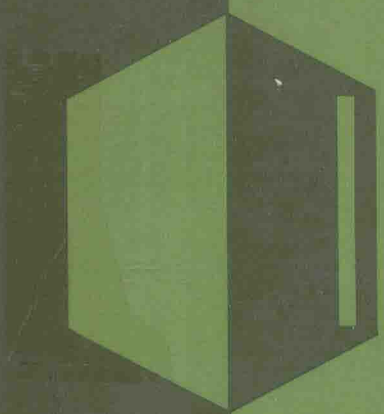
THE BIOCHEMISTRY OF THE NUCLEIC ACIDS

Ninth Edition

R.L.P. Adams, R.H. Burdon,
A.M. Campbell, D.P. Leader,
R.M.S. Smellie



The Biochemistry of the Nucleic Acids

NINTH EDITION

R. L. P. Adams et al.
R. H. Burdon
A. M. Campbell
D. P. Leader
R. M. S. Smellie

Department of Biochemistry, University of Glasgow

LONDON NEW YORK
CHAPMAN AND HALL

First published, 1950 by Methuen and Co. Ltd
Second edition, 1953 Third edition, 1957
Fourth edition, 1960 Fifth edition, 1965
Sixth edition, 1969 Seventh edition, 1972

Published by Chapman and Hall Ltd
11 New Fetter Lane, London EC4P 4EE
First published as a Science Paperback, 1972
Eighth edition, 1976 Ninth edition, 1981

Published in the USA by Chapman and Hall
in association with Methuen, Inc.
733 Third Avenue, New York NY 10017

Filmset by Enset Ltd.
Midsomer Norton, Bath

Printed in Great Britain by
Fletcher and Son Ltd., Norwich

ISBN 0 412 22680 4 (cased)
ISBN 0 412 22690 1 (paperback)

British Library Cataloguing in Publication Data

The biochemistry of the nucleic acids.—9th ed.
1. Nucleic acids
I. Title II. Adams, R.L.P.
574.87'328 QP620 80-42129

ISBN 0-412-22680-4
ISBN 0-412-22690-1 Pbk

The Biochemistry of the Nucleic Acids

Preface

When the first edition of this book was published in 1950, it set out to present an elementary outline of the state of knowledge of nucleic acid biochemistry at that time and it was the first monograph on the subject to appear since Levene's book on Nucleic Acids in 1931. The fact that a ninth edition is required after thirty years and that virtually nothing of the original book has been retained is some measure of the speed with which knowledge has advanced in this field.

'The Child's Guide to the Nucleic Acids' as it is known within the Department in Glasgow is still intended primarily as an introduction to the subject for advanced undergraduates in biochemistry and molecular biology, for graduates embarking upon studies in the field of nucleic acids, for chemists seeking to find some understanding of the more biological aspects of the subject and for biologists who require some knowledge of the chemical and molecular aspects.

The first seven editions emerged from the pen of the late J.N. Davidson who died in September 1972 shortly after completing the seventh edition. The eighth edition was revised extensively by four of his colleagues who recognized the need for a book giving a reasonably comprehensive coverage of the field at an up to date but elementary level. In doing so an attempt was made to retain something of the character and structure of the earlier editions while at the same time introducing new ideas and concepts and eliminating some of the more out-dated material.

Progress between 1976 and 1980 has been even more rapid than in the previous four years and the ninth edition has undergone very extensive revision not only in the content of individual chapters but also in general organization and layout. With a large amount of additional material to present the book has grown in size but every effort has been made to keep the increase within bounds by excluding non-essential detail. In a field in which new developments are occurring so rapidly it is inevitable that new knowledge will accumulate more quickly than it can be embodied in a new edition but we have endeavoured to incorporate into this edition material published up until the date of completion of the manuscript in September 1980.

It is a pleasure to express our thanks to those who have allowed us to

reproduce figures, diagrams and plates. In particular we should like to thank

Dr Lesley Coggins for Plates I and V
Drs D.E. Oling and A.L. Oling for Plate II
Dr P.R. Cook and Dr S.J. McCready for Plate III
Dr U. Scheer and Dr John Sommerville for Plate VI
Dr S.L. McKnight and Dr O.L. Miller for Plate VII
Dr J.A. Lake for the model used in Plate VIII

We are particularly grateful to Mr Ian Ramsden of the Department of Medical Illustration and Photography in the West Medical Building of Glasgow University and to Miss J.M. Gillies, Mrs J. Greenwood and Mrs C.M. Dow for valuable secretarial assistance.

R.L.P.A.
R.H.B.
A.M.C.
D.P.L.
R.M.S.S.

September 1980

Abbreviations and nomenclature

The abbreviations employed in this book are those approved by the Commission on Biochemical Nomenclature (CBN) of the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry (IUB).

Nucleosides

A	adenosine
G	guanosine
C	cytidine
U	uridine
ψ	5-ribosyluracil (pseudouridine)
I	inosine
X	xanthine
T	ribosylthymine (ribothymidine)
N	unspecified nucleoside
R	unspecified purine nucleoside
Y	unspecified pyrimidine nucleoside
dA	2′-deoxyribosyladenine
dG	2′-deoxyribosylguanine
dC	2′-deoxyribosylcytosine
dT	2′-deoxyribosylthymine (thymidine)

Minor nucleosides (when in sequence)

m^1A	1-methyladenosine
m_2^6A	N^6-dimethyladenosine
iA	N^6-isopentenyladenosine
m^5C	5-methylcytidine
ac^4C	N^4-acetylcytidine
m^1G	1-methylguanosine
m^2G	N^2-methylguanosine
m_2^2G	N^2-dimethylguanosine
m^1I	1-methylinosine
Cm	2′-*O*-methylcytidine
Gm	2′-*O*-methylguanosine

Um	2′-*O*-methyluridine
hU	5,6-dihydrouridine

Nucleotides

AMP	adenosine 5′-monophosphate
GMP	guanosine 5′-monophosphate
CMP	cytidine 5′-monophosphate
UMP	uridine 5′-monophosphate
dAMP	2′-deoxyribosyladenine 5′-monophosphate
dGMP	2′-deoxyribosylguanine 5′-monosphosphate
dCMP	2′-deoxyribosylcytosine 5′-monophosphate
dTMP	2′-deoxyribosylthymine 5′-monosphosphate
2′-AMP, 3′-AMP, 5′-AMP etc.	2′-, 3′- and 5′-phosphates of adenosine etc.
ADP etc.	5′-(pyro) diphosphates of adenosine etc.
ATP etc.	5′-(pyro) triphosphates of adenosine etc.
ddTTP	2′, 3′-dideoxyribosylthymine 5′- triphosphate
araCTP	1-β-D-arabinofuranosylcytosine 5′-triphosphate

Polynucleotides

DNA	deoxyribonucleic acid
cDNA	complementary DNA
mtDNA	mitochondrial DNA
RNA	ribonucleic acid
mRNA	messenger RNA
rRNA	ribosomal RNA
tRNA	transfer RNA
nRNA	nuclear RNA
hnRNA	heterogeneous nuclear RNA
snRNA	small nuclear RNA
Alanine tRNA or $tRNA^{Ala}$, etc.	transfer RNA that normally accepts alanine
Alanyl-$tRNA^{Ala}$ or Ala-$tRNA^{Ala}$ or Ala-tRNA	transfer RNA that normally accepts alanine with alanine residue covalently linked
poly(N), or $(N)_n$ or $(rN)_n$	polymer of ribonucleotide N
poly (dN) or $(dN)_n$	polymer of deoxyribonucleotide N

poly(N-N′), or r(N-N′)$_n$ or (rN-rN′)$_n$	copolymer of N—N′-N—N′ in regular, alternating, *known* sequence
poly(A)·poly(B) or (A)$_n$(B)$_n$	two chains, generally or completely associated
poly(A), poly(B) or (A)$_n$, (B)$_n$	two chains, association unspecified or unknown
poly(A) + poly(B) or (A)$_n$ + (B)$_n$	two chains, generally or completely unassociated

Miscellaneous

RNase, DNase	ribonuclease, deoxyribonuclease
P_i, PP_i	inorganic orthophosphate and pyrophosphate

Amino acids

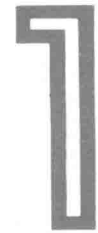
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Methionine
fMet	Formylmethionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

Further details of the recommendations of the Commission on Biochemical Nomenclature are printed in the *J. Biol. Chem.* **246,** 4894 (1971), *Biochim. Biophys. Acta,* **247,** 1 (1971), *Biochemistry,* **5,** 1445 (1966), *Arch. Biochem. Biophys.*, **115,** 1 (1966), *J. mol. Biol.*, **55,** 299 (1971), and *Progress in Nucleic Acid Research and Molecular Biology,* **22,** (1979).

In naming enzymes, the recommendations of the Commission on Enzymes of the International Union of Biochemistry (1972) are followed as far as possible. The numbers recommended by the Commission are inserted in the text after the name of each enzyme.

Contents

Introduction



The fundamental investigations which led to the discovery of the nucleic acids were made by Friedrich Miescher [1] (1844–95), who may be regarded as the founder of our knowledge of the chemistry of the cell nucleus. In early work carried out in 1868, in the laboratory of Hoppe-Seyler in Tübingen, he isolated the nuclei from pus cells obtained from discarded surgical bandages and showed the presence in them of an unusual phosphorus-containing compound that he called 'nuclein' and which we now know to have been nucleoprotein. Miescher's investigations were continued in Basel, where he spent most of his working life and where he became interested in salmon sperm as a source of nuclear material. In 1872 he showed that isolated sperm heads contained an acidic compound, now recognized as nucleic acid, and a base to which the name 'protamine' was given. It was subsequently shown that nucleic acids were normal constituents of all cells and tissues and Miescher's investigations of the nucleic acids were continued by Altman, who in 1889 described a method for the preparation of protein-free nucleic acids from animal tissues and from yeast. The work was continued later by Kossel in Heidelberg, Jones in Baltimore, Levene in New York, Hammarsten in Stockholm, Gulland in Nottingham and many others [2–7].

One of the best animal sources of nucleic acid was found to be the thymus gland, and it is not surprising therefore that much of the early work was concentrated on nucleic acid from this source. On hydrolysis it was found to yield the purine bases adenine and guanine, the pyrimidine bases cytosine and thymine, a deoxypentose and phosphoric acid. The nucleic acid from yeast on the other hand yielded on hydrolysis adenine, guanine, cytosine, uracil, a pentose sugar and phosphoric acid. Yeast nucleic acid therefore differed from thymus nucleic acid in containing uracil in place of thymine and a pentose in place of a deoxypentose. Since most nucleic acids from animal sources appeared to resemble thymus

nucleic acid, and since the only other nucleic acid which had at that time (1920) been prepared in reasonable quantities from a plant source appeared to be very similar to yeast nucleic acid, the impression grew that deoxypentose nucleic acid was characteristic of animal tissues, and pentosenucleic acid was characteristic of plant tisues [5]. Thus Jones, in 1920, stated categorically: 'we come to understand quite clearly that there are only two nucleic acids in nature, one obtainable from the nuclei of animal cells and the other from the nuclei of plant cells' [6].

It was not long before the validity of this concept was questioned. It had been known for many years that pentose derivatives were present in animal tissues. For example, the so-called β-nucleoprotein, which was originally prepared from mammalian pancreas by O. Hammarsten [7] in 1894, was known to contain a pentose sugar, and Jorpes [8] eventually prepared from this material a nucleic acid of the pentose type which he showed to resemble yeast nucleic acid and to be abundant in pancreatic tissue. The presence of pentosenucleic acids in mammary tissue was also suggested by the work of Odenius [9] and of Mandel and Levene [10]. Pentosenucleotide derivatives were also demonstrated in chick embryo pulp by Calvery [11], in spleen and the liver by Jones and Perkins [12] and by Thomas and Berariu [13], and in sea urchin eggs by Blanchard [14]. It thus appeared probable that pentosenucleic acids were normal constituents of animal tissues as well as of plant cells, and Jones and Perkins [12] expressed the view: 'the distinction between plant and animal nucleic acids will in future not be so definitely drawn'.

Final proof that ribonucleic acid is a general constituent of animal, plant and bacterial cells was not forthcoming until the early 1940's as a consequence of the ultraviolet spectrophotometric studies of Caspersson [15], the histochemical observations of Brachet [16] and the chemical analyses of Davidson [17, 18].

It took a surprisingly long time also to establish the nature of the sugars in deoxypentose and pentose nucleic acids. Eventually, however, the deoxypentose from a range of sources was shown to be D(−)-2-deoxyribose [19–23] and the pentose was shown to be D(−)-ribose [24–30] so that the two types of nucleic acid are now known as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) respectively. Looking back it seems remarkable that confirmation of the nature of these sugars was not obtained until the mid 1950's around the same time that the double-helical structure for DNA was being put forward by Watson and Crick [31].

These advances established the biology of the nucleic acids on a new foundation. The use of new techniques in cytochemistry and cell fractionation showed that DNA and RNA are normal constituents of all cells, whether plant or animal, DNA being confined mainly to the nucleus

while RNA is found also in the cytoplasm. [17, 18, 32–35.]

The development of techniques of subcellular fractionation and for the isolation of nuclei [36–41] made possible chemical measurements of the distribution of DNA and RNA amongst the subcellular fractions of various cell types, and led ultimately to the recognition of RNA in the nuclear, ribosomal and soluble fractions of cells (see Chapter 5) and to the demonstration of the constancy in the average amount of DNA per nucleus in the somatic cells of any given species [42].

The advent of isotope techniques led to the demonstration that both DNA and RNA could be synthesized *de novo* in most tissues from low-molecular-weight precursors [44] and from an early stage a correlation began to emerge between the rate of cell division in a tissue and the rate of uptake of isotopes into the DNA of that tissue [44]. So far as the synthesis of RNA was concerned, cell-fractionation studies combined with measurements of the incorporation of labelled precursors into the RNAs of different subcellular fractions [43, 44], demonstrated that there were considerable differences in the metabolic activities of the various classes of RNAs, nuclear RNA showing specially high levels of isotope incorporation.

Many of the earliest contributions to our understanding of the structure of nucleic acids arose out of the work of Levene and Jacobs [3, 45, 46] who established the presence of D-ribose, hypoxanthine and phosphorus in inosinic acid from muscle and later the presence of ribonucleotides of adenine, guanine, cytosine and uracil in yeast nucleic acid. They also recognized the occurrence of thymine in place of uracil in thymus nucleic acid.

The presence of the bases in approximately equimolar proportions led to the development of the tetranucleotide hypothesis for both DNA and RNA, in which both nucleic acids were considered to be polymeric structures containing equivalent amounts of mononucleotides derived from each of the four purine and pyrimidine bases linked together in repeating units. This concept of nucleic acid structure survived until the late 1940s despite the fact that evidence for it was not strong [47] and it was only when methods for the quantitative analysis of nucleic acids had been developed [48, 49] that the tetranucleotide hypothesis was finally abandoned as a consequence of the demonstration that the various nucleotides did not necessarily occur in equimolar proportions [50].

In the early 1950s Chargaff [51] drew attention to certain regularities in the composition of DNA, namely that the sum of the purines was equal to the sum of the pyrimidines, that the sum of the amino bases (adenine and cytosine) was equal to the sum of the keto bases (guanine and thymine) and that adenine and thymine, and guanine and cytosine, were present in equivalent amounts (Chapters 2 and 3). These observations