PINSA, **65, A**, No. 6, November 1999, pp. 633-653.

NMR OF BIO-MACROMOLECULES

R V HOSUR*, P K RADHA, S SINGH AND P K PATEL

Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005 (India)

(Received 17 June 1999; Accepted 19 July 1999)

Nuclear Magnetic Resonance has emerged during the last two decades as the single most powerful tool for determination of high resolution structures of biological macromolecules, in aqueous solutions under conditions akin to those prevalent in intracellular environments. The technique has seen radical developments, both with regard to physical concepts and with regard to instrument technologies. While the high magnetic fields have enabled greater dispersion of signals, heteronuclear multidimensional experiments have enabled spreading of the information in several dimensions making use of the different chemical shifts of the different nuclear types. This development has embraced molecular biology in all its glory for the preparation of isotopically labelled molecules in large quantities. Dealing with large amounts of data and extensive computer modelling calculations have encompassed software technologies as integral parts of NMR developments. Till date high resolution solution structures of more than 300 proteins and nucleic acid systems, having molecular weights in the range of 5-15 kDa, have been determined and the trend is to study even larger molecular systems and macromolecular complexes. The present review surveys the general trends with examples drawn largely from the work carried out in our own laboratory.

Key Words: Biomolecular Structure; Multidimensional NMR; Nucleic Acids; Proteins

1 Introduction

Nuclear Magnetic Resonance (NMR) and its applications in chemistry have been growing at an exceedingly rapid pace ever since its discovery in 1945[1,2] by two groups of researchers led by Bloch and Purcell at Stanford University and Massachussets Institute of Technology respectively. Among the variety of applications to which NMR has been put, molecular structure determination in solution media stands out prominently and continues to attract the attention of a large number of researchers world wide. Very high investments are being made for these purposes and it is a common thing to see dedicated spectrometers in individual laboratories, despite the high cost of these machines. The molecular size limits have gone up from 10-20 atoms in the early 1950's to 5000 atoms in the 1990's. The structural detail obtained by NMR today is comparable to that obtained by X-ray crystallography which has so far been the single reliable technique for molecular structure determinations. Several protein and nucleic acid structures are being determined every year by both these methods and also with comparable speeds. Thus with the availability of two reliable and independent methods now, each one of them serves as a complementary technique for the other, for comparisons, starting models etc.

Several factors have contributed to the success story of NMR described above. New concepts and techniques have been discovered. These include, chemical shift, coupling constant, spin decoupling, pulsed excitation, spin echo, nuclear Overhauser effect (NOE), Fourier transform NMR, water suppression techniques, polarization transfers, two dimensional NMR, three dimensional NMR, triple resonance experiments etc., to name a few. In parallel, spectrometer designs and performances have improved enormously over the last three decades. The magnetic field strengths have gone up from 4-5 kGauss in the 1950's to 180 k Gauss in the 1990's. This implies an approximately 250 fold increase in the sensitivity and a 40 fold improvement in the resolution in the spectra. Superconducting magnets resulted in highly stable fields and contributed to enhanced resolution and

*Author for correspondence

better line shapes in the spectra. Along with these developments were the sophistications brought about by the introduction of computers in the spectrometers and the breath-taking speed in the progress of the computer technologies.

In addition to the NMR methodology related developments, the new discoveries in the areas of chemistry and biology had their roles to play in the success story of NMR structure determination. Noteworthy among these are the genetic engineering methods, polymerase chain reaction (PCR) and new synthetic routes for peptide and nucleic acid synthesis. With the genetic engineering and PCR methods, it became possible to produce desired segments or domains of large proteins selectively and in large amounts necessary for structure determination experiments. It also became possible to isotopically label and enrich these molecules, which is a basic requirement for performing many of the present day double and triple resonance experiments. In a like manner the new synthetic routes enabled production of large amounts of pure nucleic acid and peptide segments, which paved the way for detailed structure characterizations and also provided the basic 'primers' in PCR reactions.

The present article is intended to be a trace of the excitements in the area of structural chemistry and biology by NMR. There is no intention of making it an exhaustive review and accordingly a number of examples are taken from the work carried out in our own laboratory. A few exciting illustrations are also taken from others research wherever necessary. The language and the style are deliberately maintained to be qualitative avoiding all the technical jargons of NMR, so that the material can appeal to the general readership. In the following we present a systematic description of the relevant methodological developments and this will be followed by their applications in different structure determinations. In the end we give a peep into the future following the current state of knowledge and thinking within the scientific circles.

2 Multidimensional NMR Methods

The invention of two dimensional (2D) NMR in the 1970's [3,4] caused the second major revolution in NMR, the first being the Fourier transform NMR, in the 1960's. The basic idea of the technique was to display the interactions (correlations) between the various nuclei in a given molecule on a plane, thereby enabling unambiguous assignment of the individual resonances to specific nuclei in the molecule. This in turn enabled quantification of the interactions as an essential step towards three dimensional structure determination.

Fundamentally, there are two types of interactions between nuclei which are of significance for structure determination by high resolution NMR. These are: (1) J-coupling interaction which is mediated by the electrons in the intervening bonds and (2) through space dipole - dipole interactions. These interactions cause magnetization transfers between nuclei whenever they are somehow perturbed and kept in a non-equilibrium state. The two kinds of transfers caused by the two types of interactions have distinctly different properties and hence can be separated and observed independently. This is the basis for the design of all two dimensional and higher dimensional NMR experiments.

The essential elements of one, two and three dimensional NMR experiments are sketched in Fig. 1. Conceptually, a two dimensional experiment is an extension of the one dimensional FTNMR scheme which is a time domain experiment. The data collected as a function of time following the excitation of the spin system, is Fourier transformed to generate an NMR spectrum. In the 2D experiment the above time axis is segmented by introducing additional time periods before data collection. In a generalized sense, the time axis is considered to be divided into four periods namely, Preparation, Evolution, Mixing and Detection, and the data is collected only during the Detection period. Likewise a three dimensional experiment can be conceived as an extension of a two dimensional experiment and can be treated as a combination of two 2D experiments. Higher dimensional experiments can be generated in a similar manner. In all the cases the time domain signal is actually collected during the Detection period only. Evolution periods are variable time periods in the sense that they are systematically incremented during the course of the experiment. Consequently, the time domain data is collected as a function of two or three time variables in 2D and 3D experiments respectively. Subsequent Fourier

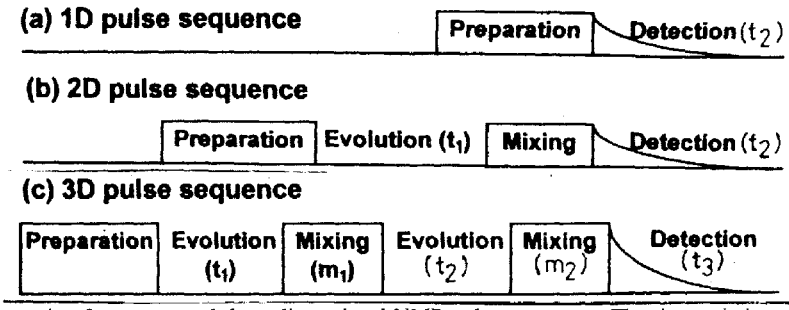

Fig. 1 Schematic of one, two and three dimensional NMR pulse sequences. The time axis is segmented into various periods as indicated. The data is actually collected only during the detection period in each case. The evolution periods serve as indirect detection periods to generate the other frequency axes. Mixing periods help in causing magnetization transfers between the spins in the network.

transformation of these data with respect to each of the time variables generates the additional frequency axes of the 2D or 3D experiments. The mixing periods following the evolution periods help in causing magnetization transfers between nuclei which are frequency labelled during evolution and thus the detected signals carry information about the history of the spin system evolutions. Consequently the evolution periods serve as indirect detection periods of the experiment. The Preparation period is used to prepare the nuclear spin system in a suitable equilibrium or a non equilibrium state depending upon the purpose of the experiment. All of these periods may simply be delays or may contain radiofrequency (RF) pulses and these distinguish the various 2D and 3D experiments.

The 2D, 3D and other higher dimensional experiments can be classified as Homonuclear or Heteronuclear, according to the types of nuclei being correlated in the experiment. In a homonuclear experiment all the frequency axes belong to the same type of nuclei, for example, proton. In contrast, in a heteronuclear experiment, one or more of the frequency axes necessarily belong to a different type of nucleus or nuclei. In most cases, the detected nucleus is ^1H for sensitivity reasons. What this means is that the frequency information of the heteronuclei evolving during the 'evolution' periods is somehow transferred to the ^1H for detection. There exist a number of approaches[5,6] for such'polarization transfers' and the most commonly used ones are INEPT (Insensitive Nuclei Enhancement by

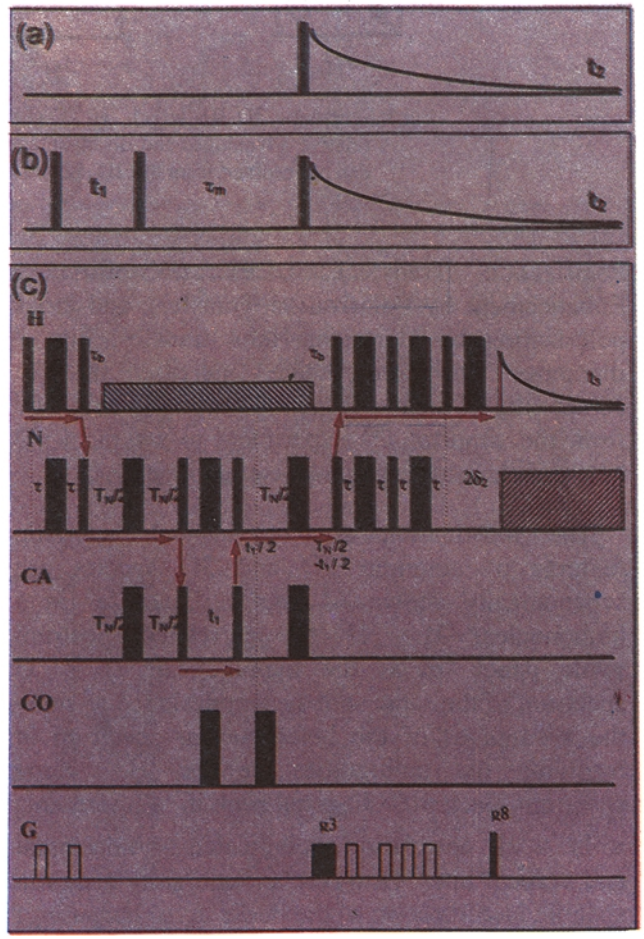

Fig. 2 Progressive increase in the complexities of the experimental pulse sequences as we go from the simplest 1D FTNMR[51] experiment through a 2D NOESY[52] pulse sequence to a 3D triple resonance HNCA[53] experiment. Each vertical thick bar represents a radiofrequency pulse. The magnetization transfer pathways in each case are indicated by coloured lines. Clearly the triple resonance experiments have many more pulses and other paramters to be optimized and demand greater stability and performance from the spectrometers. They require three independent channels. The last row in the sequence is for gradient pulses which are used for selection of transfer pathways and also for suppression of unwanted signals.

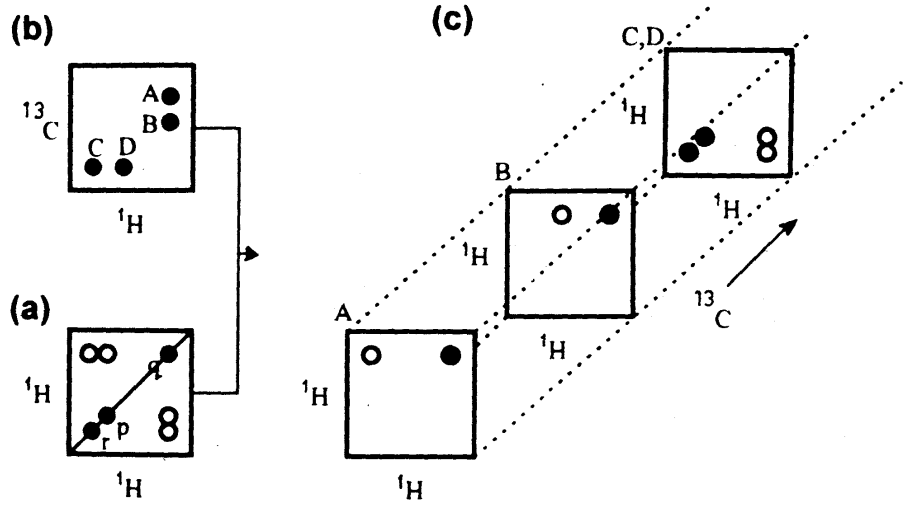

Fig.3 Schematic 2D ^{1}H-^{1}H (a), ^{1}H-^{13}C (b) and 3D ^{1}H-^{1}H-^{13}C (c) spectra to show their relative advantages and disadvantages. Ambiguities of assignments in 2D spectra can be resolved from the 3D spectra (see text).

Polarization Transfer), DEPT (Distortionless Enhancement by Polarization Transfer), and cross polarization by Hartmann-Hahn transfer. Fig. 2 illustrates the progressive increase in the complexity of the experimental sequences as we go from the simplest 1D experiment to a typical 3D heteronuclear triple resonance experiment.

The information content and characteristic features of common 2D and 3D spectra are schematically shown in Fig. 3. Panel 'a' is a homonuclear ^{1}H - ^{1}H 2D correlated spectrum while panel 'b' is the hypothetical ^{1}H - ^{13}C spectrum of the same spin system designed to show the advantages of the heteronuclear spectrum in resolving the ambiguities in the ^{1}H alone spectrum. Each off - diagonal peak in 'a' represents a ^{1}H - ^{1}H correlation as indicated by horizontal and vertical lines. However an ambiguity arises in specific identification of the peaks at '*q*' because of identical F1 chemical shifts. It remains uncertain as to whether there are two protons at '*q*' coupled to two different protons or whether there is only one proton coupled to two different protons. This is resolved in panel 'b' because of the fact that at the particular proton chemical shift two carbon chemical shifts are correlated. A converse situation arises for the other two protons for which the carbon chemical shifts are identical. Thus a combination of both these experiments would help resolve all the ambiguities. This is illustrated in the 3D spectrum shown in panel 'c' which has been generated by combining the spectral features of the two spectra in panels 'a' and 'b'. In Figs. 4 and 5 are shown illustrative homonuclear and heteronuclear, and two and three dimensional NMR spectra of some large biomolecules recorded in our laboratory. The advantages of a 3D ^{13}C resolved TOCSY spectrum over the 2D ^{1}H TOCSY spectrum are clearly evident.

The information content of a multidimensional NMR spectrum is determined by the pathway of magnetization transfer among the spins through the pulse sequence of the experiment. This is illustrated in the 3D pulse sequence in Fig. 2 by the dashed line. The transfers could be mediated either through the *J*-couplings or through the dipolar couplings between the various spins. The former is referred to as coherence transfer and the latter which results in incoherent transfer is also often described as NOE (nuclear Overhauser effect) transfer. In a particular pulse sequence there can be several steps of coherence transfers and these may also be combined with NOE transfers. While the different pathways of coherence transfer help in filtering the spectra differently, a combination of these with NOE transfer helps in studying interactions between molecules. Separation of NOE interactions according to chemical shifts of carbon, nitrogen etc. is an essential step in the study of large proteins from the structural point of view.

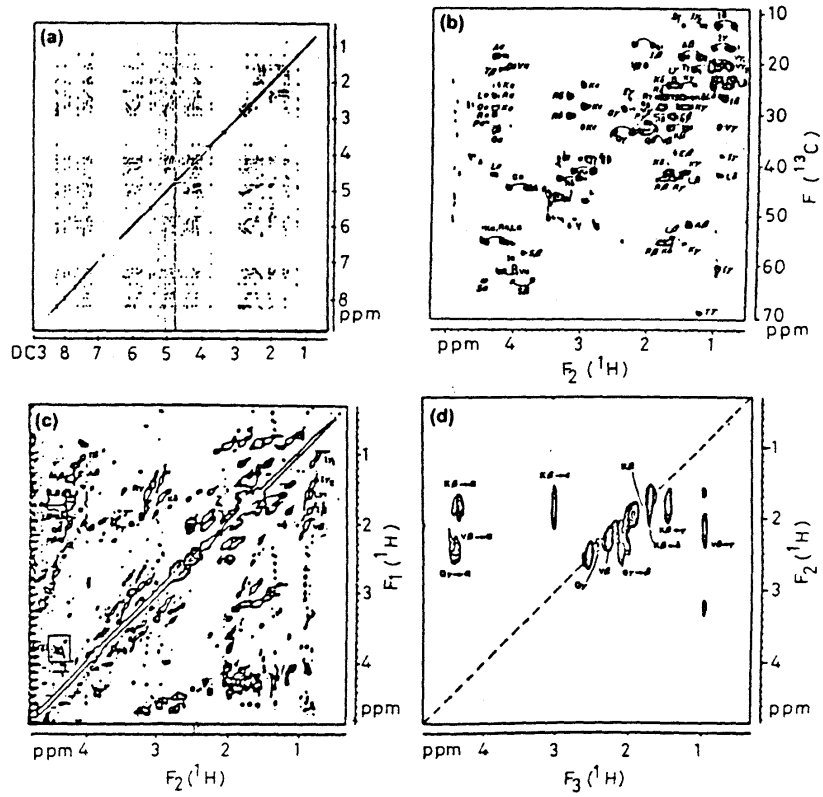

Fig.4 (a) 2D NOESY spectrum of 14 mer DNA segment . (b) 2D TOCSY relayed ^{1}H - ^{13}C HMQC[54] spectrum of DNA binding domain Myb protein (Mol. Wt. 20 kD). (c) 2D TOCSY[55] spectrum of the same protein as in 'b'. (d) A slice through the 3D ^{13}C resolved ^{1}H - ^{1}H TOCSY spectrum of the same protein[56]. The slice represents a filter through the 2D spectrum shown in 'c'. All the spectra have been recorded on a 500 MHz Bruker NMR spectrometer.

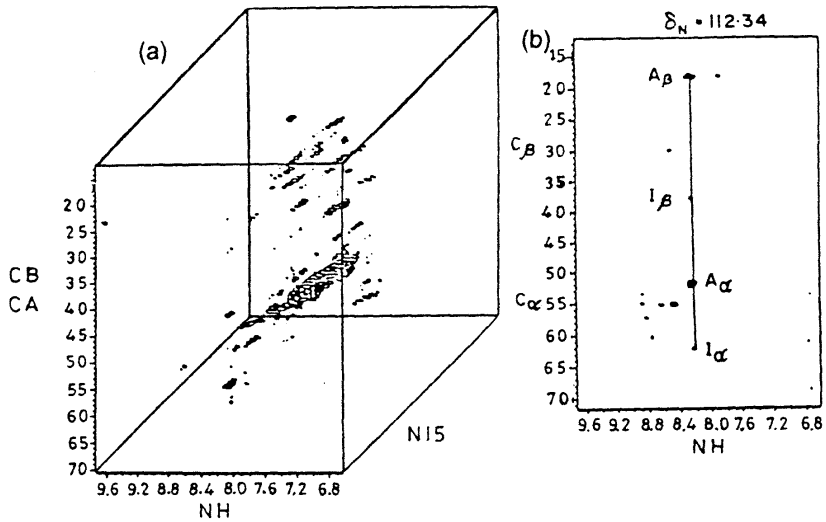

Fig.5 (a) A 3D CBCANH[57] spectrum of the same protein as in Fig. 4 recorded on a 600 MHz Varian NMR spectrometer. This contains correlations between NH protons, ^{15}N of the ith residue and C^{α} and C^{β} nuclei of the ith and (i-1)th residues along the amino acid sequence of the peptide chain. (b) a slice through the spectrum in 'a' indicating the sequential connectivity between two residues (an Ileu-Ala stretch) in the polypeptide chain.

With these ideas a large number (more than 100) of experimental sequences have been published todate. Several reviews and books have been written covering the principles of these techniques and also their applications to proteins, nucleic acids, and various types of biomolecular complexes [5-10].

3 Isotope Labelling of Proteins and Nucleic Acids

The heteronuclear multidimensional experiments on proteins depend crucially on efficient transfer of magnetization from ^{1}H to the NMR active isotopes ^{13}C and ^{15}N and transfer back to the proton for observation. At natural abundance these nuclei are available only to the extent of 1.1 % for carbon and 0.37 % for nitrogen. That means, the magnetization from only those protons attached to these nuclei will participate in the coherence transfer processes. This will clearly throw away a major portion of the proton magnetization and the sensitivity of the experiments will be extremely low. Besides, transfers between two heteronuclei would be unthinkable in such a situation; consequently complex experiments with multiple transfer steps involving heteronuclei would become formidable. Therefore, for the success of the above experiments it is absolutely necessary to have molecules with uniform isotopic enrichments. For most experiments enrichment to the level of 100 % is essential, but for some special situations partial enrichment to the extent of 50-70 % may be desirable.

Labelling by ^{13}C and ^{15}N in proteins and nucleic acids opens up possibilities of measuring heteronuclear one, two and three bond coupling constants. While, on the one hand this knowledge is crucial for the design of the experimental sequences, the three bond coupling constants are also intimately related to specific torsion angles in the molecules, on the other. These provide important and otherwise unavailable structural inputs. A substantial effort has therefore been devoted in recent years to measurement of such coupling constants in large proteins and nucleic acids and these have complemented the structural informations derived from the NOE data.

In addition to ^{13}C and /or ^{15}N labelling, many a times ^{2}H labelling has often been used in proteins. If all the nonexchangeable protons such as those on the *alpha*-carbons and those in the side chains are deuterated, the only signals observable in H_2O solutions will be the exchangeable protons such as the amides in the protein. Deuteration also eliminates one of the efficient relaxation mechanisms namely, ^{1}H-^{1}H dipole-dipole relaxation and thus causes a significant narrowing of the amide signals. This not only helps in increasing the resolution in the spectra but also permits design of experiments with longer pulse sequences. Because of the longer relaxation times of the protons the signal survives longer along the time axis and thus more steps of magnetisation transfers can be included in the pulse sequences. The disadvantage of deuterium labelling, however, is that it leads to loss of information which is contained in the different ^{1}H-^{1}H interactions. Thus it will never be possible to obtain a complete 3D structure of a protein by a fully deuterated protein alone. Deuterium labelling may also cause small changes in the chemical shifts of protons and other nuclei. Thus deuterium labelling is always used as an ancillary strategy for resolving ambiguities due to signal overlaps and often partial labelling is used to circumvent some of the problems mentioned above.

While uniform enrichment of all the carbons and nitrogens is useful for heteronuclear multi-dimensional experiments, selective labelling of a few amino acids or a group of amino acids in a large protein can be used as a selective filter. This way the amino acids located at the active site of an enzyme or at specific binding sites in a protein can be labelled and their signals can be monitored for understanding reaction mechanisms or mechanisms of specific recognition by proteins. Labelling of specific amino acids simplifies the spectra - of course by throwing away some information - and renders them easier to analyze. Quantitative analysis of the signals coming from such residues which may be spread over the whole length of a protein sequence, can help in studying protein folding pathways in large proteins.

The most common methods for isotopic enrichment of proteins and nucleic acids rely on biological techniques, namely, gene cloning, subcloning and protein engineering. The desired

proteins are produced by specially designed expression systems which carry the genes coding for the desired amino acid sequences of the proteins. The cells harboring these expression systems are grown in nutrient media containing isotopically labelled nutrients such as ^{13}C -labelled glucose and ^{15}N labelled ammonium chloride. Consequently, the expressed proteins will be uniformly isotopically enriched. Generally the expression systems are constructed in such a way that the desired protein is the major protein expressed in the cells. For deuterium labelling the cells will have to be grown in media prepared in $^{2}H_2O$ and all the nutrients supplied externally must be deuterium labelled. Similar strategies are used for production of labelled nucleic acids which employ *in-vitro* transcription systems.

Cloning, expression and labelling of proteins appears straightforward in principle, but is often beset with several uncertainties [11,12]. First of all, for a new protein, it is difficult to predict the efficacy of the expression system used. The expressed protein may turn out to be harmful to the cells and then the cells may die. The growth of the cells in isotopically enriched media - so called minimal media - is usually slower than in rich media with normal nutrients. There could also be problems of post translational modifications of the expressed proteins by the cellular processes. The expressed proteins may get associated with other proteins inside the cells and then their structures may change or it may become difficult to extract the proteins. When the cells are grown in $^{2}H_2O$ for deuterium labelling, the growth rates are much lower and some times may not even grow. The cells will have to be slowly adapted to the new solution conditions by progressively increasing the $^{2}H_2O$ content in the media. Therefore, it is clear that a lot of care has to be taken in designing the expression systems and even so, a fair amount of trial and error is involved in successful production of labelled and functionally active proteins.

4 Computational Techniques in NMR Structure Determination

Analysis of multidimensional NMR data, as will be described in the following sections, provides two types of structural inputs, namely, interproton distances and torsional angles. Then one has to construct molecular structures which satisfy these inputs treating them as some kind of constraints. An important consideration here is that the conformational space which is multidimensional has to be appropriately scanned so as to be sure that the derived structures are unique. It is also necessary to ensure that the structure derived has the lowest conformational energy and truly represents the global minimum on the potential energy surface. This becomes a particularly formidable task when the inputs derived experimentally are rather limited. Further, one has to see the effects of starting models and the effects of the environment on the structures derived. These considerations have led to the development of two types of algorithms which are briefly discussed below for the sake of completeness.

Distance Geometry

This is a general method for converting a set of distance bounds between atom pairs into a configuration of these atoms that is consistent with the bounds[13]. In this algorithm, a molecular structure is described in terms of the set of all pairwise interatom distances which is written in the form of a so-called 'distance matrix'. For a correct structure the experimentally derived distances must figure in such a matrix as precisely as possible. This is achieved by defining a 'target function' which represents the violations from the distance inputs of the actual distances in a structure. The target function is minimized so as to drive the structure towards best satisfaction of all the constraints. The tolerance in the target function satisfaction determines the precision of the structure.

The above algorithm has been used successfully in determining structures of several proteins and nucleic acid molecules. However one of the difficulties in this approach is that the mimimizer has to spend a lot of time and effort in satisfying the covalent geometry of the molecule which is conformationally invariant. To circumvent these problems algorithms have been devised which work in the torsion angle space whereby, directly bonded atom distances and bond angles are automatically fixed[14,15]. These algorithms are currently most popular because of their speed which consequently enables a better search through the conformational space.

Molecular Dynamics and Simulated Annealing

The distance geometry protocol does not explicitly include the energy of the system as a variable and thus does not ensure that the structure derived as an acceptable structure has the lowest possible conformational energy. This problem is overcome in the 'Molecular Dynamics' protocol[16]. An energy function including all different interactions is defined and the Newtons equations of motion are solved numerically to derive the trajectories of atomic motions as a function of time. These reflect, on one hand the atomic motions and help to identify the stable states in the system, on the other. From these states, energy minimization relieves the strain present in the system and helps finding the local minima on the potential energy surface. Molecular dynamics calculations allow potential barriers to be crossed, and thus enable identification of several minima on the potential energy surface. In simulated annealing the molecule is heated to a very high temperature and then slowly cooled to the temperature at which molecular dynamics calculations are to be carried out. This protocol allows a better search through the conformational space and thus can help in locating the global minimum more reliably. This also can generate several starting models for dynamics and eliminates the biases of the starting models on the final structures.

Molecular dynamics calculations can also be carried out in the presence of various types of constraints given in the form of suitable potentials. This restricts the conformational space to be searched and speeds up the computations. Further, if these constraints are derived from experimental NMR data in aqueous solutions, then the structures represent truly the solution structures of the molecules.

5 Nucleic Acid Structure and Dynamism

DNA and RNA, the two types of naturally occurring nucleic acids, are polymers consisting of a repeating sugar-phosphate backbone in which, to each of the sugar rings is attached one of the four bases, adenine (A), thymine (T) (or uracil U), cytosine (C) or guanine (G). DNA and RNA differ basically in the type of the sugar ring (deoxy ribose in DNA and ribose in RNA) and in that thymine in DNA is replaced by uracil in RNA. However, structurally and functionally, the two types of nucleic acids are widely different. RNAs exhibit a wide variety of folded structures compared to DNA and are also directly involved in a variety of functions. DNA, on the other hand is the genetic material which exists largely as a duplex wherein two polymeric chains or strands are held together by base pairing; A pairs with T with two H-bonds and G pairs with C with three H-bonds. The sequence of the bases in each strand is the genetic information which is expressed in a regulated manner by the proteins carrying out the variety of functions in the cell.

NMR spectroscopy has contributed immensely during the last two decades to our present knowledge of nucleic acid structure and function inside a cell. Atomic level details of three dimensional structures of several DNA duplexes have been determined and these have indicated that the duplex is not a regular structure as had been thought before, but rather is highly irregular with important sequence dependent variations. Thus the backbone of the DNA has some sort of a signature of its sequence and this is what enables some enzymes such as restriction enzymes and proteins such as transcription factors recognize specific sequences in a long million base DNA. Besides, the story has gone much beyond the DNA duplex which caused a major revolution in the 50's and led to the birth of the new area in science namely, 'Molecular Biology'. Although it was believed that DNA can form many other different kinds of structures, either transiently or in the presence of some other agents, a detailed characterization of several unusual forms has been achieved only during the last two decades. We will present below the NMR derived structures of these unusual forms after giving a short description of the NMR strategy of structure determination.

NMR Strategy of Nucleic Acid Structure Determination

Predictably, the first step in all such efforts is to obtain resonance assignments for all the protons in the molecule. Subsequently the interactions between the assigned protons must be quantified to derive the structural informations. By and large, nucleic acid structure determination has relied on homonuclear ^{1}H - ^{1}H 2D correlated experiments, although occasionally, heteronuclear experiments employing ^{31}P have also been used. In nucleic

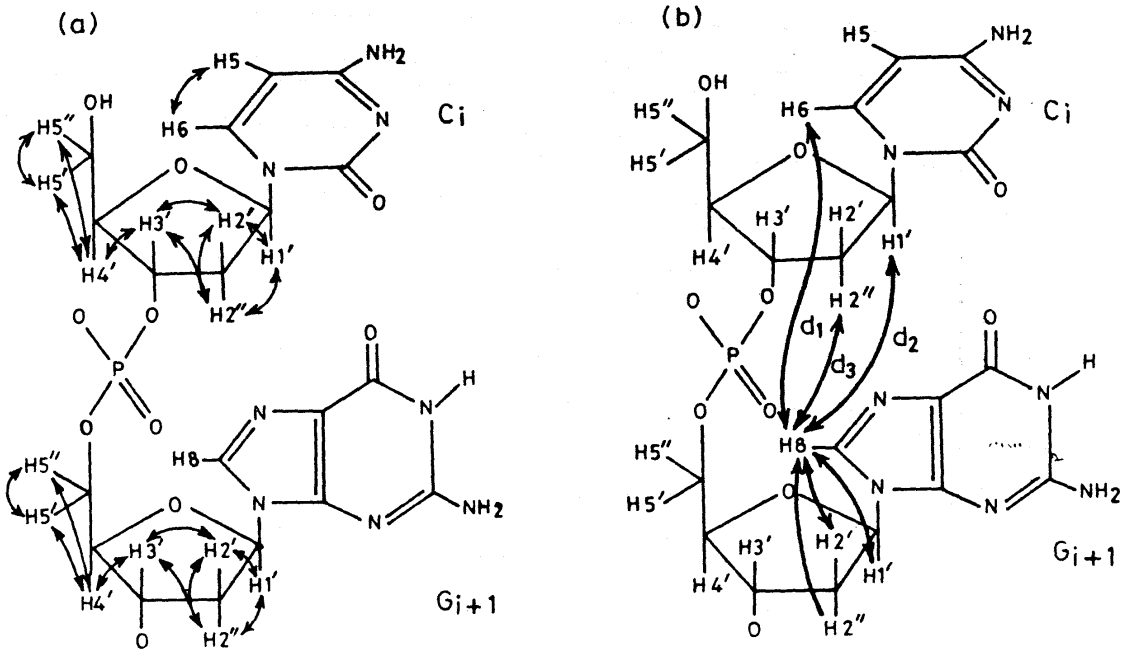

Fig.6 Observable *J*-correlations (a) and NOE correlations (b) in a duplex DNA segment. Together, these enable complete sequence specific assignments in DNA segments.

acids, two types of protons can be easily distinguished. These are: the exchangeable imino protons which resonate in a distinct region between 11-16 ppm downfield of the reference TSP (tetramethyl silyl propionate) and the non-exchangeable protons which resonate between 1-9 ppm downfield of TSP. The former are direct monitors of base pairing and indicate the symmetry, strand orientations etc.; the latter are useful for detailed characterization of the structures.

Fig. 6 shows the observable correlations in nucleic acids in 2D NMR spectra [9,17-19]. Panel 'a' displays the correlations mediated by *J*- coupling interactions between protons while panel 'b' shows the NOE correlations. The *J* coupling correlations help in identifying spin systems of the individual nucleotide units and the NOE correlations permit a sequential walk along the length of the molecule. Using these two types of correlations it is straight forward to obtain sequence specific assignments for all the protons in a given nucleic acid segment. An important point to note here is that most of the NOE correlations between nonexchangeable protons observed would be between adjacent nucleotide units in the same strand. This helps a great deal in obtaining specific assignments.

After the assignments, the NOE interactions are quantified to extract interproton distances by making use of the fact that the intensities under certain conditions are inversely proportional to the sixth power of the respective interproton distances. Thus the NOE intensities are a sensitive function of their distances and from experience it is generally observed that the observation of a NOE correlation limits the distance to less than 5-6 A°. These interproton distances constitute the major input for computational algorithms such as restrained Molecular Dynamics or Distance Geometry for reconstruction of the three dimensional structures of the molecules. The conformational space which is multidimensional is searched for identifying all those conformations which satisfy the experimental interproton distance constraints, torsion angle constraints derived from *J*-correlated spectra and also some other constraints regarding base pairing, symmetry etc. Because of the uncertainties in the estimation of the constraints, there is always a tolerance specified and at the end of the conformational search, one ends up with a family of closely related structures. For each structure, the NOE intensities are calculated by rigorous relaxation matrix methods[20,21] and then the structures are screened on the basis of quantitative fitting of calculated and experimental NOE

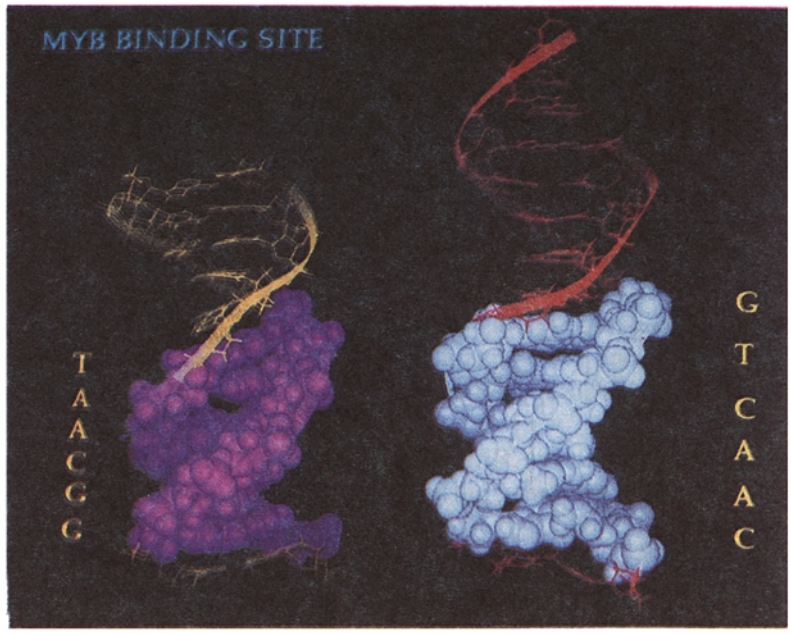

Fig.7 Solution structures of two DNA sequences ACCGTTAACGGT and ACAACTGCAGTTGT determined by NMR in our laboratory[22,23]. The space filled regions are the Myb cognate sites in the two molecules. Rest of the portion is shown in stick model to indicate the base pairing and the backbones explicitly.

intensities. It is generally observed that larger the number of the constraints, better is the convergence of the structures and that the derived structure will also be unique. Nonetheless, depending upon the extent and distribution of the constraints defining different portions of the molecules, the structures will be defined to different levels of precision at different portions. The NOE intensities also carry information regarding the dynamism in the molecules and a proper quantitative interpretation of the intensities reflects on these aspects as well.

Polymorphism in DNA

(a) Duplexes

This is the most common form of DNA formed by holding together of two strands by specific base pairs A:T and G:C. Fig. 7 shows the atomic resolution structures of two DNA duplexes, 12 and 14 units long, determined in our laboratory[22,23]. Fig. 8 shows an analysis of the local helical axis parameters in the two cases to illustrate the sequence dependent variations in the structures. Both these DNA segments contain recognition sites for the Myb protein, an oncogene product, but the recognition sequences have different flanking units and have different lengths of DNA on either sides. These factors seem to cause substantial differences in the three dimensional structures of the two molecules.

(b) Mismatches

Base pair mismatches in DNA occur due to limited fidelity of the DNA polymerase enzyme and these are then corrected by the repair machinery of the cell. Sometimes, however, the mismatches go uncorrected for a variety of reasons one of which may be that the mismatches go unnoticed, which in turn reflects on the structural features at the mismatch sites. Some mismatches are tolerated more than others indicating that all the mismatches do not distort the duplex structure to the same extent. What would be the consequences of such mismatches in specific recognition processes? We have investigated this problem in one particular case namely, that of LFB1 recognition. This homeodomain protein recognizes the sequence TTAATNATTAA where N is any nucleotide unit. We have studied the structural consequences of GG and CC mismatches at the N site in the molecule[24]. We observed that CC mismatch produces a looped structure with substantial distortion in the centre. On the other hand a GG mismatch can produce a variety of

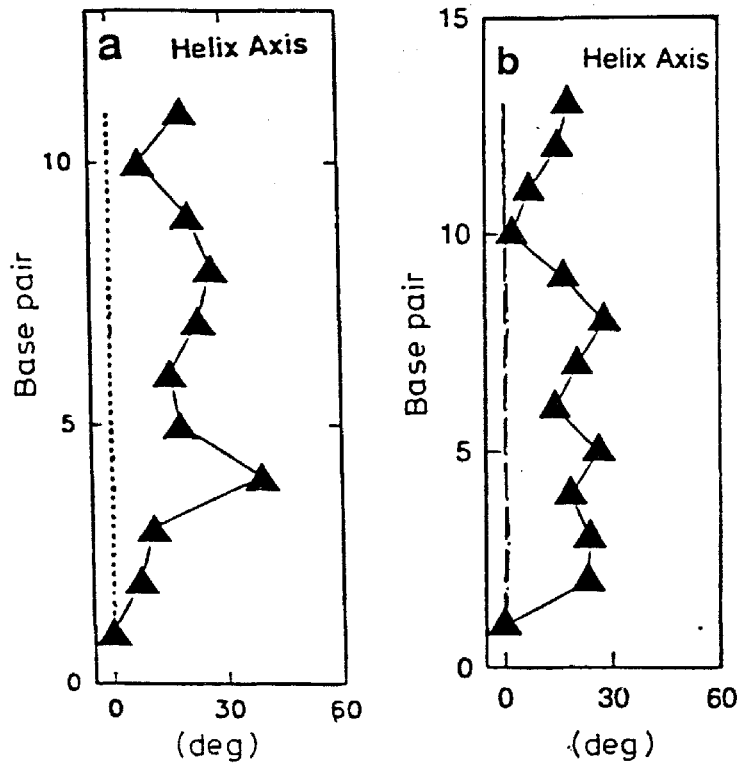

Fig.8 Progress of the helical axes in the structures of the two molecules shown in Fig. 7. The zig zag path indicates the irregularities in the structures and these are different in the two molecules.

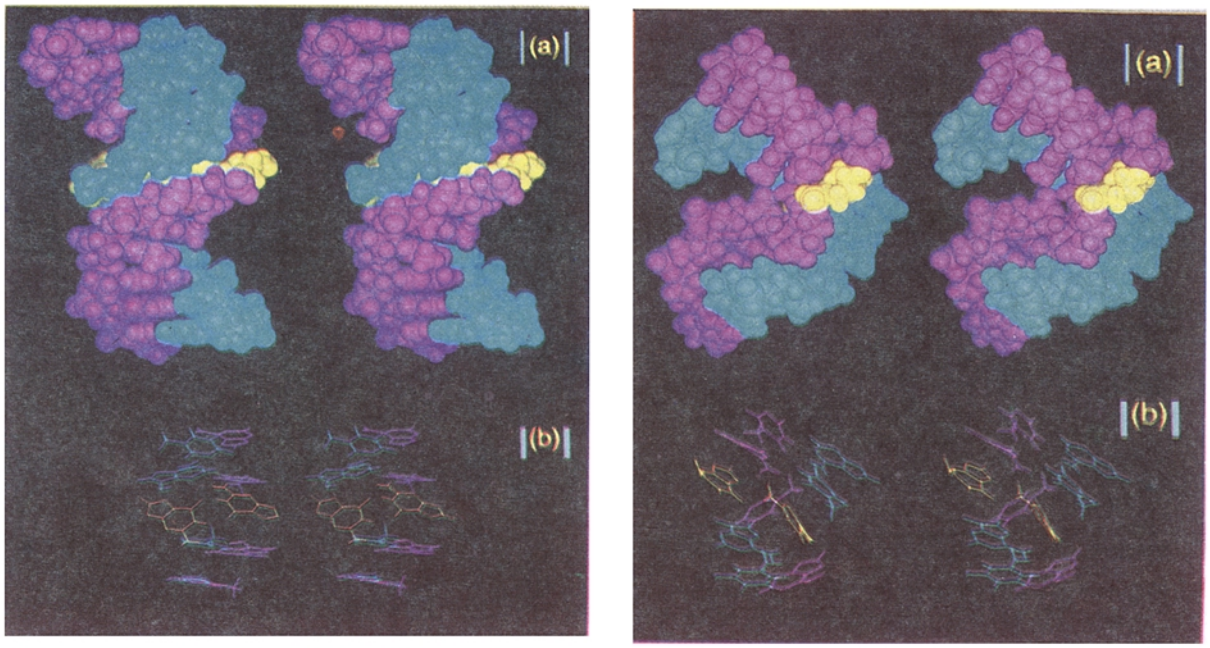

Fig.9 Structures of the sequence GTTAATGATTAAC which has a GG mismatch in the center as determined by NMR and MD simulations in our laboratory. The structure with a GG pairing (left) has slight distortions as compared to the normal B-DNA duplex, whereas the structure (right) without GG pairing shows a substantial bending and distortion in the center. Space filling models are shown on the top (a) while base pairing at the mismatch site is shown at the bottom (b).

distortions depending upon whether the two G's are paired or otherwise. A particular type of pairing causes minimal distortion in the duplex structure and such a structure may be tolerated by the LFB1 protein. Fig. 9 shows two different topologies that could be produced by a GG mismatch at the N site. It is also clear that the mismatches cause substantial dynamism in the molecules.

(c) Hairpins and Dumbbells

The genomic DNA which is a duplex has to open up for most of its functions. Then the nascent single stranded DNA can in principle undergo a number of structural transitions. Hairpins and Dumbbells belong to this category of transient structures. Fig. 10 shows the structure of a DNA dumbbell determined in our laboratory recently[25]. The stem of the dumbbell has two nicks in the center, yet it is noticed that the stem retains a good duplex structure with excellent stacking of the adjacent base pairs. The loop portion of the dumbbell exhibits a high degree of dynamism and dumbbell structures can be formed with different relative orientations of the loops.

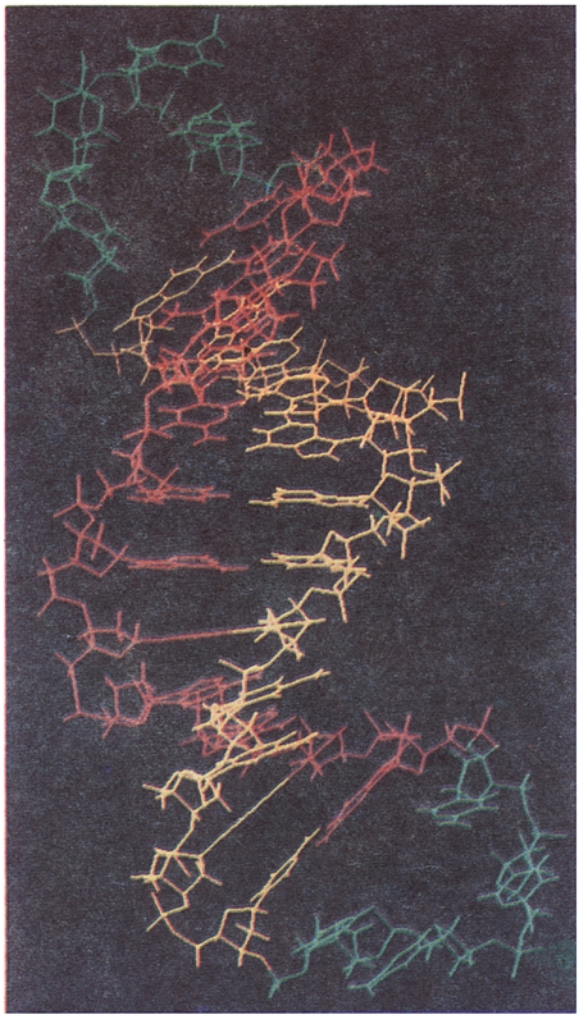

Fig.10 Solution structure of a DNA dumbell formed by association of two molecules of the sequence GATCTTCCCCCCGGAA, as determined by NMR in our laboratory[25]. The stem of the molecule has two nicks in the middle, but the stacking of the base pairs is maintained through the nicks.

(d) Triplexes

These are three stranded structures formed by two pyrimidine (T and C) strands and one purine (G and A) strand or by two purine strands and one pyrimidine strand. The base pairing is highly

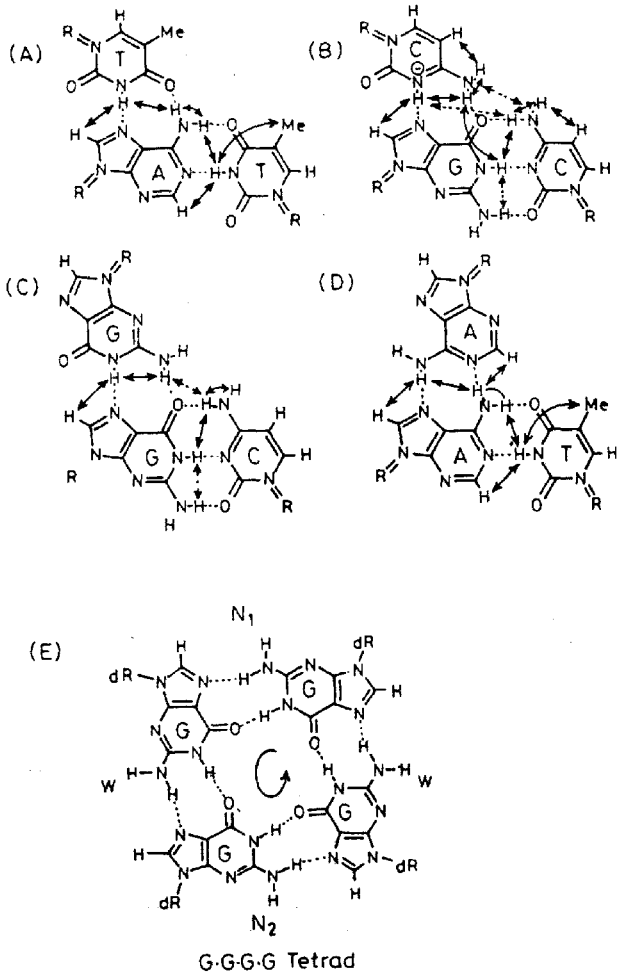

Fig.11 (A-D) Base pairing schemes in py.pu.py and py.pu.pu. types of triplexes where py represents a pyrimidine base and pu represents a purine base. The arrows indicate the short distances observable by NOE in the NMR spectra. (E) Base pairing scheme in the GGGG tetrad in G-quadruplexes.

specific and consist of T.(A:T), C^+.(G:C) triplets in the former type of triplexes and A.(A:T), G.(G:C) triplets in the latter type of triplexes (Fig.11 A-D)[26]. In either case the like strands run in opposite orientations. In certain situations some mismatch triplets such as G.(T:A) have also been observed. Fig. 12 shows the atomic resolution structure of one particular triplex molecule reported in the literature[27]. It is seen that the triplex structure is formed by placing one strand (the third strand) in the 'major' groove of the duplex formed by one purine strand and its complementary pyrimidine strand, and the three bases in the H-bonded triplet lie almost in the same plane. H-bonding is very specific and these aspects play important roles in specific recognition and have applications in gene targeting and gene therapy[28]. Triplex formation is also known to be playing a dominant role in recombination reactions. The structural details obtained by NMR provide a molecular rationale for these phenomena..

(e) G-Quadruplexes

It is known that the ends of chromosomes (so called telomeres) have G-rich sequences which are believed to be playing important roles in cell differentiation and growth. The base G has the unique property that it has two donor sites and two acceptor sites for H-bond formation. As a result it can form a very stable H-bonded network with four G bases all in the same plane (so called G-tetrad), and this is what leads to the formation of four stranded structures involving exclusively G-tetrads stacked over each other[29-32]. The H-bonding network in a G-tetrad is shown schematically in Fig. 11E. Recently it has been observed that GCGC tetrads can also be formed in some situations although this network will have fewer H-bonds and the structure will be relatively less stable[33]. Fig. 13 shows an illustrative G-quadruplex structure and the different ways such structures can be formed is schematically indicated in Fig. 14. The four strands can have different orientations and the bases in each strand can also have different orientations with respect to the sugar rings. Clearly there is a great variety in these structures and these must have important roles to play in recognition by enzymes such as telomerases, polymerases etc.

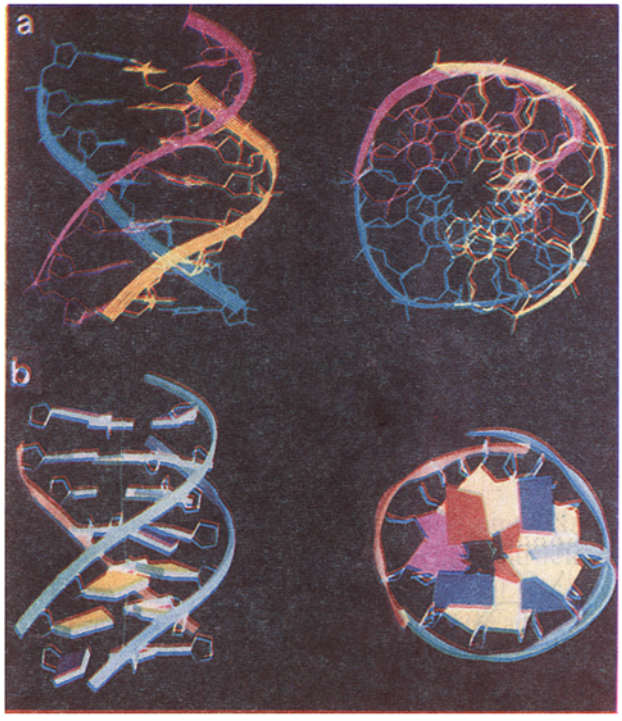

Fig. 12 Solution structure of a DNA triplex determined by NMR (reproduced with permission from ref. 26) The structure is shown both at high resolution and at gross level with ribbons and planks for bester clarity of base and backbone orientations. Two orthogonal views are shown.

(f) i-motif

This is also a four stranded structure but is

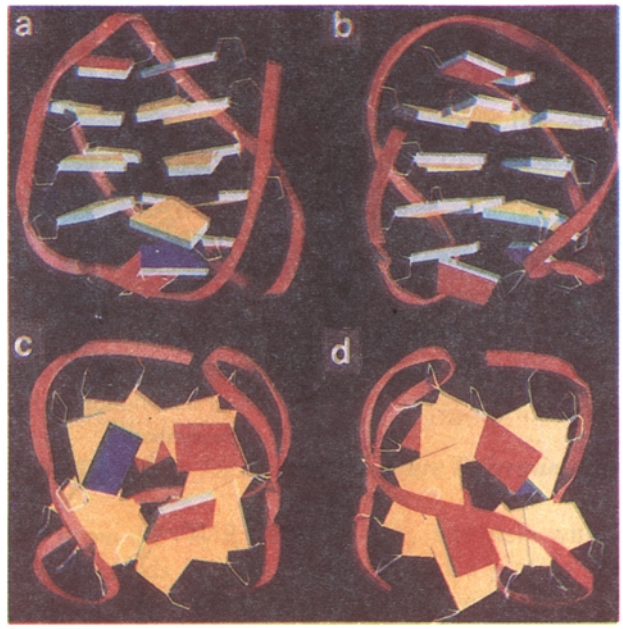

Fig.13 Solution structure of a G-quadruplex determined by NMR (reproduced with permission from ref. 29). For the sake of better clarity, the structure is drawn with planks and ribbons to represent bases and backbone respectively. Different views are shown.

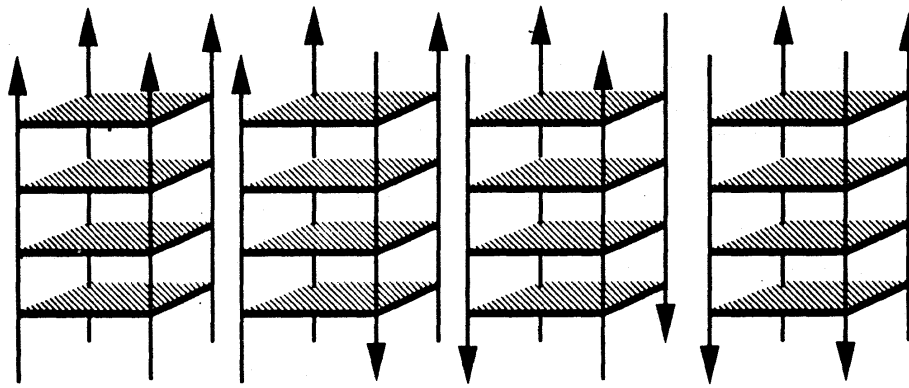

Fig.14 Schematic drawings of different possibilities of G-quadruplex formation. Each vertical line represents a G strand and the plane between the four strands represents a GGGG tetrad. The arrows indicate the orientations of the strands.

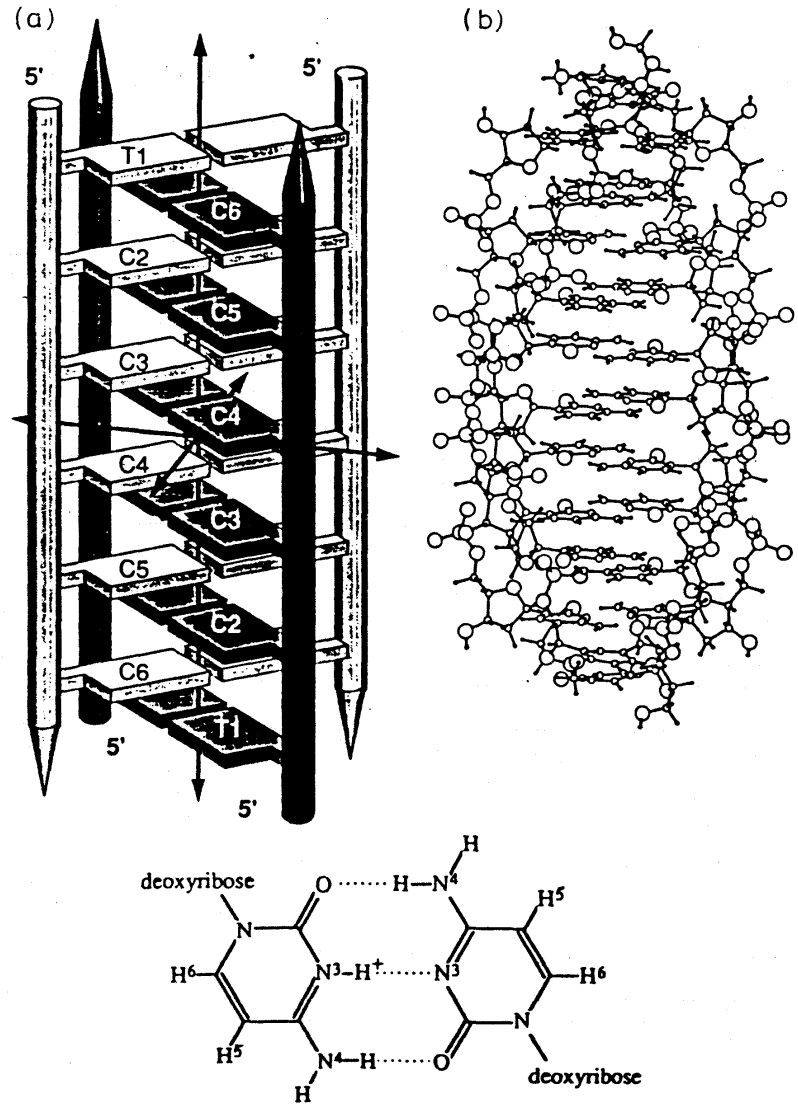

Fig.15 Schematic representation (a) and solution structure (b) of an i-motif formed by a C-stretch as determined by NMR. It consists of two parallel stranded duplexes interdigitating each other. The duplexes run in opposite directions. The C^+-C base pairing responsible for the duplex formation is also indicated (taken from ref. 34)

formed by continuous C-stretches under low pH conditions[34-36]. Two C-strands running in parallel orientation interact via $C\text{-}C^+$ pair formation and two such duplexes running in antiparallel direction interdigitate. The base pairing scheme and the NMR structure of the TC_5 sequence in which this structure was discovered are shown in Fig. 15. Both the duplexes are stretched and unwound significantly and the strands look fairly extended. Again, this kind of structures are expected to be formed at the chromosomal ends of the DNA.The low pH required for the formation of these structures may be produced by the charged groups in the proteins binding to the DNA.

In addition to the above distinct structural motifs which have been investigated extensively by several researchers, there have been occasional investigations on many other types of structures such as three way, four way and five way junctions, cruciform DNA, parallel stranded duplexes, RNA-DNA hybrids etc. (reviewed in ref 8). Thus not only that the DNA exhibits the various above type of structures, it is also observed that depending upon the sequence and experimental conditions, a particular DNA can undergo structural transitions. A classic case has been noted in our laboratory during the last few months. The molecule with the sequence 5'-GATCTTCCCCCCGGAA exists as a monomeric hairpin at low concentrations, forms a dumbbell with antiparallel duplex in the stem at higher concentrations, and goes over to a parallel stranded duplex and an i-motif structure as the pH is lowered. Thus it is clear that a given DNA segment can exhibit dynamism even at the topological level and these can be induced by the proteins in a living cell.

The number of investigations on RNA is relatively small because of practical difficulties, but currently, this is a hot area of research. Different RNA aptamers, hammerhead RNAs, tRNAs with folded structures are being investigated by multidimensional NMR experiments. Because of the greater complexities of the RNA spectra, isotopic labelling is becoming essential and heteronuclear NMR spectra are being used to resolve the spectra in different dimensions(see reviews in ref 8).

6 Protein Structure and Dynamism

Proteins constitute the most important component of the machinery of life. They are also the most complex molecules with greatest diversity in their sizes, properties, stabilities, structures etc. Consequently these are also the most widely studied molecules by a variety of physical and chemical techniques to unravel the secrets of their variety and specificities of functions. It is common knowledge now that proteins are made up of 20 different types of amino acids, and their sizes range from 50 to several thousands of amino acids in a single polypeptide chain. The sequence and composition of the amino acids largely determines the properties of a protein in terms of its 3D structure, behaviour under different experimental conditions, dynamism etc. Some proteins are made up of domains which function independently and this allows dissection of the domains for studying their properties.

The three dimensional structure of a protein molecule has well defined structural elements - so-called secondary structure elements - such as α–helices, 3_{10} helices, β-sheets (parallel and antiparallel), different types of turns, loops etc. Spatial arrangement of these units determines the 'tertiary structure' of the molecule. Each one of these elements is characterized by specific combination of torsion angles along the backbone, and in the case of helices and sheets there are also specific patterns of H-bonds between the amide protons and the carbonyl oxygens of the backbone chains. There are two torsion angles, namely, ϕ and ψ for every residue, which determine the fold of the polypeptide chain. The acceptable ranges of these angles from the point of view of avoiding short contacts and their ranges for specific secondary structures are described elegantly by the Ramachandran plots[37].

NMR Strategy for Protein Structure Determination

From the NMR point of view, a protein with 50 amino acids is a small protein and a protein with 300 amino acids is currently a large protein. The NMR strategies for protein structure determination have been evolving and the size limits may be pushed further in the years to come. While NOE remains the crucial and the final experimental input

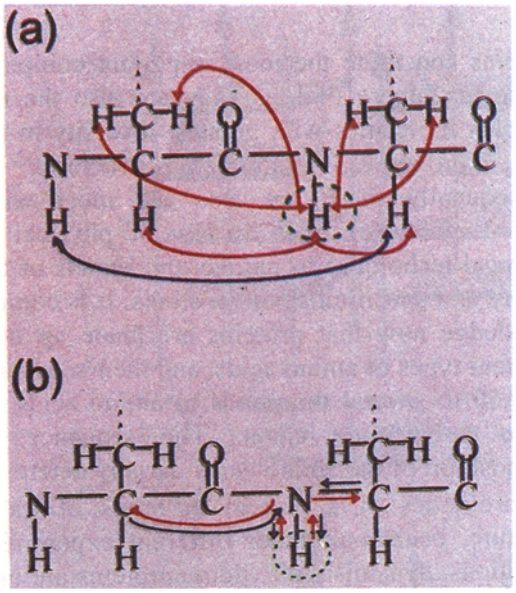

Fig.16 Strategies for obtaining sequence specific assignments in small (a) and large (b) proteins. The former rely on use of ^1H - ^1H J and NOE correlated techniques while the latter make use of triple resonance experiments involving ^1H, ^{13}C and ^{15}N correlations. The arrows in 'a' indicate the observable sequential NOE correlations and the arrows in 'b' indicate the magnetization transfer pathways used for obtaining the assignments. In 'b' the magnetization originates from the NH proton, travels to nitrogen and carbons as indicated and returns to NH for observation. The forward and backward transfers are indicated by different colours.

for structure calculations in all cases, the strategies for obtaining sequence specific assignments are different for small and large proteins[5,7,9,38-41]. In small proteins where the number of signals is small, homonuclear ^1H-^1H J and NOE correlated experiments proved sufficient. In larger proteins, however, extensive overlap of signals necessitated the use of heteronuclear and multidimensional NMR experiments. The two strategies are schematically indicated in Figs. 16a and 16b respectively. In Fig. 16a, the intraresidue connections are obtained from J-correlated experiments and the sequential walk through the polypeptide chain is performed via a set of NOEs. There are well defined NOE patterns for the regular secondary structures which help in obtaining the assignments. For example, in α helices, one observes a series of sequential NH-NH NOEs whereas in β sheets, one observes series of sequential NH_i - $C^{\alpha}H_{i-1}$ NOEs. Antiparallel β– sheets are also characterized by interstrand $C^{\alpha}H$ - $C^{\alpha}H$ NOEs. In Fig. 16b which represents the strategy for large proteins, assignments of the ^1H and heteronuclei along the backbone are obtained first using a variety of triple resonance three dimensional experiments. The arrows in the figure indicate the pathways of magnetization transfers employed for obtaining sequential connections. This is then followed by other three dimensional heteronuclear experiments for assigning the side chains of the individual amino acids. In all these experiments the various magnetization transfer steps are essentially one-bond transfers and thus the assignments are independent of the three dimensional structures of the protein molecules.

After getting sequence specific assignment of the resonances by one of the procedures described above, several short and long range NOEs between the nuclei are identified and quantified using multidimensional NOE experiments. The NOE intensities are converted into approximate upper bounds on internuclear distances to be used as constraints during structure calculations. Further, three bond coupling constants are measured from J-correlated spectra and are converted into approximate torsion angle constraints. All these constitute the basic input for the structure calculation algorithms, Distance Geometry, and Simulated Annealing /Restrained Molecular Dynamics mentioned earlier. In both, the violations of the constraints are defined in terms of some sort of a potential and this potential is minimized using one of the different minimization algorithms. Finally in the refinement step, the NOE intensities are calculated for the convergent structures by relaxation matrix methods and these are matched with the experimental NOE intensities for acceptability of the structures. In the end one generates a family of 15-20 similar structures, all of which are supposed to represent the structure of the molecule. A detailed analysis of these structures in terms of pairwise *rmsds* along the length of the polypeptide backbone indicates, first of all the uniqueness of the structure and secondly the precision of the structure derived at different portions of the molecule. It is generally observed that the backbone in the interior of the protein is well characterized but the loop portions and the terminal regions are usually poorly characterized. Fig. 17 shows a NMR protein structure determined in recent years as an illustration[42,43]. Protein structure determination by the above

strategies is labour intensive and analysis of spectra can take as much as 2-3 years for an average size protein of 150 amino-acid residues. Even so, at the present state of development of the techniques, nearly 100 protein structures in different molecular size ranges are solved every year world wide.

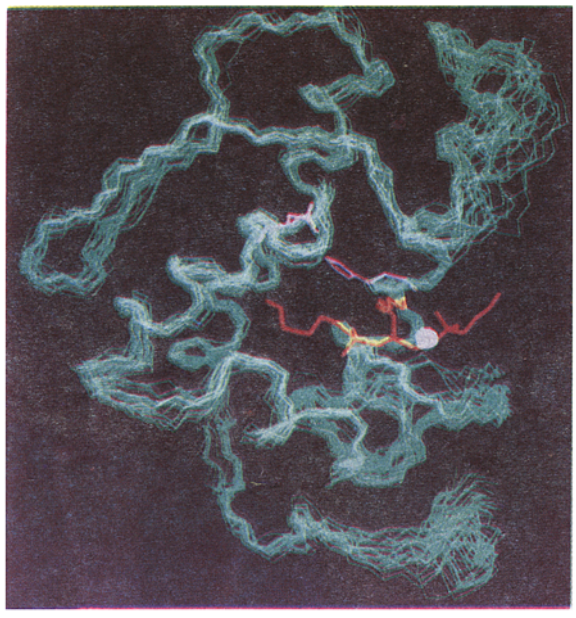

Fig.17 Solution structure of Phospholipase A_2 protein molecule determined by NMR as an illustration (reproduced with permission from ref. 42). Several equally probable structures are superimposed and the variations in these reflect the precision in the structure definition. The figure shows a backbone trace of the polypeptide chain.

The lack of precision of structure definition by NMR in any particular segment of a protein is a reflection of the local dynamism in the molecule. Segmental motions are known to occur in proteins and these result in average values for NOE intensities. A simplistic interpretation of the intensities in terms of the distances is then not appropriate and clearly no single distance is a satisfactory choice. Because of these, conservative distance estimates are given which leads to a greater variability in the structures. Additionally, dynamism in the molecule results also in poorer NOE intensities which in turn lead to fewer usable NOEs and greater errors in intensity measurements. These also contribute to the variability in the structures and hence to lower precision in the structure definition.

Protein structure and dynamism are influenced significantly by the environment. To investigate these effects it becomes necessary to compute the intrinsic conformational preferences and compare the structures with those derived from NMR under different experimental conditions. This will allow definition of the general principles and the various physical forces/interactions which govern the

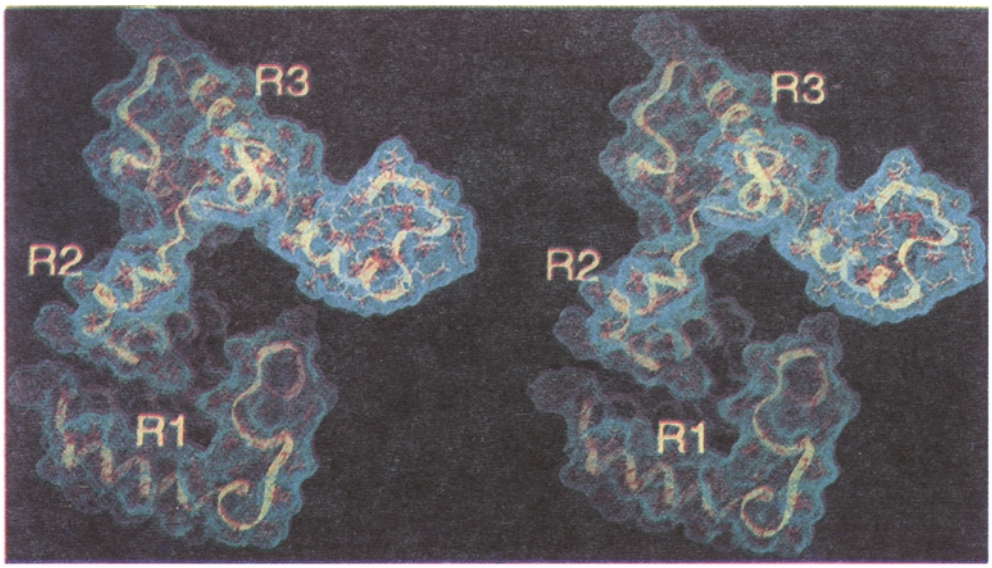

Fig.18 MD generated structure of the DNA binding domain of Drosophila Myb protein. The calculations were performed after immersing the protein in a box of water molecules and the conformational space was scanned by simulated annealing. The calculations used some constraints derived from secondary structure prediction algorithms and homology comparisons. R1, R2 and R3 are three domains in the molecule.

protein folding process. Molecular dynamics and simulated annealing are the best suited techniques for such investigations and many large molecules have thus been investigated. Calculations can be carried out in the presence and absence of water, to see the effects of solvation and they can be compared with the NMR derived results. Fig. 18 is a presentation of such an effort in which the theoretical 'solution' structure of a 160 residue DNA binding domain of Drosophila Myb protein investigated in our laboratory by simulated annealing and molecular dynamics calculations is shown. This structure differed quite significantly from the one derived when water molecules were not included in the calculations. Several tightly bound water molecules were detected and these clearly influence the solution structures. There is no NMR data on this protein yet, but some of the general features may be compared with data available on smaller segments of the protein from other sources[44]. The R1 domain showed formation of helices at the predicted positions in the absence of the constraints. The three segments R1, R2 and R3 do not have interactions between each other.

7 Protein - DNA Interactions

Recognition between nucleic acids and proteins lies at the heart of genetic expression and regulation inside the cell. Recognition of 'operator' and 'promoter' sequences by repressors, polymerases,

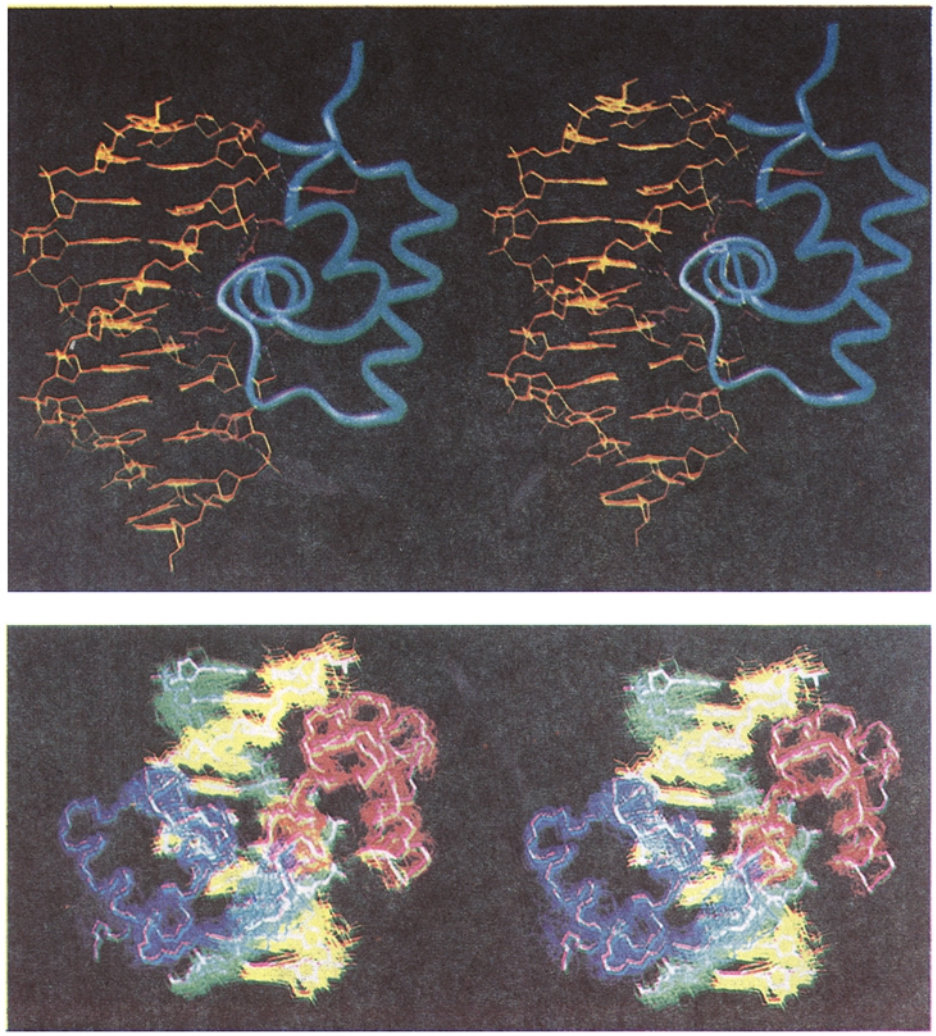

Fig.19 Solution structures of lac-DNA[46] (top) and MybR2R3 - DNA[47] (bottom) complexes determined to atomic resolution by a combination of various triple resonance techniques and using isotopically labelled protein and unlabelled DNA molecules (reproduced with permission). In the latter, a family of convergent structures is shown.

transcription factors, cleavage sites by restriction enzymes are a few examples. Understanding these molecular interactions is a top priority in present day research all over the world, and NMR has been successful in unraveling the mechanisms of these interactions in several cases. Fig. 19 shows two recent examples of DNA-protein complex structures determined by heteronuclear NMR[45-47]. By using isotopically labelled proteins and unlabelled DNA molecules and by using special filtering techniques, it is possible to observe the NMR signals originating exclusively from either of the two molecules in a complex. NOE correlations between the protons on the protein and the protons on the DNA enable identification of contact points. Such a knowledge is a forerunner for the design of new and more efficient proteins.

Most of the DNA-protein complexes investigated so far concerned DNA duplexes. The structural investigations have indicated that the DNA binding proteins interact with DNA using one of a few well defined motifs, namely, 'helix-turn-helix', 'Zinc finger', 'leucine zipper' and 'helix-loop-helix'. The details of these have been reviewed extensively[48-50]. The examples shown in Fig. 19 make use of the helix-turn-helix motif. One of the helix sits in the major groove of the duplex DNA and the other helix orients itself in an orthogonal manner. Direct interactions occur between the sidechains of the recognition helix and the bases or the phosphates. The interactions with the bases are either H-bonding interactions or stacking of the aromatic residues over the bases. The phosphates are involved mostly in charge-charge electrostatic interactions but may also participate in H-bonding interactions. In addition to these, there will also be non specific interactions between the amino acid side chains and DNA backbone which contribute to the stability of the complex. It is observed in a few cases, lac repressor for example, that on complex formation the DNA and protein undergo partial conformational changes. In others, however, Myb DNA binding domain for example, no structural changes have been observed.

8 Future Directions

No one would have predicted twenty years ago that NMR would reach the heights it has reached today, encompassing biology in the real sense of the term. Structural biology is as much an NMR branch of science as it is of X-ray crystallography. Structures of complex proteins, nucleic acids, protein-nucleic acid complexes, drug-nucleic acid complexes, protein-protein complexes etc. are being determined with breathtaking speed by both the methods. There have been instances where NMR structure has provided the starting model for X-ray structure determination and also *vice versa*. Clearly NMR has emerged as an alternative but complementary technique for macromolecular structure determination. The macromolecular concentrations used in NMR are comparable to the total protein concentration inside a living cell and the other conditions are adjustable to near physiological conditions.

Learning from past experience it would be clearly unwise and impossible to make a prediction far too long in the future as to the direction NMR research would take. The NMR methodology and technology are developing at a rapid pace with the objective of handling larger and larger molecules on the one hand - this includes complexes - and smaller concentrations of the molecules on the other, since larger molecules may not be soluble to the extent desired today and also because of the high cost of production of the labelled molecules. Assemblies of macromolecules could also be studied which provide insight into the cooperative structural changes, so crucial for a complex biological function.

During the last two decades, a large emphasis was laid on the determination of the average structures of the macromolecules in aqueous solutions. It is, however, known that the molecules are highly dynamic, exhibiting motions of different kinds. Different domains in a protein may have different motional characteristics and these would be important for adaptability of a protein molecule for a particular function. There are cases where the same molecule is responsible for a diverse set of functions. This can occur either by involvement of different domains in the molecule or in some other situations the same domain may undergo conformational changes. Thus there is increasingly greater interest today in defining the so-called dynamic structure of a molecule. In otherwords, efforts are being made to study the rates of

segmental motions, overall tumbling motions, anisotropy of motions, conformational exchanges, surface motions vs. motions in the interior of a protein, waters of hydration, ordering and trapping of water molecules in the interior of proteins, folding - unfolding equilibria in proteins, folding pathways and identification of intermediates, to name a few. Study of substrate - enzyme complexes would throw light on the mechanisms of enzyme actions.

As regards the biological systems, the primary consideration is their biological significance. While proteins continue to attract a lot of attention, RNAs have also emerged as hot favourites for many research groups because of the recent discoveries of ribozymes which are RNA catalysts, greater variety in RNAs, and greater involvement of RNAs in a number of biological functions. RNA and DNA aptamers and their complexes with the specific substrates are attracting substantial attention since these are expected to throw light on the intricate details of molecular recognitions.

The practice of NMR research, especially with reference to biomolecular NMR, has also changed dramatically. It has become highly multidisciplinary encompassing, biologists, spectroscopists, chemists, computer engineers and physicists alike. Scientific collaboration is an absolute necessity today. In many labs, molecular biology research is an integral part of NMR activity. The production of desired materials by genetic engineering methods is expensive and the machine requirements have also increased quite substantially. It has become necessary to have quick and high access to sophisticated spectrometers for competitive research. In contrast to the situation in the eighties where most machines were run as service facilities, it is common now to see dedicated high field spectrometers for a small group of scientists. The spectrometer designs are improving rapidly, magnetic field strengths are spiraling upwards, and computer powers are pushing the sophistications and applications. Thus, to conclude, it remains anybody's guess as to what further heights NMR research would reach in the years to come.

9 Acknowledgements

The authors are thankful to the National Facility for High Field NMR at TIFR for most of their research included in this article. The authors also thank the Journals and the authors who have given them permissions for reproducing some of their Figures.

References

1 F Block, W W Hansen and M E Packard *Phys Rev* **69** (1946) 127
2 E M Purcell, H C Torrey and R V Pound *Phys Rev* **69** (1946) 37
3 J Jenner *Ampere International Summer School* Bosco Polje, Yugoslavia (1971)
4 L Muller, A Kumar and R R Ernst *J Chem Phy* **63** (1975) 5490
5 J Cavanagh, W J Fairbrother, A G Palmer and N J Skelton *Protein NMR Spectroscopy: Principles and Practice* Academic Press California (1996)
6 R R Ernst, G Bodenshausen and A Wokaun *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* Clarendon Press Oxford (1987)
7 Methods in Enzymology **239** (1995) *Nuclear Magnetic Resonance and Nucleic Acids* (Ed T L James) Academic Press Inc San Diego
8 Methods in Enzymology **261** (1995) *Nuclear Magnetic Resonance and Nucleic Acids* (Ed T L James) Academic Press Inc San Diego
9 K Wuthrich *NMR of Proteins and Nucleic Acids* John Wiley & Sons NY (1986)
10 E R P Zuiderweg, S R van Doren, A V Kurochkin, R R Neubig and A Majumdar *Perspectives in Drug Discovery and Design* **1** (1993) 391
11 L P McIntosh and F W Dahlquist *Quat Rev Biophys* **23** (1990) 1
12 D M LeMaster *Prog NMR Spec* **26** (1994) 371
13 T F Havel and K Wuthrich *Bull Math Biol* **46** (1984) 673
14 W Braun and N Go *J Mol Biol* **186** (1985) 611
15 R Ajaykumar, R V Hosur and G Govil *J Biomol NMR* **1** (1991) 363
16 M Levitt (1983) *J Mol Biol* **170** (1983) 723
17 R M Scheek, N Russo, R Boelens, R Kaptein and J H Boom *J Am Chem Soc* **105** (1983) 2914
18 M Ravikumar, R V Hosur, K B Roy, H T Miles and G Govil *Biochem* **24** (1985) 7703
19 R V Hosur, G Govil and H T Miles *Magn Reson Chem* **26** (1988) 927
20 J W Keepers and T L James *J Magn Reson* **57** (1984) 404
21 A Majumdar and R V Hosur *Prog NMR Spectr* **24** (1992) 109
22 P K Radha, A Madan, R Nibedita and R V Hosur *Biochemistry* **34** (1995) 5913
23 P K Radha, R Nibedita, R Ajay Kumar and R V Hosur *Methods in Enzymology* **261** (1995) 73
24 P K Radha, A Madan and R V Hosur *J Chem Soc Chem Commun* **19** (1993) 1491

25 S Singh, P K Patel and R V Hosur *Biochemistry* (1997) (in press)
26 I Radhakrishnan and D J Patel *Structure* **2** (1994) 17
27 I Radhakrishnan and D J Patel *Structure* **1** (1993) 135
28 R D Wells, D A Collier, J C Hanvey, M Shimizu and F Wohlraub *FSSEB Jour* **2** (1988) 2939
29 Y Wang and D J Patel *Structure* **1** (1993) 263
30 P Shultze, F W Smith and J Feigon *Structure* **2** (1994) 221
31 Y Wang and D J Patel *J Mol Biol* **251** (1995) 76
32 A Keniry, G D Strahan, E A Owen and R H Shafer *Eur J Biochem* **233** (1995) 631
33 A Kettani, R Ajaykumar and D J Patel *J Mol Biol* **254** (1995) 638
34 K Gehring, J L Leroy and M Gueron *Nature* **363** (1993) 561
35 J L Leroy, K Gehring, A Kettani and M Gueron *Biochem* **32** (1993) 6019
36 S Ahmed, A Kintanar and E Henderson *Nature Struct Bio* **1** (1994) 83
37 G N Ramachandran *Conformation of Biopolymers* Academic Press New York (1967)
38 K Wuthrich, C Spitzfaden, K Memmert, H Widmer and G Wider *FEBS Lett* **285** (1994) 237
39 H Viz, M Marini, E Constantin Vorgias, K S Wilson, R Kaptein and R Boelens *Quat Magn Reson in Biol Med* **2** (1995) 107
40 A Bax and S Grzesiek *Acc Chem Res* **26** (1993) 131
41 M Ikura, L E Kay and A Bax *Biochemistry* **29** (1990) 4659
42 B Van den Berg, M Tissari, R Boelens, R Dijmman, R Kaptein, G H de Haas and H M Verheij *J Biomol NMR* **5** (1995) 110
43 B Van den Berg, M Tissari, R Boelens, R Dijkman, R Kaptein, G H de Haas and H M Verheij *Nature Struct Biol* **2** (1995) 402
44 N Jamin, O S Gabrielsen, N Gilles, P N Lirsac and F Toma *Eur J Biochem* **216** (1993) 147
45 K Ogata, S Morikawa, H Nakamura, A Sekikawa, T Inoue, H Kanai, A Sarai, S Ishii and Y Nishimura *Cell* **79** (1994) 1
46 K Ogata, S Morikawa, H Nakamura, H Hojo, S Yoshimura, R Zhang, S Aimoto, Y Ametani, Z Hirata, A Sarai, S Ishii and Y Sishimura *Nature Struct Biol* **2** (1995) 309
47 V P Chuprina, J A C Rullman, R M J N Lamerichs, J H van Boom, R Boelens and R Kaptein *J Mol Biol* **234** (1993) 446
48 C O Pabo and R T Sauer *Ann Rev Biochem* **53.**(1984) 293
49 C Branden and J Tooze *Introduction to Protein Structure* Garland NY (1991)
50 T A Steitz *Quart Rev Biophysics* **23** (1990) 93
51 R R Ernst and W A Anderson *Rev Sci Instru* **37** (1966) 93
52 J Jeener, B H Meier, P Bachmann and R R Ernst *J Chem Phys* **71** (1979) 4546
53 L E Kay, M Ikura, R Tschudin and A Bax *J Magn Reson* **89** (1990) 496
54 L Muller *J Am Chem Soc* **101** (1979) 4481
55 L Braunschweiler and R R Ernst *J Magn Reson* **53** (1983) 521
56 P K Radha, A Madan, L C Padhy and R V Hosur *Proc Indian Acad Sci (Chem Sci)* **106** (1994) 1537
57 S Grzesiek and A Bax *J Magn Reson* **99** (1992) 201