Nucleic Acid Structure

W. Guschlbauer

Nucleic Acid Structure

AN INTRODUCTION

Wilhelm Guschlbauer
Service de Biochimie
Centre d'Études Nucléaires de Saclay
91190 Gif-sur-Yvette
France

Library of Congress Cataloging in Publication Data

Guschlbauer, Wilhelm.
Nucleic acid structure.

(Heidelberg science library; v. 21)
Includes index.
1. Nucleic acids. I. Title. II. Series.
QD433.G87 547'.596 75-11796

Printed in the United States of America.

ISBN 0-387-90141-8 Springer-Verlag New York

ISBN 3-540-90141-8 Springer-Verlag Berlin Heidelberg

Preface

Teaching a course on nucleic acid structure is a hazardous undertaking, especially if one has no continuous teaching obligations. I still have done it on several occasions in various French universities, when colleagues, suffering from administrative overwork and excessive teaching obligations, had asked me to do so. This was generally done with a pile of notes and a dozen slides, and I always regretted that no small, concise, specialized book on nucleic acid structure for students at the senior or beginning graduate level existed. Every year, the lecture notes became more and more voluminous, with some key reprints intermingled.

Everything changed when, in the spring of 1973, I received an invitation to teach such a course, under the UNESCO-OAS-Molecular Biology Program at the Universidad de Chile in Santiago during October 1973. I had accepted rather enthusiastically, but soon discovered that it would be necessary to produce a photocopied syllabus for the students. This was the first premanuscript of this book.

For nonscientific reasons, the course was first canceled and then postponed until December 1973. Nearly a year later, the course, in slightly amended form, was presented at the Lemonossow-State University in Moscow.

The manuscript was designed as a nonmathematical, voluntarily phenomenological introduction to nucleic acid structure, as a complement to the normally strongly enzymological approach in molecular biology. This book is not a monograph and has, therefore, no pretention to be a rigorous treatment of the physical chemistry of nucleic acids.

Besides the fact that our understanding of the structural constraints of nucleic acids is far from being precise, it is not very useful to charge students with such specialized problems and techniques as statistical thermodynamics, fast

kinetics, or quantum mechanical treatments, for which considerably more physical background is necessary than can normally be expected from a senior student in biochemistry. Also, many of the problems treated in this book are still evolving, and it is to be expected that new results will appear before the publication of the book. It is for this reason that protein–nucleic acid interactions, for instance, have been treated generally, rather than specifically.

On the other hand, certain new results have been included, despite the limited comprehension we have of them. This was done to demonstrate the "possibilities" of nucleic acids, their fine structure, and their future. For this reason, repeating sequences, restriction enzymes, and unusual polynucleotide complexes are treated.

It is certain that the choices of the subjects treated will be criticized. This is inevitable. The same is true for the references included. In general, and if available, review and symposium articles are cited preferentially. For those, who wish to deepen their understanding of the physical chemistry of nucleic acids, the recently published *Physical Chemistry of Nucleic Acids* by Bloomfield, Crothers, and Tinoco (Harper and Row, 1974) and the three-volume series *Basic Principles of Nucleic Acid Chemistry* edited by P. O. P. Ts'o (Academic Press, 1974) are recommended.

I shall not forget the many people who have been directly or indirectly involved in the elaboration of the manuscript. In the first place, there is Professor S. Litvak (now at the Université Bordeaux II) who invited me to Chile and suggested that I teach a structure-oriented course. Professor Z. Shabarova and Dr. E. S. Gromova (Moscow) further influenced the orientation of this book. Professor P. Cohen (Paris) generously made his lecture notes available. Professor P. Fromageot (Saclay) was a constant and generous supporter of the project. Professor A. E. V. Haschemeyer (New York) was a particularly thorough and severe commentator; and Professor S. Arnott (Purdue) read large parts of the manuscript and generously furnished X-ray fiber diffraction patterns and drawings of nucleic acid structures. Dr. G. Bernardi (Paris) was the first to suggest that I write a book and he and Dr. V. Vetterl (Brno) read the manuscript and made valuable comments and useful criticisms. My collaborators, Marie-Therese Sarocchi, Danielle

Helical columns have held a strong attraction for man for a long time, as shown by this medieval candelabra in the Church of Santa Maria in Cosmedin, in Rome.

Thiele, J. F. Chantot, P. Tougard, and Tran-Dinh S., have helped at different stages. Barbara L. Haas was the constant and thorough counsel in English style and spelling. To all of them, I express my deep gratitude. They are, obviously, not responsible for the shortcomings of the book.

Finally, my wife, Marie-Pierre, is to be thanked, for her patience and moral help all along these years.

Saclay, Spring 1975 Wilhelm Guschlbauer

Contents

1 Introduction

1.1 History

During the last 25 years, two classes of natural compounds, proteins and nucleic acids, have attracted most of the attention in what has been called molecular biology. This attention is due to the key role of these two classes of compounds in all cellular processes.

As we shall see, nucleic acids are the carriers of genetic information, whereas proteins are concerned with the execution of biological processes. The near invariability and exactness of genetic information is, in part, the result of the rather rigid structure of deoxyribonucleic acid (DNA) and the virtual chemical inertness of its structure under physiological conditions. Proteins, on the other hand, are quite adaptable to their different purposes because they can undergo a large variety of chemical interactions. Their structure is generally not rigid and can be changed by the action of outside agents (substrates, cofactors).

Two main purposes of nucleic acids can be noted:

1. Storage and faultless transmission of genetic information to progeny. This process, called replication, implies the exact copying of a DNA molecule to form two identical sets of cellular DNA.
2. Transmission and coding of this information into proteins. These processes, called transcription and translation, are mediated by a second class of nucleic acids, ribonucleic acid (RNA).

We shall see that the basic principle underlying all these processes is identical for all organisms and is based on a specific pairing of certain building blocks, the nucleic acid bases. We shall also see that most "errors," called mutations, are due to accidental physical or chemical changes

in these building blocks, with resultant changes in their pairings.

The origin of our knowledge of hereditary processes and their chemistry came, independently and simultaneously, from the work of two pioneers: in 1863, Gregor Mendel discovered the basic laws of genetics, crossing garden peas of different phenotypes; in 1868, Friedrich Miescher found a macromolecular, phosphorous-containing substance in wound pus, which he called "Nuklein." It was 80 years before Avery, McLoed and McCarthy could show, in 1944, that the active principle behind these two discoveries was one and the same: deoxyribonucleic acid or DNA. It is worth noting that none of these discoveries was appreciated at their time.

It was only after World War II that the first significant advances were made, by the geneticists of the "phage group," centered around Delbrück and Luria, and by the "structuralists," of the Pauling and Astbury school. The x-ray work of British crystallographers and the chemistry of Chargaff's group led to the single most important discovery in biology of our century: the DNA double helix structure by Watson and Crick in 1953. This discovery provided, for the first time, a unified theory that took into account all the chemical, biological and physical data available at that time. The last 20 years have yielded a wealth of experimental data to confirm this theory, leading to the so-called "central dogma" of molecular biology.

1.2 Cellular Localization of Nucleic Acids

In eucaryotic cells, the largest cellular structure easily visualized in the light microscope is the nucleus. Within the nucleus the chromatin accounts for about 95% of cellular DNA. This roughly spherical structure possesses a triple membrane with numerous pores, which connect it with a cytoplasmic structure, the endoplasmic reticulum. Through these pores certain macromolecules, synthesized in the nucleus, e.g., messenger RNA (mRNA) can enter the cytoplasm where they become operational.

In the interior of the nucleoplasm, small spherical structures of high density can be localized. These are the nucleoli, in which most of the nuclear RNA occurs (about 20% of the total RNA).

On the endoplasmic reticulum, one finds the granular ribosomes, containing about equal quantities of RNA and protein, in aggregates that form polysomes or ergosomes.

Another cellular constituent, the mitochondrion, which is essentially responsible for the energetics of the cell, also contains RNA and DNA, but in small quantities (less than 1% of the protein by weight). It is probable that the mitochondrial DNA and RNA constitute the genetic equipment for the replication of these plasmids and control some of

their protein synthesis. DNA and RNA are also found in chloroplasts and plants.

Bacteria and procaryotes, in general, have a less dense nuclear region; the procaryotic chromosomes are much more diffuse and are not delineated by a membrane at any stage of growth. All DNA (apparently in a single molecule) is concentrated in the chromosome. Polysomes seen in the bacterial cell are sometimes bound to the cell wall.

In viruses, the DNA (or RNA) is very densely packed into the capsides, which are always surrounded by protein.

2 Methods and Techniques

Surveys of many of the techniques in nucleic acid chemistry are found in some recent review articles (1–6). A new book by Haschemeyer and Haschemeyer (7) is an excellent text for nearly all the biophysical techniques. In this chapter, we deal briefly with four techniques that have found rather widespread use in recent years:

1. Optical activity [optical rotatory dispersion (ORD) and circular dichroism (CD)],
2. Nuclear magnetic resonance (NMR),
3. Ultracentrifugation
4. Fiber X-ray diffraction

In any spectroscopic method, i.e., whenever electromagnetic quanta interact with matter, the fundamental equation $E = h\nu$ is applicable, where h is Planck's constant, E the energy of the quantum of radiation, and ν its frequency. In any spectroscopic analysis, the absorption of electromagnetic radiation by a population of spectroscopically identical particles (molecules, atoms, electrons) as a function of frequency will appear as a bell-shaped, generally Gaussian, curve.

The three parameters that define such a spectral band are (a) intensity, which is the surface of the band; (b) position of the band peak or the frequency of the absorption maximum; and (c) breadth of the band, generally defined as the width at half-peak height.

2.1 Absorption and Optical Activity (ORD and CD)

Any substance the electrons of which absorb electromagnetic quanta in the range between 150 and 800 nm will have an ultraviolet (UV) or a visible absorption spectrum. Generally, the absorption bands will be Gaussian, as a function of frequency for a given transition. Coupling between transitions from neighboring chromophores may reduce the

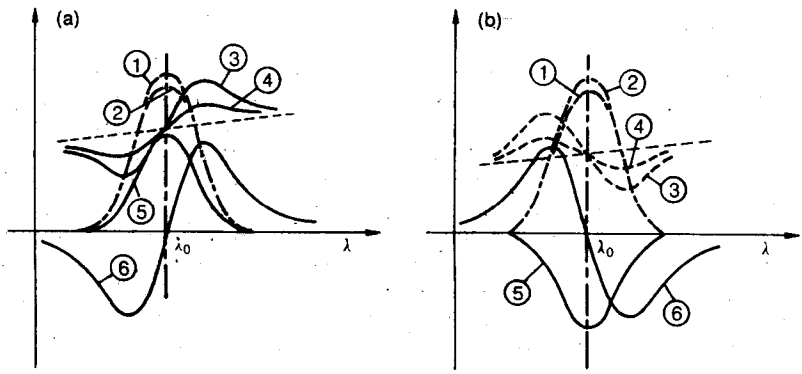
absorption observed at ε_{max} of the two isolated bands. This is the well-known hypochromism of nucleic acids, due to the interaction of the π-electrons of the bases (see Fig. 5-1).

For a substance to be optically active, it is necessary and sufficient to contain neither a plane nor a center of symmetry. In other words, the substance and its mirror image must not be superimposable.

Optical activity manifests itself in two ways: (a) by the rotation of the plane of polarization of linearly polarized light, due to the difference of refraction indices for the left- and right-handed components of linearly polarized light (see Fig. 2.1) and (b) by a difference in absorption between the left- and right-hand component of circularly polarized light at frequencies where absorption occurs, due to a difference in velocity between the two light components. In the first case, we have optical rotation, in the second, circular dichroism (CD) (Fig. 2.1). Both phenomena are produced by optically active molecules, the chromophores of which satisfy the requirement of asymmetry. This asymmetry need not be intrinsic, i.e., due to the chemical structure of the compound; it can be induced, e.g., through the binding of an optically inactive chromophore to a matrix to establish an asymmetric environment or through the polymerization of an optically inactive substance to form an asymmetric, optically active superstructure.

If an absorption band is optically active, a CD band with a shape similar to that of the absorption band will appear (Cotton effect). Optical rotatory dispersion (ORD) occurs

FIGURE 2.1. Simplified presentation of a dichroic band (Cotton effect). (a) Positive Cotton effect; (b) negative Cotton effect. Molar extinction coefficients for left- and right-hand polarized light components ① and ②; Refraction index of left- and right-hand polarized light components ③ and ④; Circular dichroism ⑤; Optical rotatory dispersion ⑥.



in regions outside the absorption band as a monotonous change of the rotatory power as a function of wavelength. In the region of the absorption band, an anomalous ORD will be found (Cotton effect) (Fig. 2.1). These spectra can be interconverted via the Kronig–Kramers transform; this amounts to the first derivative of the CD spectrum, which will be the ORD spectrum.

In principle, both CD and ORD yield similar information. There are, however, advantages and inconveniences to both techniques. The fact that ORD is observed over the whole spectral region permits the study of this phenomenon even in systems where the absorption band is outside the measurable range. Also, from ORD one can calculate using the Drude equation, certain useful optical parameters outside the experimentally accessible range. The advantage of CD is the simplicity of the spectra, which permits more exact band assignments (less band overlap). The disadvantage of ORD is the more complex band pattern, which can make it difficult to assign transitions precisely. But CD can apparently only be used within an optically active band, being useless beyond this range (see Fig. 2.1).

The following units are used in ORD measurements:

Specific rotation: $[\alpha] = \frac{100\alpha}{cd}$

Molar rotation: $[m] = \frac{M[\alpha]}{100} = \frac{\alpha}{c_m l}$

Reduced molar rotation: $[m'] = \frac{3[m]}{(n^2+2)}$

where α is the observed rotation in degrees, c the concentration in percent, d the cuvette path length in decimeters, l the cuvette path length in centimeters, M the molecular weight of the molecule (or monomer), and c_m the molar concentration.

Similarly, in CD the units used are:

Molar CD: $\Delta\varepsilon = \Delta A \cdot s/(c_m l)$

Molar ellipticity (θ or ψ): $\theta = 3300\Delta\varepsilon$

where ΔA is the measured absorbance difference, s the sensitivity of the instrument (in absorbance units).

Since the nucleic acid bases are aromatic compounds (see Chapter 3), the analogy with benzene has led several authors to correlate the absorption spectra (which also determine CD) of these two series. The region between 200 and 300 nm shows four absorption bands in benzene and other aromatic compounds. The names B_{2u} (around 265 to 285 nm) and B_{1u} (around 230 to 260 nm) have been assigned. These bands are π-π^* bands and are polarized in the plane of the base. Below 220 nm, benzene possesses a doubly

degenerate band, E_{1u}, which is also found in the purines and pyrimidines. These two bands (E_{1ua} and E_{1ub}) are also polarized in the plane of the base (π-π^*) (see Fig. 3.3).

There exists the possibility of n-π^* transitions, polarized perpendicular to the base plane. They can be distinguished by the classical McConnell criteria:

1. An electron-donating group (Cl, amino, methoxy) on an aromatic cycle displaces the π-π^* transitions toward the red and the n-π^* bands toward the blue.
2. Replacement of a nitrogen atom by a carbon has the opposite effect.
3. Hydrogen bond formation, protonation, and solvents with a high dielectric constant shift n-π^* transitions toward the blue but do not affect the π-π^* transitions. Deprotonation, breaking of hydrogen bonds, and aprotic solvents (hexane, etc.) induce red shifts.
4. In general, n-π^* transitions have molar extinctions of a few hundred; π-π^* band extinctions reach thousands and tens of thousands. This is not necessarily seen with CD, i.e. relatively large CD bands may be observed for n-π^* transitions.

2.2 Nuclear Magnetic Resonance (NMR)

Nuclear magnetic resonance (NMR) spectroscopy is based on the absorption of radiofrequency electromagnetic radiation (of the order of 100 MHz) by atomic nuclei [mainly protons: proton magnetic resonance (PMR), but also ^{13}C, ^{15}N, ^{17}O, ^{19}F, ^{23}Na, ^{31}P, ^{25}Mg, etc.] in substances placed in a very strong magnetic field [10 to 100 kiloGauss (kG)]. In NMR, the band width is usually rather small (normally a few Hertz), compared with the absorbed frequency, and the absorption peaks are single rays for a given nucleus or a group of magnetically equivalent nuclei. In many cases, the spectrum of nuclei in a single chemical group (e.g., methyl) shows a coupling of the spins with those of neighboring nuclei. From the disposition of the peaks, the spin-spin coupling constants can be evaluated. This is, however, sometimes only possible with extensive calculations, requiring theoretical analysis and computers. Empirical relations have been deduced between the coupling constants of resonating nuclei and dihedral angles. The most commonly used form for couplings between vicinal protons is the Karplus Equation (8)

$$J_{ij} = J_o \cdot \cos^2 \phi_{ij} - 0.28 \text{ Hz}$$

where J_{ij} is the observed coupling constant in Hz, J_o an empirical constant (which can vary between 8 and 16 Hz, depending on the compound, the electronegativity of its substituents, and its geometry) and ϕ_{ij}, the dihedral angle in degrees. A modified form of this relation applicable to

the ribose geometry has recently been proposed (9) in the form

$$J_{ij} = A\cos^2\phi_{ij} + B\cos\phi_{ij} + C$$

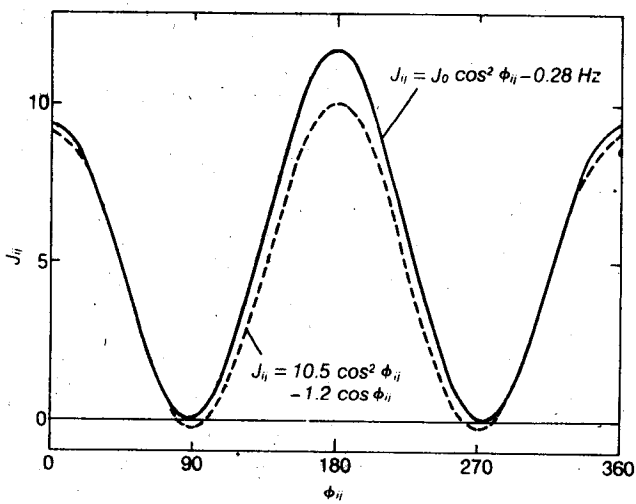
where $A = 10.5$ Hz, $B = -1.2$ Hz, and $C = 0$. Figure 2.2 shows the dependence of the coupling constant J_{ij} on the dihedral angle, using the two forms of the Karplus equation.

The principal parameters for NMR studies are chemical shifts, intensity (area) of absorption peaks, coupling constants, and relaxation times. Detailed interpretations of these spectroscopic parameters in NMR spectra depend on a number of suppositions, which, generally, have been verified.

1. The intensity of the absorption line is linearly proportional to the number of magnetically equivalent nuclei. For small molecules, comparison of intensities with chemical structure usually allows the unambiguous assignment of lines to individual atoms or groups of atoms in a molecule.

2. The resonance frequency (chemical shift, expressed as frequency relative to a standard) depends on the nature of the nucleus (^{1}H, ^{13}C, ^{19}F, etc.), the screening or shielding of the electrons surrounding the absorbing nucleus, and the spatial configuration of the nucleus. It may be influenced by permanent local fields, like unpaired electrons or ring currents from aromatic systems. Thus, protonation of a nitrogen of a nucleoside base will affect the neighboring protons greatly (Fig. 2.3). Titration of a phosphate group in a nucleotide will show displacements of the chemical shifts

FIGURE 2.2. The dependence of the proton-proton spin coupling constants J_{ij} on the dihedral angle ϕ_{ij}, using different forms of the Karplus equation.



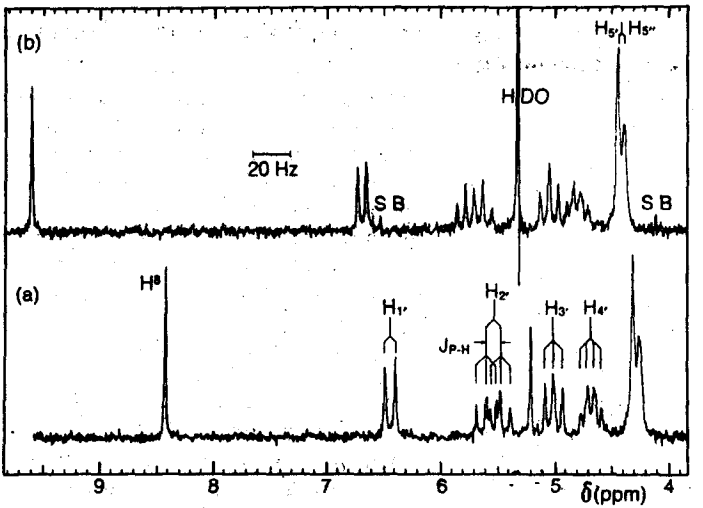


FIGURE 2.3. The 60-MHz spectra of Guo-2'-P at 25° C (0.14 M). (a) pH = 7.4; (b) pD = 1.2. The large shift of the H_8 resonance is due to protonation at N_7.

From Ref. 10. Reprinted with permission from J. Amer. Chem. Soc. © 1972 by the American Chemical Society.

of the adjacent sugar protons. Ring current effects have been used to define the geometry of stacked oligonucleotides and self-associating bases.

3. The band width will be determined by the rate of atomic motion of the resonating nucleus, which, in turn, is determined by the relaxation times t_1 and t_2. The larger these relaxation times (i.e., the faster the motion), the narrower the line widths.

4. Generally speaking, NMR permits the study of physical phenomena that involve single atoms and their immediate neighborhood. Therefore, great changes in chemical shifts can frequently be observed with minor changes in the environment; on the other hand, otherwise major changes in the system, which give rise to strong signals with other techniques, may not change the NMR significantly. This is particularly true if many absorbing nuclei are present, e.g., as in nucleic acids or proteins. It is for this reason that NMR has, until now, been used mainly for small molecules, monomers, and oligomers. On the other hand, selective deuteration (deuterium does not absorb in the proton frequency range) has been used effectively, but the technique requires very extensive synthetic work. For the same reason deuterium oxide (D_2O) has to be used instead of H_2O as solvent.

For a more detailed discussion of the theory and the biochemical applications of NMR, reviews (4, 11, 12) and books (13–14) are available.