Structures of Biomolecules by NMR Spectroscopy

Hanudatta S Atreya

NMR spectroscopy has the unique ability to probe both structure and dynamics of biomolecules with high resolution. This has rendered it a powerful tool for structural biology. In this article, an overview of the structure determination process is described. The sample and experiments required are briefly discussed.

Hanudatta S Atreya is at Indian Institute of Science currently as an Associate Professor. His research interests include development and application of new magnetic resonance methodologies for studying structure, function and dynamics of biomolecules.

1. Introduction

Each cell in a living organism consists of a number of macromolecules (called 'biomolecules'), which are required for various cellular functions and metabolism. The biological activity of these biomolecules is governed by their interactions with other molecules inside the cell. Such interactions, in turn, are governed by their three-dimensional (3D) structures in space [1]. Thus, knowledge of the 3D structure of a biomolecule is essential for a complete understanding of its function. A discipline of science concerned with the study of relationship between the 3D structure, function and interaction of biomolecules is called 'structural biology'. Structural biology is an intensely pursued area of research today and has been instrumental in furthering our understanding of the origins and finding the cures for a number of human diseases [2].

There are many well-established experimental techniques to determine the 3D structure of biomolecules at high-resolution. Of these, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the most popular ones. The former requires crystals of high purity, while the latter is used to determine biomolecular structures in aqueous solution or in the solid state. NMR provides information about 3D structures complementary to that obtained from X-ray crystallography. NMR has an edge over the X-ray method as it can be used to study biomolecules in their native state (i.e., in solution) under physi-

Keywords
Protein structures, protein dynamics, biomolecular function, isotope labeling.

A parallel development has been the progress in biosynthetic incorporation of stable NMR active isotopes, such as ^{13}C, ^{15}N and ^{2}H.

ological conditions or as in the case of proteins, even in a molten globule or a denatured state. Moreover, the freedom of motion in solution makes NMR more suitable for studying the dynamic behavior of macromolecules.

The first high resolution protein structure by NMR spectroscopy was carried out in mid-1980s [3]. Before the beginning of this millennium, NMR spectroscopy was limited to solving 3D structures of proteins with molecular masses less than ~30 kDa [4]. This was primarily due to the increased complexity and reduced sensitivity of an NMR spectrum with increase in the size of the biomolecular system. However, developments of new experimental techniques and methodologies have brought NMR spectroscopy to the forefront of structural biology [5]. Innovations in hardware and spectrometer design, data collection and analysis and computational tools have together have made NMR an extremely powerful technique for studying biological macromolecules. A parallel development has been the progress in biosynthetic incorporation of stable NMR active isotopes, such as ^{13}C, ^{15}N and ^{2}H [6]. Such advances have provided a means for characterization of structure and dynamics of large proteins at atomic resolution. Compared with the NMR structures determined ten years ago, the present day structures are determined much faster, have higher precision and are of much larger molecules.

One of the significant contributions towards speeding up of the process of biomolecular structure determination has been the access to fast computing facilities and the development of new computational tools [7]. This has emerged as a subject of pivotal interest in the era of 'structural genomics', where the goal is to unravel tens of thousands of protein structures within this decade [8]. The availability of a large number of 3D structures, in turn, has contributed to our understanding of NMR spectral data vis-à-vis protein structures, thereby aiding in the refinement of these tools. Such a wealth of data is made available to the NMR community through two major public domain databases, namely, the BioMagResBank (BMRB) (www.bmrb.wisc.edu) and the protein data bank (PDB) (www.rcsb.org), which have become an

integral part of NMR-based structural biology research. A number of computational tools are being developed which utilize such information from BMRB and PDB at various stages of the structure determination process.

2. An Overview of the Structure Determination Process

A flowchart of the different steps en-route to biomolecular structure determination by NMR spectroscopy is depicted in *Figure* 1 and discussed below.

i. Sample Preparation

An efficient structure determination requires a highly purified sample preparation. The first step in every NMR study therefore

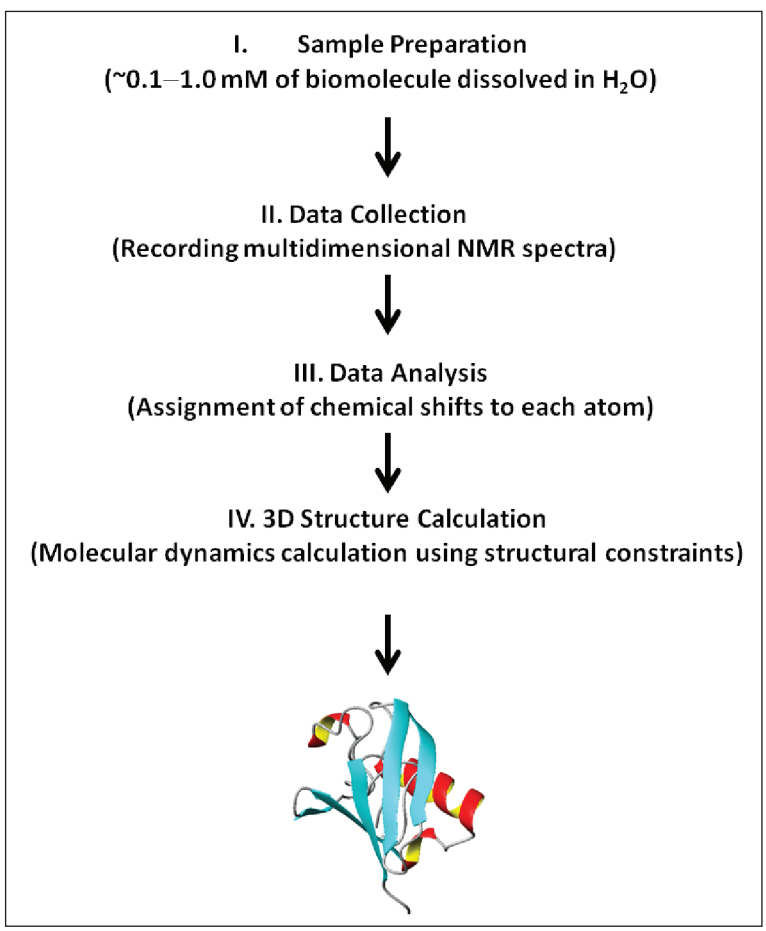

Figure 1. A schematic depiction of the work flow in NMR-based biomolecular structure determination. Regions of protein structure colored in cyan and red represent beta strands and alpha helices, respectively.

Proteins can be enriched with ^{13}C or/and ^{15}N isotopes. This methodology is also referred to as *isotope labeling*.

involves optimization of the sample conditions. The pH, ionic strength, and temperature can often be adjusted to mimic physiological conditions. The macromolecule under study should be stable in the chosen conditions until the data is recorded. The sample is prepared by dissolving the required amount of the compound under investigation in a small quantity (~600 microlitres) of buffer in water (H_2O). The concentration required to obtain a good spectrum depends on factors such as stability of the protein, the sensitivity of the NMR spectrometer and the type of NMR experiments to be recorded on the sample. Typically, a protein concentration of ~1 mM suffices in most cases; the lower limit on concentration is largely determined by the sensitivity of the NMR spectrometer being used. With highly sensitive modern-day spectrometers, equipped with what is known as cryogenic probes, one can obtain a good NMR spectrum with a concentration as low as 0.1 mM of the protein samples (i.e., 1 mg/mL for a ~100 amino acid protein).

In the case of large molecular weight proteins, there is an overlap of peaks in the NMR spectrum. In such cases, one can utilize the good resolution and sensitivity offered by ^{13}C and ^{15}N isotopes in combination with multidimensional NMR experiments. Proteins can be enriched with ^{13}C or/and ^{15}N isotopes. This methodology is also referred to as *isotope labeling* [6] (see *Box* 1).

ii. NMR Data Collection

In this step, several multidimensional spectra are recorded on the sample [3, 9]. These experiments fall in two classes. The first contains a family of experiments aimed at identification and assignment of signals or chemical shifts to the respective NMR active nuclei in the protein. The second category of experiments is used to obtain structural information. A number of experiments in this class have been proposed to date [9]. For larger proteins (mol. wt. > 25 kDa), 3D experiments do not suffice and hence four dimensional (4D) triple resonance experiments are recorded. However, the measurement time increases steeply with the number of dimensions of a NMR spectrum: typical two-dimensional

Box 1. Isotope Labeling

In nature, the natural abundance of hydrogen (^{1}H) in every molecule is nearly 100%. However, if other NMR-active nuclei such as ^{13}C and ^{15}N are considered, their natural abundance is very low (1% and 0.3%, respectively. This implies that in every hundred carbon atoms, one ^{13}C atom is encountered with the remaining being NMR-inactive ^{12}C). Thus, ^{13}C population is very low at natural abundance. To increase this population (and consequently the sensitivity), the molecules have to be 'enriched' with ^{13}C. Similarly, for carrying out ^{15}N NMR studies, the molecules have to be enriched in ^{15}N. This process is termed as 'isotope labeling'.

Proteins can be enriched with ^{13}C or/and ^{15}N isotopes, by over-expressing them in a suitable host, such as *Escherichia coli*. The host bacteria, *E. coli*, is grown in a medium containing $^{15}NH_4Cl$ and ^{13}C-glucose as the sole source of nitrogen and carbon, respectively. This is depicted in *Figure* A.

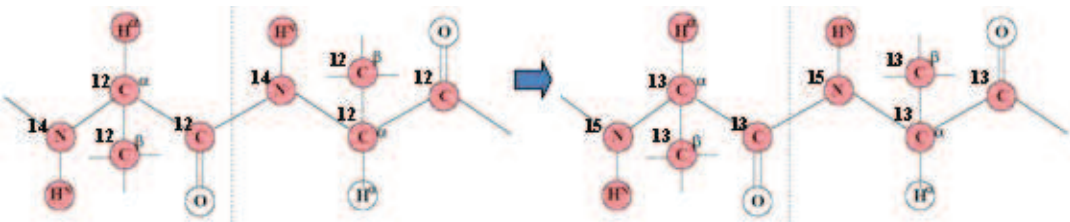

Figure A. α and β are symbols used for naming the backbone and side-chain carbon atoms in amino acids, respectively.

(2D), three-dimensional (3D) and four-dimensional (4D) NMR spectra need minutes, hours or days, respectively, for completion, whereas five- and six-dimensional (5D, 6D) experiments would take too long to be feasible. Hence, the development of new NMR techniques to speed up data acquisition is an actively pursued area of research today (reviewed in [10]).

iii. NMR Data Analysis

Structural information is obtained from features in NMR signals, such as intensity or the chemical shift. This information is useful only after the identities of the atoms that have given rise to the signals are known. Such an assignment of NMR signals to individual residues/groups in the macromolecule is termed as sequence specific resonance assignment [3] and forms a very important step towards structural characterization of a biomolecule.

Structural information is obtained from features in NMR signals, such as intensity or the chemical shift.

Until recently, at least in the case of proteins, the resonance assignment step was considered the most tedious and time-consuming. With the advent of multidimensional NMR experiments for resonance assignments, it has become increasingly clear that the information content of protein spectra can allow complete automation of resonance assignment.

iv. 3D Structure Determination and Refinement

All hydrogen atoms of a biomolecule (^{1}H) that come close in space (< 5Å) experience dipolar interactions with each other. Once chemical shifts of hydrogen atoms are assigned, such inter-proton interactions are mapped using specific NMR techniques and used in the final step to obtain a high resolution 3D structural model of the molecule. This step involves the collection of such 'conformational' constraints (i.e., those constraints that fix the distance between two atoms of the molecule). This is achieved using different computational tools.

A biomolecular structure is represented as a set of 3D co-ordinates (x, y and z) for each atom. As of today, >100,000 protein structures have been determined experimentally, which can be accessed in a database called 'protein data bank' or PDB (http://rcsb.org). Out of these, about 10% have been determined by NMR spectroscopy and the rest have been determined by other experimental techniques.

Once an initial structure of the biomolecule has been obtained, its quality is further refined using information from a wide spectrum of different NMR experiments. Such techniques provide information that is complimentary to that used for initial structure calculation and hence are suitable for refinement as well as validating the quality of the structure.

Once an initial structure of the biomolecule has been obtained, its quality is further refined using information from a wide spectrum of different NMR experiments.

3. Future Directions

The field of biomolecular NMR is progressing rapidly with new advances being made every year. Biomolecules, because of their complexity in structure, have posed challenges both from compu-

tational as well as spectroscopic point of view. Hence, new experimental and computational NMR methods will continue to be developed to aid in their 3D structural characterization.

Today, NMR spectroscopy is being used in new areas of research such as 'metabolomics' and 'systems biology'. With the increase in the availability of high field NMR spectrometers equipped with cryogenic probes, an increased emphasis will be on reducing the time required for structure determination. Thus, in future, fast NMR spectroscopy will witness new developments.

Suggested Reading

[1] B Alberts, D Bray, J Lewis, M Raff, K Roberts, J D Watson, Molecular Biology of the Cell, Garland Publishers, New York, NY, 1994.

[2] R J Epstein, Human Molecular Biology: An Introduction to the Molecular Basis of Health and Disease, Cambridge University Press, Cambridge, 2002.

[3] K Wüthrich, NMR of Proteins and Nucleic Acids, Wiley, New York, NY, 1986.

[4] K Wüthrich, The second decade - into the third millennium, *Nat Struct Biol*., Vol.5, pp.492–95, 1998.

[5] A Constans, NMR hits the big time, *The Scientist*., Vol.17, pp.45–51, 2003.

[6] H S Atreya, (Editor), Isotope labeling methods in Biomolecular NMR, Springer, Netherlands, 2012.

[7] P Güntert, Structure calculation of biological macromolecules from NMR data, *Quarterly Reviews of Biophysics*., Vol.31, pp.145–237, 1998.

[8] S K Burley, An overview of structural genomics, *Nat Struct Biol.,* Vol.7, pp.932–34, 2000.

[9] J Cavanagh, W J Fairbrother, A G Palmer, N J Skelton, Protein NMR Spectroscopy, Academic Press, San Diego, CA, 1996.

[10] H S Atreya, T Szyperski, Rapid NMR data collection, *Methods Enzymol.,* Vol.394, pp.78–108, 2005.

Address for Correspondence
Hanudatta S Atreya
NMR Research Centre
Indian Institute of Science
Bengaluru 560 012.
Email:hsatreya@gmail.com