Nuclear Magnetic Resonance (NMR) Spectroscopy: Structural Analysis of Proteins and Nucleic Acids

Milton H Werner, *The Rockefeller University, New York, USA*

Introductory article

Article Contents

doi: 10.1038/npg.els.0003100

Nuclear magnetic resonance spectroscopy (NMR) enables the determination of the three-dimensional structure of proteins and nucleic acids in solution. A set of NMR experiments identifies the residues in the molecule and determines the order of residues that comprises the primary structure. Distance and angular relationships are subsequently measured and utilized as input to the structure calculation.

Introduction

The basis of the nuclear magnetic resonance spectroscopy (NMR) experiment lies in the behaviour of an atom's nucleus when placed in a magnetic field. The nucleons within the nucleus of an atom rotate or 'spin' in a magnetic field at a rate that is characteristic of its nuclear structure and surrounding environment. 'Spin' is a source of angular momentum inherent to nuclei and other subatomic particles. Spin angular momentum provides a nucleus with a magnetic moment and, thereby, with a discrete pattern of energy levels when a nucleus is placed in a magnetic field. The goal of NMR spectroscopy is to probe these energy levels by transient excitation and measurement of the time course of relaxation back to equilibrium. The modern NMR experiment can distinguish chemically bonded nuclei from those that are not. Distinct experiments can probe the spatial arrangement of nuclei in a molecule whether or not the atoms are chemically bonded to one another. Thus, both the chemical structure and the spatial arrangement of the atoms in a molecule can be studied to construct a three-dimensional model of a molecule of interest.

To understand the information content of an NMR spectrum, it is first necessary to understand a set of commonly used terms that define the elements of the NMR spectrum and the principal means by which nuclear spin can be manipulated to determine molecular structure.

Chemical shift

Chemical shift defines the location of an NMR signal in units scaled by the frequency of the magnetic field in which the signal was measured. Thus, the NMR frequency of an isolated hydrogen atom at a magnetic field strength of 11.7 T (tesla) corresponds to 500.13 MHz (megahertz). At this field strength, the hydrogen atoms in water would be measured at 500.1323506 MHz. An easier way to express this quantity is to scale the frequency of an observed NMR signal by expressing it in parts per million (ppm), relative to the magnetic field strength in which it is measured. For example, at a field strength of 11.7 T, 0 ppm is defined as 500.13 MHz. This means that the chemical shift of the hydrogen atoms of water will be:

Resonant frequency

The resonant frequency is the rate of precession of nuclear spin as measured for a nucleus or collection of nuclei when placed in a magnetic field. The absolute value of the precession frequency is unique to each nucleus and is influenced by the environment in which the nucleus resides. Both the chemical bonding and spatial environment of a nucleus can alter the absolute value of the nuclear spin frequency.

Correlation

Correlation refers to a physical relationship that can be established between two nuclei in an NMR experiment. For example, if one measures the resonant frequencies of two atoms that are chemically bonded to one another, one can define a correlation between these atoms in terms of the resonant frequencies of the individual nuclei.

J coupling

The spin–spin coupling or '*J* coupling' is a measure of the magnetic interactions of nuclei that are chemically bonded in a molecule. The measurement of *J* coupling permits elucidation of chemical structure and the strength of the coupling is independent of the strength of the magnetic field in which it is measured. This phenomenon will become the basis for determining the primary sequence of a protein or nucleic acid in solution.

Nuclear Overhauser enhancement (NOE)

The NOE effect results from dipolar cross-relaxation between two nuclei near each other in space. Since the phenomenon is dependent on the distance between nuclei, the NOE reports on the spatial separation of the nuclei whether or not the atoms, within which the nuclei reside, are chemically bonded to one another. For proteins and nucleic acids, the NOE can be measured for atoms that reside ≤ 5 Å apart from one another. The NOE effect is extensively used to define the spatial arrangements of atoms in the three-dimensional structure of a protein or nucleic acid.

Multidimensional NMR Spectroscopy

In the early 1980 s, a second technique was developed that enabled the determination of the molecular structure of proteins and nucleic acids. In contrast to X-ray crystallography, the prevailing methodology at the time, multidimensional nuclear magnetic resonance (NMR) spectroscopy permitted the study of biomolecules under conditions that were closely related to the physiological conditions of the cell. The structures of many proteins and nucleic acids became accessible simply owing to their solubility in water, whether or not they formed an ordered crystalline array.

The chemical or spatial arrangement of atoms in a molecule can be used to measure a correlation between the resonant frequencies of different atoms, a phenomenon first recognized by Jeener and Ernst. Their two-dimensional (2D) NMR experiment mapped the correlated frequencies between two atoms in a molecule, placing the resonant frequency of each atom along the diagonal of a two-dimensional plot (**Figure 1a**, black circles). At the intersection of the resonant frequency of each atom, the correlation between them is observed as a crosspeak (**Figure 1a**, grey). The correlation represents either an NOE (i.e. through-space) or *J* coupling (i.e. through-bond) relationship between the atoms. Thus, the crosspeaks contain the structural information necessary to determine the chemical bonding or spatial arrangement between two atoms. Wüthrich and co-workers quickly recognized that this two-dimensional experiment could be used to study the molecular structure of a large number of nuclei in a single experiment, enabling the study of proteins and nucleic acids structure in solution.

Despite the power of the 2D NMR experiment, it became clear that it was limited by the number of signals that could be completely resolved in the two-dimensional spectrum. For biomolecules, this meant that the range of hydrogen atom (referred to as proton) frequencies was too narrow to permit simultaneous analysis of more than about 100 amino acids. The solution to this problem was to increase the spectral resolution by simultaneously observing a proton with an attached second nucleus whose resonant frequencies were dispersed over a wider range (**Figure 1b**). The stable isotopes of carbon and nitrogen (^{13}C and ^{15}N) provided a 5-fold to 10-fold increase in the frequency range of the NMR experiment. The development of methodologies to incorporate these stable isotopes biosynthetically into biomolecules gave rise to modern multinuclear, multidimensional NMR spectroscopy. Today, proteins, nucleic acids and their complexes can be routinely studied up to a molecular mass of ~ 40 kDa.

Isotope Labelling of Proteins and Nucleic Acids

The incorporation of stable isotopes of carbon and nitrogen into biomolecules is readily accomplished by growth of a suitable microorganism in a medium containing biosynthetic precursors for both amino acids and nucleotides. The organism of choice is *Escherichia coli*, the most common bacterium used for the overexpression of proteins. The basic approach is to use standard techniques of molecular biology to construct a plasmid, known as an expression vector, that contains the gene of interest placed behind a promoter that stimulates constitutive gene expression by the organism's RNA polymerase. Since the gene is constitutively activated on the expression vector, several tens of milligrams of the desired protein are produced by the bacterial protein synthesis machinery. Such a large quantity of the desired protein is required because the NMR experiment requires protein samples to be approximately 1 mmol L^{-1} to produce sufficient sensitivity in the experiment.

The key to incorporating ^{13}C and ^{15}N into the overexpressed protein is to feed the bacterium a restrictive diet in which the sole carbon and nitrogen source contains $\geq$ 99% ^{13}C and/or ^{15}N. The most commonly used reagents for this purpose are $^{15}NH_4Cl$ and [^{13}C] glucose, although $^{15}(NH_4)_2SO_4$, [^{13}C] acetate and $^{13}CH_3OH$ have also been used. These simple chemical reagents suffice to uniformly enrich all amino acids in a protein with ^{13}C and ^{15}N. In addition to these reagents, *E. coli* requires only a source of phosphate, NaCl, Mg^{2+}, Ca^{2+}, heavy metals (such as iron, cobalt, molybdenum), thiamin and niacin in order to syn-

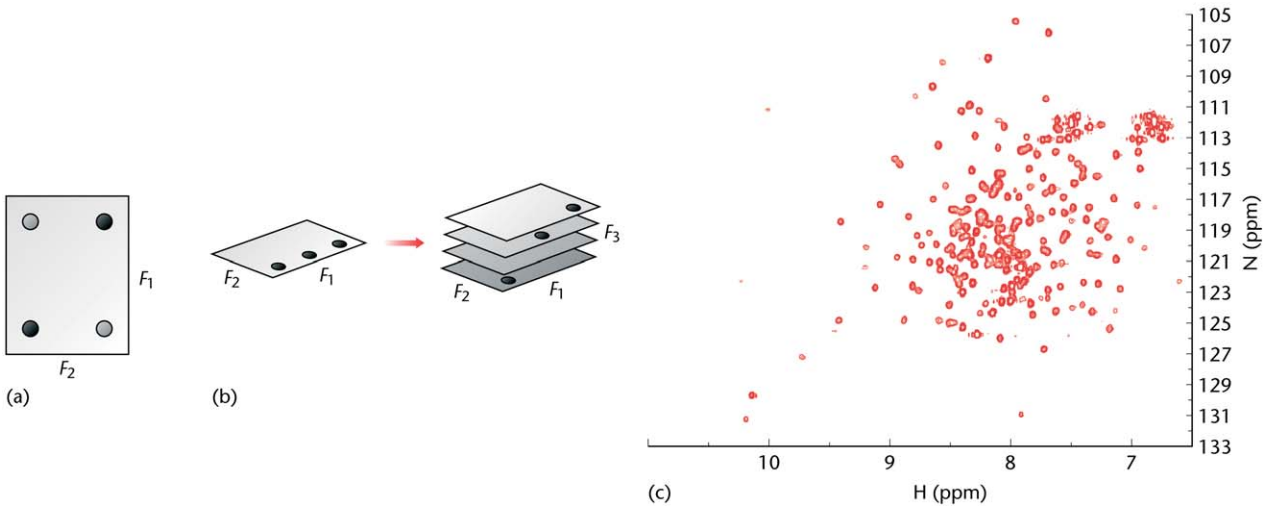


Figure 1 Schematic of multidimensional NMR. (a) A hypothetical molecule of two atoms has a one-dimensional NMR spectrum that is displayed along the diagonal of a two-dimensional spectrum (black circles). Correlation of the resonant frequencies of the two atoms leads to off-diagonal crosspeaks (grey) that represent a connectivity between the atoms. The connectivity can derive either from the nearness in space of the two atoms or the chemical bonding relationship between them. The two frequency axes, F_1 and F_2, are typically displayed in units of parts per million (ppm). (b) The transition from two to three dimensions can be appreciated when crosspeaks (black) in the two-dimensional spectrum appear at the same frequency along one of the axes in the two-dimensional spectrum. In the example shown, the three crosspeaks are resolved along one frequency axis (F_1), but are degenerate along the second frequency axis (F_2). Dispersion of the F_2 frequencies along a third axis, F_3, now resolves all three crosspeaks. Each crosspeak now has a unique F_1–F_2–F_3 position in the three-dimensional spectrum. (c) Two dimensional ^{1}H–^{15}N correlation spectrum of a protein displaying the protein backbone nitrogen and attached proton chemical shifts.

thesize all the necessary building blocks to sustain its own life. Indeed, the methods developed to sustain the growth of a bacterium on a minimal diet preceded the widespread availability of the enriched reagents by some twenty years.

A more recent development in the application of NMR to structural biology is the study of both RNA and DNA in solution (**Figure 2**). RNA has historically resisted efforts to analyse its conformation by X-ray crystallography. It is now apparent that this difficulty stemmed from the relative flexibility of an RNA molecule, which precluded the formation of an ordered crystalline array. NMR, on the other hand, is well suited to the study of poorly ordered systems if suitable quantities of the molecule to be studied can be prepared. Efficient procedures have been developed to isolate the nucleic acids from a bacterium and break the nucleic acids down into their constituent building blocks. The same methods used to label an overexpressed protein could therefore be applied to the preparation of biosynthetic precursors for enzymatic RNA synthesis. Ribonucleotide monophosphates are primarily derived from bacterial ribosomal RNA and subsequently phosphorylated using nucleotide monophosphate and diphosphate kinases available from commercial sources. The RNA nucleotides are then utilized in an *in vitro* reaction using recombinantly prepared T7 RNA polymerase and a DNA template to drive the synthesis of multiple copies of single-stranded RNA (**Figure 2a**). Despite the apparent labour involved, this approach has provided a wealth of new information on the diversity of RNA conformations that are found in nature.

An altogether different approach has been taken to the labelling of DNA for multinuclear NMR studies. For many years, DNA did not appear to require the application of isotope enrichment techniques. Large quantities of DNA have been readily available since the early 1980 s from the use of automated solid-phase synthesis. The ease of synthesis led directly to the application of two-dimensional proton NMR to the study of DNA structure in solution within a few years of the development of two-dimensional NMR itself. As larger DNAs and more complex conformations became the focus of study, chemical and enzymatic methods were pursued in an effort to prepare labelled DNA with the same relative ease as could be achieved by solid-phase chemistry. Among the most efficient routes to labelled duplex DNA are those that utilize the polymerase chain reaction (PCR) for the amplification of a tandem repeat of a desired DNA sequence (**Figure 2b**). PCR enables the synthesis of megabase-long DNAs comprising a repeating unit of designed sequence. The product of the PCR reaction is digested with a nuclease that specifically recognizes a DNA sequence built into the tandem repeat, enabling the preparation of multimilligram quantities of a specified DNA sequence with no byproducts. The inputs to the PCR reaction are deoxynucleotide triphosphates prepared from genomic DNA of a bacterium employing a strategy closely related to that for the preparation of RNA precursors.

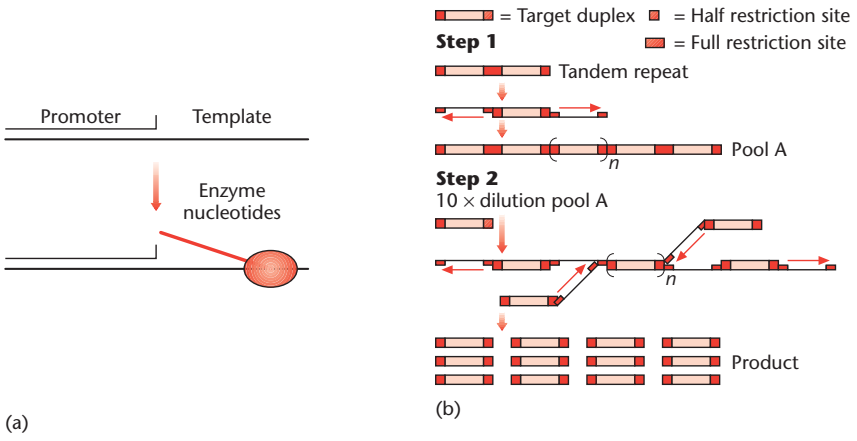


Figure 2 Two schemes for isotope labelling of nucleic acids. (a) Template-directed synthesis. The template (thin lines) is a duplex RNA or DNA molecule with an extended overhang. The overhang sequence is complementary to the DNA or RNA sequence that is desired. The short vertical line represents the transition point between the template and the desired product. During RNA synthesis, the enzyme (red oval) processively synthesizes the product RNA (heavy line) using the overhang as a template. During DNA synthesis, the enzyme is capable of making only one copy of the product per template. For DNA synthesis, the vertical transition point must be a single ribonucleotide to prime the synthesis. (b) PCR-based approach to isotope enrichment of DNA. A tandem repeat of the desired sequence is prepared with a restriction endonuclease site 5′ and 3′ to the desired sequence. The tandem repeat is then propagated by thermal amplification in the manner illustrated. Following the two-step synthesis, the product DNA is a tandem repeat of the desired sequence containing as many as 20 000 copies of the desired sequence. The product DNA is digested with the restriction endonuclease, dividing the product into the individual oligonucleotide duplexes.

Triple-resonance Techniques for Complete Spectral Assignment

Backbone assignment of proteins

The analysis of a protein structure begins with the identification of the NMR signals that belong to the individual amino acids present in the molecule and the assignment of these signals to each amino acid in the primary sequence of the protein. To identify each amino acid, a spectroscopic 'handle' is chosen to which all backbone and side-chain resonances are associated. This handle is utilized to catalogue all the chemical shifts of a given amino acid in a protein. The most convenient handle for an amino acid is the backbone amide proton and nitrogen (**Figure 3**), as each amino acid contains only one amide NH group in the backbone. The NH groups of amino acids are generally well resolved from one another in an NMR spectrum. As such, the NH group can serve as a reference point to which all other backbone and side-chain atoms of the amino acid can be linked. **Table 1** summarizes the information content of the basic NMR experiments used to sequentially assign the backbone atoms of a protein.

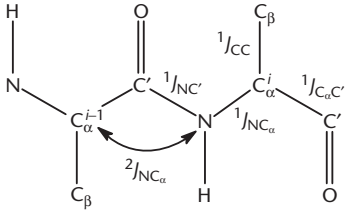
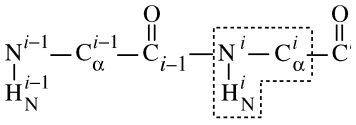
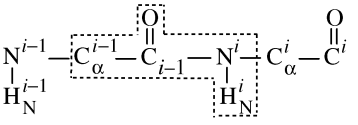
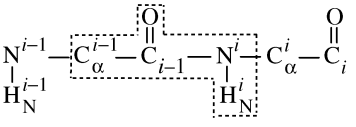
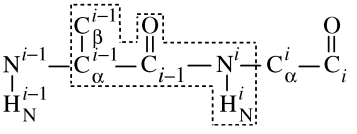
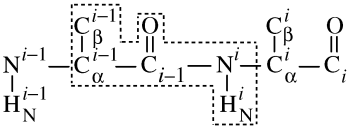
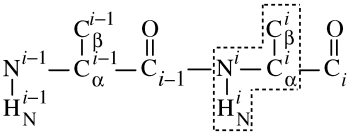


Figure 3 Sequence-specific assignment of proteins. The indicated *J* couplings provide information on the sequential connectivity of the amino acids in a protein. The relative uniqueness of dipeptide elements in a protein allows for direct sequencing of the protein from the NMR spectra.

$$(500.1323506 - 50013)\text{MHz} \times \frac{1000\ \text{Hz}}{1\ \text{MHz}} \times \frac{1\ \text{ppm}}{500.13\ \text{Hz}} = 4.699\ \text{ppm}$$

Each experiment in **Table 1** correlates a residue's N and H_N chemical shifts with the chemical shifts of either its own or the chemical shifts of the previous residue in the protein sequence. By combining the CBCA(CO)NH data with that of the HNCACB experiment, for example, the chemical shifts of the previous residue are linked to the N and H_N chemical shifts of the next residue twice (**Table 1**). For example, the Cα and Cβ chemical shifts of an amino acid and those of the previous amino acid appear in the HNCACB experiment. Thus, for every N and H_N chemical shift, four signals would be observed in the HNCACB experiment. The signals that belong to the previous residue are identified from the CBCA(CO)NH experiment, which correlates the N and H_N chemical shifts exclusively with the C_α and C_β chemical shifts of the previous residue. By combining the information from these two experiments, one can trace the chemical shifts across the peptide bond for each pair of amino acids. The trace of chemical shifts follows the polarity of a polypeptide chain, from the N-terminus to the C-terminus. Moreover, the C_α and C_β shifts

Table 1 NMR experiments for backbone sequential assignment of proteins

Experiment	Atoms observed	
HNCA	$C_{\alpha}^{i}-N^{i}-H_{N}^{i}$	
	$C_{\alpha}^{i-1}-N^{i}-H_{N}^{i}$	
HN(CO)CA	$C_{\alpha}^{i-1}-N^{i}-H_{N}^{i}$	
CBCA(CO)NH	$C_{\alpha,\beta}^{i-1}-N^{i}-H_{N}^{i}$	
HNCACB	$C_{\alpha,\beta}^{i-1}-N^{i}-H_{N}^{i}$	
	$C_{\alpha,\beta}^{i}-N^{i}-H_{N}^{i}$	

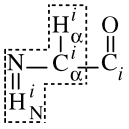
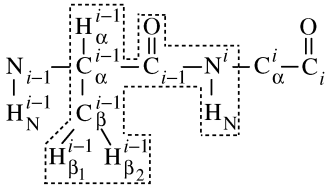
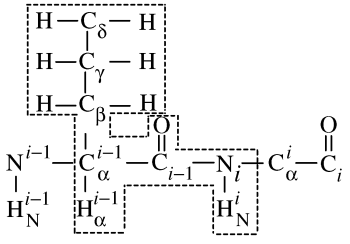
can be used to identify which amino acid type gave rise to those resonances since the chemical shifts for C_{α} and C_{β} adopt characteristic values for each amino acid. Thus, the primary sequence of a protein can literally be read from the NMR data by determining the sequential connectivity between $C_{\alpha,\beta}$ pairs and the NH group of each dipeptide element in the protein.

Side-chain assignment in proteins

Subsequent to the backbone sequential assignment, the side-chain carbon shifts are derived from two different types of experiments (**Table 2**). One experiment, the C(CO)NH experiment, correlates the previous residue's side-chain carbon shifts with the N and H_N chemical shift of the next residue, an experiment closely related to the CBCA(CO)NH experiment. A second set of experiments attempts to correlate the proton and carbon chemical shifts within a residue only. The HCCH-COSY experiment correlates the chemical shifts of CH groups separated by a single carbon–carbon bond. The complementary HCCH-TOCSY experiment permits simultaneous correlation of CH groups in a side-chain that are separated by one, two, three and even four carbon–carbon bonds. These experiments complement one another and therefore both are run to obtain complete assignments.

The C(CO)NH experiment offers an advantage over HCCH-type experiments in that the resolution of N and H_N shifts is frequently greater than that of CH groups

Table 3 NMR experiments for proton assignments of proteins

Experiment	Atoms observed	
HNHA	$H^{i}_{\alpha}-N^{i}-H^{i}_{N}$	
HBHA(CO)NH	$H^{i-1}_{\alpha,\beta}-N^{i}-H^{i}_{N}$	
H(CCO)NH	$H^{i-1}_{\alpha,\beta,\gamma,\delta,\varepsilon}-N^{i}-H^{i}_{N}$	
HCCH-COSY	$H^{i}_{\alpha}-C^{i}_{\alpha}-C^{i}_{\beta}-H^{i}_{\beta}$	
	$H^{i}_{\beta}-C^{i}_{\beta}-C^{i}_{\gamma}-H^{i}_{\gamma}$	
	$H^{i}_{\gamma}-C^{i}_{\gamma}-C^{i}_{\delta}-H^{i}_{\delta}$	
	$H^{i}_{\delta}-C^{i}_{\delta}-C^{i}_{\varepsilon}-H^{i}_{\varepsilon}$	
HCCH-TOCSY	$H^{i}_{\alpha}-C^{i}_{\alpha}-C^{i}_{\beta,\gamma,\delta,\varepsilon}-H^{i}_{\beta,\gamma,\delta,\varepsilon}$	
	$H^{i}_{\beta}-C^{i}_{\beta}-C^{i}_{\gamma,\delta,\varepsilon}-H^{i}_{\gamma,\delta,\varepsilon}$	
	$H^{i}_{\gamma}-C^{i}_{\gamma}-C^{i}_{\delta,\varepsilon}-H^{i}_{\delta,\varepsilon}$	

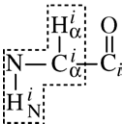
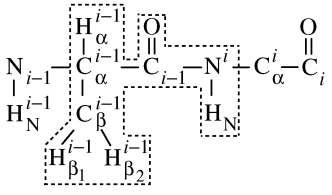
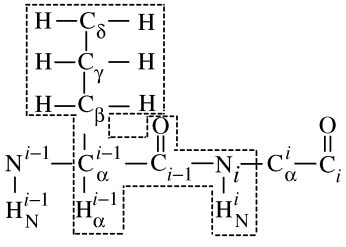
alone. Moreover, correlation of the side-chain carbon shifts of residue $i-1$ with the NH groups of residue i provides a verification of the amino acid type assignments made from $C_{\alpha,\beta}$ pairs. The number of carbon shift correlations and the absolute value of the carbon chemical shifts both contribute to the identification of the likely amino acid type for a given set of signals. For example, alanine is readily distinguished from valine owing to the presence of three carbons in the valine side-chain as compared to one carbon for that of alanine. Alanine is further distinguished from valine by the absolute value of its C_{β} shift, which is frequently 5–10 ppm upfield of the methyl groups of valine. The concept of a pattern of connectivities that can identify amino acid type is also used in the analysis of HCCH experiments wherein the number of connected CH_2 groups

can be counted as a means of identifying the possible amino acid that gives rise to the observed connectivities.

Proton assignments in proteins

The most difficult set of assignments to be made for proteins are the chemical shifts of the protons. The chemical shift dispersion of protons is relatively narrow and frequently leads to overlaps that would be impossible to resolve without simultaneous analysis of carbon and/or nitrogen chemical shifts. For this reason, the proton shifts are most easily measured by correlation with the best-resolved functional group of a protein, the NH group at the protein backbone. H_α and H_β are readily identified by a

Table 3 NMR experiments for proton assignments of proteins

Experiment	Atoms observed	
HNHA	$H^i_\alpha - N^i - H^i_N$	
HBHA(CO)NH	$H^{i-1}_{\alpha,\beta} - N^i - H^i_N$	
H(CCO)NH	$H^{i-1}_{\alpha,\beta,\gamma,\delta,\varepsilon} - N^i - H^i_N$	
HCCH-COSY	$H^i_\alpha - C^i_\alpha - C^i_\beta - H^i_\beta$	
	$H^i_\beta - C^i_\beta - C^i_\gamma - H^i_\gamma$	
	$H^i_\gamma - C^i_\gamma - C^i_\delta - H^i_\delta$	
	$H^i_\delta - C^i_\delta - C^i_\varepsilon - H^i_\varepsilon$	
HCCH-TOCSY	$H^i_\alpha - C^i_\alpha - C^i_{\beta,\gamma,\delta,\varepsilon} - H^i_{\beta,\gamma,\delta,\varepsilon}$	
	$H^i_\beta - C^i_\beta - C^i_{\gamma,\delta,\varepsilon} - H^i_{\gamma,\delta,\varepsilon}$	
	$H^i_\gamma - C^i_\gamma - C^i_{\delta,\varepsilon} - H^i_{\delta,\varepsilon}$	

combination of experiments (**Table 3**) which are related to the CBCA(CO)NH experiment. Proton assignments that cannot be derived from HNHA, HBHA(CO)NH and HC(CO)NH experiments must be derived from the same HCCH-type experiments utilized to identify carbon shifts.

Resonance assignment in nucleic acids

Resonance assignment of nucleic acids follows a similar strategy to that of proteins for the identification of residue type. The nucleotide type is identified from a combination of HSQC, Hb(C)Nb and Hs(Cs)N(Cb)Hb experiments (**Table 4**). The combination of these three experiments provides a nearly unambiguous assignment of H1' and H6/H8 resonances to one of the four nucleotide types via the N9/N1 nitrogens and C8/C6 carbons of the nucleotide base. From this point, the assignment diverges significantly from that of proteins, relying on the analysis of nuclear Overhauser enhancement (NOE) spectroscopy which provides information on the spatial arrangement of atoms in the oligonucleotide rather than information on the pattern of chemical bonds (**Figure 4**).

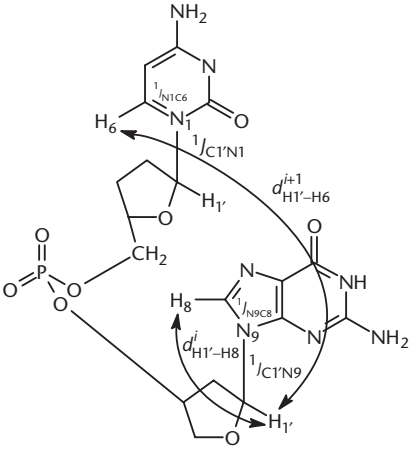


Figure 4 Sequence-specific assignment of nucleic acids. The indicated *J* couplings permit discrimination of each of the four residue types in DNA or RNA. In contrast to proteins, the sequential assignment of nucleic acids is accomplished by NOESY using the indicated distance (d) connectivities between H6/H8 and H1′ protons within and between dinucleotide steps in the sequence. Intraresidue connectivities (:) and sequential connectivities (9) permit complete assignment of the nucleotide sequence when combined with residue type assignments indicated in **Table 4**.

Generating Structural Restraints from Nuclear Overhauser Enhancement Spectroscopy (NOESY)

With the completion of the sequence-specific assignment of all ^{1}H, ^{13}C and ^{15}N chemical shifts, it is possible to begin the process of assembling a table of geometric restraints that is capable of defining the three-dimensional structure of a molecule. The principal source of geometric restraints derives from NOESY experiments, which report on the relative proximity of two atoms in space. The intensity of the NOE signal is inversely proportional to r^6, the sixth power of the distance between the atoms. As a consequence, distances between atoms separated by more than ~ 5 Å are generally not observed in NOESY. This means that the structure is constructed from a large number of nearest-neighbour relationships between atoms closely spaced in the three-dimensional structure.

The challenge to the NMR spectroscopist is to properly assign and calibrate the NOEs. The only distance relationship that can be measured in a biomolecule is that between protons. The relatively poor chemical shift dispersion of protons can often lead to multiple assignment possibilities for a given NOE; thus it is imperative to identify a subset of NOEs for which the assignment is unambiguous. To accomplish this, the carbon or nitrogen nucleus attached to a proton is often simultaneously observed so that both chemical shifts can be used to make an NOE assignment. A number of NOE data sets are therefore collected to achieve this end (**Table 5**).

The relationship between NOE intensity and distance is calibrated by choosing a known distance between two atoms in an amino acid and examining the statistics of the NOE intensity variation for each occurrence of this distance in a given data set. Because a number of experimental

Table 4 NMR experiments for the identification of residue type in nucleic acids

Experiment	Atoms observed
HSQC	Purine H8–C8 pyrimidine H6–C6
Hb(C)Nb	Purine H8–N9 Pyrimidine H6–N1
Hs(Cs)N(Cb)Hb	Purine H8–N9–H1′ purine H6–N1–H1′

Table 5 NOE experiments to generate distance restraints in biomolecules[a]

Experiment	Distance observed
3D ^{15}N-separated NOESY	H^j–$N^j \cdots H^i$
3D ^{13}C-separated NOESY	H^j–$C^j \cdots H^i$
4D ^{13}C/^{13}C-separated NOESY	H^j–$C^j \cdots C^i$–H^i
4D ^{13}C/^{15}N-separated NOESY	H^j–$C^j \cdots N^i$–H^i

[a]The NOE is observed between two residues, i and j, which are not necessarily sequentially related.

artefacts can lead to miscalibration of the NOE intensities, the NOEs are grouped into three or four distance ranges that are distinguished on the basis of the upper bounds of the range. A common scheme is to use 2.7 Å to represent the strongest NOEs and 5 Å to represent the weakest, with an intermediate intensity given an upper distance of 3.5 Å. In this way, strong, medium and weak NOEs can be loosely interpreted as representing distances that are no more than 2.7 Å, 3.5 Å or 5 Å, respectively. The lower bound to any NOE distance range is typically set to be that of the van der Waals contact distance for hydrogen atoms, 1.8 Å in practice.

Calculating Three-dimensional Structures from NMR Data

The calculation of a three-dimensional structure from NOE data is an iterative process. An initial set of NOEs whose assignment is unambiguous is the beginning of a process in which the early structures are examined in conjunction with the NOE and chemical shift data to identify as many distance relationships as possible that can define the global fold of the molecule. This initial step is laborious and requires the greatest care. Unless the molecule is known to be closely related to one whose structure has already been described, establishing the correct initial fold is the most important part of the structure calculation process.

Identification of secondary structure elements

Correct identification of elements of regular secondary structure can simplify the construction of distance restraint tables and help analyse whether the initial fold of the molecule faithfully represents the experimental data. The basic elements, α helix and β sheet, are identified from a distinct pattern of NOEs in conjunction with secondary C_α and C_β shifts. The C_α and C_β chemical shifts of amino acids adopt characteristic values for each amino acid. The chemical shift values for these atoms can be used to identify the type of amino acid in a protein as discussed above. Amino acids that reside in helical and sheet secondary structures display deviations from the expected values of C_α and C_β chemical shifts for each amino acid. The deviation, or secondary shift, is diagnostic for elements of regular secondary structure. Positive deviations for C_α and negative deviations for C_β are observed for helical residues; the opposite pattern is observed for sheet residues. NOE patterns can confirm these assignments, as outlined in Table 6.

Iterative refinement

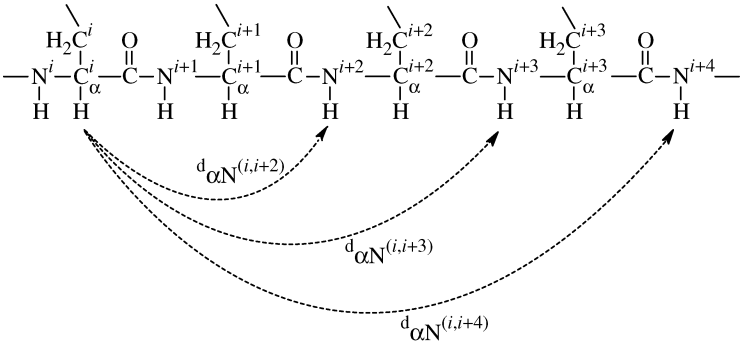
Figure 5 outlines the iterative refinement process employed in a typical structure determination. Briefly, the unambiguous NOE subset is used to generate an initial family of structures. Since there are too few NOEs utilized at this stage to uniquely define the conformation of the molecule, a family of structures having generally the same fold are generated by the molecular dynamics simulation. The initial structures are checked against the NOEs for distance violations and for accurate representations of the secondary structure elements. Those NOEs found to deviate more than 0.5 Å from the upper bound in the family of structures represent possible misassignments and/or categorical miscalibration of NOEs in a given data set. These NOEs are immediately checked in the original data, corrections are made to the assignment or data set calibration, and the calculation is repeated. Once the initial violations are removed from the structure family, the structures should be assessed by validation tools that check for torsion angle violations in the backbone. These violations can frequently provide an insight into more subtle assignment or calibration errors.

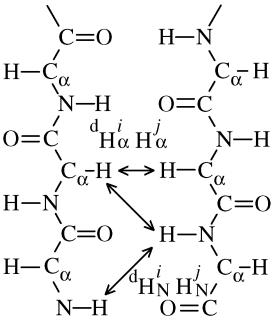
Once the initial family is constructed, it can be used to assign more NOEs by simultaneous analysis of distances measured from the initial conformers in conjunction with the chemical shifts of the NOEs. A dual search in this manner can help resolve ambiguous assignments in which the chemical shifts alone do not permit a clear identification of the atoms that gave rise to the observed NOE. For example, an NOE may appear to belong to two possible proton pairs based on the observed chemical shifts; however, one of these pairs may be 9 Å apart in the initial structures while the second pair may be only 4 Å apart. This suggests that the latter distance is the more likely assignment. Of course, the presumption made here is that the initial conformer from which a distance estimate is made approximately represents the true structure. For this reason, great care must be taken in assessing the validity of the initial fold. It is often advisable to challenge the observed conformers by looking for NOEs which may not have been used in the initial calculation but should be clearly visible in the NMR spectrum. Such challenges can act as a check as to whether the right elements of secondary structure are oriented near each other in the three-dimensional structure.

Higher resolution can be achieved in the iterative process if angular restraints are measured from *J* coupling constant data. The *J* coupling can report on the orientation of atoms that is established by the chemical bonding network. The most useful angular information involves atoms separated by three or four covalent bonds. When combined with NOE data, iterative refinement leads to a higher degree of convergence for a family of structures.

Table 6 NOE patterns that define regular secondary structure elements in proteins

NOE observed	Secondary structure element
$H_{\alpha}^{i}-H_{N}^{i+2}$	Helix
$H_{\alpha}^{i}-H_{N}^{i+3}$	Helix
$H_{\alpha}^{i}-H_{N}^{i+4}$	Helix
$H_{\alpha}^{i}-H_{N}^{j}$	Sheet
$H_{N}^{i}-H_{N}^{j}$	Sheet
$H_{\alpha}^{i}-H_{\alpha}^{j}$	Sheet





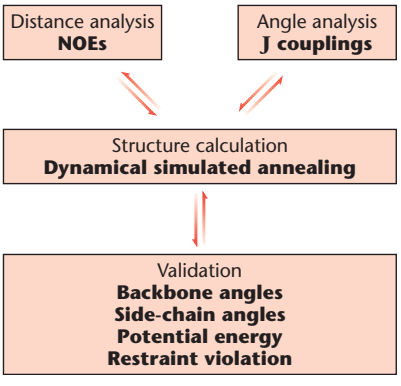


Figure 5 Iterative refinement procedure for NMR structure calculation. Distance and angle restraints are used as inputs to a calculation protocol termed dynamical simulated annealing. The simulated annealing calculation uses a target function that models the potential energies of chemical bonds, nonbonded energies such as van der Waals forces, angular relationships and electrostatic energies in addition to the experimental distance and angle restraints. Validation is done at each step in a cycle to check for violations in experimental restraints and deviations from chemical bond lengths and angles commonly found in proteins and nucleic acids. The iterative cycle attempts to refine the structure by inputting more and more experimental restraints until all possible experimental data are utilized in the calculation.

Summary

NMR spectroscopy permits the analysis of molecular conformation in solution. The procedure leading to a three-dimensional structure begins with the identification of the resonant frequencies of each atom and their assignment to specific locations in the primary sequence of the nucleic acid or protein. Subsequent analysis of distance and angular relationships between atoms enables the construction of a family of three-dimensional models that represent the conformation of the molecule in an aqueous environment.

Further Reading

Cavanagh J, Fairbrother WJ, Palmer AG and Skelton JJ (1996) *Protein NMR Spectroscopy: Principles and Practice*. New York: Academic Press.

Krishna NR and Berliner LJ (eds) (1998) *Biological Magnetic Resonance, vol. 16–17, Modern Techniques in Protein NMR*. New York: Kluwer Academic/Plenum.

Wüthrich K (1986) NMR of Proteins and Nucleic Acids. New York: Wiley Interscience.