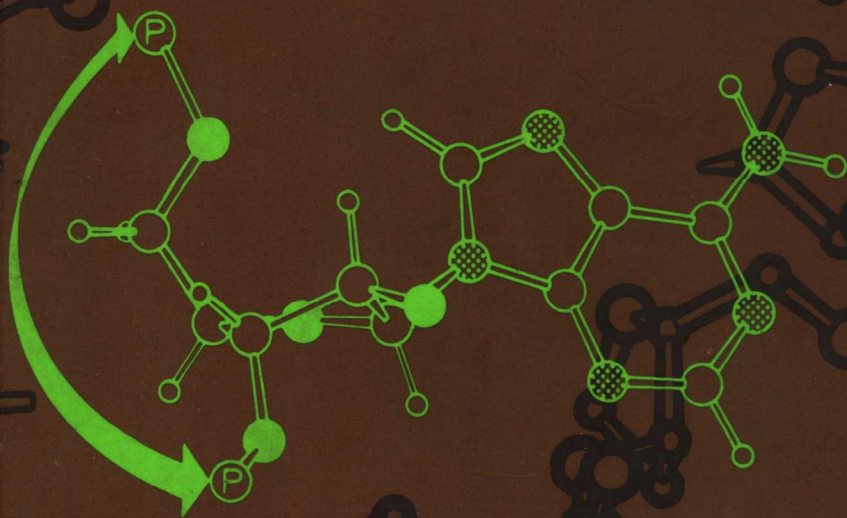


Wolfram Saenger

Principles of Nucleic Acid Structure



Springer-Verlag

New York Berlin Heidelberg Tokyo

Wolfram Saenger

Principles of Nucleic Acid Structure

With 227 Figures



Springer-Verlag
New York Berlin Heidelberg Tokyo

Wolfram Saenger
Institut für Kristallographie
Freie Universität
1000 Berlin 33
Federal Republic of Germany

Series Editor:

Charles R. Cantor
Department of Human Genetics and Development
College of Physicians & Surgeons of Columbia University
New York, New York 10032 U.S.A.

Library of Congress Cataloging in Publication Data
Saenger, Wolfram.

Principles of nucleic acid structure.

(Springer advanced texts in chemistry)

1. Nucleic acids. I. Title. II. Series.

QD433.S24 1983 547.7'9 82-19445

© 1984 by Springer-Verlag New York Inc.

All rights reserved. No part of this book may be translated or reproduced in any form without written permission from Springer-Verlag, 175 Fifth Avenue, New York, New York 10010, U.S.A.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Typeset by Progressive Typographers, Emigsville, Pennsylvania.
Printed and bound by R.R. Donnelley & Sons, Crawfordsville, Indiana.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-90762-9 Springer-Verlag New York Heidelberg Berlin (Hardcover Edition)
ISBN 3-540-90762-9 Springer-Verlag Berlin Heidelberg New York (Hardcover Edition)
ISBN 0-387-90761-0 Springer-Verlag New York Heidelberg Berlin (Softcover Edition)
ISBN 3-540-90761-0 Springer-Verlag Berlin Heidelberg New York (Softcover Edition)

Series Preface

New textbooks at all levels of chemistry appear with great regularity. Some fields like basic biochemistry, organic reaction mechanisms, and chemical thermodynamics are well represented by many excellent texts, and new or revised editions are published sufficiently often to keep up with progress in research. However, some areas of chemistry, especially many of those taught at the graduate level, suffer from a real lack of up-to-date textbooks. The most serious needs occur in fields that are rapidly changing. Textbooks in these subjects usually have to be written by scientists actually involved in the research which is advancing the field. It is not often easy to persuade such individuals to set time aside to help spread the knowledge they have accumulated. Our goal, in this series, is to pinpoint areas of chemistry where recent progress has outpaced what is covered in any available textbooks, and then seek out and persuade experts in these fields to produce relatively concise but instructive introductions to their fields. These should serve the needs of one semester or one quarter graduate courses in chemistry and biochemistry. In some cases the availability of texts in active research areas should help stimulate the creation of new courses.

New York

CHARLES R. CANTOR

Preface

This monograph is based on a review on polynucleotide structures written for a book series in 1976. The respective volume was, however, never published and the manuscript was kept in the drawer until Springer-Verlag decided to accept it as part of the Advanced Texts in Chemistry. When I looked at my 1976 manuscript again in 1980, I found that many views had changed. Left-handed DNA, above all, had been introduced and changed the notion of the occurrence of exclusively right-handed double-helices. As a matter of fact, the whole story had to be rewritten.

But even then, during the past two years many new discoveries were made and it often appeared to me that I was trying to catch a running train. Research on DNA structure is now booming, caused in part by the availability of synthetic oligonucleotides of defined sequence. They can be crystallized either per se or as complexes with certain metal ions, drugs, and proteins binding to them. And with the proteins again a new field has opened. We can now look at the three-dimensional structures of quite a number of DNA- and RNA-binding proteins. Work is in progress with more complicated systems involving the respective complexes, made possible by recent advances in crystallographic techniques.

I cannot close this short preface without thanking those who have contributed enthusiastically to this book. There are, above all, Meinhard Heidrich and Manfred Steifa who operated the Cambridge Data File to retrieve atomic coordinates. These were used in a combination of programs SCHAKAL (Ref. 1376) and ORTEP (Ref. 1377) to do most of the artistic plots of molecules displayed in this book. Drawings which could not be prepared on a computer came from the hand of Ludwig Kolb who did a magnificent and superb job on the rough drawings I gave him.

If writing a book, one has to have available a good library and people

associated with it. The library at the Max-Planck-Institut für Experimentelle Medizin is one of the best I have encountered and Mrs. M. Henschel, B. Küttner and Mr. W. Gellert helped to make available the more exotic journals. Finally, writing and rewriting many versions of the manuscript was done patiently by several secretaries, Chantal Celli, Jenny Glissmeyer and Petra Große contributing predominantly.

Last, but certainly not least, I would like to thank my scientific teacher Professor Friedrich Cramer for his guidance and support during the past 18 years.

Göttingen, August 1983

WOLFRAM SAENGER

Acknowledgments

Acknowledgment is gratefully made for permission to reproduce material as follows:

Figures 6-3, 6-15, 6-16, 6-17, 6-18, 11-4, 11-5, 11-6, 12-1, 12-3, 12-4, 12-8, 13-14, 15-2, 15-3, 16-3, 17-6, 17-7, and 18-11 are reprinted by permission from the *Journal of Molecular Biology*, copyright Academic Press Inc. (London) Ltd.

Table 3-2 is reprinted by permission from the *Journal of Theoretical Biology*, copyright Academic Press Inc. (London) Ltd.

Reprinted with permission from the *Journal of the American Chemical Society* are Figure 6-10, copyright 1967 American Chemical Society; Figures 2-8 and 15-10, copyright 1972 American Chemical Society; Figures 3-8 and 7-6, copyright 1977 American Chemical Society; Figures 4-6 and 13-13, copyright 1978 American Chemical Society.

Reprinted by permission from *Nature* are Figure 18-7, copyright 1955 Macmillan Journals Ltd; Figure 10-5, copyright 1971 Macmillan Journals Ltd; Table 6-10, copyright 1973 Macmillan Journals Ltd; Figure 4-24, copyright 1977 Macmillan Journals Ltd; Figure 4-29a and Table 4-4, copyright 1978 Macmillan Journals Ltd; Figure 17-10, copyright 1979 Macmillan Journals Ltd; Figures 11-3, 16-7, and 16-8, copyright 1980 Macmillan Journals Ltd; Figures 18-17 and 18-18, copyright 1981 Macmillan Journals Ltd; Figures 8-5 and 17-12, copyright 1982 Macmillan Journals Ltd.

Reprinted by permission from *Science* is Figure 1-3, copyright 1969 by the American Association for the Advancement of Science.

Reprinted by permission of the authors from *Proceedings of the National Academy of Science, U.S.A.* are Figures 3-5, 6-5, 6-6, 14-2, 14-3, 14-5, 16-6, 18-3, 18-5, 18-12, 19-2, and 19-7.

Figures 4-27, 6-14, 6-20, and 16-5 are reprinted by permission from *Biopolymers*, copyright John Wiley and Sons. Figure 15-1 is reprinted by permission from *Advances in Enzymology*, copyright John Wiley and Sons.

Figures 19-4, 19-6, 19-8 are from "The Nucleosome" by Roger D. Kornberg and Aaron Klug, copyright 1981 by Scientific American, Inc. All rights reserved. Box 19-3 is from "Supercoiled DNA" by William R. Bauer, F.H.C. Crick, and James H. White, copyright 1980 by Scientific American, Inc. All rights reserved.

Figure 19-1 is reprinted by permission from *Cell*, copyright M.I.T. Press.

Contents

Chapter 1

Why Study Nucleotide and Nucleic Acid Structure?	1
--	---

Chapter 2

Defining Terms for the Nucleic Acids	9
--------------------------------------	---

2.1 Bases, Nucleosides, Nucleotides, and Nucleic Acids—Nomenclature and Symbols	9
2.2 Atomic Numbering Scheme	13
2.3 Torsion Angles and Their Ranges	14
2.4 Definitions of Torsion Angles in Nucleotides	16
2.5 Sugar Pucker Modes: The Pseudorotation Cycle	17
2.6 <i>syn/anti</i> Orientation About the Glycosyl Bond	21
2.7 Orientation About the C ₄ —C _{5'} Bond	23
2.8 Helical Parameters: Hydrogen Bonding Between Bases	24
Summary	27

Chapter 3

Methods: X-Ray Crystallography, Potential Energy Calculations, and Spectroscopy	29
---	----

3.1 Crystal Structure Analysis of Small Molecules	29
3.2 Potential Energy Calculations	38
3.3 Crystallography of Macromolecules	42
3.4 Fiber Structure Determination	44
3.5 Spectroscopic Methods	45
Summary	50

Chapter 4

Structures and Conformational Properties of Bases,
Furanose Sugars, and Phosphate Groups

51

4.1	Geometry of Bases	51
4.2	Preferred Sugar Puckering Modes	55
4.3	Factors Affecting Furanose Puckering Modes	61
4.4	Bond Distances and Angles in Furanoses	65
4.5	<i>syn/anti</i> Conformation	69
4.6	The <i>high anti(-sc)</i> Conformation	73
4.7	Factors Affecting the <i>syn/anti</i> Conformation: The Exceptional Guanosine	76
4.8	The Orientation About the C ₄ –C _{5'} Bond	79
4.9	Factors Influencing the Orientation about the C ₄ –C _{5'} Bond	80
4.10	The “Rigid Nucleotide”	81
4.11	The Phosphate Mono- and Diester Groups and the Pyrophosphate Link: Bonding Characteristics and Geometry	82
4.12	Orientation About the C–O and P–O Ester Bonds	88
4.13	Correlated Rotations of Torsion Angles in Nucleotides and in Nucleic Acids	96
4.14	Helical or Not Helical—and if, What Sense?	97
	Summary	101

Chapter 5

Physical Properties of Nucleotides: Charge Densities,
pK Values, Spectra, and Tautomerism

105

5.1	Charge Densities	105
5.2	pK Values of Base, Sugar, and Phosphate Groups: Sites for Nucleophilic Attack	107
5.3	Tautomerism of Bases	110
	Summary	115

Chapter 6

Forces Stabilizing Associations Between Bases:
Hydrogen Bonding and Base Stacking

116

6.1	Characterization of Hydrogen Bonds	116
6.2	Patterns of Base–Base Hydrogen Bonding: The Symmetry of a Polynucleotide Complex	119
6.3	Detailed Geometries of Watson–Crick and Hoogsteen Base Pairs	122
6.4	The Stability and Formation of Base Pairs as Determined by Thermodynamic, Kinetic, and Quantum Chemical Methods: Electronic Complementarity	126
6.5	Patterns of Vertical Base–Base Interactions	132
6.6	Thermodynamic Description of Stacking Interactions	134

6.7 Forces Stabilizing Base Stacking: Hydrophobic Bonding and London Dispersion	137
6.8 Formation and Breakdown of Double-Helix Structure Show Cooperative Behavior	141
6.9 Base-Pair Tautomerism and Wobbling: Structural Aspects of Spontaneous Mutation and the Genetic Code	149
Summary	158

Chapter 7

Modified Nucleosides and Nucleotides; Nucleoside Di- and Triphosphates; Coenzymes and Antibiotics	159
7.1 Covalent Bonds Bridging Base and Sugar in Fixed Conformations: Calipers for Spectroscopic Methods	159
7.2 Cyclic Nucleotides	167
7.3 Nucleosides with Modified Sugars: Halogeno-, Arabino-, and α -Nucleosides	173
7.4 Modified Bases: Alkylation of Amino Groups (Cytokinins) and of Ring Nitrogen, Thioketo Substitution, Dihydrouridine, Thymine Dimers, Azanucleosides	178
7.5 The Chiral Phosphorus in Nucleoside Phosphorothioates	189
7.6 The Pyrophosphate Group in Nucleoside Di- and Triphosphates and in Nucleotide Coenzymes	193
7.7 Nucleoside Antibiotics: Puromycin as Example	197
Summary	199

Chapter 8

Metal Ion Binding to Nucleic Acids	201
8.1 Importance of Metal Ion Binding for Biological Properties of Nucleic Acids	201
8.2 Modes of Metal Ion Binding to Nucleotides and Preferred Coordination Sites	203
8.3 Platinum Coordination	210
8.4 Coordination of Metal Ions to Nucleoside Di- and Triphosphates: Nomenclature of Bidentate Λ/Δ and of Tridentate Λ/Δ endo/exo Chelate Geometry	211
Summary	219

Chapter 9

Polymorphism of DNA versus Structural Conservatism of RNA: Classification of A-, B-, and Z-Type Double Helices	220
9.1 Polymorphism of Polynucleotide Double Helices	221
9.2 The Variety of Polynucleotide Helices with Right-Handed Screw	

Classified into Two Generically Different Families: A and B	228
Summary	241
 Chapter 10	
RNA Structure	242
10.1 A-RNA and A'-RNA Double Helices Are Similar	242
10.2 RNA Triple Helices Simultaneously Display Watson-Crick and Hoogsteen Base-Pairing	243
10.3 A Double Helix with Parallel Chains and Hoogsteen Base-Pairs Formed by Poly(U) and 2-Substituted Poly(A)	246
10.4 Mini-Double Helices Formed by ApU and GpC	247
10.5 Turns and Bends in UpAH ⁺	251
Summary	252
 Chapter 11	
DNA Structure	253
11.1 A-DNA, The Only Member of the A Family: Three Crystalline A-Type Oligonucleotides d(CCGG), d(GGTATACC), and d(GGCCGGCC)	256
11.2 B-DNA Structures Exhibited by Polymeric DNA and by the Dodecanucleotide d(CGCGAATTCGCG): Introduction to B-Family Duplexes	261
11.3 "Alternating B-DNA" and the Tetranucleotide d(pATAT); d(TpA), Dinucleoside Phosphate Mimicking Double Helical Arrangement	266
11.4 The Conformationally Stiff Unique Poly(dA)·Poly(dT) Double Helix and Its Transformation into Triple Helix	271
11.5 C-DNA Double Helix Formed by Natural and Synthetic DNA	273
11.6 D-DNA Is Only Formed by Synthetic DNA with Alternating A, T-Sequence and by Phage T2 DNA	275
11.7 DNA-RNA Hybrids Restricted to RNA-Type Double-Helices: A and A'. Polymers and r(GCG) d(TATACGC). The B-DNA Form of Poly(A)·Poly(dT).	277
Summary	281
 Chapter 12	
Left-Handed, Complementary Double Helices— A Heresy? The Z-DNA Family	283
12.1 Crystal Structures of Oligo(dG-dC) Display Left-Handed Double Helix	284
12.2 Extrapolation from Oligo- to Polynucleotides. The Z-DNA Family: Z-, Z _I -, Z _{II} , and Z'-DNA	290

12.3	Left-Handed Z-DNA Visualized in Fibers of Three Alternating Polydeoxynucleotides	293
12.4	Factors Stabilizing Z-DNA	294
12.5	Does Z-DNA Have a Biological Significance?	295
	Summary	296

Chapter 13

	Synthetic, Homopolymer Nucleic Acids Structures	298
13.1	Right-Handed, Base-Stacked Single Helix Revealed for Poly(C) and the O ₂ -Methylated Analog	299
13.2	Bases Turned "in" and "out" in Nine- and Twofold Single-Stranded Helices of Poly(A)	302
13.3	A Double Helix with Parallel Strands for Poly(AH ⁺)-Poly(AH ⁺) Forms under Acidic Conditions. Helix, Loop, and Base-Pair Stacks in ApAH ⁺ pAH ⁺ Dimer	305
13.4	The Deoxydinucleotide d(pTpT) Suggests Single-Stranded Poly(dT) Helix with Nonstacked Bases Turned "out"	310
13.5	The Antiparallel, A-RNA-Type Double Helices of Poly(U), Poly(s ² U) and poly(X)	311
13.6	Sticky Guanosine-Gel Structure of Guanosine and Guanylic Acid: Quadruple Helix Formed by Poly(G) and Poly(I)	315
	Summary	320

Chapter 14

	Hypotheses and Speculations: Side-by-Side Model, Kinky DNA, and "Vertical" Double Helix	321
14.1	Side-by-Side Model—An Alternative?	321
14.2	Does DNA Fold by Kinking?	324
14.3	κ - and β -Kinked DNA: Breathing with the Speed of Sound	326
14.4	Bends in DNA at Junctions of A- and B-Type Helices	327
14.5	"Vertical" Double Helix for Polynucleotides in <i>high-anti</i> Conformation	330
	Summary	330

Chapter 15

	tRNA—A Treasury of Stereochemical Information	331
15.1	Primary and Secondary Structure of tRNA: The Cloverleaf	331
15.2	Folding of the Cloverleaf into Tertiary Structure: The L Shape	333
15.3	Stabilization of tRNA Secondary and Tertiary Structure by Horizontal and Vertical Base-Base Interactions	334
15.4	Change in Sugar Pucker, π Turn, and Loop with Phosphate-Base Stacking: Structural Features of General Importance	338

15.5	Some Stereochemical Correlations Involving Torsion Angles χ, δ and α, ζ	340
15.6	Metal and Polyamine Cation Binding to tRNA	343
15.7	Anticodon Preformed to Allow Rapid Recognition of Codon via Minihelix	346
	Summary	349
Chapter 16		
	Intercalation	350
16.1	General Phenomena of Intercalation into DNA and RNA Double Helices	350
16.2	Stereochemistry of Intercalation into DNA- and RNA-Type Dinucleoside Phosphates	353
16.3	Improving the Model: The Daunomycin-d(CpGpTpApCpG) Complex	360
16.4	Model Building Studies Extended to A- and B-DNA	361
16.5	DNA Saturated with Platinum Drug Unwinds into a Ladder to Produce L-DNA	362
16.6	Actinomycin D: An Intercalator Specific for the GpC Sequence	365
	Summary	365
Chapter 17		
	Water and Nucleic Acids	368
17.1	Experimental Evidence for Primary and Secondary Hydration Shells around DNA Double Helices	368
17.2	Different Hydration States Associated with A-, B-, and C-DNA	370
17.3	Solvent Accessibilities in A- and B-DNA	372
17.4	Theoretical Considerations	376
17.5	Hydration Schemes in Crystal Structures of A-DNA Tetramer and B-DNA Dodecamer Suggest Rationale for A \rightarrow B Transition	377
17.6	Water Pentagons in Crystalline Dinucleoside Phosphate Intercalation Complex: The Generalized Concept of Circular Hydrogen Bonds and of Flip-Flop Dynamics	380
	Summary	384
Chapter 18		
	Protein-Nucleic Acid Interaction	385
18.1	General Considerations about Protein-Nucleic Acid Interactions	386
18.2	Model Systems Involving Nucleic Acid and Protein Constituents	395

18.3	Model Systems Combining Nucleic Acids and Synthetic Polypeptides or Protamines	399
18.4	Nucleotides and Single-Stranded Nucleic Acids Adopt Extended Forms When Binding to Proteins	404
18.5	Nature of Protein-Nucleotide and Nucleic Acid Interaction and Recognition	405
18.6	Proteins Binding to DNA Double Helix and Single Strands	420
18.7	Prealbumin-DNA Interaction: A Hypothetical Model	430
	Summary	431

Chapter 19

Higher Organization of DNA 432

19.1	DNA Condensed into ψ -Form, Supercoils, Beads, Rods and Toroids	432
19.2	Lamellar Microcrystals Formed by Fragmented DNA	435
19.3	DNA in Cells in Organized in the Form of Chromosomes	436
19.4	Structure of the Nucleosome Core	440
19.5	Organization of Nucleosomes into 100 Å and 300 Å Fibers: The Super-Superhelix or Solenoid	446
19.6	Organization of Chromatin in Chromosomes: A Glimpse at Transcription	448
19.7	Topological Problems in Circularly Closed, Supercoiled DNA	449
	Summary	457

References 459

Index 541

Chapter 1

Why Study Nucleotide and Nucleic Acid Structure?

Before embarking on a description of nucleotide and nucleic acid structures, let us examine the biological importance of this class of molecules and find out why their structural principles should be known at the atomic level.

Nucleotides have many functions in living organisms. The hereditary material, deoxyribonucleic acid (DNA), is a linear polymer built up of monomeric units, the nucleotides. Even if these units were not constituents of DNA, they would nevertheless be among the most important molecules in biology.

A nucleotide consists of three molecular fragments: sugar, heterocycle, and phosphate. The sugar, ribose or deoxyribose, is in a cyclic, furanose form and is connected by a β -glycosyl linkage with one of four heterocyclic bases to produce the four normal nucleosides: adenosine, guanosine, cytidine, and thymidine (uridine in ribonucleic acid, RNA). If the 3'- or 5'-hydroxyl group of sugar is phosphorylated, we have a nucleotide. This unit, the nucleotide, is not only the building block of the polynucleotides DNA and RNA but it also exhibits independent functions.

For example, with adenosine derivatives displayed in Figure 1-1, we can show that, depending on chemical modifications, adenosine adapts to several, dramatically different biochemical roles in life. As di- and triphosphates, adenosine acts as an energy pool for many enzymatic processes and for muscle work. The importance of adenosine triphosphate is demonstrated by its turnover rate in humans: about one body weight per day per person. The 3', 5'-cyclic phosphate of adenosine is the "second hormonal messenger," controlling and mediating the activities of peptide hormones. In the form of puromycin, adenosine is a potent inhibitor of protein biosynthesis and as arabino- or 8-azaderivatives, adenosines display antibiotic activities. Adenosine diphosphate, equipped at the terminal phosphate with certain biological molecules, is a constituent of both coenzymes A and NAD⁺ and is essential for the proper functioning of enzymes which require these cofactors.

To understand the biological function of a nucleotide, we must know its structural features. With the adenosine derivatives mentioned above, we know the *chemical structure* and we know the function. But why and how do they function? Why, for instance, is 8-azaadenosine an antibiotic even though, from a chemical point of view, the isoelectronic substitution of

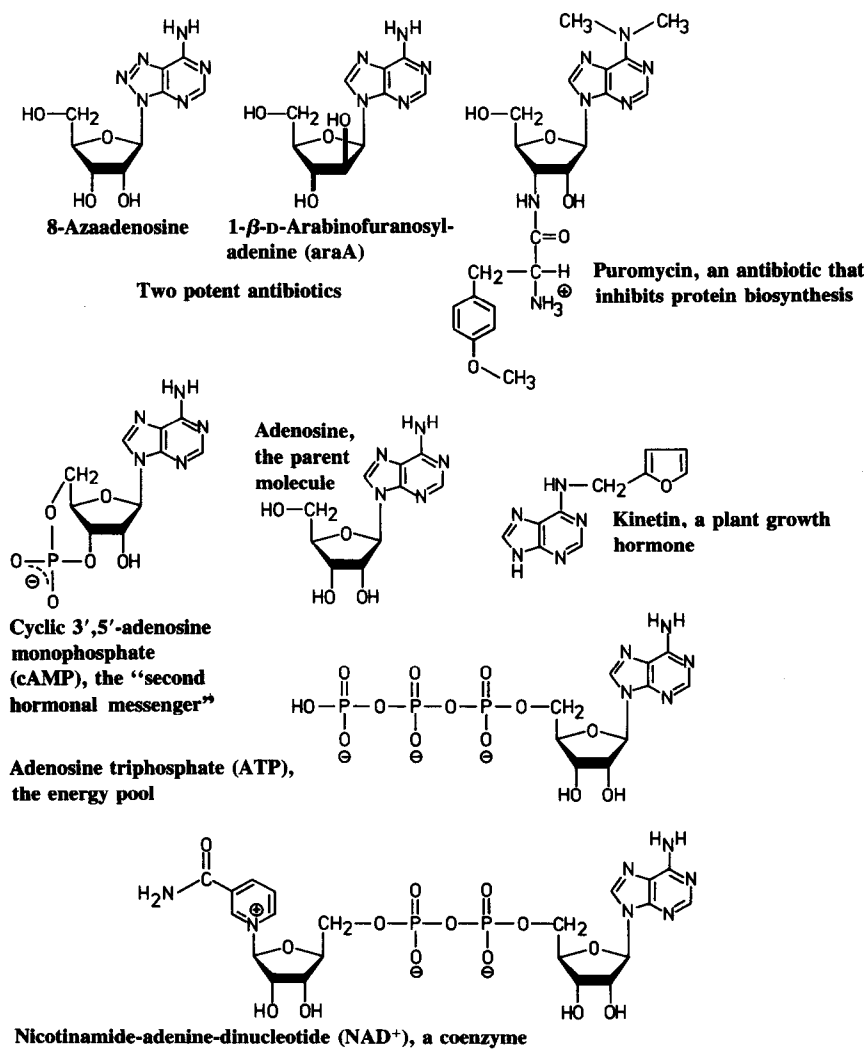


Figure 1-1. The properties of a biological molecule can change dramatically with only minor chemical modifications. This scheme displays adenosine (color) and some of its derivatives which exhibit very different biochemical behavior.

8-CH by an 8-aza group is expected to have only a modest influence on its chemistry? In this case it is not the electronic but rather the *three-dimensional structure* or conformation of the nucleoside which is changed and which explains 8-azaadenosine's particular properties.

Because nucleosides and nucleotides interact with proteins in all their metabolic or control operations, specific, mutual recognition of the two (or more) reactants is required. This presupposes that the partners involved have well-defined three-dimensional structures which, if we desire