Z. Shabarova, A. Bogdanov

Advanced Organic Chemistry of Nucleic Acids

Weinheim · New York
Basel · Cambridge · Tokyo

This Page Intentionally Left Blank

Z. Shabarova, A. Bogdanov

Advanced Organic Chemistry of Nucleic Acids

Distribution:
VCH, P.O. Box 101161, D-69451 Weinheim (Federal Republic of Germany)
Switzerland: VCH, P.O. Box, CH-4020 Basel (Switzerland)
United Kingdom and Ireland: VCH (UK) Ltd., 8 Wellington Court,
Cambridge CB1 1HZ (England)
USA and Canada: VCH, 220 East 23rd Street, New York, NY 10010–4606 (USA)
Japan: VCH, Eikow Building, 10-9 Hongo 1-chome, Bunkyo-ku, Tokyo 113 (Japan)

ISBN 3-527-29021-4 (VCH, Weinheim)

Z. Shabarova, A. Bogdanov

Advanced Organic Chemistry of Nucleic Acids

Weinheim · New York
Basel · Cambridge · Tokyo

Zoe A. Shabarova
Moscow State University
Korpus M, apt. 139
Moscow 117234
Russia

Alexey A. Bogdanov
Moscow State University
Korpus K, apt. 102
Moscow 117234
Russia

Published jointly by
VCH Verlagsgesellschaft mbH, Weinheim (Federal Republic of Germany)
VCH Publishers, Inc., New York, NY (USA)

Translator: Vladimir Vopyan
Editorial Director: Dr. Thomas Kellersohn
Production Manager: Dipl.-Ing. (FH) Hans Jörg Maier

Library of Congress Card No. applied for.

British Library Cataloguing-in-Publication Data: A catalogue record for this book is available from the British Library.

Deutsche Bibliothek Cataloguing-in-Publication Data:
Šabarova, Zoe A.:
Advanced organic chemistry of nucleic acids / Z. Shabarova ; A. Bogdanov. [Transl.: Vladimir Vopyan]. - Weinheim ; New York ; Basel ; Cambridge ; Tokyo : VCH, 1994
ISBN 3-527-29021-4
NE: Bogdanov, Aleksej A. :

Composition: Fa. Hagedornsatz GmbH, D-68519 Viernheim. Printing: betz-druck gmbh, D-64291 Darmstadt. Bookbinding: IVB, D-64646 Heppenheim. Cover design: Graphik & Text Studio Zettlmeier-Kammerer, D-93164 Laaber-Waldetzenberg.
Printed in the Federal Republic of Germany.

Dedicated to

Professor Mikhail A. Prokofiev,

the Teacher

This Page Intentionally Left Blank

Preface

In the early sixties, we started teaching an advanced course of chemistry of nucleic acids to Moscow University chemistry majors already with a solid organic and physical chemistry background. To teach this particular subject was most exciting at the time when virtually every year was marked by stunning discoveries in the field of nucleic acids. We still derive a great deal of pleasure and satisfaction from teaching the course.

Our main difficulty throughout the years has been the absence of a suitable textbook on nucleic acid chemistry. Of course, publication of Michelson's, Khorana's and other scientists' books, especially the monograph written by Kochetkov and coworkers, did help, but these books had been written with professionals, rather than students, in mind. Therefore, putting doubts and wavering aside, we decided to write a textbook ourselves. In making this decision we were wholeheartedly encouraged and supported by Professor M. A. Prokofiev, founder of nucleic acid chemistry in our country, to whom this book is dedicated.

The Russian edition of the textbook was published in 1978 by "Khimiya" (Moscow) under the title "The Chemistry of Nucleic Acids and Their Components". To the best of our knowledge, for many years it had been the only (and still is one of the few) textbook on the subject worldwide. However, access to the book has been limited outside our country. This is why we accepted with enthusiasm VCH's proposal to prepare the English edition.

Naturally, the English edition of a textbook to be published almost 15 years after the Russian predecessor could by no means be a mere translation of the latter – too many important events have taken place in DNA and RNA science not to leave an imprint on the chemistry of nucleic acids. Our greatest efforts have been spent in revising the chapters concerned with determination of the primary structure of nucleic acids, their synthesis, macromolecular structure, and chemical modification. Many chapters, including all of Chapter 10 dealing with ribozymes, have entirely been written anew. However, at the same time (much to our surprise), many fundamentals of nucleic acid chemistry are still as valid as 15 years ago, which is yet another proof of the great strides made in the field already in the seventies.

We are convinced that the chemistry of nucleic acids has always formed and still provides the basis for elaboration of methods as well as key concepts of molecular and cell biology. It should be remembered that, having emerged from organic and physical chemistry, the chemistry of nucleic acids is now exerting major influence on the two fields of knowledge. This gives us every hope that this book will be of use to all those who wish to become versed in both sciences.

Chapters 1 through 6, most of Chapter 9, and Chapter 11 have been written by Z. A. Shabarova, whereas A. A. Bogdanov is author of Chapters 7, 8, part of Chapter 9, and Chapter 10.

We should like to express our gratitude to Professor D. Soll (of Yale University) for giving us the impetus to embark on the English edition. We are thankful to our colleagues and co-workers from the Department of Chemistry of Natural Compounds, Moscow State University, for their assistance and support. In particular, we gratefully thank D. Chernov, I. Kozlov and N. Naryshkin for preparation of all formulas and some figures. Special thanks should be expressed to N. Naryshkin and I. Kozlov for their help with correction of the manuscript and proof-reading. And we also thank V. G. Vopian, our translator, for his cooperation and long patience.

Z. Shabarova
A. Bogdanov

Contents

1	**Structure of Nucleosides**	1
1.1	Introduction	1
1.2	Pyrimidine and Purine Bases	2
1.2.1	Pyrimidines	3
1.2.2	Purines	4
1.2.3	Nomenclature of Pyrimidines and Purines	4
1.2.4	Abbreviations	4
1.3	Carbohydrate Moieties of Nucleosides	5
1.4	Bonding Between Carbohydrate Moiety and Heterocyclic Base	7
1.4.1	Purine Nucleosides	7
1.4.2	Pyrimidine Nucleosides	9
1.5	Size of the Oxide Ring in the Sugar	10
1.6	Configuration of the Glycoside (Anomeric) Center	12
1.7	Nomenclature and Abridged Formulas of Nucleosides	15
1.8	Minor Nucleosides	18
1.9	Pseudouridine	23
1.10	Nucleoside Antibiotics	26
1.10.1	Purine Nucleosides	26
1.10.2	Pyrimidine Nucleosides	30
1.11	Other Nucleosides	31
	References	32
2	**Properties of Nucleosides**	33
2.1	Heterocyclic Bases	33
2.1.1	General Concepts	33
2.1.2	Tautomerism	35
2.1.3	Reactions with Electrophilic Reagents	39
2.1.4	Reactions with Nucleophilic Reagents	44
2.1.5	Addition Reactions	47

2.1.6 Reactions Involving Exocyclic Amino Groups 50
2.2 Reactions at the Carbohydrate Moiety 52
2.2.1 Substitution for Hydrogen Atoms in Hydroxyl Groups 52
2.2.2 Oxidation 56
2.3 Reactions Involving Heterocyclic Bases and the Carbohydrate Moiety 58
2.4 Stability of *N*-Glycosidic Bonds 60
2.4.1 Effect of the Heterocyclic Base Species 60
2.4.2 Effect of Substituents in the Carbohydrate Moiety 62
2.4.3 Mechanism of Hydrolysis of *N*-Glycosidic Bonds 63
2.5 Properties of Pseudouridine 66
References 70

3 **Structure of Nucleotides** 71
3.1 Introduction 71
3.2 Nomenclature and Isomerism 73
3.3 Structure of Nucleotides 77
3.3.1 Nucleoside 5′-Phosphates 77
3.3.2 Nucleoside 3′- and 2′-Phosphates 83
3.3.3 Nucleoside Cyclic Phosphates 86
3.3.4 Nucleoside 3′(2′),5′-Diphosphates 88
3.4 General Comments Regarding the Structure of Monomer Units in Nucleic Acids 89
References 92

4 Properties of **Nucleotides** 93
4.1 Acid-Base Behavior 93
4.1.1 Ionization of Bases 93
4.1.2 Ionization of Hydroxyl Groups in Pentose 96
4.1.3 Ionization of the Phosphate Group 96
4.2 Formation of Derivatives with Metals 98
4.3 Reactions at Heterocyclic Bases and Pentose 99
4.4 Some Properties of the Phosphate Group (General Concepts) . 105
4.4.1 Structure of the Phosphate Group and the Mechanism of Nucleophilic Substitution at the Phosphorus Atom 106
4.4.2 Catalysis of Nucleophilic Substitution at the Phosphorus Atom 116
4.4.3 Hydrolysis of Alkyl Phosphates 119
4.4.4 Cyclic Phosphates 125
4.4.5 β-Elimination Reactions 131
4.5 Reactions of Nucleotides Involving the Phosphate Group . . . 132
4.5.1 Chemical and Enzymatic Dephosphorylation 132
4.5.2 Migration of the Phosphate Group 134

4.5.3	Alkylation of the Phosphate Group	134
4.5.4	Activation of the Phosphate Group in Nucleotides. Synthesis of Some Derivatives with Respect to the Phosphate Group	135
4.5.5	Acylation of the Phosphate Group	144
4.6	Properties of Nucleotide Derivatives with Substituents in the Phosphate Group	145
4.6.1	Nucleoside Cyclic Phosphates	146
4.6.2	Alkyl Esters of Nucleotides	152
4.6.3	Mixed Anhydrides of Nucleotides	161
4.6.4	Amides of Nucleotides	169
4.7	Hydrolysis of *N*-Glycosidic Bonds	178
	References	180
5	**Primary Structure of Nucleic Acids**	181
5.1	Introduction	181
5.2	Major Types of Nucleic Acids, Their Isolation and Characteristics	182
5.2.1	DNA	182
5.2.2	RNA	183
5.3	Nature of Internucleotide Linkages	184
5.3.1	Internucleotide Linkage in DNA	185
5.3.2	Internucleotide Linkage in RNA	187
5.4	Nomenclature, Abridged Formulas and Abbreviations	193
5.5	Nucleotide Composition	196
5.5.1	Composition of DNA	197
5.5.2	Composition of RNA	198
5.6	Sequence of Nucleotide Units	198
5.6.1	Internucleotide Bond–Breaking Enzymes (Nucleases)	200
5.6.1.1	Ribonucleases	201
5.6.1.2	Deoxyribonucleases	204
5.6.1.3	Nonspecific Nucleases	206
5.6.2	Methods for Determining the Nucleotide Sequence in Oligonucleotides	208
	References	222
6	**Determination of the Primary Structure of Nucleic Acids**	223
6.1	Introduction	223
6.2	DNA Mapping	225
6.3	Restriction Endonucleases	227
6.4	Controlled Chemical DNA Cleavage Method (Chemical Sequencing)	230
6.4.1	Basic Principle of the Method	230

6.4.2	Chemical Methods for Specific Cleavage of Polydeoxyribonucleotide Chain	232
6.4.3	Obtaining Individual DNA Fragments and ^{32}P-Labeling	240
6.4.4	Polyacrylamide Gel Electrophoresis: High-Resolution System for Separating Oligo(poly)nucleotides According to Chain Length	245
6.4.5	Direct Reading of Nucleotide Sequence from Autoradiogram	246
6.4.6	Double-stranded DNA Sequencing Strategy	248
6.5	DNA Sequencing by Polymerase Copying Method (Enzymatic Sequencing Method)	252
6.5.1	Basic Principle of the Method. Priming and Termination of Enzymatic Synthesis of DNA Copies	252
6.5.2	Strategy for Sequencing Large Single-stranded DNAs by the Enzymatic Method	257
6.6	Comparison of Chemical and Enzymatic DNA Sequencing Methods	266
6.7	Determination of Nucleotide Sequences in RNA	268
6.7.1	Direct Chemical Sequencing of RNAs Labeled at the 3′ End	268
6.7.2	Direct Enzymatic Sequencing of RNA	274
6.7.3	Sequencing of High-Molecular Weight RNAs by Polymerase Copying with the Aid of Reverse Transcriptase	279
6.8	Solid-Phase Sequencing of DNA and RNA	281
6.9	Polymerase Chain Reactions in the Analysis of the Primary Structure of Nucleic Acids	283
6.10	Automation of the Nucleic Acid Sequencing Process	287
6.11	Computers in Nucleic Acid Sequencing	292
	References	301
7	**Conformation of Nucleic Acid Components. Macromolecular Structure of Polynucleotides**	303
7.1	Conformation of Nucleosides and Nucleotides	304
7.1.1	Heterocyclic Bases	304
7.1.2	Ribose and Deoxyribose	305
7.1.3	Orientation of the Heterocyclic Bases Relative to the Sugar	307
7.2	Intermolecular Interactions Between the Heterocyclic Bases	312
7.2.1	Complementary Interactions	314
7.2.2	Stacking Interactions	317
7.3	Conformation of Single-Stranded Oligonucleotides and Polynucleotides	319
7.4	Structure of Double- and Multiple-Stranded Polynucleotide Complexes	325

7.5	Complexes of Polynucleotides with Mono- and Oligonucleotides	331
	References	334
8	**Macromolecular Structure of DNA and RNA**	335
8.1	DNA	335
8.1.1	The Watson and Crick Model	335
8.1.2	Polymorphism of the Double Helix of DNA	337
8.1.3	Single-Crystal X-Ray Structures of DNA	341
8.1.4	Denaturation and Renaturation of DNA	343
8.1.5	Some Aspects of DNA Behavior in Solution	350
8.1.6	Supercoiling of DNA	351
8.1.7	Unusual DNA Structures	354
8.1.8	Interaction of Ligands with Double Helices of DNA	357
8.2	RNA	363
8.2.1	Secondary Structure of RNA. Fresco-Alberts-Doty Model	363
8.2.2	Elements of Secondary Structure	366
8.2.3	Macromolecular Structure of Transfer RNAs	371
8.2.4	Three-Dimensional Structure of High-Molecular Weight RNAs	381
	References	388
9	**Chemical Properties of Polynucleotides. Modification of Nucleic Acids**	389
9.1	Introduction	389
9.2	Reactions of Heterocyclic Bases of Polynucleotides	391
9.2.1	General Concepts	391
9.2.2	Modification of Heterocyclic Bases	398
9.2.2.1	Reactions at Carbon Atoms	398
9.2.2.2	Reactions at Pyridine Nitrogens	412
9.2.2.3	Reactions at Pyrrole Nitrogens and Exocyclic Amino Groups	417
9.3	Hydrolysis of *N*-Glycosidic Bonds	426
9.3.1	Direct Hydrolysis Methods	426
9.3.2	Indirect Hydrolysis Methods	427
9.4	Reactions of Carbohydrate Moieties	428
9.4.1	Substitution for Hydrogen Atoms in Hydroxyl Groups	428
9.4.2	Oxidation of 3′-Terminal *cis*-Glycol Group in RNA	433
9.5	Reactions of Nucleotide Phosphate Groups	434
9.5.1	Cleavage of Internucleotide Linkages	435
9.5.2	Cleavage of the Sugar Phosphate Backbone with the Aid of "Chemical Nucleases"	445
9.5.3	Reactions with Internucleotide Linkages Remaining Intact	448
9.6	Reactions of Terminal Phosphate Groups	449

9.7	Chemical Probing of Nucleic Acid Structure	455
9.7.1	Location of Modified Nucleotides	455
9.7.2	Systematic Chemical Probing of RNA Secondary and Tertiary Structure	458
9.7.2.1	Transfer RNA	458
9.7.2.2	Ribosomal RNA	462
9.7.2.3	Searching for Unusual RNA Structures	463
9.7.3	Studying Conformational Changes in DNA	465
9.7.4	Evaluation of Four-stranded G4-DNA	468
	References	469
10	**Catalytic Activity of Nucleic Acids**	471
10.1	RNA subunit of Ribonuclease P	471
10.2	*Tetrahymena* Ribozyme	474
10.3	Ribozymes Based on Self-Cleaving RNAs	483
	References	489
11	**Synthesis of Nucleic Acids**	491
11.1	Introduction	491
11.2	Historical Background	493
11.3	General Principles of Chemoenzymatic Synthesis of Duplex DNAs	495
11.3.1	Implementation of the Principle of Complementarity of Heterocyclic Bases in Nucleic Acids	495
11.3.2	Ways to Produce Duplex DNAs	498
11.3.3	Enzymatic Methods for Construction of Duplex DNAs	500
11.3.4	Chemical Methods for Assembling Duplex DNAs	501
11.4	Methods for Chemical Synthesis of Oligodeoxyribonucleotides	505
11.4.1	Nucleoside and Nucleotide Components for Synthesis by the Phosphodiester and Phosphotriester Methods	506
11.4.1.1	Blocking of the Hydroxyl Groups of Pentose and Amino Groups in Heterocyclic Bases	506
11.4.1.2	Protection of Phosphate Groups	509
11.4.1.3	Preparation of the Nucleoside and Nucleotide Components	509
11.4.1.4	Mechanisms of Internucleotide Linkage Formation	512
11.4.2	Chemical Synthesis of Oligodeoxyribonucleotides in Solution	515
11.4.2.1	Phosphodiester Method	515
11.4.2.2	Phosphotriester Method	518
11.4.2.3	Phosphite triester Method	527
11.5	Solid-Phase Synthesis of Oligodeoxyribonucleotides and Its Automation	529
11.5.1	Basic Principle of Solid-Phase Synthesis	531

11.5.2 Polymer Supports and Immobilization of the First Monomer Thereon 532
11.5.3 Phosphoramidite-Triester Method 534
11.5.4 Hydrophosphoryl Method 537
11.5.5 Phosphotriester Method 540
11.5.6 Automation of Synthesis 541
11.5.7 Multiple Simultaneous Synthesis on Polymer Segments 542
11.5.8 Separation and Purification of Synthetic Oligonucleotides . . . 545
11.6 Synthetic Genes and Their Cloning 546
11.6.1 Methods of Duplex DNA Construction 546
11.6.2 Synthesis of Genes and Their Functionally Important Segments 546
References 573

Index 575

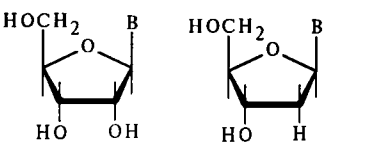
This Page Intentionally Left Blank

1 Structure of Nucleosides

1.1 Introduction

The term "nucleoside" was introduced in 1909 by Levene and Jacobs to denote carbohydrate derivatives of the purine bases isolated from yeast nucleic acid hydrolysates. Later, the term was expanded to additionally cover compounds containing pyrimidines and other heterocyclic bases. Today, it is applied to a broad class of naturally occurring and synthetic compounds that are essentially *N*- and *C*-glycosides – derivatives of various carbohydrates and heterocyclic compounds.

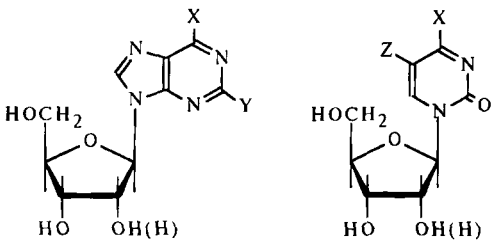
Nucleic acids contain nucleosides of two types: derivatives of D-ribose, known as ribonucleosides, and those of 2-deoxy-D-ribose, known as deoxyribonucleosides or, sometimes, deoxynucleosides.



ribonucleoside deoxyribonucleoside

Such classification of nucleosides stems from the nature of their constituent sugar. According to the structure of the heterocyclic base (B), which is the other component of nucleosides, the latter are also divided into pyrimidine and purine bases (see Tables 1-1 through 1-4)[1].

[1] If X and Y = OH, then the base exists in the oxo form.



X = OH or NH_2; Y = H,OH or NH_2; Z = H or CH_3

Nucleosides can be isolated from nucleic acids (NA) after chemical or enzymatic hydrolysis. Ribonucleic acids (RNA) are hydrolysed to nucleosides when boiled with aqueous pyridine, with a diluted ammonia solution or in an ammonium formiate buffer at pH 4.0. Nucleotides are yielded as intermediates:

$$\text{RNA} \rightarrow \text{Ribonucleotides} \rightarrow \text{Ribonucleosides}$$

Deoxynucleosides are produced by the enzymatic hydrolysis of deoxyribonucleic acids (chemical hydrolysis being accompanied by some side processes). To this end, use is normally made of snake venom containing the enzymes phosphodi- and phosphomonoesterase.

Some nucleosides occur in a free state and can be isolated by direct extraction.

The isolation and identification of nucleosides usually involve ion-exchange, thin-layer and gas–liquid chromatography as well as UV spectroscopy and mass spectrometry.

Acid hydrolysis of the nucleosides present in nucleic acids yields a heterocyclic (pyrimidine or purine) base and pentose. Nucleosides do not react at the aldehyde group, the implication being that the linkage between the sugar and bases is via a glycosidic bond. To establish the structure of each nucleoside present in a nucleic acid one must determine: (1) the structure of the constituent base; (2) the structure of the constituent sugar; (3) the type of the bond linking the two (the site at which the sugar is attached to the base); (4) the size of the oxide ring in the carbohydrate moiety; and (5) the configuration of the glycosyl (anomeric) moiety.

1.2 Pyrimidine and Purine Bases

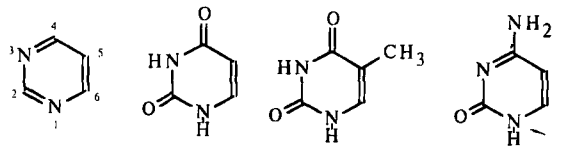
Pyrimidines and purines, first isolated from hydrolysates of nucleic acids (1874–1900), were identified using classical methods of organic chemistry (see Table 1-1). An important contribution was made by Emil Fischer who must

be credited with the earliest synthesis of purines (1897). Today, pyrimidines and purines can be identified rapidly and reliably by chromatographic and spectrophotometric techniques.

1.2.1 Pyrimidines

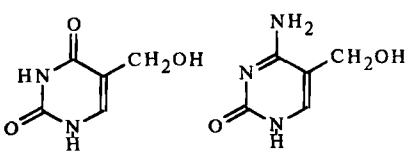
Pyrimidines, also known as *meta*-diazines, are structurally akin to benzene and pyridine. Pinner (1885), who had noticed this analogy, coined a new term from the words "pyridine" and "amidine", emphasizing thereby that apart from aromaticity pyrimidines also exhibit properties inherent in amidines. Pyrimidines are numbered according to the Chemical Abstracts Service registry.

Given below are the formulas of the pyrimidines whose compounds are nucleic acid constituents, such as uracil, thymine, and cytosine (see Table 1-1).



pyrimidine (*m*-diazine) uracil thymine cytosine

Thymines and cytosines are usually present in DNA, while those of uracil and cytosine are found in RNA. The DNAs of some phages are exceptional (see Table 1-4) in that they contain 5-hydroxymethylcytosine or its glycosides associated with the 5-hydroxymethyl group instead of cytosine (even-numbered T phages) and 5-hydroxymethyluracil (phage SP8) or uracil (phage PBS1) instead of thymine.



5-hydroxymethyluracil 5-hydroxymethylcytosine

Uracil, thymine, and cytosine are ubiquitously present in nucleic acids in the form of the corresponding nucleosides in significant amounts (each accounting for at least 5 % of the total bases in nucleic acids).

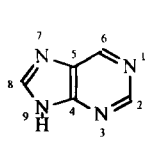
These compounds are usually referred to as the major pyrimidine bases of nucleic acids. In certain DNA and RNA species, some uracil and cytosine derivatives (usually *N*-alkyl ones) have been found. Such pyrimidine bases are referred to as rare or minor. Their structural formulas (as part of the corresponding nucleosides) are given in Tables 1-5 and 1-6. Minor pyrimidine

bases do not occur in all nucleic acids, and the content of each base is usually below one or two per cent.

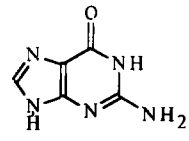
1.2.2 Purines

Purines are heterocyclic systems consisting of a pyrimidine and an imidazole condensed at the 4–5 bond. The term "purine" (from "purum" and "uricum") was introduced in 1898 by Emil Fischer. Purines are also numbered according to the Chemical Abstracts Service registry.

Both DNA and RNA contain two major purine substituents – adenine and guanine:


purine


adenine


guanine

Minor purines differing from adenine or guanine by the presence of alkyl (more commonly, methyl), acyl and other groups have also been isolated from some nucleic acids. The structural formulas of the minor purines present in nucleosides are presented in Tables 1-5 and 1-6.

1.2.3 Nomenclature of Pyrimidines and Purines

In the case of natural pyrimidines and purines, the above-mentioned common names are widely used. As regards synthetic bases and various analogues or modifications of natural pyrimidines and purines, specialists resort to the nomenclature usually applicable to heterocyclic bases, with appropriate numbering of atoms in the pyrimidine or purine ring. For example, thymine is called 2,4-dioxo-5-methyltetrahydropyrimidine or 2,4-dihydroxy-5-methylpyrimidine, whereas guanine is called 2-amino-6-oxodihydropurine or 2-amino-6-hydroxypurine.

Since the plane of symmetry in unsubstituted pyrimidine passes through C^2 and C^5, positions 4 and 6 are equivalent.

1.2.4 Abbreviations

Bases are very often designated by abbreviations (symbols) when writing structural formulas of nucleosides and their derivatives. Table 1-1 lists the names and symbols of the major pyrimidine and purine bases constituting RNA and DNA.

Table 1-1. Names and Abbreviated Symbols of Major Pyrimidines and Purines Constituting Nucleic Acids.

Name	Symbols one-letter	three-letter
Uracil (2,4-dioxotetrahydropyrimidine)	U	Ura
Thymine (2,4-dioxo-5-methyltetrahydropyrimidine)	T	Thy
Cytosine (2-oxo-4-aminodihydropyrimidine)	C	Cyt
Adenine (6-aminopurine)	G	Gua
Guanine (2-amino-6-oxodihydropurine)	A	Ade
Pyrimidine	–	Py*
Purine	–	Pu*

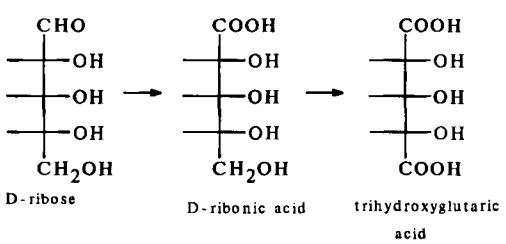
* Use is made of two-letter symbols, although three-letter ones (Pyr, Pur) can also be used.

The symbolic notation of the formulas in this and other chapters follows the recommendations of the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry (IUB) and has been approved by the 3rd All-Union Working Session on Nucleotide Chemistry.

1.3 Carbohydrate Moieties of Nucleosides

The nuclesides isolated from nucleic acids contain only two simple sugars, D-ribose and 2-deoxy-D-ribose, which, as has already been mentioned, determine the nucleic acid type (RNA and DNA).

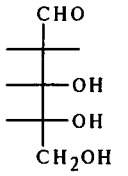
As far back as 1891, Albrecht Kossel pointed out that hydrolysis of RNA (this name was non-existent at that time) yielded a carbohydrate which was isolated in a crystalline state 18 years later and identified as a yet unknown sugar D-ribose. When oxidized under certain conditions, this simple sugar underwent conversion first into D-ribonic acid and then into optically inactive trihydroxyglutaric acid.



D-ribose D-ribonic acid trihydroxyglutaric acid

The structure of D-ribose was confirmed by its synthesis. Synthetic D-ribose turned out to be identical with the crystalline simple sugar isolated from RNA nucleosides.

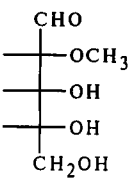
Serious difficulties had to be overcome in establishing the nature of the sugar contained in the nucleosides that are the building blocks of DNA (i. e. deoxyribonucleosides). It was found that the monosaccharide was more labile than D-ribose. Another finding was that it formed a hydrazone readily soluble in water, but no osazone, which rendered isolation of its hydrolysates difficult. It was long believed that this sugar was a hexose because hydrolysis of nucleosides isolated from DNA yielded levulinic acid. In 1929, however, it became possible to isolate, through mild hydrolysis, a crystalline carbohydrate from the guanine nucleoside of DNA, which turned out to be deoxypentose. Its structure was established beyond any doubt after the synthesis of 2-deoxy-L-ribose. Both preparations exhibited similar porperties as well as identical absolute, but opposite in sign, values of optical rotation. The conclusion that followed was that the sugar found in deoxyribonucleosides was 2-deoxy-D-ribose.



2-deoxy-D-ribose

It was also established that mild acid hydrolysis of DNA in the presence of benzyl mercaptan yielded benzyl mercaptal of 2-deoxy-D-ribose, which can be isolated in crystalline form. As was demonstrated in the fifties by chromatographic analysis, the only sugar in DNAs of plant, animal and bacterial origin is 2-deoxy-D-ribose. In 1954, a full description of 2-deoxy-D-ribose isolated from different types of DNA was provided by comparison with its synthetic analogue (melting point of mixed sample, optical rotation, synthesis of derivatives of known structure).

In the late fifties, an unusual carbohydrate was found in some preparations of RNA extracted from plants and animals, which was identified as 2-*O*-methyl-D-ribose.



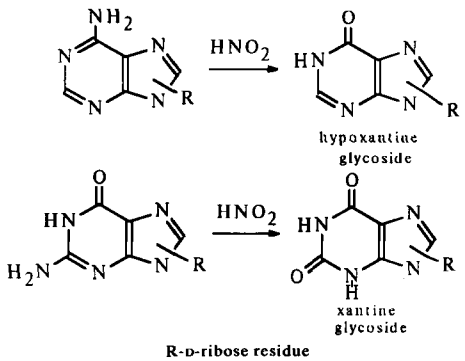
2-*O*-methyl-D-ribose

As was established later, 2-*O*-methyl-D-ribose is present in very small amounts in certain RNAs, which is why this sugar is referred to as a rare or minor component of RNA.

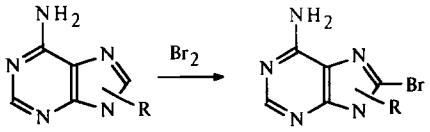
1.4 Bonding Between Carbohydrate Moiety and Heterocyclic Base

1.4.1 Purine Nucleosides

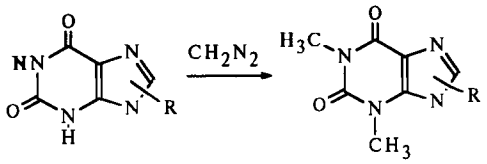
It is logical to assume that adenines and guanines in nucleosides are linked to the carbohydrate moiety in the same manner. This is why bonding at the oxygen atom is impossible in the case of guanine. The ribosides of adenine and guanine readily lend themselves to deamination in the presence of nitrous acid with the linkage between the base and carbohydrate remaining intact.



Consequently, the possibility of bonding at amino groups is also ruled out. Thus, the sugar moiety may be linked to adenines and guanines at C^8 (C-C bond) or at one of the nitrogens (C-N bond) of the pyrimidine (N^1 or N^3) or imidazole (N^7 or N^9) rings. These two types of bonds must differ markedly in stability under conditions of acid hydrolysis, since it is known that *N*-glycosides undergo hydrolysis rather easily, whereas *C*-glycosides are extremely stable. The ease of hydrolysis of purine nucleosides attests to their being *N*- rather than *C*-glycosides. This is also corroborated by the fact that the adenine and guanine nucleosides can be converted into C^8-substituted derivatives (the substituent at the carbon in the imidazole ring is absent). For example, adenine riboside is readily brominated at the imidazole ring carbon:

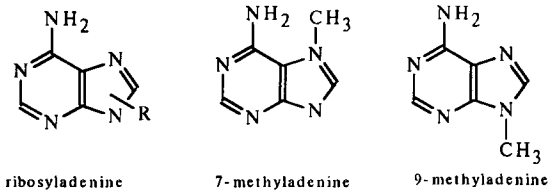


Out of the four nitrogens of the purine base those in the pyrimidine ring (N^1 and N^3) are excluded because methylation of the xanthine riboside yielded, as has already been mentioned, by deamination of guanine riboside gives theophylline (1,3-dimethyl xanthine) riboside.

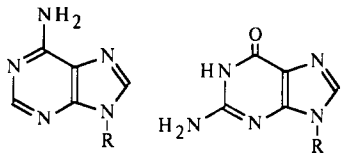


The nitrogens at positions 7 and 9 can be regarded as possible binding sites.

The question as regards the site of sugar to base binding was answered once and for all by Gulland in the forties. He was the first to take advantage of the fact that the UV spectra of the purine base containing an alkyl radical and a carbohydrate moiety at the same nitrogen atom are identical and markedly dependent on the position of the substituted nitrogen in the heterocyclic nucleus. Comparison of the UV spectra of adenine riboside with those of 7- and 9-methyladenine showed that the adenine nucleoside has a UV spectrum that is virtually identical with that of 9-methyladenine and bears no resemblance to that of 7-methyladenine.



Similarly, the UV spectrum of guanine riboside looks like the spectrum of 9-methylguanine and differs from that of 7-methylguanine. These findings clearly indicate that purine nucleosides are 9-ribosylpurines. The structures of adenine and guanine deoxyribosides are very much the same.

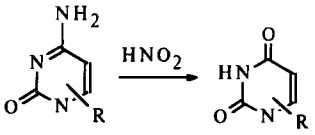


R - D-ribose or 2-deoxy-D-ribose residue

Rigorous proof that the sugar moiety in purine nucleosides is at position 9 was provided by the synthesis of the respective nucleosides, conducted by Todd and coworkers.

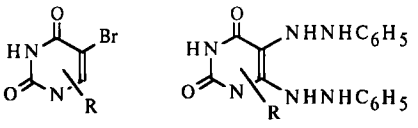
1.4.2 Pyrimidine Nucleosides

In the case of pyrimidine nucleosides, the involvement of the substituent at position 4 in the glycosidic bond should be ruled out in view of the fact that deamination of ribosylcytosine may yield ribosyluracil:



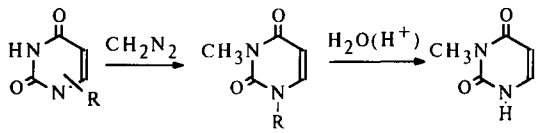
R - monosaccharide residue

The greater stability of pyrimidine nucleosides during acid hydrolysis, as compared to *O*-glycosides, suggests that the oxygen at position 2 does not participate in the formation of a bond with the carbohydrate moiety either. The same applies to the carbons at positions 5 and 6, since the substitution products associated with C^5 and C^6 [5-bromo- and 5,6-di(phenylhydrazyl)-nucleosides] have been derived from a natural uracil nucleoside.



R-D-ribose residue

Consequently, only N^1 and N^3 of the pyrimidine ring can be linked to the sugar. Methylation of ribosyluracil gives mono-*N*-methylribosyluracil (uracil is converted into 1,3-dimethyluracil under the same conditions) which breaks down to 3-methyluracil during acid hydrolysis:



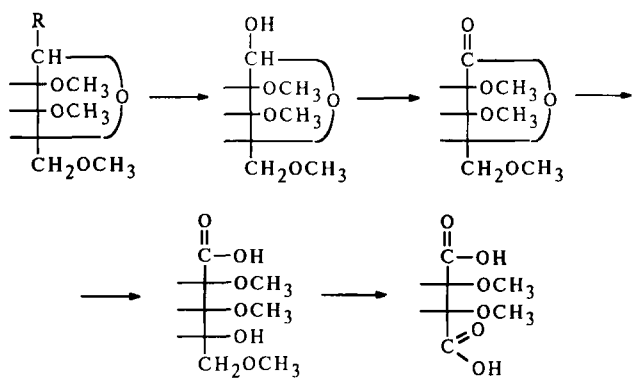
This is indicative of the fact that uracil nucleoside, just like cytosine nucleoside, is 1-riboside.

The similarity between the UV spectra of cytosine riboside and deoxyriboside, as well as the isolation of 3-methylthymine from DNA methylation products, attest to linkage between the sugar of deoxynucleosides and N^1 of the pyrimidine ring.

The conclusions as to the site of ribose binding in nucleosides have been corroborated by X-ray structure analysis.

1.5 Size of the Oxide Ring in the Sugar

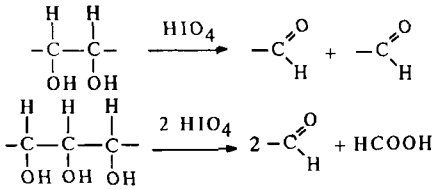
The structure of the oxide ring in ribonucleosides has been established by the methylation method. 2,3,5-*O*-Trimethyl-D-ribofuranose has been isolated from an acid hydrolysate of methylated nucleosides, its oxidation yielding at first trimethyl-D-ribonolactone and then *meso*-dimethoxysuccinic acid:



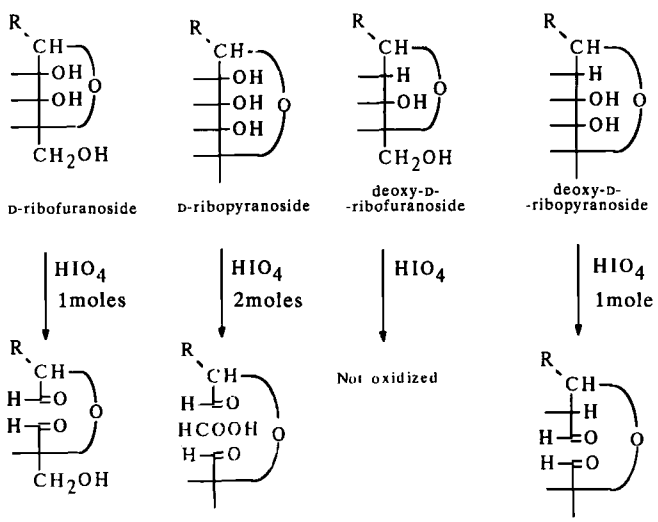
The furanose structure of ribonucleosides is also borne out by their response to triphenylchloromethane which, as is well known, reacts predominantly with the primary hydroxy groups of the carbohydrate, giving triphenylmethyl (trityl) esters. All nucleosides isolated from DNA and RNA form 5′-*O*-trityl derivatives with this reagent.

A convenient and simple way to determine the size of the ring in the sugar is through oxidation with periodic acid. This method which was developed for *O*-glycosides was applied to nucleosides by Todd and coworkers. It is based

on the possibility of oxidative cleavage of the C-C bond in 1,2-diols and similar compounds in the presence of periodic acid:



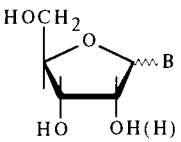
The reaction proceeds quantitatively in aqueous solutions. Measurement of the spent amount of periodic acid by iodometric titration may give an indication whether the glycoside is a furanoside or a pyranoside. Given below are some data concerning the composition of the oxidation products and the oxidizing agent requirements for furanosides and pyranosides of D-ribose and 2-deoxy-D-ribose:



The experimental data on oxidation of ribonucleosides and deoxyribonucleosides indicate that the only structure these compounds can have is that of furanose. The oxidation of nucleosides with periodic acid has provided the basis for a micromethod to determine the size of the carbohydrate ring with chromatographic analysis of the dialdehyde reduction (and hydrolysis) products. The method permits the carbohydrate ring size to be determined at very low concentrations of the substances under investigation.

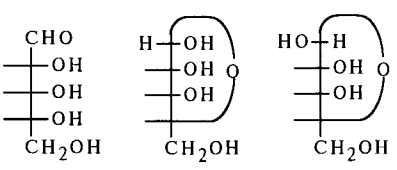
1.6 Configuration of the Glycoside (Anomeric) Center

After it had been definitively demonstrated (in the early fifties) that nucleosides have the furanose structure, the way their formulas used to be written was revised. Fischer's formulas for carbohydrates are too cumbersome and fail to adequately represent the three-dimensional configuration of their molecules. Instead, use was made of so-called "perspective" formulas proposed by Haworth, in which the convex portion of the carbon chain projects upward from the page, while the oxygen of the furanose ring lies in the background. The molecule is pictured in compliance with the laws of perspective, as is shown below (when the ribofuranose ring is represented in this fashion, it is conventionally assumed to be two-dimensional.



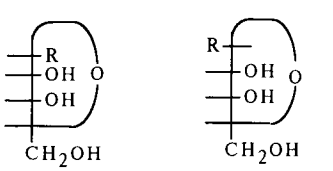
It would now be appropriate to consider the interrelation between Fischer's projection formulas and Haworth's perspective ones, which is of paramount importance for writing the formulas of nucleosides in the context of stereochemistry of the glycoside (anomeric) center.

The emergence of a new asymmetric carbon (C^1) when the ribofuranose ring of D-ribose is closed gives rise to two stereoisomeric forms: α-anomer with *cis*-configuration at C^1 and C^2 and β-anomer with *trans*-configuration at the same carbons.



α-D-ribose (α-anomer) β-D-ribose (β-anomer)

Corresponding to these anomers are two series of glycoside derivatives, namely:



α-D-ribofuranoside β-D-ribofuranoside

R-aglycone residue