NUCLEIC ACIDS
Structures, Properties, & Functions

NUCLEIC ACIDS

STRUCTURES, PROPERTIES, AND FUNCTIONS

Victor A. Bloomfield
UNIVERSITY OF MINNESOTA

Donald M. Crothers
YALE UNIVERSITY

Ignacio Tinoco, Jr.

WITH CONTRIBUTIONS FROM

John E. Hearst
UNIVERSITY OF CALIFORNIA, BERKELEY

David E. Wemmer
UNIVERSITY OF CALIFORNIA, BERKELEY

Peter A. Kollman
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Douglas H. Turner
UNIVERSITY OF ROCHESTER

University Science Books
Sausalito, California

University Science Books
55D Gate Five Road
Sausalito, CA 94965

Fax: (415) 332-5393
www.uscibooks.com

Production manager: *Susanna Tadlock*
Manuscript editor: *Jeannette Stiefel*
Designer: *Robert Ishi*
Compositor: *Eigentype*
Printer & Binder: *Maple-Vail Book Manufacturing Group*

This book is printed on acid-free paper.

Library of Congress Cataloging-in-Publication Data

Bloomfield, Victor A.
Nucleic acids : structures, properties, and functions / by Victor A. Bloomfield, Donald M. Crothers, Ignacio Tinoco, Jr. ; with contributions from John E. Hearst ... [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 0-935702-49-0 (hardcover)
1. Nucleic Acids. I. Crothers, Donald M. II. Tinoco, Ignacio. III. Title.
QP620.B64 1999
572.8—dc21 98-45268
CIP

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

NUCLEIC ACIDS

STRUCTURES, PROPERTIES, AND FUNCTIONS

Preface

In this book we present a comprehensive account of the structures and physical chemical properties of the nucleic acids, with emphasis on implications for biological function. The level of presentation assumes that the reader has knowledge of physical chemistry and molecular biology that would be obtained from introductory courses in these subjects.

We have three intended audiences: molecular biologists, physical biochemists, and physical chemists. Molecular biologists and nucleic acid biochemists have made remarkable strides in defining the major classes of nucleic acids, determining their sequences, and proposing connections between sequence and biological function. Such connections pertain not just to the direct coding of amino acid sequences in proteins, but also to regulation of transcription, translation, replication, and molecular evolution. The next step, one that requires physical chemical approaches, is to understand the mechanism by which a sequence of nucleotides exerts its coding or regulatory function. For this, one must know how the structure and dynamics of that sequence differ from those of its neighbors, and how those properties are affected by interaction with other molecules. One of our goals is to present to the molecular biologist the results that have been obtained on these issues, the experimental basis of these results, and some idea of what may be expected from even closer collaboration between biologists and physical chemists in the future.

For those physical biochemists already working on nucleic acids, or beginning a research career in this area, our goal is to provide a comprehensive treatment of the major experimental and theoretical approaches to the structures and physical properties of the nucleic acids.

Physical chemists have developed a powerful set of experimental and theoretical techniques for determining the structural and dynamical properties of molecules. New methods in spectroscopy and diffraction, and new theoretical approaches to complex molecular systems are producing ever more sensitive and detailed insights into nucleic acid behavior. We hope in this book to give physical chemists an overview of the biological context in which their contributions are being used, and some perspectives on the challenges that remain.

To reach three such different audiences with the same book requires careful organization. We have adopted a uniform plan for each chapter, leading off with a statement of the biological significance of each topic, and following with a clear presentation of the basic physical ideas and major results. The quantitative details, which are important to the physical chemists and physical biochemists, are developed in special sections or appendices.

The main logic of organization is through a systematic consideration of techniques used in the study of nucleic acid structure and properties. This should maximize the book's utility as a textbook in physical biochemistry, and as a reference for techniques that may be encountered in the literature or in one's own work. Examples of the uses of the various techniques go from the simpler techniques to the more complex. Readers interested in applications to particular molecules will find cross-references in the index.

We hope that the book will serve a variety of functions: as a textbook in physical biochemistry and biophysical chemistry classes; as an aid to molecular biologists or biochemists trying to gain a better understanding of physical techniques they use or encounter in the literature; and as a reference work for physical chemists and physical biochemists.

We are grateful to our colleagues who have sent us reprints and original figures for reproduction, and who have commented on the various versions of this book. We give special thanks to David Draper, whose detailed comments notably improved our efforts.

NUCLEIC ACIDS

STRUCTURES, PROPERTIES, AND FUNCTIONS

Contents

CHAPTER 1

Introduction

1. BIOLOGICAL ROLES OF THE NUCLEIC ACIDS

Nucleic acids are central molecules in the transmission, expression, and conservation of genetic information. The role of DNA as the carrier of genetic information has been amply demonstrated beginning with the classic experiments of Avery et al. (1944) and Hershey and Chase (1952). The classic example of how biological function follows from biomolecular structure comes from the elucidation by Watson and Crick (1953) of the structure of DNA as a double helix, using the X-ray fiber diffraction patterns generated by Franklin, Wilkins, and their associates (Franklin and Gosling, 1953; Wilkins et al., 1953) and the chemical evidence on base complementarity of Chargaff (1950). It was immediately obvious how information could be passed from one generation to the next by synthesizing a complementary strand for each of the parent strands and pairing with the parental complement. This mode of semiconservative replication was verified by Meselson and Stahl (1957) using the newly developed technique of density gradient ultracentrifugation.

Chemical analysis showed that there was a great deal of RNA in cells; most of the various types of RNA and their biological functions were understood in the 1950s and 1960s. Messenger RNA (mRNA) is the product of transcription of DNA into the carrier of the genetic code. It, in turn, is translated into proteins. Ribosomal RNA (rRNA) is complexed with ribosomal proteins to form ribosomes, the organelles on which protein synthesis occurs. Transfer RNA (tRNA) serves as the adaptor molecule between the amino acid and the genetic code triplet on the mRNA. Recently, we have

learned that RNA plays even wider and unexpected roles. Genetic information can be carried in RNA viruses and retroviruses, and copied into DNA by reverse transcriptase. Complexes with proteins called small nuclear ribonucleoproteins (snRNPs) have been discovered that splice out intron sequences from transcripts of genomic DNA (Sharp, 1987). Ribonucleic acid itself can have enzymatic activity (such molecules are termed ribozymes) (Cech, 1987) and may have been the primordial enzyme, a key idea in considerations of prebiotic evolution (Gesteland and Atkins, 1993). A laboratory demonstration of RNA-catalyzed RNA polymerization has been achieved by Doudna and Szostak (1989).

The monomers of nucleic acids, the nucleotides and nucleosides, serve a diversity of roles. Some of these have been known over much of the span of modern biochemistry, such as the role of ATP as the energy currency in cells, and of the role of adenosine-containing cofactors nicotinamide adenine dinucleotide (NAD) and NADH as coenzymes in oxidoreduction reactions. More recently, cyclic adenosine - monophosphate (AMP) and cyclic guanosine monophosphate (GMP) have been shown to be crucial second messengers in controlling a wide variety of cellular processes.

The reader is referred to some of the basic texts and historical monographs listed at the end of this chapter for more information on these topics.

2. PHYSICAL CHEMISTRY AND NUCLEIC ACID FUNCTION

When the precursor to this book (Bloomfield et al., 1974) was written, most of these biological functions were known (but not all—splicing and RNA enzymatic activity were yet to come). What was not nearly so well known then as now was the detailed structural basis for these biological activities. A few of the crucial developments in the last 15 years are listed below (references will be given in later chapters).

- Discovery of Z-DNA, and determination of the X-ray structure of a B-DNA dodecamer, both of which demonstrated that DNA structure was highly variable at the atomic level, providing a concrete foundation for thinking about the structural basis of recognition and regulation.
- Discovery of naturally bent DNAs, which also demonstrated that particular sequences had defined and unique molecular structures.
- Recognition of the ubiquity and importance of supercoiling as a way of exerting long-range control of DNA.
- Recognition that ions and water can greatly influence the structure and interactions of DNA.
- Determination of the many non-Watson–Crick base pairings, and specific ribose-base–phosphate interactions present in folded RNA molecules.

Much of this progress has come from the development of new methods and new approaches to data analysis.

- Development of gel electrophoresis as a way to analyze the size and structure of nucleic acids and their complexes with proteins, permitting rapid and inexpensive analysis of complex molecules, and likely leading to sequencing of the human genome within a few years after this book is published.
- Ability to prepare synthetic nucleic acids, or purify natural ones, in adequate quantity for physical studies.
- Multidimensional NMR as a way to determine molecular structure in solution.
- Accumulation of data on a wide variety of nucleic acids, permitting recognition of regularities and suggesting structural generalizations.
- Random synthesis of nucleic acid sequences and selection for binding ability or catalytic activity, allowing identification of novel properties of nucleic acids, and revealing the range of sequences that can perform the same function.
- Computer analysis methods that have pointed to similarities between functionally and evolutionarily related molecules, suggesting common structures in RNAs that may regulate protein synthesis. (This topic, because of its mathematical and computer basis, may have considerable attraction for quantitatively oriented scientists. We do not discuss it in this book, but interested readers are directed to the references on computer analysis listed at the end of this chapter).
- Thermodynamic data bases for predicting secondary structure, loops, bulges, and so forth.

As an example of the important biological questions that physical chemistry may help us to answer, consider protein binding to a regulatory sequence of DNA. What structural features cause this specific binding site to be recognized? How can the protein search so much intracellular DNA to find that particular sequence? How much energy is required to deform the DNA and protein for optimal binding? How can the effects of sequence mutations on binding strength be understood? How does change in ionic strength affect binding? How does protein binding affect the structure and activity of distant DNA sequences, including the binding of other proteins? In this book, we have tried to show how answers to such questions may be obtained.

3. OUTLINE OF THIS BOOK

The first part, Chapters 2–8, treats the properties of the nucleic acids mainly at the level determined by atomic and molecular structure.

Chapter 2 (Bases, Nucleosides, and Nucleotides) discusses the physical and chemical properties of the monomeric building blocks of nucleic acid structure: the bases, nucleosides and nucleotides. Seemingly recondite monomer structural features such as sugar pucker and glycosidic bond conformation are shown later to be key determinants of helix geometry. Electron distributions in the bases influence their hydrogen-bonding and base stacking capabilities, as well as their spectroscopic behavior.

Chapter 3 (Chemical and Enzymatic Methods) discusses chemical and photochemical reactivity, important for structural determinations and mutagenesis. Much of the recent rapid progress in our understanding of nucleic acid sequence and structure, and of the binding of ligands, comes from increasingly sensitive and specific reactions. It also notes some of the naturally occurring base modifications that influence specificity and recognition.

Chapter 4 (Nucleic Acid Structures from Diffraction Methods) presents the basic ideas and results of diffraction techniques applied to nucleic acids. It begins with fiber diffraction, which gave the earliest insights into double helix structure and whose results are still important in delineating helical parameters for long sequences that cannot be crystallized. It continues with high-resolution X-ray crystallography of DNA oligonucleotides and tRNA, whose structures can be determined at atomic resolution. Results on A, B, and Z forms of DNA have had enormous impact on our thinking about the variability and flexibility of nucleic acid conformation. This chapter concludes with results of crystallographic studies on complexes with proteins, metal ions, water, and drugs—the sorts of complexes through which DNA exerts its biological functions.

Chapters 5 (Structure and Dynamics by NMR) and 6 (Electronic and Vibrational Spectroscopy) approach structure and dynamics through the NMR and optical spectroscopic techniques that have yielded so much of our current information about nucleic acids in solution. Nuclear magnetic resonance (NMR) has been the tool that provided the basic understanding of the structure of mononucleotides in solution. Now it is being used to determine the three-dimensional solution structures of tRNAs, rRNAs, and ribozymes. Atomic resolution studies of oligonucleotides are being used to probe the structure and dynamics of loops and bulges caused by base mismatches, and to determine the rates at which bases fluctuate between paired and unpaired states. Ultraviolet (UV) spectroscopy has for decades been one of the standard techniques for determining nucleic acid concentration and helix–coil transition behavior. In more sophisticated forms, it can elucidate the orientation of the bases along the helix, and give information on DNA conformation in packaged intracellular forms such as chromatin. Circular dichroism (CD) is the prime method for distinguishing the main helical forms of double-stranded nucleic acids, and was the first technique to show the existence of left-handed Z-DNA. Infrared (IR) and Raman spectroscopy are sensitive measures of backbone and base geometry; they can thus provide critical information on the conformation of DNA inside viruses.

Chapter 7 (Theoretical Methods) provides a theoretical reprise of many of the preceding topics. Developments in computer hardware and software, and in understanding of intermolecular forces, are making possible highly detailed modeling of the structural, energetic, and reactive properties of the nucleic acids and their complexes. Reliable predictions of the structural and dynamic effects of base substitutions, of conformational changes induced by ligand binding, and of the relative affinities of intercalating drugs are clearly on the horizon.

Chapter 8 (Conformational Changes) deals with conformational transitions, helix–coil, and helix–helix, which are crucial to understanding the forces stabilizing helices, to calculating the likelihood of finding biologically significant alternative structures, and to assessing the complexity of genomic sequences. It has strong emphasis on

base stacking and double helix formation in single-stranded polynucleotides, matters that are crucial to understanding the way in which folding of RNA influences protein binding, and transcription and translation.

There follow two chapters on the size and shape of nucleic acids considered as macromolecules.

Chapter 9 (Size and Shape of Nucleic Acids in Solution) surveys the major experimental methods for characterizing the molecular weight, size, and shape of nucleic acids chains. These include the gel electrophoretic and density gradient centrifugation techniques that are so widely used in molecular biology laboratories, as well as other hydrodynamic, scattering, and microscopy techniques. It also considers size and shape from the theoretical point of view, including such matters as conformations of single-stranded molecules, the rigidity of double-stranded ones, and the occurrence of naturally bent DNA.

Chapter 10 (Supercoiled DNA) deals with the key topic of closed circular, supercoiled DNA. It discusses how the various facets of supercoiling—linking, twisting, and writhing—can be described and interrelated. It then shows how the considerable energy stored in supercoiled DNA can be utilized to drive reactions and conformational transitions.

The next four chapters (Chapters 11–14) consider the noncovalent interactions of nucleic acids with the many types of molecules that affect stability or regulate function.

Chapter 11 (Interaction of Nucleic Acids with Water and Ions) considers the most basic interactions with water and ions. Water is a fundamental determinant of nucleic acid structure, as recognized since the 1950s when the A, B, and C forms of DNA were found to be interconvertible as a function of relative humidity. The basis for water stabilization of nucleic acid structure is now being revealed by X-ray and computer modeling studies. Since DNA and RNA are highly charged, their properties are strongly influenced by ions in solution. Ions affect the relative stability of double and single helices, the bending rigidity of DNA, and the condensation of DNA into compact particles. Perhaps the most dramatic effect of ions is on the binding of proteins. A two-fold increase in salt can produce a 100-fold or more decrease in binding constant.

Chapter 12 (Interactions and Reactions with Drugs) surveys binding of small ligands such as drugs. Drug binding is not a passive event; it requires changes in nucleic acid structure, and can thus be used as a probe of structural changes. The structural, thermodynamic, and kinetic principles governing binding affinity are discussed here.

Chapter 13 (Protein–Nucleic Acid Interactions) deals with the important interactions between nucleic acids and proteins, both regulatory and structural. These complex systems stretch the limits of biophysical approaches. But a variety of experimental methods, coupled with computer analysis of sequences and recognition of motifs characterizing binding domains (such as zinc fingers and leucine zippers) have given remarkable insight.

Chapter 14 (Higher Order Structure) concludes the book with a discussion of the higher order structures and mechanisms in the packaging of DNA in viruses and chromatin. These, it has become increasingly apparent, must be understood if we are to comprehend much of the biological functioning of the nucleic acids.

4. OBTAINING NUCLEIC ACIDS FOR PHYSICAL CHEMICAL STUDY

There are reciprocal influences between physical biochemistry and molecular biology. Not only does molecular biology provide the problems for the physical chemistry of nucleic acids, it also may help to provide the nucleic acids themselves. The need in physical studies is for macromolecules to be as pure in size and composition, and in as large quantities, as possible. In addition, there should be a range of sizes, compositions, and sequences.

Traditionally, DNA has been isolated from readily available sources such as calf thymus, or from bacterial viruses. Once the importance of avoiding shear breakage was recognized, and delicate handling techniques were devised, bacteriophages gave a good source of monodisperse DNA. However, viral DNA is sufficiently large (2–110 million molecular weight) that it gives information mainly on average and long-range conformational properties.

For many types of studies, it is desirable to have short segments of DNA. Sequence and compositional variations can be more extreme and better controlled than in longer molecules. Local stiffness and bending can be analyzed more directly, since the DNA conformation approximates a rigid rod rather than a random coil. And small nucleic acids, with only a few dozen base pairs (bp), are most suitable for high-resolution NMR and X-ray structural studies. Large DNA molecules can be broken into shorter ones in a variety of ways. Unless extreme care is taken, high molecular weight DNA will be broken by the shear generated by standard laboratory manipulations, such as pouring and pipeting, into fragments of about 5 million molecular weight. Further breakage, down to about 1 million (1500 bp) or somewhat below, can be achieved by more vigorous mechanical shearing, for example, by repeated passage through a hypodermic syringe, or processing in a homogenizer or blender. Even smaller fragments, down to 50–100 bp, can be obtained by sonication for several hours. All of these procedures, however, produce a broad range of molecular sizes, and separation to obtain narrow length fractions is difficult.

To obtain more monodisperse preparations, many workers have turned to nuclease cleavage of chromatin. As described in Chapter 14, the central structural element of chromatin is the nucleosome, consisting of about 150 bp of DNA wrapped around a histone octamer. An additional several dozen bp of DNA, complexed with histone H1, constitute a spacer region between nucleosomes. By a carefully regulated sequence of enzymatic digestions and deproteinizations (Wang et al., 1990), it is possible to obtain hundreds of milligrams, or even grams, of 150 bp mononucleosomal DNA from conveniently available tissues rich in chromatin, such as calf thymus or chicken erythrocytes. The range of lengths in these preparations is quite narrow, about 150 ± 10 bp. This length is useful for many physical studies, since it corresponds to about one persistence length of DNA, and therefore exhibits measurable but not overwhelming flexibility (Chapter 9). Unfortunately, multiples of this fundamental length are much harder to purify in useful quantities, and the mononucleosomal DNA is quite random in base composition and sequence.

If molecules with different though defined lengths or with known sequences are desired, workers generally resort either to large-scale production of plasmids for DNAs

in the range around 1 kbp, or to chemical synthesis for the range 10–50 bp. Plasmids may be engineered by recombinant DNA technology to contain desired sequences, and then overexpressed in *Escherichia coli* to yield a few milligrams of pure DNA (Lewis et al., 1985). Density gradient ultracentrifugation, which is discussed in Chapter 9, is an important technique for plasmid purification. The plasmids may either be used intact, or the sequences of interest may be excised by digestion with a restriction enzyme. In the latter case, though, it is not always straightforward to separate the desired sequence from the other fragments.

To obtain relatively short sequences, such as those used in X-ray crystallographic or NMR studies, it is generally most direct to use phosphoramidite solid-phase oligonucleotide synthesis (Itakura et al., 1984). A reactive phosphoramidite group on the 3′ position of deoxyribose is coupled to the 5′-OH of the next monomer unit. The amino groups of the bases are protected. At the end of the synthesis the protected bases must be deblocked and the products purified. Using commercial synthesizers, it is easy to produce milligram quantities of oligonucleotides one hundred bp long.

The polymerase chain reaction (PCR) is beginning to be used to produce physical quantities of DNA. This technique was originally developed to amplify genomic sequences for detailed analysis (Mullis and Faloona, 1987). Specific synthesis of DNA *in vitro*, via a polymerase catalyzed chain reaction, can amplify a particular sequence up to 1 million-fold. It employs two synthetic oligonucleotides as primers for enzyme-catalyzed DNA synthesis, using as templates the complementary strands in the region between the priming sites. Exponential increase of the sequence of interest is achieved by repeated cycles of thermal denaturation of the duplex, annealing of the primers, and polymerase-catalyzed synthesis. The key to practical success of the method is a highly processive and thermostable DNA polymerase from *Thermus aquaticus*, known as the *Taq* polymerase. It can produce microgram amounts from a single copy. To scale up to milligrams requires substantial development, including use of many reaction vessels in parallel; but the scale-up is technically feasible.

To obtain physical quantities of RNA is either easier or harder than for DNA, depending on the material desired. Certain naturally occurring RNAs, such as rRNAs and the various tRNAs, can be readily isolated in large quantities and are commercially available. To produce oligonucleotides, the same phosphoramidite solid-state synthetic chemistry is employed as for DNA, though the protection is more difficult because of the 2′-OH. Synthetic homopolymers with a random length distribution may be synthesized using polynucleotide phosphorylase. Large quantities of monodisperse RNAs with a desired sequence in the size range from a dozen to a few hundred bases can be obtained by RNA polymerase transcription from a synthetic DNA template (Milligan et al., 1987).

5. THE PERIODICAL LITERATURE OF NUCLEIC ACIDS

In such an active area, keeping up with the literature is difficult but necessary. The major journals that carry work related to physical chemistry of nucleic acids are *Biochemistry, Biophysical Journal, Biopolymers, Journal of Biological Chemistry, Journal of*