victor a. bloomfield
donald m. crothers
ignacio tinoco, jr.

physical chemistry of nucleic acids

16

physical chemistry of nucleic acids

VICTOR A. BLOOMFIELD
University of Minnesota
DONALD M. CROTHERS
Yale University
IGNACIO TINOCO, Jr.
University of California, Berkeley

Harper & Row, Publishers
New York, Evanston,
San Francisco, London

Sponsoring Editor: John A. Woods
Project Editor: Lois Wernick
Designer: Michel Craig
Production Supervisor: Valerie Klima

PHYSICAL CHEMISTRY OF NUCLEIC ACIDS

Library of Congress Cataloging in Publication Data
Bloomfield, Victor A
Physical chemistry of nucleic acids.
Includes bibliographies.
1. Nucleic acids. I. Crothers, Donald M., joint author. II. Tinoco, Ignacio, joint author. III. Title. [DNLM: 1. Chemistry, Physical. 2. Nucleic acids. QU55 B655p 1974]
QD433.B44 547'.596 73-8373
ISBN 0-06-040779-4

preface

The nucleic acids and their constituents are key molecules in the storage and transmission of genetic information, in the readout of this information during protein synthesis, and as substrates, cofactors, or allosteric effectors in a host of other metabolic processes. Because of this central importance, the nucleic acids have been subject to extraordinarily intensive scrutiny by investigators using a wide range of physical and chemical techniques.

We have aimed in this book to present the methods and results of the physical chemical study of nucleic acids, in a fundamental and comprehensive fashion. We have attempted to lay particular stress on general principles, to cite particular results only when they seem firmly established (not an easy matter to assess in a rapidly changing field), and to provide, by fairly detailed discussion of the theoretical basis of the major experimental techniques, a firm foundation for future work.

The book is directed at graduate students and research scientists who are directly concerned with nucleic acids or who, in neighboring fields of physical chemistry, polymer chemistry, biochemistry, biophysics, and molecular biology, may be interested in how their own studies may illuminate, or be illuminated by, the physical chemical properties of nucleic acids. We have assumed that the reader has a basic knowledge of physical chemistry and of the biochemistry of the nucleic acids.

We have attempted to reconcile the often conflicting demands, in a book of finite size, of an advanced research-level monograph and of a textbook that will be useful (when properly supplemented) in graduate courses on biophysical chemistry and on the chemistry of nucleic acids. To this end, we have provided a substantial amount of pedagogical and mathematical detail on topics which have not been extensively covered in standard textbooks. With other more standard topics, particularly spectroscopy and some aspects of polymer physical chemistry, we have contented ourselves with a brief discussion of basic principles before proceeding to applications and have directed the reader to appropriate references for a more detailed treatment of fundamentals.

It is inevitable, and not undesirable, that the authors of a book such as this will emphasize topics close to their own research interests. It is fortunate, therefore, that the research interests of the present authors complement each other to a fair extent–hydrodynamics, light scattering, and dimensional statistics of polynucleotides (Bloomfield); helix-coil transitions and ligand binding reactions (Crothers); and spectroscopic properties and thermodynamics (Tinoco)–so that the book as a whole is, we hope, reasonably comprehensive and balanced. Although each author took primary responsibility for the material indicated above, all chapters were scrutinized and revised by each author to provide coordination and avoid overlap.

No book like this can be written without the active assistance of many colleagues. Professor Charles Cantor and Dr. Gary Felsenfeld read the entire book in manuscript, and Professor James Wang read most of it. Each made many valuable suggestions. Peter and Susan Berget prepared the index. We are also grateful to the many people who sent us preprints and original photographs and who graciously granted permission for use of published figures and tables. We finally acknowledge our students and co-workers, whose comments often improved our understanding of the material in this book.

V. A. B.
D. M. C.
I. T.

contents

Preface ix

Chapter 1 INTRODUCTION 1

Chapter 2 MONOMERS 9
- I. Configuration and Conformation 10
- II. Electronic Properties 19
- III. Thermodynamic Properties 66

Chapter 3 OLIGONUCLEOTIDES AND SINGLE-STRANDED POLYNUCLEOTIDES 83
- I. Introduction 84
- II. Dinucleotides 84
- III. Trinucleotides 93
- IV. Polynucleotides 95

Chapter 4 DOUBLE-STRANDED POLYNUCLEOTIDES 103
- I. Introduction 104
- II. Chemical Structure 105
- III. Physical Structure Determination 108

Chapter 5 MOLECULAR WEIGHT AND LONG-RANGE STRUCTURE 151
- I. Introduction 152
- II. Conformational Statistics 153

III. Direct Visualization 171
IV. Scattering of Radiation 183
V. Sedimentation Equilibrium in Density Gradients 202
VI. Transport Phenomena 215
VII. Results 255

Chapter 6 CONFORMATIONAL CHANGES 293
I. Introduction 294
II. Experimental Methods 295
III. Stoichiometry of Helix Formation 322
IV. Compositional Effects on Helix Stability 328
V. Environmental Effects on Helix Stability 332
VI. Molecular Weight Dependence of Helix Stability 338
VII. Oligomer-Polymer Interactions 339
VIII. Oligomers and Natural RNA 343
IX. Mechanisms of DNA Melting 354
X. Strand Separation and Renaturation 356

Chapter 7 BINDING OF SMALL MOLECULES 373
I. Hydration of Nucleic Acids 375
II. Polyelectrolyte Behavior 389
III. Thermodynamics of Binding Reactions 406
IV. Statistical Mechanics of Binding Reactions 415
V. Coordinate Binding of Metals to Nucleic Acids 420
VI. Modes of Binding Drugs to DNA 429
VII. Experimental Methods 431
VIII. Binding of Specific Substances 456

Chapter 8 TRANSFER RNA 477
I. Primary Structure 478
II. Secondary Structure 482
III. Tertiary Structure 486

Author Index 495
Subject Index 507

chapter 1
introduction

The key insight into the biological function of nucleic acids at the molecular level is claimed by Watson[1] when in 1952 he taped up a paper sheet above his desk saying:

$$\text{DNA} \longrightarrow \text{RNA} \longrightarrow \text{protein}$$

This slogan represented the transfer of genetic information from the sequences of nucleotides in DNA molecules to the sequences of amino acids in proteins. The details of the processes symbolized by the arrows have since been extensively studied. A slightly more explicit description of the chemical reaction involved is:[2]

Transcription:

$$\left\{\begin{matrix}\text{nucleotide}\\ \text{triphosphates}\end{matrix}\right\} \xrightarrow[\left\{\begin{matrix}\text{DNA dependent}\\ \text{RNA polymerase, etc.}\end{matrix}\right\}]{\text{DNA} \atop +} \left\{\begin{matrix}\text{messenger RNA(mRNA)}\\ \text{transfer RNA(tRNA)}\\ \text{ribosomal RNA(rRNA)}\\ \text{other RNA}\end{matrix}\right\}$$

Translation:

$$\text{amino acids} \xrightarrow[\left\{\begin{matrix}\text{ribosomes}\\ +\text{ tRNA}\\ +\text{ enzymes, etc.}\end{matrix}\right\}]{\text{mRNA} \atop +} \text{proteins}$$

TABLE 1-1 THE GENETIC DICTIONARY[a]

First base	Second base				Third base
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met or start	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val or start	Ala	Glu	Gly	G

[a] The structures and full names of the amino acids are given in Table 1-2.

The four ribonucleotide triphosphates are polymerized on the DNA templates by RNA polymerase to form many different RNA molecules. Only one strand of the DNA molecule is copied. A single DNA molecule contains the templates for many different molecules. The DNA base sequences which specify the beginning and end of the RNA molecules are not known at this time. Apparently protein molecules are bound to those sites and control the action of RNA polymerase[3,4]. The code for transcribing DNA into RNA is simple

DNA ⟶	RNA
adenine (A)	uracil (U)
cytosine (C)	guanine (G)
guanine (G)	cytosine (C)
thymine (T)	adenine (A)

The translation step is more complicated. Up to 20 amino acids are polymerized to form proteins. The sequence of amino acids is governed by the sequence of base triplets on the messenger RNA. All but 3 of the 64 possible base triplets are read by transfer RNA molecules, each of which carries an amino acid. The three special triplets specify termination of the protein chain; they are recognized by specific proteins.[5] The genetic dictionary for translating from RNA base sequences to protein amino acid sequences is given in Table 1-1. Table 1-2 gives the dictionary for translating

TABLE 1-2 THE GENETIC DICTIONARY[a]

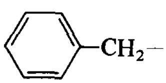
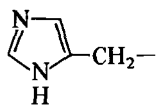
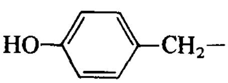
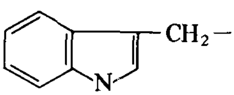
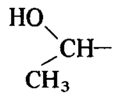
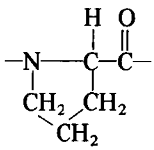
Amino Acid	Triplet	Amino Acid	Triplet
$H-$ glycine	GGX	$^{-}O(O{=})CCH_2-$ glutamic acid	GAPy
CH_3- alanine	GCX	$^{-}O(O{=})CCH_2CH_2-$ aspartic acid	GAPu
$(CH_3)_2CH-$ valine	GUX	$H_2N(O{=})CCH_2-$ glutamine	AAPy
$(CH_3)_2CHCH_2-$ leucine	CUX UUPu	$H_2N(O{=})CCH_2CH_2-$ asparagine	CAPu
$CH_3CH_2(CH_3)CH-$ isoleucine	AUPy AUA	$H_2NCH_2CH_2CH_2CH_2-$ lysine	AAPu
 phenylalanine	UUPy	 histidine	CAPy
 tyrosine	UAPy	$H_2NC(=NH_2)NHCH_2CH_2CH_2-$ arginine	CGX AGPu
 tryptophan	UGG	$HOCH_2-$ serine	AGPy UCX

TABLE 1-2 – *continued*

Amino Acid	Triplet	Amino Acid	Triplet
$HSCH_2-$ cysteine	UGPy	 threonine	ACX
 proline	CCX	$CH_3SCH_2CH_2-$ methionine	AUG
Start ≡ formyl methionine	AUG GUG	Stop	UAPu UGA

[a]The structure of the side chain of each amino acid is given. X = any base, Pu = purine, Py = pyrimidine.

from amino acids to base triplets. The structures of the side groups of the amino acids are also shown in Table 1-2. Protein synthesis begins at an AUG or GUG which is recognized by a tRNA carrying *N*-formyl methionine.[6] As these triplets also code for amino acids in the interior of the polypeptide, it is clear that additional starting signals are present in messenger RNA.[7] The formyl methionine is subsequently removed from the completed polypeptide chain.

The translation of the genetic information takes place on ribosomes. The messenger RNA, the transfer RNAs and all the enzymes and protein factors interact with the ribosomes and each other to produce a polypeptide. More than one region of the messenger RNA is being translated at the same time so more than one ribosome is attached to the messenger RNA.[8] An electron microscope picture of these polyribosomes is shown in Fig. 1-1.

Both DNA and RNA must replicate in order to store and propagate the genetic information. Furthermore, RNA is sometimes transcribed into DNA.[9] Therefore, a more complete mnemonic for the molecular biology of nucleic acids is:

The physical chemistry of nucleic acids includes knowledge of the structures of all the polynucleotides involved. The conformations, or secondary and tertiary structures, at each step in the various reactions are of

FIGURE 1-1 The transcription of DNA into RNA and the translation of the RNA into proteins. [From O. L. Miller, Jr., B. A. Hamkalo, and C. A. Thomas, Jr., *Science*, **169**, 392 (1970). Reprinted with permission.]

particular importance. These structures determine what reactions occur and how fast they happen. The interactions of the nucleic acids with each other, with proteins, and with small molecules control the process.

In the succeeding four chapters we have presented the fundamentals of the determination of structure in nucleic acids. These include X-ray and light scattering, spectroscopic techniques, and hydrodynamic methods. There follow two chapters on the thermodynamics, statistical mechanics, and kinetics of reactions of nucleic acids. The last chapter is a detailed description of physical chemical studies on one type of ribonucleic acid: transfer RNA.

We have attempted to stress general principles and to cite particular results only when they seem firmly established.

The plan of this book has been to proceed from the simple to the complex; from monomers to oligomers to polymers; from nucleic acids in their native, unperturbed states to nucleic acids whose conformations have been perturbed by some chemical or physical agent; from "pure" nucleic acids to nucleic acids interacting with other molecules.

It is worthwhile to list some of the important topics we have covered only briefly, or not at all. These include the organic chemical, as opposed to the physical chemical, behavior of the nucleic acids, and the biological functioning and metabolism of the nucleic acids. To learn about these subjects and to obtain a more balanced view, the reader is referred to earlier books on nucleic acids. These include the three-volume set entitled *The Nucleic Acids* edited by E. Chargaff and J. N. Davidson (Vols. I, II, 1955; Vol. III, 1960, Academic Press); *The Chemistry of Nucleic Acids* by D. O. Jordan (1960, Butterworth); *Polynucleotides* by R. F. Steiner and R. F. Beers, Jr. (1960, Elsevier) and *The Chemistry of Nucleosides and Nucleotides* by A. M. Michelson (1963, Academic Press). A short and very useful summary of the biological function of nucleic acids at the molecular level is *The Biochemistry of Nucleic Acids* by J. N. Davidson (6th ed., 1970, John Wiley). To obtain a good historical perspective, *Nucleic Acids* by P. A. Levene and L. W. Bass (1931, Chemical Catalog Co.) is recommended.

The physical chemistry of nucleic acids combined with proteins and other macromolecules in functional complexes such as chromosomes, ribosomes, viruses, and mitochondria and membrane-bound replication complexes are not discussed in this book. These topics are very important and interesting. However, their study is just beginning; we leave them for the future. To keep abreast of these developments one must avidly read the daily newspapers and the weekly magazines such as *Science, Nature*, and *Time*. The more lasting scientific contributions are reviewed in such annual publications as *Progress in Nucleic Acid Research and Molecular Biology* edited by J. N. Davidson and W. E. Cohn (Academic Press). The annual *Cold Spring Harbor Symposia on Quantitative Biology* have often had very timely collections of articles on a particular aspect of nucleic acids.

REFERENCES

1 J. D. Watson, *The Double Helix*, Atheneum, New York (1968), p. 153.
2 J. D. Watson, *The Molecular Biology of the Gene*, W. A. Benjamin, New York (1965).
3 A. A. Travers and R. R. Burgess, *Nature,* **222**, 537 (1969).
4 J. W. Roberts, *Nature,* **224**, 1168 (1969).
5 E. M. Scolnick and C. T. Caskey, *Proc. Nat. Acad. Sci., U.S.,* **64**, 1235 (1969).
6 H. P. Ghosh, D. Söll, and H. G. Khorana, *J. Mol. Biol.,* **25**, 275 (1967).
7 J. A. Steitz, *Nature,* **224**, 957 (1969).
8 J. R. Warner, A. Rich, and C. E. Hall, *Science,* **138**, 1399 (1962).
9 H. M. Temin and S. Mizutani, *Nature*, **226**, 1211 (1970); D. Baltimore, *Nature,* **226**, 1209 (1970).

chapter 2
monomers

I CONFIGURATION AND CONFORMATION 10
A Introduction 10
B Bases 10
C Nucleosides 12
D Nucleotides 17

II ELECTRONIC PROPERTIES 19
A Introduction 19
B Charge Distributions 19
C Ultraviolet Optical Properties 22
D Infrared Spectroscopy and Raman Scattering 52
E Magnetic Resonance 56

III THERMODYNAMIC PROPERTIES 66
A Introduction 66
B Interactions in Water 66
C Ionization 75

I CONFIGURATION AND CONFORMATION

A. Introduction

The distinction between configuration and conformation made by the organic chemist is the one we will use here. Configuration refers to the covalent bonding in a molecule; conformation refers to the three-dimensional structure of a molecule. The configuration is constant for a particular molecule; the conformation depends on the environment (temperature, solvent, etc.) of the molecule. As a final distinction we remark that organic chemists and biochemists are usually more interested in configuration, while physical chemists are interested in conformation.

The large amount of chemical work necessary to establish configuration of the nucleic acid components is described in Volume I of *The Nucleic Acids.*[1] The conformations of the nucleosides and nucleotides are still not very well known. We will discuss what is known at present and describe the methods used to study the problem.

B. Bases

We will consider the four bases in DNA and RNA which are responsible for coding genetic information. These are adenine, cytosine, guanine, and