Multidimensional NMR Methods for Protein Structure Determination

Voula Kanelis,1,2 Julie D. Forman-Kay,1,2 and Lewis E. Kay2,3,4,5
1Program in Structural Biology and Biochemistry, Hospital for Sick Children, 555 University Ave., Toronto, Ontario M5G 1X8, Canada
2Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada
3Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Ontario M5S 1A8, Canada
4Department of Chemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada
5Protein Engineering Network Centres of Excellence, National Business Centre, 750 Heritage Medical Research Centre, Edmonton, Alberta T6G 252, Canada

Summary

Structural studies of proteins are critical for understanding biological processes at the molecular level. Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for obtaining structural and dynamic information on proteins and protein-ligand complexes. In the present review, methodologies for NMR structure determination of proteins and macromolecular complexes are described. In addition, a number of recent advances that reduce the molecular weight limitations previously imposed on NMR studies of biomolecules are discussed, highlighting applications of these technologies to protein systems studied in our laboratories.

Keywords
Deuteration; dipolar couplings; multidimensional heteronuclear NMR; TROSY.

INTRODUCTION

NMR spectroscopy is a powerful technique for studying protein structure and dynamics. NMR studies of proteins and protein complexes with molecular weights up to approximately 25 kDa are now, in many cases, routine. This has largely been the result of a number of important improvements in technology, including 1) the increase in dimensionality of experiments from two dimensions to three and four, providing improved resolution for complex spectra of large macromolecules; 2) incorporation of NMR-active nuclei such as 15N and 13C into the molecule of interest and the development of NMR experiments that transfer magnetization between these nuclei via through bond (scalar coupling) or through space (dipolar coupling) interactions; and 3) significantly improved radio frequency (RF) electronics, increased field strengths and the development of commercially available hardware, such as pulsed field gradients and cryo-probes.

Recently, further advances have facilitated NMR studies of proteins larger than 25 kDa. For example, labeling strategies involving the incorporation of deuterium (2H) in 15N, 13C-enriched proteins have been developed to remove pathways that cause rapid decay of the NMR signal, termed spin relaxation, resulting in spectra with increased sensitivity and resolution (1, 2). Further, attenuation of relaxation effects by new spectroscopic means have resulted in significant sensitivity gains in many classes of NMR experiments (3). Finally, the incorporation of global restraints that provide information regarding relative orientations of discrete structural elements has increased the precision of NMR-derived structures, especially those with a limited number of distance restraints, often the case for large molecules (4). In this review, we first briefly describe the utility of triple resonance multidimensional NMR for structural studies of proteins, focusing on studies of a small complex comprised of a WW domain from the ubiquitin ligase Nedd4 (50 residues) and a target peptide (17 residues) (5). Subsequently, advances in methodology that have led to structural studies of larger molecules are described. Developments and applications from work in our laboratories are highlighted. For further details the interested reader is referred to a large number
of reviews that have been published on multidimensional NMR spectroscopy of biological macromolecules in the past several years (1, 2, 6, 7).

**Chemical Shift Assignment**

As with any spectroscopy it is first necessary to assign the frequencies at which energy absorption occurs to specific sites in the molecule. The unique chemical environment for each NMR spin probe means that each NMR nucleus of a given type has a distinct absorption frequency. The frequency of absorption is related to the chemical shift and the goal of every NMR spectroscopist at the start of a project is to obtain the chemical shift of each NMR active spin, a process termed chemical shift assignment. Once such assignments are available, experiments that encode structural or dynamic information at every site in the macromolecule can be analyzed. In applications to proteins, where it is not uncommon to find well over 500 protons per molecule, it is often the case that more than one proton absorbs energy at a given frequency. This leads to overlap of resonance lines in spectra and can significantly complicate interpretation of data. A particularly important advance in chemical shift assignment has been the development of multidimensional NMR experiments. The overlap in one- and two-dimensional NMR spectra can be significantly reduced by ‘spreading out’ the information content in these spectra into a cube to generate a three-dimensional data set. Extensive reviews of multidimensional NMR have been published previously (see previous references) and we consider here a small number of experiments important for chemical shift assignment.

**Triple Resonance Multidimensional NMR Experiments**

A robust approach for obtaining chemical shift assignments involves the uniform labeling of proteins with $^{15}$N and $^{13}$C and recording a series of three-dimensional experiments that correlate information between three nuclei at a time, generally $^{15}$N, $^{13}$C, and $^1$H spins (hence the designation triple resonance). Proteins can be readily generated with $^{15}$N and $^{13}$C labels by over-expression in bacteria grown in minimal media supplemented with $^{15}$N-$\text{NH}_4\text{Cl}$ and $^{13}$C-glucose as the sole nitrogen and carbon sources, respectively. Triple resonance experiments make use of the large J-couplings (through bond interactions) between $^{15}$N and $^{13}$C nuclei and between these nuclei and their attached protons for efficient information (magnetization) transfer. Chemical shift assignment of a protein begins with the backbone and many experiments have been developed that correlate different backbone nuclei. These experiments are often analyzed in pairs with one experiment recording both intra- and interresidue correlations and the second recording only interresidue correlations. Thus, experiments have been developed which correlate the $^{1}$HN, $^{15}$N chemical shifts of residue $i$ with the $^{13}$CO (carbonyl) shift of residue $i-1$ or $i$, the $^{1}$HN, $^{15}$N chemical shifts of residue $i$ with the $^{13}$Ca or $^{13}$Cb shifts of residue $i-1$ or the $^{13}$Ca or $^{13}$Cb shifts of both residues $i-1$ and $i$, or the $^{1}$HN, $^{15}$N chemical shifts of residue $i$ with carbon or proton side chain chemical shifts of the preceding residue, to name but a few (8).

The increased dimensionality in these experiments (i.e., three vs. two dimensions in two-dimensional $^1$H-only spectra) coupled with isotopic enrichment leads to a reduction in resonance degeneracy. For example, although the probability that any pair of chemical shifts is unique to a particular site in the protein might not be so high (two-dimensional NMR), the probability that a triplet of shifts is unique (three-dimensional NMR) is much higher. Thus, a larger number of resonance assignments can be made unambiguously, leading to a greater number of probes of structure and dynamics. In the case of protein structure determination, the increased number of assignments translates to more structural restraints and hence increased accuracy and precision of NMR-derived structures.

As an illustration of the utility of triple resonance NMR approaches for assignment, we focus our discussion here on two of the most important experiments for backbone assignment, the 3D HNCa$^\beta$ (9) and 3D (H$\beta$)C$^\beta$Ca(CO)NNH (10) experiments. The HNCa$^\beta$ experiment correlates the $^{1}$HN and $^{15}$N chemical shifts of residue $i$ with the $^{13}$Ca and $^{13}$Cb shifts of residues $i-1$ and $i$ (Fig. 1A, top left), while the (H$\beta$)C$^\beta$Ca(CO)NNH experiment provides correlations linking $^{1}$HN and $^{15}$N chemical shifts of a residue with $^{13}$Ca and $^{13}$Cb shifts of the previous residue (Fig. 1A, top right). Therefore in the HNCa$^\beta$ experiment, cross peaks are observed at $(\Omega_{\text{Ca},i-1}/\Omega_{\text{Cb},i-1}, \Omega_{N,i}, \Omega_{\text{HN},i})$ and at $(\Omega_{\text{Ca},i-1}/\Omega_{\text{Cb},i-1}, \Omega_{N,i}, \Omega_{\text{HN},i})$, where $i$ denotes residue number and $\Omega_X$ is the resonance frequency of a particular nucleus. In contrast, the (H$\beta$)C$^\beta$Ca(CO)NNH experiment provides correlations of the form $(\Omega_{\text{Ca},i-1}/\Omega_{\text{Cb},i-1}, \Omega_{N,i}, \Omega_{\text{HN},i})$, allowing facile separation of the intra- and interresidue correlations in the HNCa$^\beta$. One major advantage of these experiments is that the $^{13}$C$^\beta$ chemical shift is recorded alongside that of the $^{15}$N nuclei (red vs. black in Fig. 1B), thus eliminating any ambiguity in assigning the $^{15}$N chemical shifts and are easily distinguished from other residues. Of note, in the HNCa$^\beta$ experiment, cross peaks are observed at $(\Omega_{\text{Ca},i-1}/\Omega_{\text{Cb},i-1}, \Omega_{N,i}, \Omega_{\text{HN},i})$ allowing facile separation of the intra- and interresidue correlations in the HNCa$^\beta$. Other amino acids, including Gly and Ser, also display distinct chemical shifts and are easily distinguished from other residues. Of note, in the HNCa$^\beta$ experiment $^{13}$C$\beta$ signals are of opposite phase to those originating from $^{13}$Ca nuclei (red vs. black in Fig. 1B), thus eliminating any ambiguity in assigning $^{13}$Ca and $^{13}$C$\beta$ spins.

Fig. 1A (bottom) shows portions of the three-dimensional HNCa$^\beta$ data set recorded on a sample consisting of the C-terminal WW domain from the ubiquitin ligase Nedd4, uniformly labeled with $^{15}$N and $^{13}$C, complexed with an unlabeled peptide derived from the $\beta$ subunit ($\beta$P2) of the epithelial sodium
channel (ENaC). Structural studies of this complex are of interest because mutations in the βP2 region of ENaC abrogate binding to Nedd4 WW domains (12) and cause Liddle syndrome (13–16), a hereditary form of hypertension resulting from increased ENaC activity, thus implicating Nedd4 as a suppressor of channel activity (12). In the representation shown in Fig. 1 strips from the 3D HNCαCβ cube at 1H N and 13C N chemical shifts of each of the residues displayed along the x-axis are illustrated with the 13C N and 15C β chemical shifts indicated along the y-axis. Assignment of the HNCαCβ data was accomplished with assistance from the (HβCβCα(CO)NNH experiment (data not shown) to separate intra- from interresidue correlations. The interresidue sequential connectivities are easily identified and are represented by dotted lines. As an example of the utility of 13C β shifts in resonance assignment, the 13C β chemical shift of Thr 485 is significantly downfield with respect to other amino acids, facilitating identification of this residue type. In some cases, the HNCαCβ and (HβCβCα(CO)NNH experiments are sufficient to obtain a complete chemical shift assignment of 1H N, 15N, 13C N, and side chain 13C β nuclei. However, for most proteins other experiments are necessary.

Once backbone chemical shifts have been obtained, the assignment of side chain chemical shifts is straightforward. In a similar manner to the (HβCβCα(CO)NNH, the H(C)(CO)NH-TOCSY (11, 17) and (H(C)(CO)NH-TOCSY (17, 18) experiments provide correlations linking the 13N N and 1H N shifts of residue i with all aliphatic 1H or 13C shifts in the previous residue, respectively (data not shown). Because correlations contain the previously assigned 15N N and 1H N nuclei, reading off chemical shifts for aliphatic resonances is straightforward and in cases where distinct chemical shifts exist for aliphatic (α, β, γ, etc.) positions, assignments are easily made. In some cases additional spectra must be recorded in which pairs of proximal protons are labeled with the chemical shift of the heteroatom (15N or 13C; so called 15N or 13C-edited data sets) to which they are attached. A particularly important experiment in this category is one that is recorded on uniformly 15N, 13C-labeled samples in H2O (27).

In this simultaneous 15N, 13C-edited three-dimensional data set cross peaks are observed at chemical shifts of (ΩH i, ΩX j, ΩH j) and (ΩH i, ΩX k, ΩH k), where i and j are proximal group; and X indicates an 15N or 13C nucleus. The intensity of each correlation is related to the inverse sixth power of the distance between the proton spins i and j, which is the source of the distance information in the NOE experiment. Cross-peaks of the form (ΩH N i, ΩC j, ΩH C j) are observed, connecting proximal proton spins HN i and HC j, along with a symmetry-related correlation at (ΩH C k, ΩH N j, ΩH N k). Carbon-bound protons, k and l, that are close in space would give rise to correlations at (ΩH C k, ΩC j, ΩH C j) and (ΩH C k, ΩC l, ΩH C l), yet proximal amide protons, m and n, yield cross-peaks at (ΩH N m, ΩH N n, ΩH N l) and (ΩH N n, ΩH N m, ΩH N l). The 13C or 15N chemical shift that is recorded in each of the correlations is critical for the identification of the correct pair of neighboring protons because there is often a large amount of overlap in two-dimensional homonuclear NOE spectra.

Fig. 1B shows strips from the 15N N- and 13C-edited NOESY recorded on the Nedd4 WW domain—ENaC βP2 peptide complex described previously. Each of the strips is extracted from the three-dimensional data set at the 15N or 13C frequency of the nucleus indicated at the top of the panel. Note the symmetry related peaks linking Trp 465 Hδ3 to Phe 477 Hβ2 and Ile 482 HN to Lys 483 Hσ indicated with the dashed horizontal lines. In addition, because NOEs originating from any proton in the protein (or complex) can be observed, intermolecular NOEs between the βP2 peptide and the WW domain are also seen in this experiment (denoted by *). Specifically, correlations are observed between Tyr 618 Hδ in the βP2 peptide and Ile 482 HN and Lys 483 Hσ in the WW domain, as well as between Tyr 618 Hδ

**Structural Information**

Structural information in NMR can be obtained from a large variety of different parameters. For example, chemical shifts can be used to obtain restraints for the backbone dihedral angles φ and ψ and, as discussed later, to yield orientational information (22, 23). Hydrogen bond restraints can be obtained from a number of different approaches, ranging from simple experiments in which lyophilized protein is dissolved in D2O and the rate of disappearance of amide proton signals are monitored, to more sophisticated experiments in which magnetization is transferred between the 15N and 13CO nuclei across the hydrogen bond (24, 25). In addition, there are a large number of experiments for measuring scalar couplings, J, which can, in turn, be related to dihedral angles, θ, providing that suitable J vs. θ parameterizations exist (26). However, the most important class of restraints in NMR structure determination is provided by nuclear Overhauser effect (NOE) experiments that establish correlations between protons close in space (within 5–6 Å). In a folded protein, a given proton is potentially surrounded by as many as 15 proximal protons and thus, many NOE correlations are possible. NMR structure determination is heavily dependent on the measurement of large numbers of NOEs to give interproton distance restraints because the number of NOEs unambiguously assigned, or mapped to a specific pair of protons, has a large effect on the precision of the structure. Therefore, analysis of NOE spectra is critical to the NMR-based structure determination process and, not surprisingly, a significant effort has been expended in the development of robust experiments for measuring these restraints. As with the triple resonance experiments discussed before for chemical shift assignment, NOE experiments have also enjoyed the benefits of increased dimensionality. Typically, experiments are recorded in which pairs of proximal protons are labeled with the chemical shift of the heteroatom (15N or 13C; so called 15N or 13C-edited data sets) to which they are attached. A particularly important experiment in this category is one that is recorded on uniformly 15N, 13C-labeled samples in H2O (27).
Figure 1. NMR structure determination of proteins and protein complexes. (A) Sequential assignment of the Nedd4 WW domain. Schematic representations of dipeptides showing nuclei in residues $i$ and $i-1$, where $i$ is the residue number, that are correlated (circled) in the HNCα/β (top left) and (Hβ)CβCa(CO)NNH experiments (top right). The $^{13}$Cβ nuclei observed in the HNCαCβ are colored red, denoting the opposite phase of signals arising from these spins relative to phases of $^{13}$Cα signals. Strip plots from the HNCαCβ experiment are at the $^1$HN($i$) and $^{15}$N($i$) chemical shifts for residues His 480 to Glu 488 of the Nedd4 WW domain (bottom). The negative $^{13}$Cβ signals are represented as red contours. Correlations between sequential $^{13}$Cα/$^{13}$Cβ resonances are indicated by dotted lines. The asterisks (*) in the His 480 strip identify peaks with increased intensity on another plane. This spectrum was recorded at 500 MHz ($^1$H frequency) on a 1-mM $^{15}$N/$^{13}$C-labeled Nedd4 WW domain bound to the unlabeled ENaC β2 peptide in 10 mM sodium phosphate, 90% H$_2$O, 10% D$_2$O, pH 6.5 at 30°C. (Continued)
Hβ2 and Lys 483 Hα. (ENaC peptide residues are marked with a “°” to distinguish them from WW domain residues.) Obtaining intermolecular NOEs is critical to the structure determination of molecular complexes and experiments have been specifically designed for selecting only intermolecular NOEs (28, 29). By labeling one component of the complex with 15N and 13C, with the second molecular player unlabeled (as in the present example), it is possible to observe only the signal that has been transferred selectively from protons that are not directly coupled to 15N or 13C spins (i.e., from the unlabeled molecule) to protons that are directly attached to NMR active spins (i.e., to the labeled protein).

Distance restraints derived from NOEs (1,745 intramolecular and 54 intermolecular), 28 dihedral angle restraints from chemical shifts and scalar couplings along with 14 hydrogen bond restraints have been used to calculate an ensemble of structures of the Nedd4 WW domain—ENaC βP2 peptide complex (Fig. 1C). The structures produced have very few violations of the input experimental data and satisfy known covalent geometry (5). The models obtained are of high precision, with backbone and side chain atomic root mean squared deviations (r.m.s.d.) from the mean structure of 0.23 Å and 0.62 Å, respectively, for residues Leu 461 to Arg 493 in the WW domain and residues Pro 615 to Leu 621 in the βP2 peptide. It is clear from this example and from many others in the literature that well-defined structures of large complexes can be obtained by NMR. In the following sections we describe methods for studying larger systems where, in general, it is not possible to measure as many structural restraints.

APPLICATIONS TO LARGER COMPLEXES

Labeling Strategies

Multidimensional, multinuclear NMR experiments are comprised of building blocks that transfer magnetization from one spin to the next, exploiting the large one- and two-bond scalar couplings that manifest between the spins. The transfer mechanism can be schematized as,

$$A \rightarrow B \rightarrow C \rightarrow \ldots \rightarrow Z$$

[1]

where A, B, C, ..., Z are NMR active spins and the arrows denote information transfer relying on through bond couplings. These couplings are molecular size invariant and it might be assumed, therefore, that the transfer efficiency would also be independent of molecular weight. In fact, this is far from the case, because during the transfer process itself the signal decays and this decay rate scales with the size of the molecule. The process of signal decay is termed relaxation and the relaxation of magnetic nuclei is caused by the magnetic fields that each NMR spin (effectively a bar magnet) helps to create. These fields fluctuate in time, because the NMR spins are attached to molecules that tumble in solution, and these fluctuating fields cause transitions that restore the equilibrium distribution of magnetization and hence destroy the signal that is meant to be transferred according to equation [1]. The size of these fields scale with the gyromagnetic ratios, γ, of the nuclei that produce them, so that fields due to proton spins are four-fold larger than those generated by carbon spins, for example, since γH/γC = 4. Thus, the triple resonance methods described previously often fail for large protonated proteins (typically on the order of 25 kDa or larger).

It is possible to remove a large portion of the fluctuating magnetic fields that are so deleterious for many of the NMR experiments by substituting deuterons for carbon-bound protons, because the gyromagnetic ratio of a deuteron is approximately a factor of 6.5 times less than that of a proton (I). Highly deuterated, 15N, 13C-labeled samples can be prepared by overproducing the protein of interest in media where D2O is used exclusively as the solvent (I). Typically, the deuterons attached to backbone and side chain nitrogen positions are then exchanged with protons by dissolving the protein in H2O (or unfolding and subsequently refolding the protein if necessary), so that a small, albeit important, complement of proton spins remains. The significance of maintaining a high level of amide protons is that many NMR experiments start with and subsequently detect signal on protons, for reasons of sensitivity (i.e., A and Z in scheme [1] would be 1H).

The substitution of deuterons for protons leads to a decrease in the relaxation rates of 13C spins and for experiments that rely on magnetization transfer through 13C nuclei, such as the backbone

Figure 1. (Continued) (B) Strips from the 150-ms 13C/15N-edited NOESY-HSQC spectrum, recorded on the sample described in (A) illustrating NOEs to H3 of Trp 465, Hβ of Phe 477, HN of Ile 482, and Hα of Lys 483. The chemical shifts of each of the protons indicated above are at the bottom of the appropriate traces with the directly attached heteroatom (C or N) labeled at the top. Symmetry-related peaks linking Trp 465 H3 to Phe 477 Hβ2 and Ile 482 HN to Lys 483 Hα are shown with a dotted line. The diagonal in each strip is indicated by H3, Hβ2, HN, and Hα. Asterisks (*) represent intermolecular NOE correlations between Ile 482 HN and Tyr 618 Hδ and between Lys 483 Hα and Tyr 618 Hβ2, Hδ. (C) Solution structure of the Nedd4 WW domain—ENaC βP2 peptide complex. A superposition of the final 15 structures (PDB code 115H) is displayed (left) alongside a schematic ribbon diagram of the lowest energy structure (right). The backbone of the WW domain is colored blue for the β-strands and gray for loops and termini. The side chains of the canonical WW domain tryptophan and proline residues are shown in pink/purple. The side chains of additional residues involved in the hydrophobic cluster and in peptide binding are shown in cyan. The βP2 peptide backbone is in orange and selected side chains are shown in yellow. Residue numbers and termini for the βP2 peptide are denoted with a prime ('). Fig. 1C is reproduced with permission from Kanelis et al. (5).
and side chain triple resonance experiments discussed above, the slowed relaxation translates into increased sensitivity and resolution. Not surprisingly, therefore, many of the experiments originally developed for application to fully protonated samples have been modified for studies of highly deuterated proteins (1). In addition to the decreased relaxation rates of the carbon spins, the remaining protons, for example the labile backbone and side-chain HN protons, also benefit from slower decay rates because many relaxation pathways normally involving these protons are attenuated. The decreased amide proton linewidths lead to dramatic improvements in NOESY spectra that measure HN→HN correlations relative to data sets obtained on protonated samples (30, 31). Further, the flow of magnetization in these experiments is restricted to fewer protons than would be the case in a protonated molecule and as a result longer distances can be measured and, in addition, these distances are determined more accurately.

Unfortunately, the benefits associated with perdeuteration in terms of chemical shift assignment and recording HN→HN NOE connectivities do come with a cost. The removal of the majority of the protons in the protein eliminates many of the distance probes that would otherwise form the basis of structural studies with NOE experiments. As described before, the quality of NMR-derived structures depends heavily on the number of unambiguous distance restraints that are available and often only crude structures can be obtained on the basis of HN→HN restraints exclusively (32). With this in mind, we have developed a protocol in which highly deuterated proteins are prepared with proton labeling at key positions in the molecule. The goal is to produce molecules that retain a high enough level of deuteration so that assignment methodologies will be efficient, while maintaining a sufficient number of distance probes to facilitate structure determination. Methyls are attractive targets for selective protonation as they are enriched in hydrophobic cores and protein-protein interfaces (33). Furthermore, NMR spectroscopic properties of methyls are also favourable as a result of well-resolved $^{13}$C→$^2$H correlations and narrow linewidths due to rapid rotation about the methyl symmetry axis. We have developed an efficient and cost-effective method for selective protonation of the methyls of valine, leucine, and isoleucine ($\delta$1 only) of proteins overexpressed in E. coli within the context of a deuterated background (34). In this approach, $[3$-$^2$H],$^{13}$C-ketoisovalerate is used to supplement $^{13}$C,$^2$H-glucose, $^{15}$NH$_4$Cl, $^2$H$_2$O media to produce $^{15}$N,$^{13}$C,$^2$H-labeled proteins protonated at the methyl groups of valine and leucine. In addition, $[3,3$-$^2$H],$^{13}$C-ketobutyrate can be added, resulting in protonation of the isoleucine $\delta$1 methyls. High levels of protonation are obtained (>90%) and importantly the only protonated methyl groups that are generated are of the CH$_3$ variety (i.e., $^{13}$CHD$_2$ and $^{13}$CH$_3$D isotopomers are not obtained). Fig. 2A illustrates the distribution of protons within a sample of maltose binding protein (MBP, 370 residues) prepared using the labeling scheme described before, with labile amide protons indicated in white and methyl protons in gray. It is clear that methyl protonation leads to a significant increase in the numbers of distance probes that are available. In the case of MBP, the uniformly high level of deuteration at the C$\alpha$ and C$\beta$ positions in the protein has allowed the collection of triple resonance data with excellent resolution and sensitivity, facilitating rapid assignment of over 95% of backbone $^1$HN, $^{15}$N, $^{13}$C, and side-chain $^{13}$C$\beta$ nuclei (35). Finally, the high level of protonation at methyl positions of valine, leucine, and isoleucine ($\delta$1 only) residues enabled the site specific assignment of these groups using sensitive experiments optimized for methyl-protonated, highly deuterated proteins (35, 36).

**Structures of Large Proteins**

As described above methyl protonation of highly deuterated proteins greatly increases the number of distance restraints that are available for structural analyses relative to the case with perdeuterated molecules. In principle, HN→HN, HN-methyl, and methyl-methyl NOE correlations can be measured from methyl-protonated, $^{15}$N, $^{13}$C, $^2$H-labeled samples, with the quality of the resulting structures dependent on the number of NOEs, the topology and secondary structure content of the protein and the location and distribution of methyl groups in the protein (32). A number of NOE experiments have been developed that exploit the specific labeling pattern that is generated in these methyl protonated proteins. One such three-dimensional experiment takes advantage of the resolution in the $^{13}$C methyl dimension of spectra to produce an NOE data set in which cross-peaks are observed at (C$_\text{M}_{\text{CB}}$, C$_\text{M}_{\text{CM}}$, C$_\text{M}_{\text{HM},\text{M}}$), where protons of methyls i and j are proximal (37). This experiment also correlates methyls with nearby amide protons, since correlations of the form (C$_\text{M}_{\text{CM}}$, C$_\text{M}_{\text{CM}}$, C$_\text{M}_{\text{HM}}$) are generated as well. A second three-dimensional experiment gives rise to NOE correlations at (C$_\text{M}_{\text{CM}}$, C$_\text{M}_{\text{CB}}$, C$_\text{M}_{\text{HM}}$), where CB is the carbon atom adjacent to the methyl group carbon CM in residue i (38). These two experiments are supplemented with additional NOE data sets for measuring HN→HN (30, 31) and CH→HN (39) distances. Fig. 2B displays backbone traces of the 10 lowest energy structures (out of 100) of MBP in complex with a cyclic heptasaccharide, $\beta$-cyclodextrin, calculated on the basis of 1943 NOE, 464 dihedral angle, and 48 hydrogen bond restraints (40). Of the 1943 distance restraints, 769 and 348 were of the CH→HN (30, 31) and CH→CH$_3$ type, respectively. Methyl NOEs were instrumental in defining the structure of MBP as they comprise a significant number (77%) of long-range restraints, those between residues greater than four amino acids apart in primary sequence, and were useful in linking different elements of secondary structure (37, 40). It is clear, even at this level of structure refinement, that the protein is comprised of two domains, with the overall topology of each of the domains established. The N and C domains are reasonably well defined, with pairwise r.m.s.d. values of 2.3 Å and 3.3 Å, respectively, although the relative orientation of these domains is not well defined as indicated by the global pairwise r.m.s.d. of 3.8 Å.
Addition of Global Restraints

A significant limitation of dihedral angle-based structural restraints and many of the NOE restraints as well is that they provide information only on local molecular features. Dihedral angle restraints orient bonds proximal in sequence and NOEs are observed between two protons that are within approximately 5 Å of one another. However, many macromolecules and molecular complexes contain discrete domains or structural elements that are separated by distances greater than 5 Å and NOEs are therefore not available to connect these elements together.
In the case of the MBP-β-cyclodextrin complex, for example, few interdomain NOEs were observed and the relative orientation of the domains could not be well determined (40). Recently, new classes of restraints have been introduced, not strictly local in nature, that can be used to provide orientational information for two degree structural elements and domains in multidomain proteins in a manner that is independent of the distance between the modules.

One method for obtaining orientational information in proteins is based on measuring residual dipolar couplings between pairs of NMR-active nuclei (41, 42). The dipolar splitting that results from the interaction of two spin-1/2 nuclei i and j (i.e., 1H, 13C, or 15N in isotopically enriched proteins) is given by,

$$D_{ij} = \frac{\gamma_i \gamma_j \hbar}{2\pi r_{ij}^3} \langle 3 \cos^2 \theta - 1 \rangle$$  \[2\]

where $\gamma_i (\gamma_j)$ is the gyromagnetic ratio of spin i(j), $\hbar$ is Planck’s constant divided by $2\pi$, $\theta$ is the bond vector connecting the nuclei and the external magnetic field and the angular brackets $\langle \rangle$ indicate averaging over all allowed orientations of the bond vector. In cases where the molecule has no preferential orientation in solution (i.e., every orientation of the i-j bond vector is equiprobable) the factor $\langle 3\cos^2 \theta - 1 \rangle$ averages to zero and dipolar couplings are not observed. Proteins that possess nonzero magnetic susceptibility anisotropies will orient in an external magnetic field, producing a nonzero value for $\langle 3\cos^2 \theta - 1 \rangle$ that results in small splittings in spectra of the dipolar coupled nuclei. For proteins with magnetic susceptibilities of zero, molecular alignment can be imparted by dissolving the sample in an anisotropic medium, which itself orients in the magnetic field and causes alignment of the solute protein through steric or electrostatic effects (Fig. 2C) (4). This again leads to the observation of dipolar splittings (Fig. 2D). Many liquid crystalline systems have been introduced that induce molecular alignment of macromolecules. One of the most widely adopted, due to its ease of use, is Pf1 filamentous bacteriophage (43). The arrangement of the coat protein as a helical structure parallel to the long axis of the phage, which coats the phage DNA, contributes to the large anisotropic magnetic susceptibility, producing alignment of phage particles in the magnetic field (Fig. 2C). The degree of alignment of macromolecules dissolved in phage or in any other liquid crystalline media is tunable simply by increasing or decreasing the concentration of the alignment-inducing component. The desired degree of alignment is one that results in observable dipolar couplings only between nuclei close in space. Typically, an alignment of roughly 1 part in 1,000 is employed. Too much alignment results in an exponential increase in the number of observed dipolar splittings with a concomitant large decrease in sensitivity.

The most straightforward approach for incorporating dipolar couplings in structure calculations is by direct refinement of the orientation of bond vectors against the measured dipolar couplings. This approach has been successful in improving accuracy and precision of structures when used in conjunction with nearly complete NOE sets, J couplings, and chemical shift data (44). Improvements in structures have also been observed for several small proteins with well-defined topologies and limited NOE sets (45). Unfortunately, however, for each individual dipolar coupling there is an infinite number of orientations that are consistent with the data leading to a complex energy surface over which refinement occurs. Nevertheless, we have recently succeeded in developing a protocol within the molecular mechanics program CNS (46) that produces structures of proteins from minimal NOE restraint sets and dipolar couplings with good convergence properties. Using this approach we have calculated structures of MBP with dipolar couplings measured for the one-bond 1H–15N, 15N–13CO, and 13CO–13Cα interactions (47).

Orientational restraints can also be derived from the changes in chemical shift that occur upon alignment (23). Chemical shifts are a function of electronic environments that create local magnetic fields around a particular nucleus. As the molecule tumbles in solution, these fields change so that the chemical shift is an average over all molecular orientations. For molecules that are fractionally aligned certain orientations are more probable and the local fields are therefore averaged differently from what they would be in an isotropic solution, leading to changes in chemical shifts. Changes in chemical shifts can be incorporated into structure calculations in a manner analogous to the incorporation of dipolar couplings (47). Fig. 2E shows the ensemble of the 10 lowest energy structures of β-cyclodextrin-loaded MBP calculated with NOE, dihedral angle, and hydrogen bond restraints and with direct refinement against 1H–15N, 15N–13CO, 13CO–13Cα dipolar couplings and 13CO chemical shift restraints (47). The improvement in precision compared with structures generated from NOE, dihedral angle and hydrogen bond restraints only is apparent (Fig. 2, compare B and E). The pairwise root mean square deviations in NMR-derived structures of the N domain, the C domain, and the overall protein are reduced to approximately 1.5 Å, 2.2 Å, and 2.1 Å, respectively. In addition, there is a reduction in the r.m.s.d. between the N and C domains of the mean structure and the X-ray structure of MBP loaded with β-cyclodextrin (48) from about 3 to 2 Å, as well as a reduction in the global backbone r.m.s.d. between X-ray and NMR structures from 5 to 2.8 Å (47).

It is of interest that the agreement with the crystal structure is still better for the individual domains than for the entire protein. A major difference between the solution and X-ray structures is that the N- and C-domains in the calculated solution conformations of MBP are rotated towards each other by an additional 10° (47). Notably, the solution structures have been obtained using NMR-derived information exclusively. It is also possible, however, to start from the X-ray structure and to use the dipolar coupling to reorient the N- and C-domains, keeping the structures of the domains fixed. This process assumes that the solution and X-ray conformations of the individual domains
are unchanged, which is an excellent approximation in the case of MBP (35) and seems likely to hold for many other proteins as well. Using this procedure, we again find that there is $10^\circ$ closure of the $\beta$-cyclodextrin loaded form of MBP relative to the X-ray structure (49). In contrast, when this process is repeated for the apo and maltotriose-bound states of MBP very similar domain orientations are observed in solution and X-ray-derived structures (50).

An accurate description of the interdomain structure in multidomain proteins is of critical importance in understanding the biochemical processes that these molecules regulate. In the case of MBP, for example, the existence of an open (apo state) and
closed (maltodextrin-bound) conformation is critical for its role in the signaling cascade that regulates maltotriose uptake and chemotaxis \((48, 51)\) and only in the closed conformation is the protein able to bind its cognate receptor. Dipolar couplings have also been used to determine relative domain orientations in a number of different forms of bacteriophage T4 lysozyme and differences were also found between the solution and crystal structures \((52)\). These studies highlight the utility of dipolar coupling data in the determination of relative domain orientations in multidomain proteins and illustrate the importance of alternative methods to X-ray crystallography for obtaining this information.

Transverse Relaxation Optimized Spectroscopy, TROSY

The major limitation with extending NMR methods to increasingly larger molecules is the concomitant growth in the rate of signal decay. As discussed above the rate of signal loss can be decreased through the substitution of deuterons for protons which minimizes the amplitude of fluctuating local magnetic fields that, in turn, cause relaxation in the first place. These fluctuating magnetic fields can add or subtract from each other; signals that evolve in fields that are derived from the addition of all of the local magnetic interactions will relax more quickly than those experiencing smaller fields. In the TROSY approach, only the signal that relaxes slowly is selected, (i.e., the signal that sees the smaller net field, derived from subtraction of the local fields) providing significant improvements in spectral quality for large proteins \((53)\).

We have exploited the TROSY principle in the development of 3 four-dimensional experiments for backbone chemical shift assignment of high molecular weight proteins \((54, 55)\). These four-dimensional experiments correlate four chemical shifts at once, such as the \(^{13}C\alpha\) \(^{13}CO\), \(^{15}N\), and \(^1HN\) shifts of residue \(i\) [4D TROSY-HNC\(\bar{CO}\)], the \(^{13}C\alpha\) and \(^{13}CO\) shifts of residue \(i–1\) with the \(^{15}N\), \(^1HN\) chemical shifts of residue \(i\) [4D TROSY-HNCO\(\bar{C}α\)], and the \(^{13}CO\) shift of residue \(i–1\) with the \(^{13}C\alpha\), \(^{15}N\), and \(^1HN\) shifts of residue \(i\) [4D TROSY-HNCO\(\bar{C}α\)]. An illustration of the use of these experiments for the sequential assignment of proteins is presented in Fig. 3A. Starting with the 4D TROSY-HN cancellation provided a basis for excluding the proposed allosteric model for regulation of p53 \((57)\). Chemical shift assignment using the 4D TROSY experiments provided a basis for excluding the proposed allosteric model for regulation of p53 \((57)\).

These experiments were used to obtain an almost complete (>95%) backbone \(^{1HN}, \(^{15}N, \(^{13}CO, \(^{13}C\alpha\) and side chain \(^{13}Cβ\) chemical shift assignment of an \(^{15}N, \(^{13}C, \(^2H\)-labeled sample of a 67-kDa dimeric form of p53 consisting of the DNA-binding and oligomerization domains of the molecule \((56)\). (Dimeric p53 is created by mutations in the oligomerization domain and has similar DNA-binding activity to tetrameric p53.) p53 is a tumor suppressor that regulates cell arrest or apoptosis in response to DNA damage by acting as a transcription factor that activates genes involved in cell cycle control \((57)\). Mutations in p53 are found in over half of human cancers, thus prompting structural studies of wild-type and mutant p53. Although structures have been determined for the individual DNA-binding \((58)\) and oligomerization \((59, 60)\) domains, an understanding of the regulation of p53 function by the N-terminal transactivation and C-terminal regulatory domains requires studies on larger fragments.

To investigate the potential interactions between regions of p53, HNCO\(\bar{C}α\) spectra of the dimeric p53 fragment with (latent p53) and without (active p53) the C-terminal regulatory domain were compared. Negligible chemical shift differences were observed between these molecules, proving that the C-terminal regulatory domain does not affect the conformation of the DNA-binding domain, the oligomerization domain or the intervening linker. In addition, comparison of the \(^1HN\)–\(^{15}N\) correlation spectrum of active p53 with spectra of the oligomerization or the DNA-binding domains (Fig. 3B, C) confirms the modular and independent nature of these domains \((61)\). Chemical shift assignment using the 4D TROSY experiments provided a basis for excluding the proposed allosteric model for regulation of p53 \((57)\).

CONCLUSIONS

NMR spectroscopy is a powerful method for obtaining structural information on macromolecules and molecular complexes. Technological, spectroscopic, and labeling advances have greatly improved the quality of NMR data that can be recorded on biomolecules, have extended the molecular weight limit, and have significantly increased the type of information that is available. Continued advancements will ensure that NMR remains well poised to make further important contributions in structural biology.

ACKNOWLEDGEMENTS

V.K. is a recipient of an Ontario Graduate Fellowship. J.D.F.-K. and L.E.K. acknowledge support from the National Cancer Institute of Canada (NCIC) and the Canadian Institutes of Health Research (CIHR), respectively. L.E.K is an International Research Scholar of the Howard Hughes Medical Institute.

REFERENCES


