNA circle to make it duplex. With a few more clues, the organization and control of replication and gene expression will be understood well enough to explain how cells develop, differentiate, and die.

1-2 Primary Structure[11]

The two kinds of nucleic acid—the ribonucleic acids, RNA, and the deoxyribonucleic acids, DNA—are polymers of nucleotides.

A nucleotide (Fig. 1-1) has three components: (i) a purine or pyrimidine *base*, linked through one of its nitrogens by an N-glycosidic bond to (ii) a 5-carbon cyclic *sugar* (the combination of base and sugar is called a nucleoside) and (iii) a *phosphate*, esterified to carbon 5 of the sugar. Nucleotides occur also in activated di- and triphosphate forms, in which one or two phosphates are linked to the nucleotide by phosphoanhydride (pyrophosphate) bonds.

In each of the two main kinds of nucleic acids there are generally four types of nucleotides. These are distinguished by their bases: adenine (A), guanine (G), uracil (U), and cytosine (C) in RNA, and adenine, guanine, thymine (T), and cytosine in DNA. The bases and their nucleoside and nucleotide forms are listed in Table 1-1. The major distinction in base composition between RNA and DNA is that RNA contains uracil whereas DNA contains thymine (5-methyluracil). Exceptions to this rule are the presence of thymine in transfer RNA, and of uracil in the DNAs of certain phages as well as transiently in most DNAs. RNA also contains a variety of other bases. The significance of the structures of the various bases becomes apparent when one considers the secondary structure of nu-

11. Sober, H. A. (ed.) (1970) *Handbook of Biochemistry, Selected Data for Molecular Biology*, 2nd ed., The Chemical Rubber Co., Cleveland (Section G).

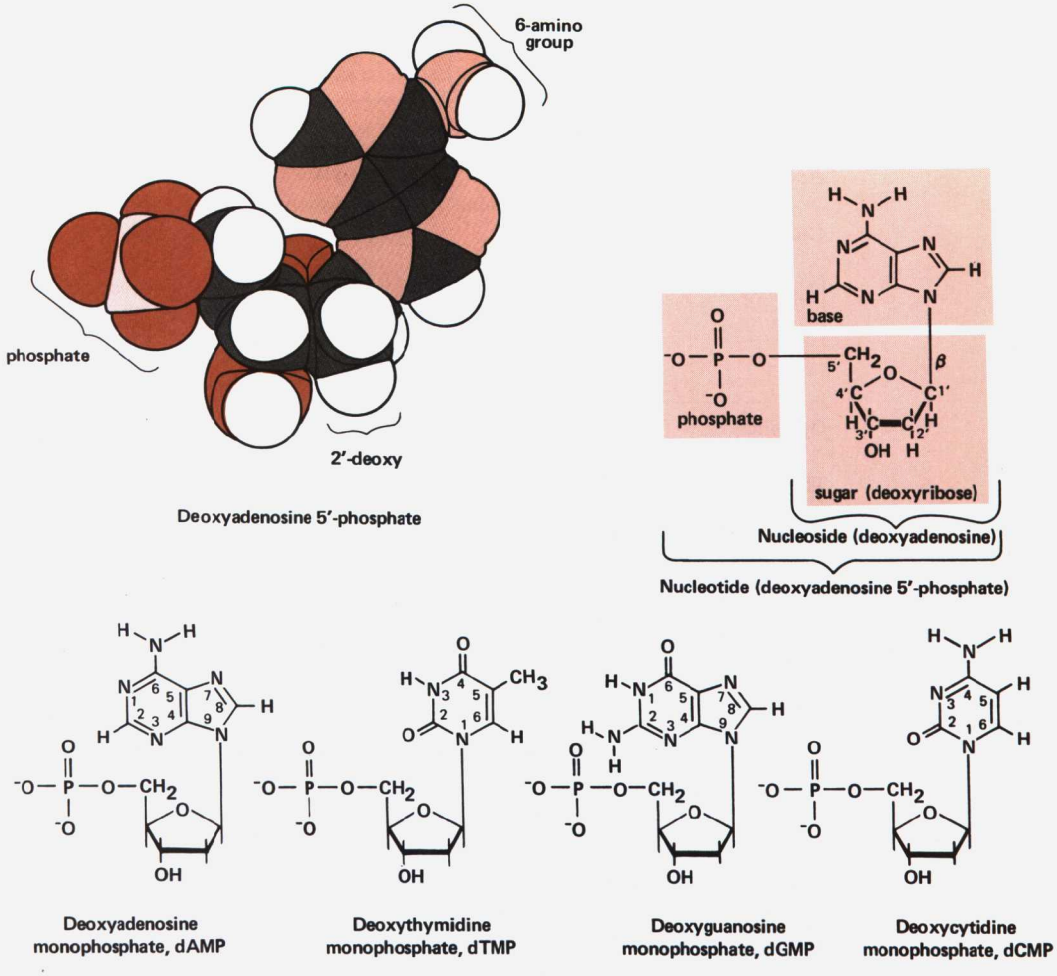


FIGURE 1-1
Deoxynucleotides of DNA; space-filling model of dAMP; generic nucleotide structure.

cleic acids (see next section). The truly major distinction between RNA and DNA is in their sugar-phosphate backbones: RNA contains only ribose; DNA contains only 2-deoxyribose.

Two advantages are apparent in the distinctive structure of DNA: The 2'-deoxy-containing backbone is more resistant to hydrolysis than the ribo form (Ch. 17-1). Thymine also ensures stability of the genetic message; chance deaminations of cytosine to uracil result in excision-repair rather than in retention of the uracil and consequent mispairing (Ch. 16-2).

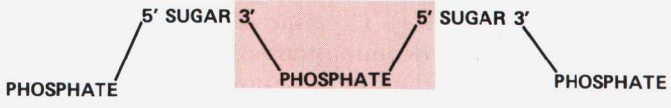
TABLE 1-1
Nucleic acid nomenclature

Base	Nucleoside[a]	Nucleotide[a]	Nucleic acid
		PURINES	
Adenine	adenosine	adenylate	RNA
	deoxyadenosine	deoxyadenylate	DNA
Guanine	guanosine	guanylate	RNA
	deoxyguanosine	deoxyguanylate	DNA
[Hypoxanthine	inosine	inosinate	precursor of adenylate and guanylate]
		PYRIMIDINES	
Cytosine	cytidine	cytidylate	RNA
	deoxycytidine	deoxycytidylate	DNA
Thymine	thymidine or deoxythymidine	thymidylate or deoxythymidylate	DNA
Uracil	uridine	uridylate	RNA

[a]*Nucleoside* and *nucleotide* are generic terms which include both ribo and deoxyribo forms. Deoxyribonucleosides and deoxyribonucleotides are designated as deoxynucleosides and deoxynucleotides, respectively, to make the names less cumbersome. Hypoxanthine and its derivatives are bracketed because they are precursors of the purine nucleotides but are found only in tRNA.

Designating deoxyribo-containing nucleosides and nucleotides with the prefix deoxy- or deoxyribo- has been accepted in all cases except for those containing thymine. Because thymine was originally thought to occur only in DNA, it seemed redundant to use the prefix, and the terms *thymidine, thymidine monophosphate,* and *thymidylate* were accepted. However, now that the natural occurrence of ribothymidylate has been recognized as one of the unusual nucleotides in transfer RNA and the thymine ribonucleotide and nucleoside have been made available by chemical and enzymatic syntheses, confusion in nomenclature does arise. In current practice, the terms *thymidine* and *deoxythymidine* are used interchangeably, as are *thymidylate* and *deoxythymidylate*; the ribonucleoside and ribonucleotides of thymine bear the prefix *ribo*.

Polynucleotide chains are long, unbranched polymers formed by bridges between the 5′-phosphate of one nucleotide and the 3′-hydroxyl of the sugar of the next (Fig. 1-2). The schematic diagram of nucleotide chains (Fig. 1-2, lower right) is frequently useful.



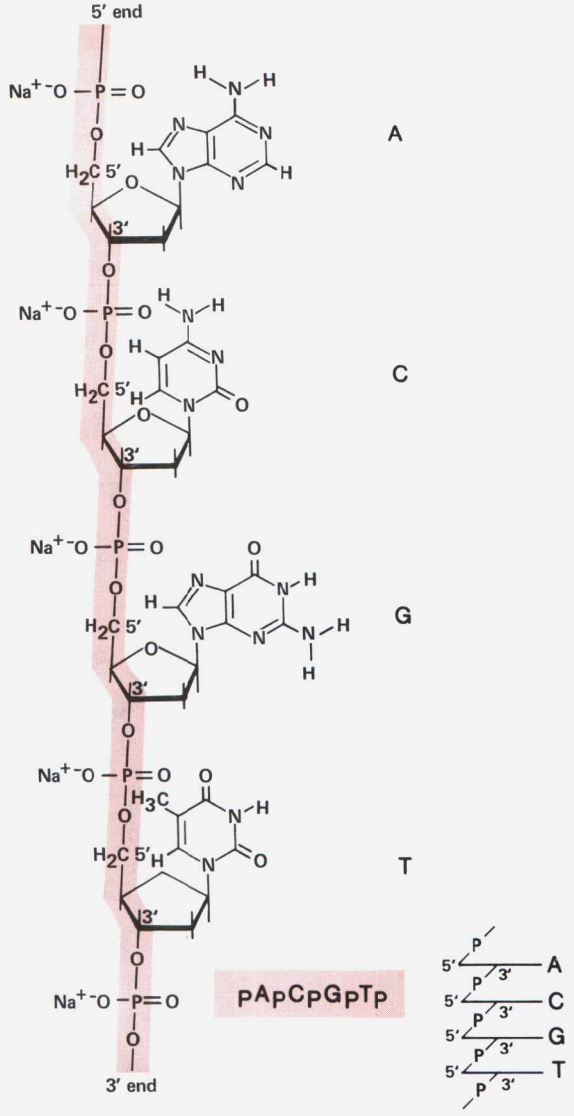


FIGURE 1-2
Segment of a polydeoxynucleotide (Na^+ salt).

The important linkage is the 3′,5′-phosphodiester bridge. This linkage is especially vulnerable to hydrolytic cleavage, chemically and enzymatically. It should be clear that such cleavage, depending on which of the phosphate ester bonds is broken, may yield either the biosynthetic intermediate 5′-phosphonucleoside or its 3′ isomer (Fig. 1-3).

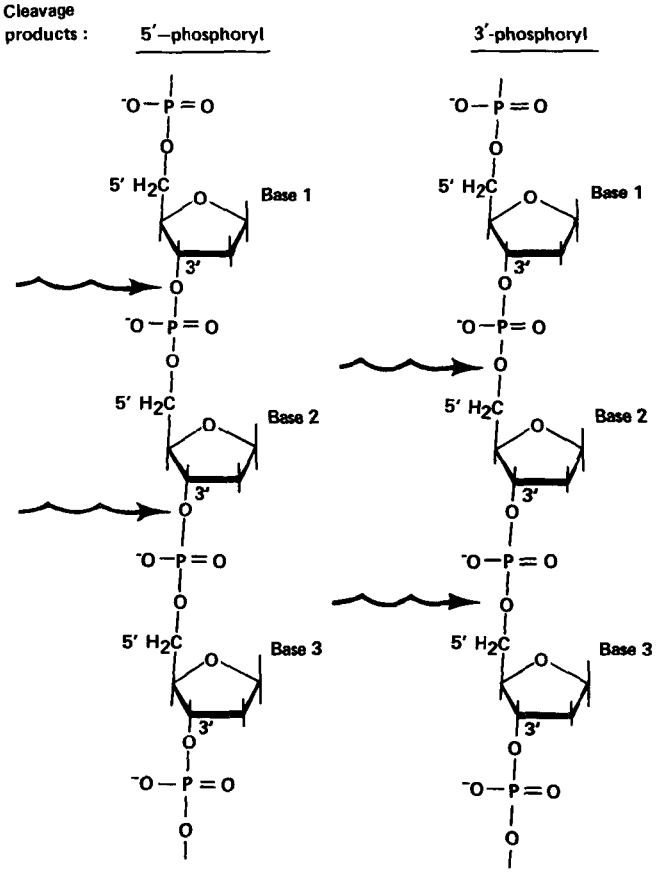


FIGURE 1-3
Cleavages of polynucleotide chains yielding 5′-phosphoryl termini or 3′-phosphoryl termini.

Alkaline treatment hydrolyzes RNA to mononucleotides but does not affect the DNA backbone. Because of the proximity of the 2′-hydroxyl group to the phosphodiester bond in RNA, a cyclic nucleoside 2′:3′-phosphate intermediate is formed, which is then hydrolyzed to a mixture of nucleoside 2′- and 3′-phosphates. Breaks in

mitochondrial DNA when exposed to alkaline pH have been attributed to interspersed ribonucleotides (Ch. 15-8). Evidence has been offered of alkali-sensitive linkers in nuclear DNA of kidney cells at a low, but significant, frequency of one per 150,000 base pairs.[12]

Because of free rotation about many of its bonds, the polynucleotide chain was originally considered to be highly flexible and to assume essentially random conformations. However, detailed studies[13] indicate that the main degrees of rotational freedom are limited to the two O–P bonds in the phosphodiester linkage and to the glycosyl bond of base to sugar. Thus, there are numerous constraints imposed on the nucleic acid backbone that permit it relatively few conformations and that make it quite stiff. To regard the single-strand polynucleotide chain as a completely random coil is therefore unwarranted.

1-3 A Double Helical Structure[14,15]

Base Pairing

Two distinctions among the nitrogenous bases of the nucleotides are crucial to the secondary structure of nucleic acids. One rests on the presence of keto and amino groups that provide opportunities for hydrogen bonding. On this basis T or U, both keto compounds, can pair with A, an amino compound, by a hydrogen bond. G and C, each having both keto and amino groups, can form two hydrogen bonds. In fact, an additional hydrogen bond can be formed between the ring nitrogens in the AT pair and also in the GC pair (Fig. 1-4).

The second important distinction among the bases is that they come in two sizes: the pyrimidines T or U and C are smaller than the purines A and G. However, the base pairs AT and GC prove to be identical in size: A pair of pyrimidines would be much smaller, a pair of purines much larger. It turns out that the AT and GC base pairs have not only the same size but the same shape!

These base-pairing characteristics of the nucleotides are responsible for the pairing of two chains to form a relatively stiff, strongly stabilized duplex (Figs. 1-5, 1-6). When two chains align themselves in this way, they assume double-helical structures, as worked out in detail by x-ray diffraction studies[15] of DNA fibers and by model building.

---

12. Filippidis, E., and Meneghini, R. (1977) *Nat.* 269, 445.
13. Levitt, M. (1972) *Polymerization in Biological Systems*, Ciba Foundation Symposium 7, Elsevier, New York, p. 147.
14. Watson, J. D., and Crick, F. H. C. (1953) *Nat.* 171, 737.
15. Arnott, S., Wilkins, M. H. F., Hamilton, L. D., and Langridge, R. (1965) *JMB* 11, 391.