NMR 20

Girjesh Govil Ramakrishna V. Hosur

Conformation of Biological Molecules

New Results from NMR

Springer-Verlag
Berlin Heidelberg New York

G. Govil R. V. Hosur

Conformation of Biological Molecules

New Results from NMR

With 92 Figures

Springer-Verlag
Berlin Heidelberg New York 1982

Prof. Dr. Girjesh Govil
Dr. Ramakrishna V. Hosur
Tata Institute of Fundamental Research
Homi Bhabha Road
Bombay 400 005, India

ISBN 3-540-10769-X Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-10769-X Springer-Verlag New York Heidelberg Berlin

Library of Congress Cataloging in Publication Data. Govil, G. (Girjesh), 1940–. Conformation of biological molecules. (NMR, basic principles and progress; 20) Bibliography: p. Includes index. 1. Macromolecules. 2. Conformational analysis. 3. Nuclear magnetic resonance spectroscopy. 4. Molecular biology. I. Hosur, R. (Ramakrishna), 1953–. II. Title. III. Series. QC490.N2 vol. 20 [QP801.P64] 538.3s 81-8976. ISBN 0-387-10769-X (U.S.) [574.19'282] AACR2

Typesetting: Elsner & Behrens, Oftersheim
Printing and bookbinding: Brühlsche Universitätsdruckerei, Giessen
2152/3140-543210

Contents

1 General Theory

1.1 Introduction

The determination of the three-dimensional structure of a biological molecule is the starting point in the understanding of molecular mechanisms involved in its complex biochemical reactions. The molecular architecture of multimolecular systems such as membranes and chromosomes provides the key to the fascinating field of molecular biology. Stereochemical details of biological macromolecules and their interactions with pharmacological agents form the basis for drug design. Naturally, the study of the structure and function of biological molecules has aroused tremendous interest and investigations in this area are being carried out in a large number of laboratories. The techniques used for this purpose include both experimental methods (X-ray and neutron diffraction measurements, study of NMR, ESR, vibrational and electronic spectra, ORD, CD and dipole moment measurements, biochemical modifications etc.) and theoretical methods (quantum mechanical and classical potential energy calculations, Monte-Carlo simulations and molecular graphics).

For several years now, X-ray diffraction [1] has served as our only source of information on the three-dimensional arrangements of atoms in biopolymers. Fiber-diffraction of DNA led to the proposal of the DNA double helix. Fibers of long-chain polymers show ordering in the direction of the fibre-axis but not in the transverse plane. Accurate estimates of the dimensions of helical structures can be made using techniques on the basis of which models of biopolymers can be constructed. On the other hand, X-ray diffraction patterns from single crystals of proteins, nucleic acids and other biomolecules, contain information on the three-dimensional arrangement of atoms. With accurate data, positions of all heavy atoms can be obtained with a high degree of precision. Structures of almost 50 macromolecules of biological interest are known with a fair degree of accuracy, not to speak of the wealth of information on small biomolecules. However, when relating these results to biological functions, one has to be sure that the crystal structure is maintained in aqueous solutions under biological conditions of pH and ionic concentrations. Secondly, X-ray crystallography gives a static picture of the molecule constrained by lattice forces and averaged over disorders and thermal motions. As a result, biologists often look at biomolecules as if these are rigid structures though it is now known that parts, or the whole molecule, may exhibit considerable degree of internal motions and flexibility.

During the last decade, NMR has proved to be one of the most powerful tools for the study of biological structures in time and space. With the existing state of art, NMR does not provide the absolute coordinates of atoms in molecules. The advantage of NMR lies in the fact that one can observe molecular structures under actual biological conditions and even monitor structural and biochemical changes in intact cells. Several books and reviews [2–10] have appeared covering various aspects of NMR of biological mole-

cules. In addition, the Chemical Society (London) regularly reviews the subject of biological macromolecules in its periodic reports [11].

This review is devoted to one particular aspect of biological molecules, namely their three-dimensional structure in solutions. To the best of our knowledge, there have been no previous reviews on this specific subject although different classes of molecules have been covered from time to time. The emphasis of this article is on basic principles and methodology and a review of the literature published during the period 1972–80 is given. We have assumed that the reader of this series is familiar with the techniques of NMR and present developments in this area. Therefore, only brief mention has been made of techniques which have particular significance in the context of the structure of biological molecules. Further, since the reader may not be familiar with the terminology used in biomolecular structures, a brief introduction to conformational analysis has been given. A computer search procedure was used to retrieve literature from Chemical Abstracts data files. This has been supplemented by cross references. In a review containing almost 1,250 references, some omissions are inevitable. We express our sincere apologies to authors whose work could not be retrieved by the methods mentioned above.

1.2 What is Conformation?

Let us consider a linear molecule consisting of four atoms A–B–C–D. The three-dimensional structure of such a molecule requires the knowledge of three bond lengths (A–B, B–C, C–D), two bond angles (A–B–C and B–C–D) and one torsion angle ϕ (the dihedral angle between the plane ABC and BCD). In general, for a molecule consisting of N atoms, the structural parameters are N-1 bond lengths, N-2 bond angles and N-3 torsion angles (also called conformational angles or rotational angles). Thus, the structure of the molecule can be completely defined by the above 3N-6 structural elements though sometimes it is more convenient to use the 3N atomic Cartesian coordinates of the molecule with respect to an externally fixed coordinate system. Of these parameters, the bond length and bond angles show relatively small variations when their values in similar molecular fragments are compared. The reason for this is, that the force-constants involved in the stretching of chemical bonds or deformation of bond angles are very large. On the other hand, the potential energies for rotation around chemical bonds are generally of the order of a few kcal/mol. Thus, molecules have a high degree of freedom in rotation around chemical bonds, and the problem of fixing the three-dimensional structure of the molecule often reduces to the knowledge of the N-3 torsion angles. It may be noted that the torsion angles of atoms branching out from the same atom (e.g. the three hydrogen atoms in the methyl group) are related to one another. Molecular symmetry may further reduce the number of independent parameters. These considerations are very helpful in NMR investigations, since in most cases the observables leave the structural details undetermined.

The molecule may exhibit stable structures for more than one value of the dihedral angle. Different spatial arrangements of atoms within a single classical organic structure produced by simple rotation (and twisting but not breaking of chemical bonds) are called conformations [12]. The terms rotational isomers, rotamers, conformers and secondary structures are synonyms of conformation. The term conformation should be distinguished

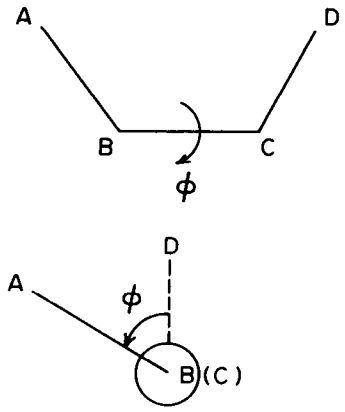


Fig. 1.2.1. The Newman projection diagram and definition of the torsion angle in the four-atom fragment A–B–C–D

from configuration (e.g. D- and L-forms of amino acids), where one can go from one structure to the other only by breaking one or more chemical bonds. In biochemistry, the term "tertiary" structure is used to denote molecular folding stabilized by long-range interactions within the same polymeric chain and "quaternary" structures to denote multichain organizations stabilized by intermolecular interactions.

The Newman projections are one of the most convenient ways to represent molecular conformations around various bonds in a molecule (Fig. 1.2.1). Several conventions exist for measuring the torsion angles. The most general one and the one followed in this review is to measure ϕ using a right-handed rotation: while looking along the bond B–C, the front atom A moves in an anticlockwise fashion relative to the far atom D. This is equivalent to saying that the far atom moves clockwise relative to near atom. The zero value of the torsion angle corresponds to the situation where the bonds A–B and C–D are *eclipsed*, and the angles are measured from 0 to 360 degrees. Such a convention has been followed in literature for most biomolecules. However, the ϕ, ψ angles in peptides are usually measured from −180 to +180 degrees (Sect. 4.1).

In organic chemistry a nomenclature which defines a family of conformations instead of precise values of torsion angles is generally used (Table 1.2.1). Each conformational state in this description covers a range of torsion angles. In solution, molecules have a greater degree of freedom for internal rotations and such a description has an obvious advantage over the specification of precise values of torsion angles.

Table 1.2.1. Classical conformational states and corresponding torsion angles

State	Abbreviations	Range of torsion angles in degrees
syn		0 ± 90
anti		180 ± 90
cis planar	*cis*	0 ± 10
trans planar	*t* (extended)	180 ± 10
gauche	g^+	60 ± 30
gauche'	g^-	300 (− 60) ± 30

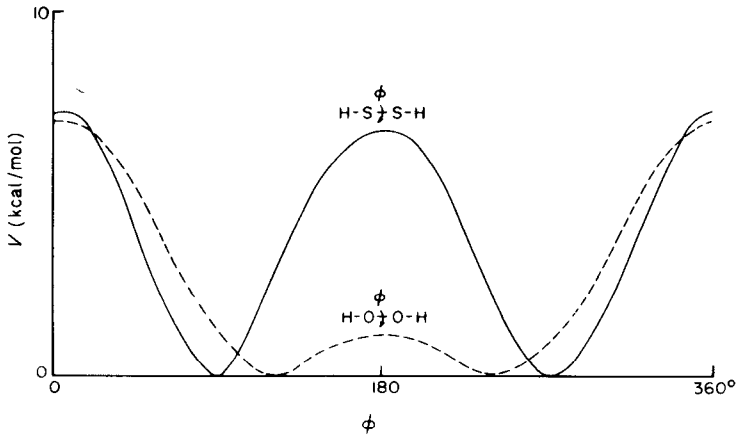


Fig. 1.2.2. Potential energy curves of internal rotation around the central bond in H–O–O–H and H–S–S–H

We can illustrate the concepts discussed above with the help of potential energy curves of two simple and related molecules H–S–S–H and H–O–O–H. The torsion energy behavior of these two molecules [13, 14] is illustrated in Fig. 1.2.2. Two minimum energy structures for each molecule are observed which are identical due to molecular symmetry. Not only the torsion angle for the minimum energy conformations but the whole potential energy curves for the two molecules are significantly different, in spite of the fact that both sulfur and oxygen belong to the same group in the Periodic Table. In fact, the conformation with torsion angles around 90° for S–S bonds are observed in proteins and peptides containing disulfide bridges [15].

1.3 Conformational Theory

It may be useful to introduce some basic principles of conformational theory which will prove useful in our later discussion. The behavior of hydrocarbon chains plays an important role in the conformational behavior of lipids, glycolipids and lipoproteins. We therefore choose *n*-butane as an example. The potential energy as a function of the torsion angle (C–C–C–C) in butane is shown in Fig. 1.3.1. There are three stable conformers for the butane molecule, g^+, t, and g^-, with the two *gauche* conformations being equivalent. The conformational thermodynamics is defined by ΔG (the difference in energy of the g and t isomers) and by ΔG^*, the energy barrier for the transition from the t to the g state. These two quantities are related to the equilibrium populations P_g and P_t of the two conformational states and the rate of interconversion $k(t,g)$ by the following expressions:

$$P_g/P_t = 2 \exp(-\Delta G/RT) \tag{1.3.1}$$

$$k(t,g) = (kT/h) \exp(-\Delta G^*/RT) \tag{1.3.2}$$

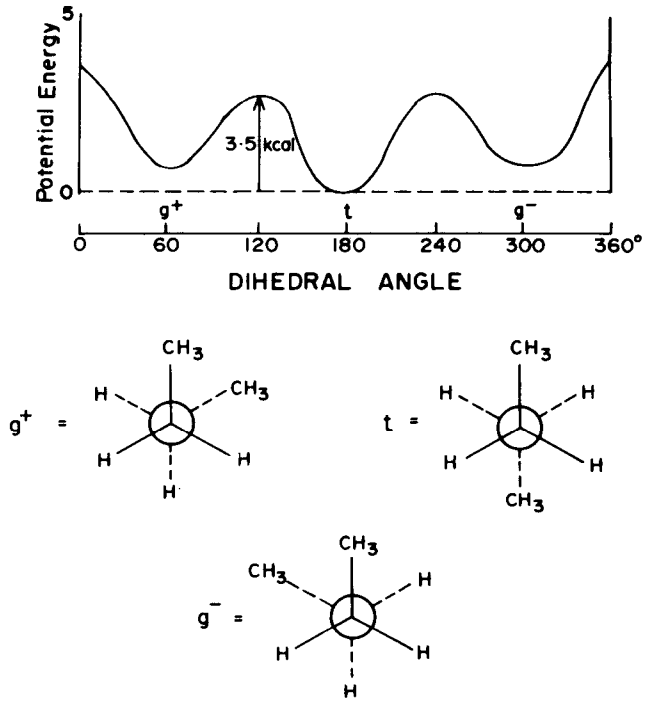


Fig. 1.3.1. Potential energy curve of rotation around the C–C–C–C bond in butane and Newman projection diagram of the minimum energy conformations g^+, t and g^-

For butane, ΔG is approx. 0.8 kcal/mol giving a value of 1 : 4 for $P_g : P_t$ and ΔG^* is 3.5 kcal/mol [16] which corresponds to a rate constant of $10^9\ s^{-1}$ at room temperature. Thus, g and t conformers interconvert at a very fast rate. For most physical techniques, the time-scale of measurements is much longer than the time for interconversion between the conformational states of butane. In such cases, one measures a time-average behavior, and the observed physical properties of the system (O) are given by:

$$\langle O \rangle = \int O\{\phi\} P\{\phi\} d\{\phi\} \tag{1.3.3}$$

with $P\{\phi\} = \exp[-V\{\phi\}/RT] / \int P\{\phi\} d\{\phi\}$.

Here $\{\phi\}$ denotes the set of torsion angles which define the structure of the molecule, $O\{\phi\}$ is the observable under consideration (NMR chemical shift, coupling constant, relaxation time etc.), $V\{\phi\}$ the torsion potential energy, and $P\{\phi\}$ the probability of the conformational state with torsion angles $\{\phi\}$.

Both O and V are functions of torsion angles. It may be noted that what is a fast coversion rate for one physical technique may be a slow rate on the time-scale of another method. Thus, at room temperatures, one can get seperate vibration bands from t and g isomers of butane while in NMR, a time-averaged behavior is observed. The picture of the three-dimensional structure of n-butane which emerges from the above con-

siderations is that the molecule is in a dynamic equilibrium and flips from one conformational state to another at a relatively fast rate. This freedom for internal motions is present even in several large molecules where part of the molecule may be rigid but other parts may show segmental motions because of "fast" conformational transitions. The NMR technique is ideally suited to the observation of such motions.

For NMR, the time-scale in measurements lies in the range 10^{-2} to 10^{-6} s. Under usual temperature ranges of NMR measurements, a time-averaged spectrum is observed for all chemical bonds with ΔG^* less than 15 kcal/mol. The barriers for double bonds are, however, larger than 20 kcal/mol and in these cases, spectra from individual rotamers can be resolved. In the intermediate range of ΔG^*, the NMR line shapes are sensitive to temperature variation due to transitions from "slow" to "fast" rates of interconversion. The line shapes can be analyzed to obtain the barriers for internal rotation in molecules [17].

The integration in Eq. (1.1.3) can be replaced by summation over a finite number of low-energy states. This approximation is known as rotational isomeric state model [18]. Such an assumption is usually made in the analysis of NMR data under conditions of "fast" exchange. Thus,

$$\langle O \rangle = \sum_j P_j\{\phi\} O_j\{\phi\} / \sum_j P_j\{\phi\}. \tag{1.3.4}$$

In the case of butane, the finite number of states(j) can be chosen as the g^+, t and g^- conformations of the molecule. The approximation is valid when the barrier to internal rotation is > 3 kcal/mol. However, if ΔG^* is of the order of 0–2 kcal/mol, then appreciable contribution to the observed properties may occur from states other than those corresponding to the minimum energy. If the functional form of $O\{\phi\}$ is known, then Eq. (1.3.4) and the observed NMR parameters $\langle O \rangle$ can be used to determine the relative conformer populations.

To extend the arguments to longer hydrocarbon chains, we consider the conformational behavior of *n*-pentane. This is described by two torsion angles ϕ and ψ (Fig. 1.3.2). On the basis of the conformational structure of butane, we may predict nine stable rotational isomeric states for *n*-pentane:

$$g^+g^+, g^+t, g^+g^-, tg^+, tt, tg^-, g^-g^+, g^-t, \text{and } g^-g^-.$$

These states can be divided into four groups, the members in each group being identical due to molecular symmetry.

$$tt;\ g^+g^- = g^-g^+;\ g^+t = g^-t = tg^- = tg^+;\ g^+g^+ = g^-g^-$$

A molecular model of the g^+g^- or g^-g^+ structure shows that the two terminal methyl groups come very close to each other. Such structures are clearly unstable due to short ("hard") contacts between the terminal methyl groups. This information can be diagrammatically represented by a two-dimensional isoenergy map as shown in Fig. 1.3.2. The energies in this map are from measurements based on Raman spectroscopy [19]. The *tt* conformation has the lowest energy and is called the global minimum. The four states where one of the two bonds has a *trans* conformation and the other one a *gauche* cor-

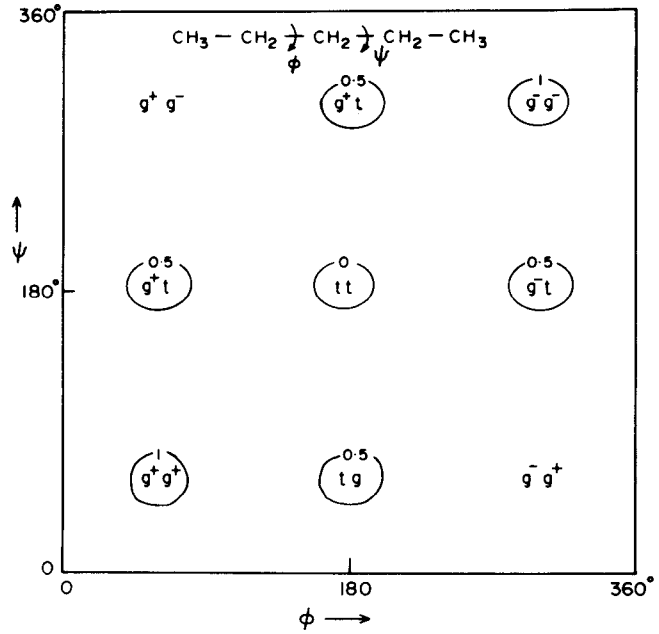


Fig. 1.3.2. Two-dimensional isoenergy map for *n*-pentane. The numbers on the map denote relative energies in kcal/mol relative to the *tt* state which corresponds the global minimum

respond to 0.5 kcal/mol and are called local minima. The g^+g^+ and g^-g^- states have energies of 1 kcal/mol while the states g^+g^- and g^-g^+ have even higher energies.

Clearly, more complex structures require the knowledge of potential energy curves in multidimensional conformational space. Fortunately, the conformational energies are dominated by nearest neighbor interactions and effects beyond two bonds become relatively unimportant except when the molecule has a possibility of folding back on itself. For example, the properties of higher alkanes and even polymethylene can be predicted fairly accurately on the basis of the isoenergy map of *n*-pentane [18, 19].

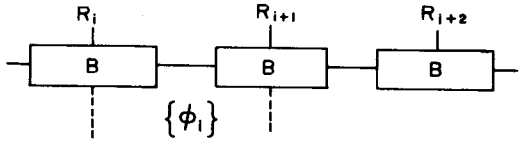
Isoenergy maps are extremely useful in the theoretical analysis of allowed conformations of biological molecules. Such maps take into account neighboring bond interactions and are thus more suitable representations of polymer conformations than the one-bond potential energy curves. They do not consider long-range interactions which are important when a long-chain molecule folds back on itself. Such polymeric structures may be stabilized further by intrachain hydrogen bonds or electrostatic interactions. A large amount of theoretical effort has gone into calculations [20–21] of such maps for all classes of biomolecules and these have proved useful in selecting the correct conformation out of the several possibilities suggested by NMR analysis.

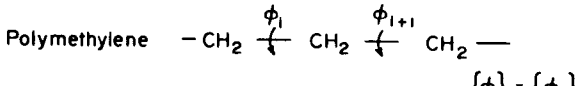
1.4 Structure of Long-Chain Polymers

Majority of biological molecules are linear polymers (Fig. 1.4.1). They are formed by condensation between monomeric units, such as amino acids, nucleotides, and sugars. Biopolymers generally consist of a repeating backbone (B) and variable side chains (Ri). Biological polymers can be classified into three categories: proteins which contain repeating peptide units, nucleic acids (where the repeating unit is a nucleotide) and polysaccharides (repeating sugar units). Each individual molecule in these categories is thus characterized by the sequence of side chains (Ri) which determines the primary structure of the macromolecule. In addition to the side chain angle (χ_i), the three-dimensional structure of these molecules is fixed by a set of backbone torsion angles $\{\phi_i\}$ (e.g. ϕ_i, ψ_i, ω_i for polypeptides). Since a typical biopolymer consists of 100–10,000 residues, the number of structural parameters is very large. Three situations may occur:

(a) Helix: When the values of the backbone torsion angles $\{\phi_i\}$ are independent of the side chain Ri and each residue in the chain rigidly acquires the same set of values, the polymer chain describes a helix. The structure of each residue in a helical polymer is the same as that of its neighbor. Clearly, the number of independent structural parameters is much less, in spite of the large number of atoms in these molecules. Examples of such structures are helical nucleic acids and polysaccharides. Helical structures are generally stabilized by intra- or inter-chain hydrogen bonds.

(b) Ordered structures: A more general case of ordered structures is seen in globular proteins and enzymes. Here, the values of set of torsion angles $\{\phi_i\}$ depend on the nature of side chains (Ri) but do not fluctuate in time. In these cases, each peptide unit has its own unique conformation and the three-dimensional structure of the polymer is several orders of magnitude more complex than the highly symmetrical structure of helical polymers.





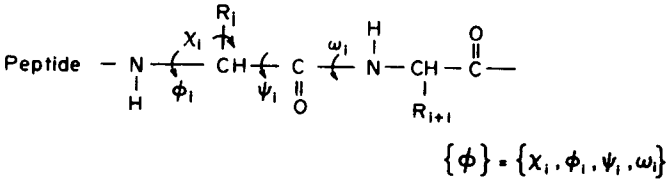


Fig. 1.4.1. Schematic representation of linear polymers. B denotes the backbone, R_i are the side chains and $\{\phi_i\}$ is the set of torsion angles which determine the conformation of the ith unit

(c) Random coil: Under the influence of heat or chemical perturbations like *p*H, solvent or ionic concentration, both types of structures can "melt" into quasirandom structures (random coils or coils). In a random coil, the structure no longer has a spatial meaning since it fluctuates both in space and time. However, a short-range order is present and the properties of a random coil are defined by time and space averages similar to Eq. (1.3.3). Thus, the observed properties of a polymer in its random coil is, to the first order, a sum of the behavior of its monomeric units. The NMR spectrum of a randomly coiled biopolymer can thus be predicted fairly accurately by summing spectra of the individual components.

In ordered (nonhelical) structures, each monomer unit has its own microenvironment and though the primary structure of a segment of the molecule may be same, small differences in chemical shifts and coupling constants may occur, due to secondary and tertiary structures of the molecule. The spectra of such molecules therefore have a large number of poorly resolved lines, particularly in the region corresponding to the backbone and the more abundant side chains. On the other hand, each unit in a helical structure or random coil has the same average environment as its neighbor and the spectra are relatively simple.

1.5 Problems in NMR Studies of Biological Molecules

In principle, each magnetic nucleus in a biomolecule can give its characteristic resonance. The resonances most commonly used in NMR studies are ^{1}H, ^{13}C, ^{31}P, ^{15}N and, to a lesser extent, the resonances of ions when they form part of the biological structures. Use has also been made of labelling the molecule chemically or biosynthetically with ^{2}H and ^{19}F and then observing these resonances. Fortunately, in most cases, labelling with these two isotopes does not alter the biological properties of the molecules. In each case, one can obtain chemical shifts (δ), coupling constants (J), relaxation times (T_1 and T_2) and, whenever necessary, the line shapes (g_v). These parameters provide information on the microenvironment of each nuclear site under study. Alterations as a result of perturbations such as temperature, solution *p*H, metal ions, and substrates can be monitored using difference spectroscopy [2]. Table 1.5.1 summarizes properties and key applications of the above-mentioned nuclear resonances in conformational analysis of biomolecules.

For low molecular weight biomolecules, NMR investigations pose as little a problem as that for a similar organic molecule. However, in the study of a biopolymer having a well-defined structure, problems of sensitivity, resolution and assignment have to be solved before an attempt to determine biomolecular structure from NMR parameters can be made.

1.5.1 ^{1}H-NMR

Even a small protein has a fairly large number of protons. These protons resonate in a narrow spectral region and therefore give rise to an envelope of overlapping lines in the 0–10 δ region of proton NMR. The problem is further complicated by the fact that due

Table 1.5.1. Nuclear magnetic probes in the study of biomolecular conformations

Isotope	I	Relative sensitivity[a]	Resonance frequency[b]	Conformation-dependent parameter and information
^{1}H	1/2	100	100.0	δ: Position of anisotropic groups, hydrogen bonds, charged groups J: Dihedral angles, conformer populations T_1: Internal motions, paramagnetic sites Line shapes: Conformational dynamics, transition rates between rotational states
^{2}H	1	0.1	15.36	D_q: Order parameter, orientation of X–D bond T_1: Conformational dynamics
^{13}C	1/2	1.7	25.14	δ: Electronic environment T_1: Conformational flexibility J: Dihedral angles
^{15}N	1/2	0.1	10.13	δ: Molecular environment J: Dihedral angles
^{19}F	1/2	83	94.08	External probe: Environment of labelled positions
^{31}P	1/2	6.6	40.48	δ: Conformation of phosphate group Line shape: Direction of chemical shift tensor J: Dihedral angles

[a] Relative sensitivity is at constant field for equal number of nuclei
[b] Resonance frequency in MHz at a field of 2.35 tesla

to slow tumbling rates of large molecules, the correlation times are longer and the lines are broader. For this reason, one generally relies on the less crowded ^{1}H spectral region such as hydrogen bonded N–H, histidine or tyrosine protons or paramagnetically shifted resonances such as in heme proteins or metalloenzymes. The advent of superconducting NMR magnets has greatly promoted proton NMR of large molecules by providing spectral dispersion. High-resolution NMR magnets with fields up to 10 Tesla are presently available and attempts at producing even higher fields are being made.

To get meaningful results one generally works with 10 mM solutions in continuous wave (CW) proton NMR. For a 1 ml solution of a protein containing 100 amino acid residues, this corresponds to about 100 mg of sample for each experiment. Gram amounts of material are needed for less sensitive nuclei such as ^{13}C. It is difficult to obtain such large quantities of biological materials. Thus, till a few years ago, protein NMR was a difficult proposition. The advent of Fourier Transform (FT) techniques provided a big boost to NMR in biological research by reducing the minimum solution concentration to 0.1 mM for ^{1}H- and 10 mM for ^{13}C-NMR. The impact of modern NMR techniques can be judged from the growth of NMR literature (Fig. 1.5.1) on biomolecules. In addition, the more recent literature has concentrated on larger molecules such as small proteins, t-RNA, their interactions with drugs and substrates and other complicated systems.

Finally, assignment of resonances has been a major problem in getting results of biological interest. For example, the assignment of the four resonances from histidine