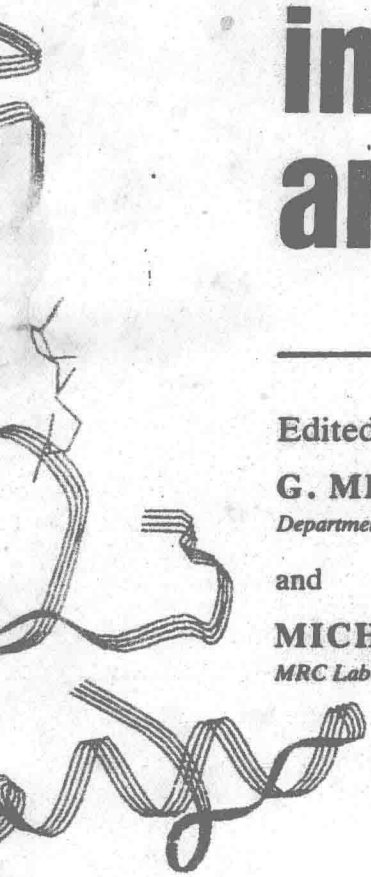
# Nucleic Acids in Chemistry and Biology

Edited by

G. MICHAEL BLACKBURN
and
MICHAEL J. GAIT

# Nucleic Acids in Chemistry and Biology



Edited by

**G. MICHAEL BLACKBURN**
*Department of Chemistry, University of Sheffield*

and

**MICHAEL J. GAIT**
*MRC Laboratory of Molecular Biology, Cambridge*

IRL PRESS
at
OXFORD UNIVERSITY PRESS
Oxford New York Tokyo

1990

*Oxford University Press, Walton Street, Oxford OX2 6DP*

*Oxford New York Toronto*
*Delhi Bombay Calcutta Madras Karachi*
*Petaling Jaya Singapore Hong Kong Tokyo*
*Nairobi Dar es Salaam Cape Town*
*Melbourne Auckland*
*and associated companies in*
*Berlin Ibadan*

*Oxford is a trade mark of Oxford University Press*

*Published in the United States*
*by Oxford University Press, New York*



*British Library Cataloguing in Publication Data*
*Blackburn, G. M.*
*Nucleic acids in chemistry and biology.*
*1. Nucleic acids*
*I. Title II. Gait, M. J. (Michael J.)*
*547.79*
*ISBN 0-19-963120-4*
*ISBN 0-19-963121-2 pbk*

*Library of Congress Cataloging-in-Publication Data*
*Nucleic acids in chemistry and biology/edited by G. Michael Blackburn and Michael J. Gait.*
*Includes bibliographical references and index.*
*1. Nucleic acids. I. Blackburn, G. Michael. II. Gait, M. J. (Michael J.)*
*QD433.N83 1990 547.7'9-dc20 90-14187*
*ISBN 0-19-963120-4 (hardback)*
*ISBN 0-19-963121-2 (paperback)*

*Set by Promenade Graphics, Cheltenham*
*Printed in Great Britain by*
*Butler & Tanner Ltd, Frome, Somerset*

# Preface

Nucleic acids dominate modern molecular science. They have vital roles that are fundamental for the storage and transmission of genetic information within cells. It follows that an accurate and detailed knowledge of their structure and function is of prime importance for molecular scientists of all descriptions. Just as significantly, the genius of biologists and chemists working together has made contemporary research into nucleic acids a rich source of discovery and invention that is dramatically transforming and improving the human condition.

Our own teaching and research experience, shared in discussions with colleagues around the world, has convinced us of the need to fill a significant gap in the modern science library by creating a broad-based yet concise and readable book on nucleic acids. Our single volume is designed to provide a compact, molecular perspective of this great subject. To that end, we have used it unashamedly to emphasize chemical and structural aspects of nucleic acids at all points. In it, we have surveyed a very broad field—up to the point where the frontiers of current studies in nucleic acids are only attainable by reading the latest issues of key journals! In particular, we have strengthened its production by drawing on the talents of an international group of co-authors whose expertise has extended the authority of this book from cover to cover. At the same time, we have tried to keep it selective and simple so as to make it widely accessible to students. This has meant that, of necessity, some sections have focused more on key concepts rather than on fine detail.

We have tried to provide a radically fresh and unified approach. This book builds on a general introduction to the chemistry and biology of the nucleic acids in order to reach out to some of the most significant modern developments of this subject. It is couched in an easily readable style and in a language which, while technically accurate, can yet be grasped quickly by those with a basic scientific background.

We begin with a brief historical perspective designed to point out the significance of later progress. We next provide an outline of the essential features of DNA and RNA structure, highlighting the new subtle insights which have been obtained by detailed analysis of crystals of synthetic oligonucleotides of defined sequence. The next four chapters are the core of the book. The first concentrates on modern chemistry applied to the synthesis of biologically important nucleosides, nucleotides, and oligonucleotides. Then comes a discussion of the biosynthesis of nucleotides, which is given a fresh presentation to emphasize how anti-cancer and anti-viral agents interfere with biosynthetic processes. The core is completed by two chapters which deal with the basic molecular biology of DNA and of RNA, showing how information stored in the form of nucleotide sequence is transmitted into cellular activity. Recent exciting developments in the auto-catalysis of RNA, ribozymes, are especially featured.

Three rather more specialized chapters then focus on the covalent and physical interactions of nucleic acids with small molecules, especially with mutagens and carcinogens and the relevant repair processes, and on their physical interactions with proteins. These important topics are at the forefront of much present research and typify the success of creative symbioses between chemistry and biology. The final chapter contrasts the *in vivo* rearrangements which DNA experiences with the *in vitro* techniques of manipulation of DNA sequences that are the essence of experimentation in cloning and mutagenesis.

This is not a textbook on the molecular biology of nucleic acids. From the outset, we have aimed this book especially at the needs of students and new research workers with a chemical or biochemical background. We hope that molecular biologists and more senior chemists and biochemists alike will find their knowledge of nucleic acids broadened through the special perspectives this book offers.

*Sheffield and Cambridge* G.M.B.
November 1989 M.J.G.

# Acknowledgements

Both Mikes express their sincere appreciation of the efforts of all who have supported the production of this book. Principally, our unqualified thanks go to those eight expert and understanding co-authors without whose contributions this book would not have been possible. We have taken considerable liberties with their manuscripts, perhaps more than is the norm with such a multi-author book, because we wanted to blend all of their contributions into a homogeneous final product. They have responded to these efforts with equanimity and understanding and have co-operated superbly in the numerous revision processes required for the production of the finished work.

We are very grateful to the following of our colleagues and fellow-scientists who have read and commented on portions of the text during its formative stages, namely Tom Brown, Chris Christodoulou, Bernard Connolly, David Hornby, Paul Kong, Christian Lehmann, Mick McLean, Daniella Rhodes, Ian Willis, and also to Joachim Engels for supplying most of the definitions in the glossary. We particularly wish to thank Lord Todd, Gobind Khorana, and Dan Brown for suggestions and comments on the early part of Chapter 1 which have, we believe, given us a genuine feeling for the key events in those seminal early years of nucleic acid studies.

The final production of this book has been supported by many able individuals. We are particularly indebted to Colin Yeomans for a substantial amount of the original artwork, to Peter Artymiuk (Sheffield University), Simon Phillips (Leeds University), and Tom Steitz (Yale University) for access to original graphics for nucleic acids and protein structures, and above all to the staff of OUP for establishing new standards for the style and presentation of this volume.

Finally, it is inevitable in a book of this breadth that omissions, occasional errors, and lapses in the accuracy of interpretation will have escaped the detection of even the most assiduous proof-readers. We hope that any such mistakes are both minor and minimal and we accept full and exclusive responsibility for them. We shall be grateful to receive your help in their identification for future rectification.

# Contributors

G. MICHAEL BLACKBURN,
Department of Chemistry, University of Sheffield, Sheffield, S3 7HF, UK

MICHAEL J. GAIT,
MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK

GORDON C. BARR,
Department of Biochemistry, Medical Sciences Institute, University of Dundee, Dundee, DD1 4HN, UK

ANDREW J. FLAVELL,
Department of Biochemistry, Medical Sciences Institute, University of Dundee, Dundee, DD1 4HN, UK

ADRIAN GOLDMAN,
Waksman Institute of Microbiology, State University of New Jersey, Rutgers, P.O. Box 759, Piscataway, New Jersey, 08854–0759, USA

DAVID L. OLLIS,
Departments of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, 2153 Sheridan Drive, Evanston, Illinois, 60201, USA

UTTAM L. RAJBHANDARY,
Departments of Biology and Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts, 02138, USA

DIETER SÖLL,
Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 6666, New Haven, Connecticut, 06511, USA

RICHARD T. WALKER,
Department of Chemistry, University of Birmingham, P.O. Box 363, Birmingham, B15 2TT, UK

W. DAVID WILSON,
Department of Chemistry, State University of Georgia, Atlanta, Georgia, 30303, USA

# Nomenclature

The nomenclature for nucleic acids and their constituents used in this book is derived from the following recommendations:

IUPAC–IUB Joint Commission on Biochemical Nomenclature. Abbreviations and symbols for nucleic acids, polynucleotides and their constituents. Recommendations 1970. (1970) *Biochemistry* **9**, 4022–7.

IUPAC–IUB Joint Commission on Biochemical Nomenclature. Abbreviations and symbols for the description of conformations of polynucleotide chains. Recommendations 1982. (1983) *Eur. J. Biochem.*, **131**, 9–15.

Definitions and nomenclature of nucleic acid structure parameters. (1989) *EMBO J.*, **8**, 1–4.

### Stereodiagrams

The stereo-pair figures used in this book are for parallel viewing, i.e., left diagram to left eye and right to right. They can be viewed either unaided (a little practice helps) or with the help of a simple convex-lens viewer.

# Contents

# Introduction and overview

# 1

## 1.1 The biological importance of DNA

From the beginning, the study of nucleic acids has drawn together, as though by a powerful unseen force, a galaxy of scientists of the highest ability. Striving to tease apart its secrets, these talented individuals have brought with them a broad range of skills from other disciplines while many of the problems they have encountered have proved to be soluble only by new inventions. Looking at their work, one is constantly made aware that scientists in this field appear to have enjoyed a greater sense of excitement in their work than is given to most. Why?

For over 60 years, such men and women have been fascinated and stimulated by their awareness that the study of nucleic acids is central to a knowledge of life. Let us start by looking at Fred Griffith, who was employed as a scientific civil servant in the British Ministry of Health investigating the nature of epidemics. In 1923, he was able to identify the difference between a virulent, *S*, and a non-virulent, *R*, form of the pneumonia bacterium. Griffith went on to show that this bacterium could be made to undergo a permanent, hereditable change from non-virulent to virulent type. This discovery was a bombshell in bacterial genetics.

Oswald Avery and his group at the Rockefeller Institute in New York set out to identify the molecular mechanism responsible for the change Griffith had discovered, now technically called **bacterial transformation**. They achieved a breakthrough in 1940 when they found that non-virulent *R* pneumococci could be transformed *irreversibly* into a virulent species by treatment with a pure sample of high molecular weight DNA. Avery had purified this DNA from heat-killed bacteria of a virulent strain and showed that it was active at a dilution of 1 part in $10^9$.

Avery concluded that '**DNA is responsible for the transforming activity**' and published that analysis in 1944, just three years after Griffith had died in a London air-raid. The staggering implications of Avery's work turned a searchlight on the molecular nature of nucleic acids and it soon became evident that ideas on the chemistry of nucleic acid structure at that time were wholly inadequate to explain such a momentous discovery. As a result, a new wave of scientists directed their attention to DNA and discovered that large parts of the accepted tenets of nucleic acid chemistry had to be set aside before real progress was possible. We need to examine some of the earliest features of that chemistry to appreciate fully the significance of later progress.

## 1.2 The origins of nucleic acids research

Friedrich Miescher started his research career in Tübingen by looking into the physiology of human lymph cells. In 1868, seeking a more readily available material, he began to study human pus cells which he obtained in abundant supply from the bandages discarded from the local hospital. After defatting the cells with alcohol, he incubated them with a crude preparation of pepsin from pig stomach

and so obtained a grey precipitate of pure cell nuclei. Treatment of this with alkali followed by acid gave Miescher a precipitate of a phosphorus-containing substance which he named '**nuclein**'. He later found this material to be a common constituent of yeast, kidney, liver, testicular, and nucleated red blood cells.

After Miescher moved to Basel in 1872, he found the sperm of Rhine salmon to be a more plentiful source of nuclein. The pure nuclein was a strongly acidic substance which existed in a salt-like combination with a nitrogenous base which Miescher crystallized and called protamine. In fact, his nuclein was really a nucleoprotein and it fell subsequently to Richard Altman in 1889 to obtain the first protein-free material, to which he gave the name '**nucleic acid**'.

Following William Perkin's invention of mauveine in 1856, the development of aniline dyes had stimulated a systematic study of the colour-staining of biological specimens. Cell nuclei were characteristically stained by basic dyes, and around 1880 Walter Flemming applied that property in his study of the rod-like segments of chromatin (so called because of their colour-staining characteristic) which became visible within the cell nucleus only at certain stages of cell division. Flemming's speculation that the chemical composition of these **chromosomes** was identical with that of Miescher's nuclein was confirmed in 1900 by E. B. Wilson who wrote:

> Now chromatin is known to be closely similar to, if not identical with, a substance known as nuclein which analysis shows to be a tolerably definite chemical compound of nucleic acid and albumin. And thus we reach the remarkable conclusion that inheritance may, perhaps, be affected by the physical transmission of a particular compound from parent to offspring.

While this insight was later to be realized in Griffith's 1928 experiments, all of this work was really far ahead of its time. We have to recognize that, at the turn of the century, tests for the purity and identity of substances were relatively primitive. Emil Fischer's classic studies on the chemistry of high molecular weight, polymeric organic molecules was under question until well into the twentieth century. Even in 1920, it was possible to argue that there were only two species of nucleic acids in nature: animal cells were believed to provide **thymus nucleic acid** (DNA), whilst nuclei of plant cells were thought to give **pentose nucleic acid** (RNA).

## 1.3 Early structural studies on nucleic acids

Accurate molecular studies on nucleic acids essentially date from 1909 when Levene and Jacobs began a reinvestigation of the structure of **nucleotides** at the Rockefeller Institute. Inosinic acid, which Liebig had isolated from beef muscle in 1847, proved to be hypoxanthine-riboside 5′-phosphate. Guanylic acid, isolated from the nucleoprotein of pancreas glands, was identified as guanine-riboside 5′-phosphate (Fig. 1.1). Each of these nucleotides was cleaved by alkaline hydrolysis to give phosphate and the corresponding **nucleosides**, inosine and guanosine respectively. Since then, all nucleosides are characterized as the condensation pro-

ducts of a pentose and a nitrogenous base while nucleotides are the phosphate esters of one of the hydroxyl groups of the pentose.

Pentose nucleic acid was available in plentiful supply from yeast and on mild hydrolysis with aqueous ammonia it gave the four pentose-nucleosides adenosine, cytidine, guanosine, and uridine. These were identified as derivatives of the four bases adenine, cytosine, guanine, and uracil (Fig. 1.1).

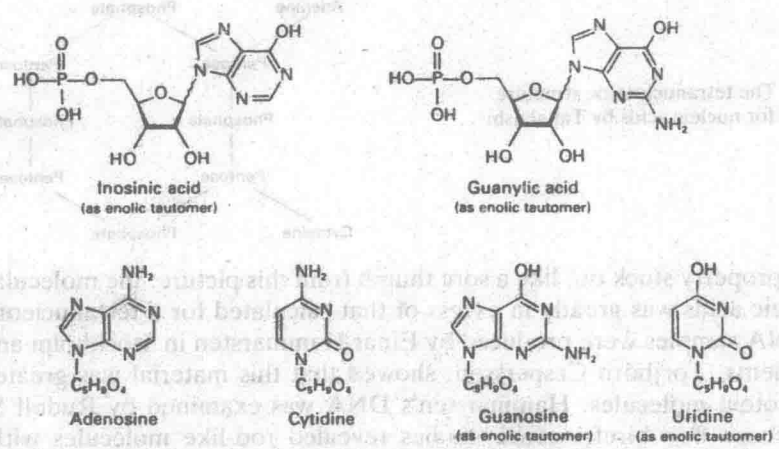



**Fig. 1.1** Early nucleoside and nucleotide structures (using the enolic tautomers originally employed).

Thymus nucleic acid, which was readily available from calf tissue, was found to be resistant to alkaline hydrolysis. It was only successfully degraded into deoxynucleosides in 1929 when Levene adopted enzymes to hydrolyse the deoxyribonucleic acid followed by mild acidic hydrolysis of the deoxynucleotides. He identified its pentose as the hitherto unknown 2-deoxy-D-ribose. These deoxynucleosides involved the four heterocyclic bases, adenine, cytosine, guanine, and thymine, with the latter corresponding to uracil in ribonucleic acid.

Up to 1940, most groups of workers were convinced that hydrolysis of nucleic acids gave the appropriate four bases in **equal relative proportions**. This erroneous conclusion probably resulted from the use of impure nucleic acid or from the use of analytical methods of inadequate accuracy and reliability. It led, naturally enough, to the general acceptance of a '**Tetranucleotide hypothesis**' for the structure of both thymus and yeast nucleic acids, which materially retarded further progress on the molecular structure of nucleic acids.

Several of these tetranucleotide structures were proposed. They all had four nucleosides (one for each of the bases) with an arbitrary location of the two purines and two pyrimidines. They were joined together by four phosphate residues in a variety of ways, among which there was a strong preference for phosphodiester

linkages. In 1932, Takahashi showed that yeast nucleic acid contained neither pyrophosphate nor phosphomonoester functions and so disposed of earlier proposals in preference for a neat, cyclic structure which joined the pentoses exclusively using phosphodiester units (Fig. 1.2). It was generally accepted that these bonded 5′- to 3′-positions of adjacent deoxyribonucleosides, but the linkage positions in ribonucleic acid were not known.

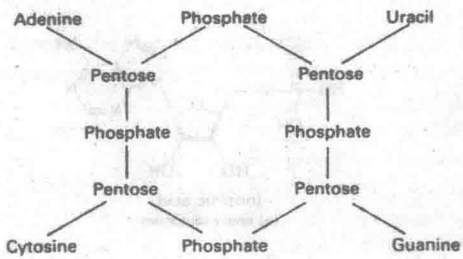



**Fig. 1.2** The tetranucleotide structure proposed for nucleic acids by Takahashi (1932).

One property stuck out like a sore thumb from this picture: the molecular weight of nucleic acids was greatly in excess of that calculated for a tetranucleotide. The best DNA samples were produced by Einar Hammarsten in Stockholm and one of his students, Torjbörn Caspersson, showed that this material was greater in size than protein molecules. Hammarsten's DNA was examined by Rudolf Signer in Bern whose flow-birefringence studies revealed rod-like molecules with a molecular weight of 0.5–1.0 × $10^6$ Daltons (Da). The same material provided Astbury in Leeds with X-ray fibre diffraction measurements that supported Signer's conclusion. Finally, Levene estimated the molecular weight of native DNA at between 200 000 and one million based on ultracentrifugation studies.

Scientists compromised. In his Tilden Lecture of 1943, Masson Gulland suggested that the concept of nucleic acid structures of polymerized, uniform tetranucleotides was limited, but he allowed that they could 'form a practical working hypothesis'.

This then was the position in 1944 when Avery published his great work on the transforming activity of bacterial DNA. One can sympathize with Avery's hesitance to press home his case. Levene, in the same Institute, and others were strongly persuaded that the tetranucleotide hypothesis imposed an invariance on the structure of nucleic acids which denied them any role in biological diversity. By contrast, Avery's work showed that DNA was responsible for completely transforming the behaviour of bacteria. It demanded a fresh look at the structure of nucleic acids.

## 1.4 The discovery of the structure of DNA

From the outset, it was evident that DNA exhibited greater resistance to selective chemical hydrolysis than did RNA. So, the discovery in 1935 that DNA could be

cut into **mononucleotides** by an enzyme doped with arsenate was invaluable. Using this procedure, Klein and Thannhauser obtained the four crystalline deoxyribonucleotides whose structures (Fig. 1.3) were later put beyond doubt by total chemical synthesis by Alexander Todd and the Cambridge school he founded in 1944. Todd established the β-configuration of the glycosidic linkage for ribonucleosides in 1951, but found the chemical synthesis of the 2′-deoxyribonucleosides more taxing. The key to success for the Cambridge group was the development of methods of phosphorylation, illustrated in Fig. 1.4 for the preparation of the 3′- and 5′-phosphates of deoxyadenosine (Fig. 1.4).

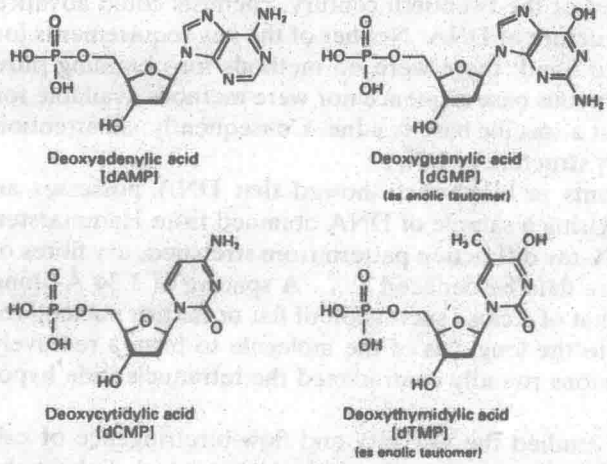



**Fig. 1.3** Structures of 5′-deoxyribonucleotides (original tautomers for dGMP and dTMP).

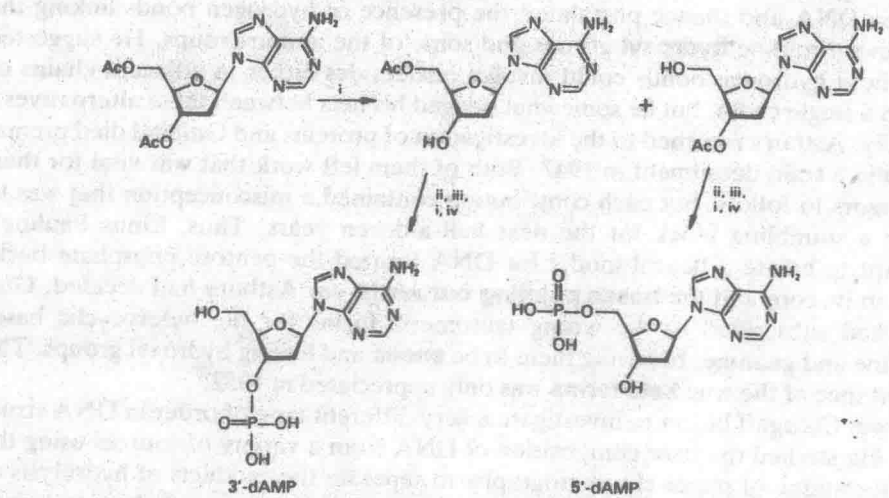



**Fig. 1.4** Todd's syntheses of deoxyadenosine 3′- and 5′-phosphates (Hayes, D. H., Michelson, A. M., and Todd, A. R. (1955). *J. Chem. Soc.*, 808–15).
**Reagents:** (i) MeOH, $NH_3$ (ii) $(PhO)_2P(O)OP(H)(O)OCH_2Ph$ (iii) *N*-chlorosuccinimide (iv) $H_2$/PdC

All the facts were now available to establish the primary structure of DNA as a **linear polynucleotide** in which each deoxyribonucleoside is linked to the next by means of a 3′- to 5′-phosphodiester (Chapter 2, Fig. 2.15). The presence of only diester linkages was essential to explain the stability of DNA to chemical hydrolysis, since phosphate triesters and monoesters, not to mention pyrophosphates, are more labile. The measured molecular weights for DNA of about one million meant that a single strand of DNA would have some 3000 nucleotides. Such a size was much greater than that of enzyme molecules, but entirely compatible with Staudinger's established ideas on macromolecular structure for synthetic and natural polymers. But by the mid-point of the twentieth century, chemists could advance no further with the primary structure of DNA. Neither of the key requirements for sequence determination was to hand: there were no methods for obtaining pure samples of DNA with homogeneous base sequence nor were methods available for the cleavage of DNA strands at a specific base residue. Consequently, all attention came to focus on the secondary structure of DNA.

Two independent experiments in biophysics showed that DNA possesses an ordered secondary structure. Using a sample of DNA obtained from Hammarsten in 1938, Astbury obtained an X-ray diffraction pattern from stretched, dry fibres of DNA. From the rather obscure data he deduced '. . . A spacing of 3.34 Å along the fibre axis corresponds to that of a close succession of flat or flattish nucleotides standing **out** perpendicularly to the long axis of the molecule to form a relatively rigid structure.' These conclusions roundly contradicted the tetranucleotide hypothesis.

Some years later, Gulland studied the viscosity and flow-birefringence of calf thymus DNA and thence postulated the presence of hydrogen bonds linking the purine–pyrimidine **hydroxyl** groups and some of the amino groups. He suggested that these hydrogen bonds could involve nucleotides either in adjacent chains or within a single chain, but he somewhat hedged his bets between these alternatives.

Sadly, Astbury returned to the investigation of proteins and Gulland died prematurely in a train derailment in 1947. Both of them left work that was vital for their successors to follow, but each contribution contained a misconception that was to prove a stumbling block for the next half-a-dozen years. Thus, Linus Pauling's attempt to create a helical model for DNA located the pentose-phosphate backbone in its core and the **bases pointing outwards**—as Astbury had decided. Gulland had subscribed to the wrong tautomeric forms for the heterocyclic bases thymine and guanine, believing them to be **enolic** and having hydroxyl groups. The importance of the true **keto forms** was only appreciated in 1952.

Erwin Chargaff began to investigate a very different type of order in DNA structure. He studied the base composition of DNA from a variety of sources using the new technique of paper chromatography to separate the products of hydrolysis of DNA and employing one of the first commercial ultraviolet spectrophotometers to quantify their relative abundance. His data showed that there is a variation in base