

Solution Nuclear Magnetic Resonance Spectroscopy

James J. Chou and Remy Sounier

Abstract

Solution nuclear magnetic resonance (NMR) spectroscopy has come a long way in characterizing the structure and function of biological molecules since the first one-dimensional spectrum of protein was recorded about 30 years ago. To date (September 1, 2012), there are 9,521 solution NMR structures in the Protein Data Bank, compared to 74,009 determined by crystallographic methods. Unlike X-ray and electron microscopy (EM) methods, which are based on the concepts of Fourier optics and image reconstruction, structure determination by NMR involves measuring structural restraints and finding structural solutions that satisfy the restraints. Although the NMR approach is much less direct in a physical sense, it has proven itself over the years to be capable of de novo structure determination at high precision. Moreover, the method is highly versatile and can be used in a variety of ways for addressing mechanistic questions. NMR measurements of protein internal dynamics and protein–protein or protein–ligand interaction are directly relevant to function in vivo because the molecules are often in physiological buffer conditions. The method can also be applied to investigate protein-folding intermediates, conformational changes, as well as intrinsically unfolded proteins. Recently, along with X-ray and EM, solution NMR has entered a state of rapid growth for structural studies of membrane proteins, already demonstrating its feasibility in de novo structure determination of membrane-embedded ion channels and receptors. As the hardware advances rapidly, especially in cryogenic probes that have much higher sensitivity, the sample concentration required for solution NMR investigation is decreasing, hopefully soon to a concentration level at which nonspecific protein aggregation is no longer an issue. After three decades of improvement in spectrometer technology, NMR pulse experiments, isotope labeling schemes, and structure determination software, we believe that solution NMR will truly enter the production phase in the next decade to answer biological questions of high impact, and to become more versatile than ever in complementing X-ray and EM in investigating protein structure and function.

Key words: NMR, Biomolecular application, Protein structure, Protein dynamics

1. Introduction

1.1. Solution NMR in Structure Determination of Biological Molecules

After Bloch and Purcell first observed the resonance spectrum of paraffin in solid phase (1, 2) and water in liquid phase (1, 2), liquid-state nuclear magnetic resonance (NMR) spectroscopy rapidly became an indispensable tool for characterizing small molecules in

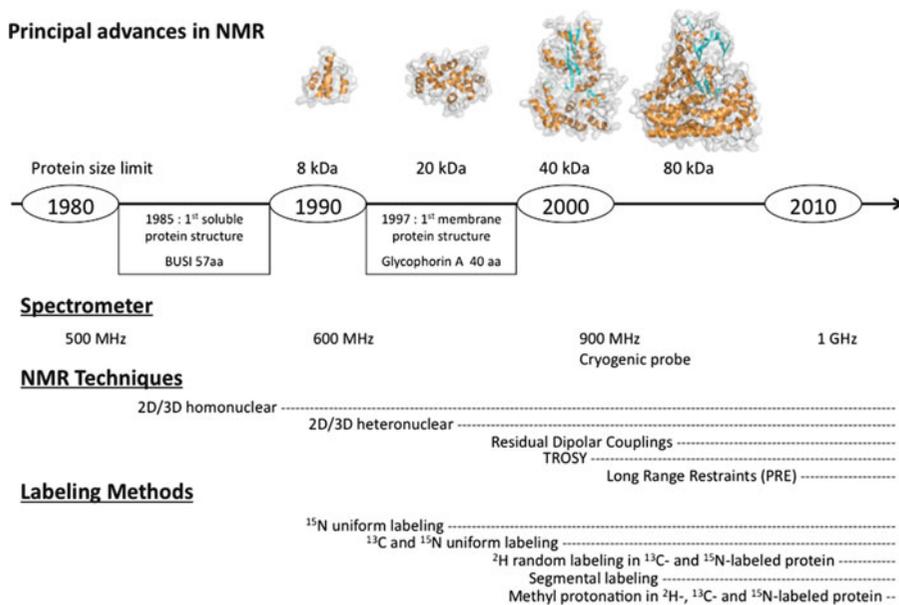


Fig. 1. A timeline summarizing the advances in NMR hardware, spectroscopic methods, and isotope labeling methods.

chemistry laboratories, and is now universally used in structural characterization of larger biomolecules such as proteins and nucleic acids as well as their complexes. A timeline that briefly describes the milestones in biomolecular NMR is shown in Fig. 1.

The obvious advantage of solution NMR is that biomolecules tumble freely in the more native aqueous environment, not having to form crystals. But the requirement for molecular tumbling in solution also happens to pose the major limitation of solution NMR. In NMR spectroscopy, the instruments generally excite spin $\frac{1}{2}$ nuclides and detect their chemical shift evolution, with proton spin being the most popular one owing to its natural abundance and large gyromagnetic ratio (γ). For other biological atoms like carbon, nitrogen, and oxygen, the natural abundance of spin $\frac{1}{2}$ nuclides is extremely low and isotope labeling is usually required for them to be NMR visible. The nuclear spins all have magnetic dipole moments and can thus interact with each other when they are close in space. The interaction between a pair of dipole moments is measured as the dipolar coupling constant, which is proportional to $3\cos^2\theta - 1$, where θ is the angle between the internuclear vector and the static magnetic field. When molecules are in solid phase, a spin experiences very strong dipolar couplings from others in the molecule (e.g., 0.1–50 kHz for ^1H), resulting in multiple levels of resonance splitting—the cause of low-resolution spectra of solids. These dipolar couplings, however, completely cancel out if the molecules in solution tumble fast enough. Proteins of average size (~ 40 kDa) reorient randomly on the nanosecond timescale, so fast that chemical-shift evolution, which occurs on the millisecond

timescale, does not feel the dipolar interactions. Therefore NMR peaks of proteins in solution are sharp and generally have the perfect Lorentzian line-shape. Although dipolar coupling is averaged to zero on the timescale at which NMR peaks are recorded, the physical quantity of dipole–dipole interaction is still present and has a profound effect on the decay of spin coherence (generally described by the transverse relaxation rate R_2). The coherence relaxes faster for larger or slower tumbling molecules, thus setting a theoretical size limit of molecules that can be studied by solution NMR.

It is difficult to predict what exactly is the size limit because new methods are constantly being developed to push the size limit. For example, protein deuteration, or replacing the non-exchangeable aliphatic protons with deuterium, can dramatically slow down the transverse relaxation because deuterium has a sixfold smaller dipole moment than a proton. Furthermore many relaxation-optimized NMR experiments have been developed, such as the ^1H – ^{15}N TROSY HSQC (3) and the methyl ^1H – ^{13}C TROSY HMQC pulse schemes (4). The size of proteins amenable to solution NMR has increased from 8 kDa in 1990 to 82 kDa today (Fig. 1), and few in the field 20 years ago could have predicted such rapid progress. In addition to the size limit, resonance complexity is also a serious problem. Typical chemical shift dispersion of a protein is about 6, 30, and 20 ppm for $^1\text{H}_\text{N}$, ^{15}N , and methyl ^{13}C , respectively. Increasing the number of peaks will inevitably result in greater resonance overlap that prohibits unambiguous assignment of protein resonances.

Structure determination by NMR uses an approach that is fundamentally different from that of diffraction methods such as X-ray and EM. Whereas the diffraction methods involve reconstructing images from diffraction data, NMR spectroscopy measures structural constraints (distances or angles) and finds a structural solution that is consistent with those experimental constraints. Therefore, the accuracy and precision of NMR structures depend strongly on the number of unambiguous restraints that can be collected for each residue. In general, the number of NMR restraints measurable for larger or more complex systems is smaller due to greater resonance overlap and lower spectral intensity, and thus precision (or resolution) is lower. The same is true for proteins that yield low-quality NMR spectra. Therefore, for proteins or nucleic acids that readily form well-diffracting crystals, it is usually more efficient to determine the structure by X-ray or electron crystallography.

1.2. Solution NMR in Functional Investigation of Biological Molecules

There are advantages of establishing an NMR system for a macromolecule even if its high-resolution structure has been determined by X-ray or electron crystallography, for further investigation of molecular interactions and dynamics. For example, once sequence-specific assignment of NMR resonances is achieved for a protein, it

is very convenient to investigate its interaction with ligands or other proteins. By comparing the chemical shift of the assigned protein resonances with and without the bound ligand, one can quickly map the binding site onto the protein of known structure. This technique is commonly known as the *chemical shift perturbation* assay. Solution NMR can also be used to detect very weak interactions (with $K_D \sim 1$ mM), so weak that protein–protein or protein–ligand complexes cannot be isolated by gel filtration for crystallization trials. In those cases of weak interaction, the binding site can still be mapped using methods such as the *saturation transfer difference* method (5), in which nuclear spin states of the ligand are selectively saturated while observing dipolar cross relaxation that affects the resonances of the protein. Alternatively, one can introduce a paramagnetic label to the ligand and measure paramagnetic broadening of the protein resonances, or vice versa. These techniques are now routinely used for studying protein–protein and protein–ligand interactions (6–8).

Solution NMR is uniquely suited for characterizing the internal dynamics of a protein that is related to its function. Although high-resolution crystal structures also contain temperature factors that can be used to infer dynamics, an aspect of dynamics only visible to NMR is the timescale of protein internal motion. As described above, relaxation of NMR signals depends strongly on protein dynamics in solution. The dynamics are a combination of the overall molecular tumbling and the internal motion of the structural segments, and motions of various frequencies contribute differently to different types of relaxation processes. Longitudinal relaxation rate (R_1) is the rate at which the excited spin-state population returns to Boltzmann equilibrium and is dominated by the fast motions (nanosecond timescale; e.g., the rotational correlation time of a 30 kDa globular protein at 25°C is around 25 ns). The transverse relaxation rate (R_2) is the rate of dephasing of spin coherence and it is dominated by the slow motions. R_2 also has a strong contribution from chemical exchange (R_{ex}). If two conformational or chemical states have different chemical shift, exchange between them on a millisecond to microsecond timescale usually leads to substantial dephasing of spin coherence, thus effectively increasing R_2 . Since most protein functional switches occur in this timescale, measurement of R_{ex} has become the most important aspect of dynamics studies by NMR.

One of the best examples of NMR dynamics measurements that led to answering important mechanistic question is the R_1 and R_2 measurement of calmodulin, a ubiquitous calcium (Ca^{2+}) sensor protein in cells. The crystal structure of Ca^{2+} -bound calmodulin shows a dumbbell-like structure in which the N- and C-terminal Ca^{2+} -binding domains are connected by a rigid helix (known as the central helix) (9), but the structure of the Ca^{2+} -calmodulin bound to the peptide from the smooth muscle myosin light-chain kinase

(smMLCK) shows that the central helix is a loop, and that the N- and C-domains come together to wrap around the smMLCK helix (10). Dissolving the central helix to accommodate a ligand is an energy costly process and it was not clear why evolution selected such a mechanism. Measurement of ^{15}N R_1 and R_2 of Ca^{2+} -calmodulin showed that the central helix observed in the crystal is largely a flexible loop in solution (11). Therefore, the rigid central helix in the crystal structure likely represents only a small population of the conformer that was stabilized by crystal packing. More recently, NMR dynamics measurements of side-chain methyl groups have been demonstrated even for very large protein machineries such as the 650 kDa 20S core-particle proteasome, which identified dynamic residues inside the antechamber that facilitate the movement of substrates to the sites of proteolysis (12). The conventional R_1/R_2 measurement and analysis can only provide information on the very fast dynamics (nanosecond to picosecond), but many proteins undergo conformational switching on the millisecond to microsecond timescale. Chemical or conformational exchange in this frequency range strongly affects chemical shift evolution of nuclear spins (known as exchange broadening of NMR resonances), and this effect has been utilized by spectroscopists for extracting millisecond to microsecond internal dynamics of interesting protein systems. The exchange contribution to R_2 can be measured using the relaxation-compensated Carr–Purcell–Meiboom–Gill (CPMG) experiment (13, 14). A good example of this application is characterizing the rate of conformational switching of the enzyme cyclophilin A (15, 16). In another interesting study, the CPMG method was used to observe the opening or “unlocking” of the channel gate during proton conduction by the M2 proton channel of influenza virus at low pH (17).

2. NMR Structural Restraints and Structure Determination

2.1. Distance Restraints from ^1H – ^1H NOE

Distance restraints derived from the ^1H – ^1H nuclear overhauser effect (NOE) are the key restraints used in structural studies by solution NMR. The Overhauser effect was named after the physicist Albert Overhauser, who showed theoretically that spin polarization of electrons could be transferred to the nuclear spins via dipole–dipole cross relaxation through space (18). This effect is the basis of the dynamic nuclear polarization (DNP) technology that aims to increase the sensitivity of NMR signals by passing the strong electron spin magnetic moment to nuclear spins (19–21). In solution NMR, the NOE is generally detected between different ^1H nuclear spins in a molecule. After inverting two spin populations A and B so that they are away from equilibrium, the two spin populations do not relax toward equilibrium (governed by

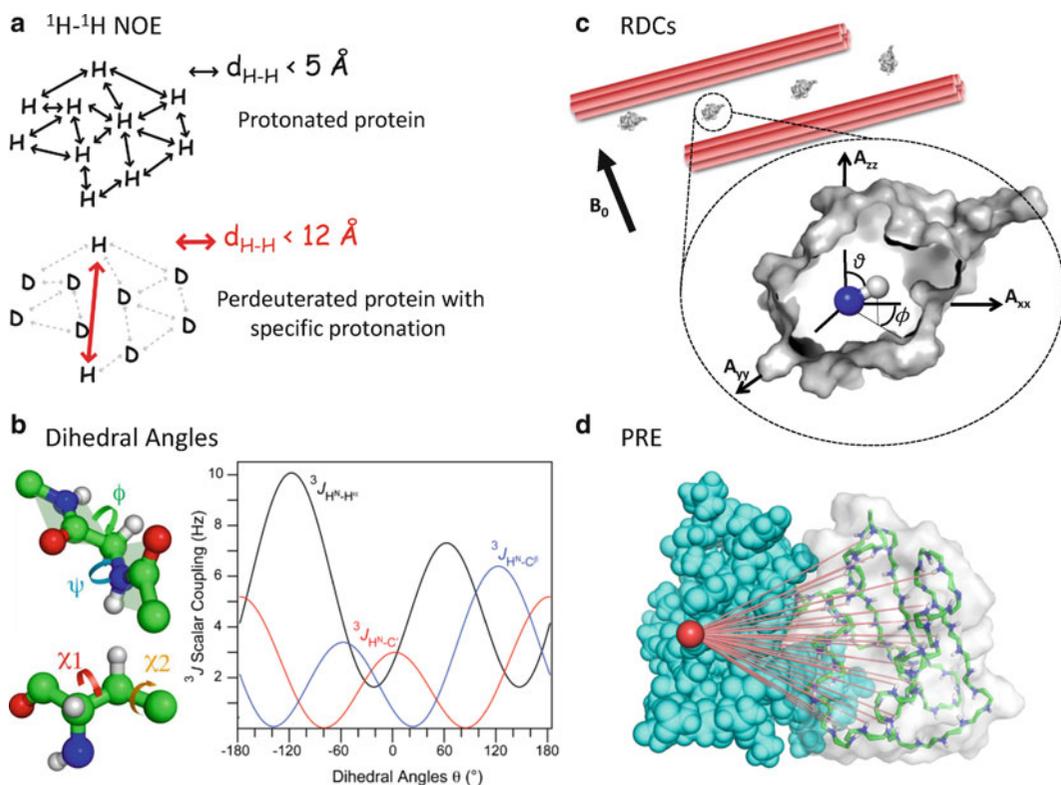


Fig. 2. Schematic illustration of the key structural restraints derived from NMR measurements, including (a) inter-proton distances from NOEs, (b) backbone and side chain dihedral angles from 3-bond J couplings, (c) internuclear vector orientation from dipolar couplings, and (d) long-range distances from PREs.

R_1) independently. If spins A and B are close in space, their magnetic dipoles interact with each other, giving rise to a cross relaxation term that transfers spin polarization between the two spin populations. This cross relaxation rate is dependent on the interatomic distance (r), and is proportional to $1/r^6$. The amount of NOE transfer depends on both the cross relaxation rate and the time of NOE mixing applied in the NMR experiment. Therefore, similar to the fluorescence resonance energy transfer (FRET), NOE is a very steep function of the distance. Using normal NOE mixing time (~ 100 ms), it is only detectable when two protons are within 5 \AA (Fig. 2a). Using a long mixing time makes the NOE distance measurement less quantitative because other protons nearby can mediate spin diffusion. Longer distance NOE can however be achieved by selectively protonating two nuclear spin probes in an otherwise perdeuterated environment (because deuteration largely removes spin diffusion) (Fig. 2a). This approach has been used to measure NOEs between protons that are separated by as much as 12 \AA (22).

NOEs can be measured using a variety of multidimensional experiments. One of the most powerful NOE experiments in protein NMR spectroscopy is the 3D or 4D ^{15}N -edited NOESY used for selectively observing NOEs between backbone amide protons and between amide protons and aliphatic protons (23, 24). Despite the upper bound of $\sim 5 \text{ \AA}$ for conventional NOE detection, the distance restraints from this experiment are very effective in determining protein secondary structures. For example, in a β -sheet, the inter-strand $\text{H}_\text{N}-\text{H}_\text{N}$ and $\text{H}_\text{N}-\text{H}_\alpha$ distances are $< 4 \text{ \AA}$ and yield measurable NOEs. Within an extended β strand, there is also very strong NOE between H_N of residue i (H_N^i) and H_α of residue $i-1$ (H_α^{i-1}) ($\sim 2.2 \text{ \AA}$). The distances that give rise to characteristic NOEs in an α helix are between H_N^i and H_α^{i-3} , and between H_N^i and H_α^{i-1} (both at $\sim 3.5 \text{ \AA}$). Assigning tertiary distance restraints, e.g., those between two helices, is much more challenging. The inter-helical distances between backbone H_N and aliphatic protons are significantly longer and therefore it is usually necessary to assign NOEs between amino acid side chains such as the methyl and aromatic groups. For larger proteins, assignment of side chain resonances can be challenging due to higher spectral complexity. Table 1 lists the type of NOE restraints important in various types of protein structures.

2.2. Dihedral Angles from Chemical Shift and Scalar Coupling Constants

Chemical shift, or the frequency at which nuclear spins evolve under local magnetic field, depends on the local electronic and structural environment of the molecule. A complete understanding of how structure determines chemical shifts would in principle allow accurate prediction of macromolecular structures based on chemical shift values alone. In protein NMR spectroscopy, assigning residue-specific chemical shift of backbone ^{15}N , $^{13}\text{C}_\alpha$, $^{13}\text{C}'$, $^{13}\text{C}_\beta$, $^1\text{H}_\alpha$, and $^1\text{H}_\text{N}$ is relatively straightforward, even for proteins as large as 82 kDa (25). Thus the notion of structure determination using chemical shift values is highly attractive and has been intensely pursued (26–29). Unfortunately, this approach is not yet feasible because the chemical shifts of backbone $^1\text{H}_\text{N}$, ^{15}N , and $^{13}\text{C}'$ are very sensitive to buffer conditions such as pH and ionic strength, and to structural factors such as hydrogen bonding geometry, hydration, and intrinsic dynamics of molecules. However, the chemical shifts of some nuclides, such as $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, and $^1\text{H}_\alpha$, are not so sensitive to the above factors, and are largely determined by the local structure. Statistical analysis showed that their chemical shifts departing from the random coil values are strongly correlated to backbone dihedral angles of known secondary structural elements (30, 31). The empirical relation between chemical shift and backbone torsion angle has been implemented in programs such as TALOS/TALOS+ (26, 32) and DANGLE (33), which are widely used today in protein NMR for identifying regions of secondary structures. The procedure for assigning the secondary structure of proteins

Table 1
Typical NOEs observed in structure determination

NOE restraints	Distance (Å)	Types of structure
$H_N^i-H_N^j$	2.3	β -strand
$H_N^i-H_\alpha^j$	3.2	β -strand
$H_\alpha^i-H_\alpha^j$	3.2	β -strand
$H_N^i-H_N^{i-1}$	2.8/4.3	α -helix/ β -strand
$H_N^i-H_N^{i-2}$	4.2	α -helix
$H_N^i-H_\alpha^{i-1}$	3.5/2.2	α -helix/ β -strand
$H_N^i-H_\alpha^{i-2}$	4.4	α -helix
$H_N^i-H_\alpha^{i-3}$	3.4	α -helix
$H_N^i-H_\alpha^{i-4}$	4.2	α -helix
H_N-H_{aromatic}	1.8–6	Long-range
H_N-H_{methyl}	1.8–6	Long-range
$H_{\text{methyl}}-H_{\text{methyl}}$	1.8–12	Long-range
$H_{\text{methyl}}-H_{\text{aromatic}}$	1.8–6	Long-range

based on chemical shift differences with respect to random coil values is generally known as Chemical Shift Index (CSI).

A more quantitative method of obtaining dihedral angles in NMR is measuring the 3-bond scalar coupling constants (J). The J coupling is indirect dipole–dipole interaction between two nuclear spins that is mediated by the bonding electrons. For two atoms separated by three chemical bonds, the J coupling between them (3J) is a function of the dihedral angle and is given by the Karplus relation ${}^3J = A \cos^2\phi + B \cos\phi + C$, where ϕ is the dihedral angle and A, B, and C are constants (sometimes referred to as the Karplus parameters) (Fig. 2b). The A, B, and C constants have been characterized for many different types of dihedral angles in proteins and nucleic acids; they are usually determined semi-empirically by fitting experimental 3J values to ϕ in high-resolution crystal structures (reviewed in ref. 34).

In the early days of protein NMR spectroscopy, the 3J between backbone 1H_N and ${}^1H_\alpha$ (${}^3J_{HNH\alpha}$) and between the backbone ${}^1H_\alpha$ and the side chain ${}^1H_\beta$ (${}^3J_{H\alpha H\beta}$) were commonly measured for the ϕ and χ_1 dihedral restraints, respectively (35, 36). Although these coupling constants could be conveniently measured for small proteins with slow R_2 , they are difficult to obtain for larger proteins due to the requirement for a long lifetime of spin coherences in the

J-coupling experiments. For larger proteins, the backbone ϕ and ψ are commonly estimated using the CSI method, whereas side chain χ rotamers are extracted from 3J between heavy atoms with a favorable relaxation property. For example, the side chain χ_1 rotamers of Val, Ile, and Thr can be extracted from $^3J_{C'_{\alpha}C_{\beta}}$ and $^3J_{NC_{\beta}}$ coupling constants (37, 38). The χ_2 rotamers of Leu and Ile residues can be extracted from $^3J_{C_{\alpha}C_{\delta}}$ coupling constants (39). These coupling constants can be measured using the protein methyl groups as the NMR readout probes because methyl groups give strong signal and generally have slower R_2 compared to other groups in the protein. For amino acids that do not have methyl groups at γ or δ position, $^3J_{NC_{\gamma}}$ couplings can be measured using deuterated proteins using the backbone amide as the readout nuclide (40, 41).

2.3. Orientation Restrictions from Residual Dipolar Couplings

In a macromolecule marginally oriented in a magnetic field, measurable dipolar coupling between a pair of spin 1/2 nuclides encodes orientations of the internuclear vector connecting the two atoms (42). The orientations are defined relative to a common reference frame, known as the principal axes of the molecule's alignment tensor (43) (Fig. 2c). Hence, dipolar couplings provide the powerful global orientation restraints that had been absent in the traditional NOE-based structure determination. In solution NMR, dipolar interactions are essentially cancelled by random rotational diffusion of macromolecules on the nanosecond timescale, though there are a few isolated cases, e.g., paramyoglobin, in which the protein's own susceptibility can give rise to a few Hz of 1H - ^{15}N dipolar couplings (44). Most proteins however do not have large enough magnetic susceptibility anisotropy to be preferentially oriented by the magnetic field, and therefore must be forced to align by either physical interaction with large oriented particles or by fusion with a weakly aligned paramagnetic tag. The selection criteria in choosing an alignment medium are stringent. It must be compatible with most biological molecules in water. It must be highly ordered in the magnetic field and must be large enough so that collisions with proteins do not alter its orientation. Finally, for introducing sizable dipolar couplings for structure determination without introducing too much 1H - 1H dipolar broadening of NMR resonances, macromolecules typically need to acquire an order parameter of $\sim 10^{-3}$ (equivalently, a molecule is aligned for 0.1% of time).

About 10 years ago, it was first demonstrated for proteins that such order can be achieved in ~ 4 – 5 wt% DMPC/DHPC liquid crystalline medium, and that in the case of ubiquitin, optimal-sized 1H - ^{15}N residual dipolar couplings (RDCs) of about ± 20 Hz can be measured (45). Over the past decade, spectroscopists have developed a number of different liquid crystalline media; each of them is suitable for aligning a particular type of proteins. These media include filamentous phage virus particles that form liquid crystals

above a certain concentration (46, 47), the ternary mixtures of cetylpyridinium chloride (CPCl) or bromide (CPBr), hexanol, and NaCl or NaBr in water (48, 49), and the binary mixture of alkylpoly(ethylene glycol) and hexanol in water (50). Among them, the rodlike filamentous phage Pfl is the most widely used because its uniform length and rigid assembly yield stable liquid crystal in concentration as low as 12 mg/ml.

Liquid crystals have delicate phase transitions and usually disintegrate under extreme sample conditions, such as high temperature, low pH, and presence of anionic detergent. A method of weakly orienting biological molecules without the use of liquid crystals is Strain-induced Alignment in polyacrylamide Gel (SAG) in polyacrylamide Gel (SAG) (51, 52). The advantage of polyacrylamide gel is obvious because the chemically cross-linked polymers can withstand very harsh sample conditions. The SAG method has been successfully applied in studies where the above liquid crystals are not applicable. These studies include measuring RDCs for Staphylococcal Nuclease partially denatured in 8 M Urea for studying folding intermediates (53) and measuring RDCs of membrane-associated proteins in the presence of high concentrations of detergent (54, 55). However, a fundamental problem of the SAG method is that it is difficult to soak high concentrations of protein into the gel because the gel pore size has a broad normal distribution. For larger protein-detergent or protein-protein complexes, only a small fraction of the pores can accommodate the macromolecules without obstructing their rotational diffusion. It is therefore more challenging to collect high-quality RDCs using the SAG method. The problem of measuring RDCs for membrane proteins is partially solved owing to the emergence of DNA-based liquid crystals that are resistant to detergents. One liquid crystal is formed with fabricated DNA nanotubes (56). The other medium is the liquid crystal of G-Tetrad DNA (57).

RDCs are extracted by subtracting J couplings acquired with the regular sample from the $J + D$ couplings of the weakly aligned sample. These couplings are typically measured using ^{15}N -, ^{13}C -, and 85% ^2H -labeled protein for decreasing ^{15}N and $^{13}\text{C}^\alpha$ R_2 and for reducing adverse ^1H - ^1H dipolar interactions in the aligned sample. For medium-sized proteins smaller than 50 kDa, it is usually possible to acquire high-quality RDCs for backbone chemical bonds H_N -N, C' - C_α , and C' -N using the HNC0 triple-resonance experiments (58-60). Probably the biggest advantage of RDC restraints is that they are completely quantitative and unambiguous. Being quantitative means the one-bond RDC value has a clean mathematical relation to the orientation of the chemical bond. RDC values are unambiguous because their assignment is based on the sequence-specific assignment of backbone resonances, which is usually very reliable in modern protein NMR. Therefore, RDC values can be used as numerical data for structure determination,

similar to diffraction in X-ray crystallography. An active question is what is the most effective way to use RDCs in structure determination. Another shortcoming of RDCs is that they do not provide translational information of the corresponding chemical bonds and thus need to be used in combination with distance restraints for de novo structure determination.

2.4. Long-Range Distance from Paramagnetic Relaxation Enhancement Measurements

A good complement to the short-range ^1H - ^1H NOEs is the long-range distance restraint derived from paramagnetic relaxation enhancement (PRE) measurements (61). Paramagnetic centers typically consist of one or more unpaired electrons, which have a very strong magnetic dipole moment. The interaction between the strong electron dipole and other nuclear spin dipoles in the molecule can strongly affect the relaxation rates of the nuclear spin populations. Similar to the rate of spin polarization transfer via NOE, the PRE of nuclear spin is proportional to $1/r^6$, where r is the distance between the nuclear spin and the paramagnetic center (Fig. 2d). The most commonly used paramagnetic labels in structural studies of proteins and nucleic acids are nitroxide compounds that have one unpaired electron. For proteins, the nitroxide compound often used is *S*-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL), which can be attached at cysteine positions in the protein by reacting with the thiol group of cysteine. The protein should only have one cysteine at a time to ensure that PRE-derived restraints are unambiguous. Similar nitroxide compounds have been used for spin labeling of DNA and RNA. For example, the MTSL-like compound 3-Iodomethyl-1-oxyl-2,2,5,5-tetramethylpyrroline can be attached to a phosphorothioate that is substituted at a specific backbone position during chemical synthesis of DNA or RNA (62).

The strong electron magnetic dipole moment ($\gamma_e \sim 660$ times that of ^1H) can broaden NMR line-width of nuclides much farther away (up to ~ 25 Å), but can also wipe out resonances of nuclides nearby (< 12 Å). Therefore, the range of PRE-derived distances that can be quantified is 12–25 Å. Probably the biggest advantage of PRE over NOE, if the sequence-specific resonance assignments are known, is that the measurement and analysis are both simple and unambiguous. PRE measurement typically involves measuring (using either 2D or 3D experiments) ^1H $R_2 + R_2^{\text{para}}$ for ^1H resonances that are broadened by a particular spin label and ^1H R_2 after reducing the unpaired electron with ascorbic acid. R_2^{para} is then used to derive distance restraints based on the known calibrations (61). For resonances that are completely wiped out by the spin label, the distance is set to < 12 Å. Therefore PRE restraints are unambiguous if the sequence-specific resonance assignments are unambiguous. The weakness of PRE is also glaring. The nitroxide spin label is ~ 8 Å long and is flexible like the arginine side chain, and thus PRE restraints have very large uncertainty, typically ± 5 Å.

It is therefore not useful for determining local or secondary structures. Moreover, introducing an unnatural spin label could affect protein structure or function. Nonetheless, a large number of unambiguous PRE restraints can compensate for their low precision. With existing knowledge of the local structures from NOE and RDCs, PRE restraints are useful in providing a low-resolution global fold that can facilitate further NOE assignments. Another application of PRE mentioned above is mapping protein–protein or protein–ligand interactions that are too weak or complex to be measured by intermolecular NOEs. If the structures of two interacting proteins were known, it would be easy to identify positions in the proteins for attaching spin labels.

2.5. Structure Calculation

The objective of structure calculation in NMR is to find structural solutions that are consistent with all experimental restraints while not violating the standard covalent geometry of protein or nucleic acid polymers. All NMR restraints have experimental uncertainties and are included in structure calculation as ranges of allowed values (e.g., lower and upper bounds). Some NMR restraints are even ambiguous. It is thus important to calculate an ensemble of structures to account for the errors and ambiguity in the restraints. A number of methods have been developed for this purpose. In the early days of protein NMR, when computers were not so powerful, algebraic methods such as Distance Geometry (DG) were used. The first step of the DG method is extrapolating, using geometric inequality limits, a complete set of lower and upper limits on all the interatomic distances from the sparse set of experimental distance restraints. The next step is to choose a random distance matrix from within the complete set of experimental and extrapolated distances, and fit a set of atomic coordinates to it using the EMBED algorithm. By repeating this procedure with different random distance matrices, one obtains an ensemble of conformations that is consistent with experimentally derived restraints. The first *de novo* structure determination by NMR was achieved using the DG method (63). The advantage of the DG method is that it is rather deterministic. For example, in the EMBED algorithm, coordinates that are a best-fit to the distance matrix are calculated by eigenvalue methods, completely avoiding the local minima problem. Probably the biggest weakness of the DG method is that the approach relies almost completely on distance restraints and thus is difficult to be implemented to include other NMR data such as dipolar couplings, which encode bond orientation information, and chemical shifts, which encode dihedral angle information.

A more generally used structure calculation method is the restrained Molecular Dynamics (rMD) method (64, 65). Although this method is more computationally intensive, it can be used to minimize variable target functions of any type and is thus readily applicable for refining structures against different types of

NMR-derived restraints. The Molecular Dynamics (MD) part of the method involves numerically solving Newton's equations of motion for a many particle system in which the total potential energy is the sum of physical potentials (such as those for chemical bonds, angles, van der Waals, etc.) and pseudo potentials from experimental restraints (such as those for interatomic distances, dihedral angles, bond orientations, etc.). For macromolecules, the refinement energy landscape becomes very complicated, posing serious local minima problems. Therefore, the rMD calculation is done in combination with simulated annealing (heating the system followed by gradual cooling), an effective procedure for "jumping" out of false local minima (66, 67).

3. Solution NMR Studies of Membrane Proteins

Structure determination of membrane proteins by solution NMR is still in the exploratory phase. In principle, the NMR protocols that have been established for water-soluble proteins should be directly applicable to membrane proteins. These protocols however need to be tailored to account for the fundamental physical chemical differences between membrane proteins and water-soluble proteins and the imperative for use of a model membrane media. There are a number of issues that complicate solution NMR studies of membrane proteins. (1) Membrane proteins need to be solubilized in detergent micelles or detergent/lipid bicelles. We do not understand exactly how various types of detergents assemble a micelle around membrane proteins and thus cannot predict the effective size of a protein-micelle complex. Furthermore, the presence of very high concentrations of detergent requires methods to suppress NMR signals arising from the detergent. (2) Amino acid sequences of membrane proteins have been optimized in nature in the membrane environment, and thus it is unclear whether detergent micelles fully mimic the lateral lipid pressure that a protein experiences in a true membrane. For α -helical membrane proteins, insufficient lipid pressure in detergent micelles could result in weaker helix-helix packing or increased internal "breathing" in solution. The internal dynamics would pose a problem for measuring long-range NOEs. (3) In general, membrane proteins contain many more methyl-bearing amino acids than water-soluble proteins. Moreover, most of the hydrophobic residues are exposed to the dynamic detergent, unconfined to a unique chemical environment. These properties of membrane proteins result in a much smaller chemical shift dispersion of the methyl groups as compared to water-soluble proteins (in which hydrophobic residues are strongly packed in the protein core). The poor chemical shift dispersion poses a big

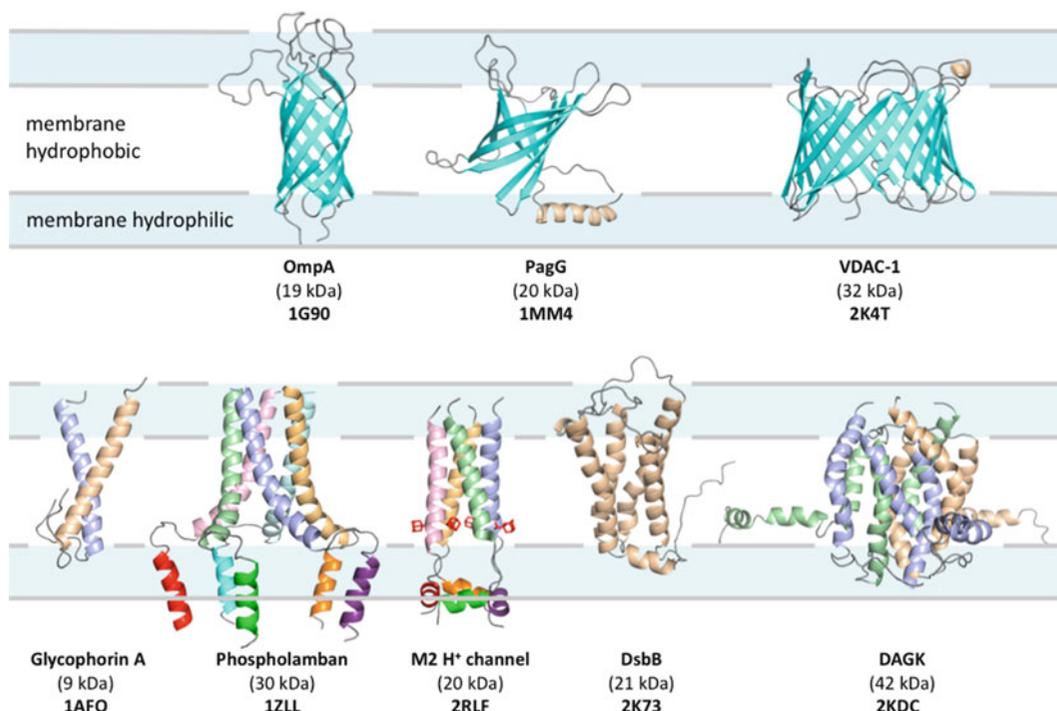


Fig. 3. Representative membrane protein structures determined using solution NMR spectroscopy. Below each ribbon structure is the protein name, molecular weight, and PDB code. On a scale relative to protein, the hydrophobic and hydrophilic regions of the membrane are ~ 30 Å and ~ 15 Å thick.

problem in assigning inter-helical NOEs, which typically requires unambiguous assignments of the methyl resonances.

Despite the above issues, individual laboratories have managed to solve new membrane protein structures by solution NMR in the past two decades, although these projects typically took longer than 4 years to complete. Figure 3 shows a number of membrane protein structures determined by solution NMR that had a high impact on the biological community, providing a rough assessment of the current capability of the technique. Although the number of structures solved is small, they cover a rather large range of membrane protein fold space, including oligomeric helical, polytopic helical, and β -barrel structures, with a substantial fraction of them revealing new structural features. The structures in Fig. 3 also span a wide range of sizes from 6 to 42 kDa. Therefore, the state-of-the-art technology in solution NMR is already capable of generating high-resolution structures of small- to medium-sized membrane proteins. The challenge in the field is to increase the speed and accuracy of structure determination.

There is a good opportunity for solution NMR to immediately contribute to the overall structural database for membrane proteins. Genomic analysis of the distribution of number of trans-membrane helices (TMHs) in membrane protein families estimated

that about 2/3 of the proteins span the membrane less than seven times (68). Moreover, the number of crystal structures of membrane proteins with less than 7 TMHs is roughly equal to those with greater than 7 TMHs, implying that crystallization of large and small proteins is equally challenging. The structures shown in Fig. 3 imply that it is possible to establish a robust solution NMR protocol for studying membrane proteins of sizes up to 50 kDa, which can cover the region of structural space up to 7 TMHs (assuming that the proteins are monomeric). Another motivation for establishing an NMR system for a membrane protein even if its crystal structure is available is that the system can be employed in studies to address function and mechanism (e.g., ligand binding of receptors or dynamics of ion transport).

3.1. Major Technical Challenges to Overcome

Probably the biggest challenge in solution NMR of membrane proteins is finding a solubilization condition that both supports the native fold of the protein and yields workable NMR spectra. Because integral membrane proteins are hydrophobic in nature, it is usually difficult to keep the proteins mono-dispersed in solution at concentrations higher than 0.5 mM—roughly the minimum concentration requirement for a full-scale NMR structure determination. Establishing an NMR system for membrane protein thus involves extensive screening of various detergents, lipids, and buffer conditions. Detergent selection is usually not based on rationale, although empirically detergents with phospho-head groups work better than those with sugar-head groups. A rather simple view of the problem of membrane protein solubilization is to find a detergent or detergent/lipid system that can form a tight micelle around the protein to prevent aggregation while still allowing the protein to tumble fast enough to yield good NMR spectrum. In general, detergents with a lower critical micelle concentration (c.m.c) are more stable. In some cases, doping the micelles with natural lipids could increase the overall stability of the protein–micelle complex. Introducing an anionic detergent to the micelles may reduce protein aggregation that is driven mainly by hydrophobic interactions. For example, LMPG (14:0 lyso phosphoglycerol) has become a popular detergent for solution NMR studies of membrane proteins, probably because it has a very low c.m.c and is negatively charged. It is obvious that there will not be a universal detergent/lipid system that will solve all the problems. As we experiment more with various combinations of detergents and lipids, we believe that there will soon be an optimized set of detergent/lipid systems for solution NMR studies of membrane proteins.

Another challenge widely recognized by the NMR community is finding sufficiently long-range NOEs for helical membrane proteins. The amino acid distribution of membrane proteins is very different from that of water-soluble proteins. Large hydrophobic residues show no preference for the protein core, and in β -barrels,

the preference is opposite to that of water-soluble proteins (69). Therefore, the number of close methyl–methyl and methyl–aromatic contacts ($<5 \text{ \AA}$) is much smaller than in water-soluble proteins. Instead, residues like alanine, glycines, prolines, serines, and threonines are often found in the helix–helix interface and the protein core. It would thus be important to develop new isotope labeling strategies for assigning NOEs based on these small amino acids. The difficulty in finding long-range NOEs could also be due to dynamics. As mentioned above, membrane proteins in detergent micelles may adopt increased internal “breathing” because detergent micelles do not fully exert the lateral lipid pressure that a protein experiences in true membrane. Since the NOE is proportional to $1/r^6$, internal motion of TMHs relative to each other can substantially reduce inter-helical NOEs while not having a significant effect on intra-helical NOEs. One way to compensate for the lack of tertiary NOEs is measuring PRE restraints and this approach has been successfully used in the structure determination of DAGK (70) and OmpA (71).

Finally, structural investigation of membrane proteins in lipid bilayers has been a long-sought goal of structural biologists. Solid-state NMR (ssNMR) spectroscopy has been the much-anticipated technique for reaching this goal. While the technology has been progressing rapidly in the past decade, its capability is still far from *de novo* structure determination at high resolution—there is not yet a single example of *de novo* structure determination of a membrane protein at high resolution by ssNMR. The biggest problem of ssNMR is still poor spectral quality. Unlike in solution, the conformational inhomogeneity in the solid phase is not averaged out on the NMR timescale, giving rise to inhomogeneous peaks and low spectral resolution. Is solution NMR study of membrane proteins in a lipid bilayer possible? One approach is to reconstitute the protein in bicelles, which is a disc of lipid bilayer surrounded by a ring of detergents. The ratio of lipid to detergent (q) determines the size of the bicelles (72). Although at large q (>0.8) the lipid region of the bicelles is an excellent representation of the lipid bilayer, at smaller q (<0.5), the lipid bilayer is really a mixture of lipid and detergent. An intriguing alternative is the use of nanodiscs, which are self-assembled patches of lipid bilayer surrounded by a ring of amphipathic membrane scaffold protein such as apolipoprotein (73). The advantage of the nanodisc system is that there is no detergent involved. The large size of nanodiscs ($\sim 150 \text{ kDa}$) has, in the past, discouraged NMR spectroscopists. This system, however, has recently been revisited and showed promising NMR spectra for medium-sized membrane proteins (74). This result, though still preliminary, shows the potential of solution NMR to at least study membrane protein interactions in a true membrane environment.

4. The Future of Solution NMR

4.1. Isotope Labeling

The history of progress in biomolecular NMR tells us that isotope labeling and NMR techniques must go hand in hand in order to push the envelope of the technology farther. In the last two decades, two isotope-labeling schemes have fundamentally changed modern protein NMR spectroscopy. One is deuteration as described above in Subheading 1, which dramatically reduces R_2 by reducing dipolar interactions from protons. The other is selective methyl group protonation in deuterated protein, which enables high-resolution methyl spectroscopy in large proteins (75, 76). Growing *E. coli* in a deuterated medium can achieve protein deuteration (77). Selective protonation of leucine, valine, and isoleucine methyl sites can be done by adding selectively protonated α -keto acid precursors to a perdeuterated medium (78). This labeling scheme allowed full resonance assignment of a 723-residue single polypeptide protein, malate synthase G (79). Recently, stereospecific methyl labeling of valines and leucines has also been achieved (80). In addition to the branched methyl-bearing amino acids, new protocols for labeling methyl groups have been developed for other amino acids, including methionine (81, 82) and alanine (83, 84). Moreover, the carbons of the protein backbone can be selectively labeled. For example, the metabolic precursors 1,3- $^{13}\text{C}_2$ -glycerol and 2- $^{13}\text{C}_1$ -glycerol can be used to produce protein that is selectively isotopic enriched at C' , C_α , H_α , and N positions, while the C_β position is selectively ^{13}C depleted and the adjacent protons are deuterated. This labeling scheme allows recording of a spectrum with very high resolution in the $^{13}\text{C}_\alpha$ dimension by removing the $^{13}\text{C}_\alpha$ - $^{13}\text{C}_\beta$ coupling (85, 86). Owing to the complete characterization of the major metabolic pathways of amino acid synthesis in *E. coli*, it is possible to introduce various labeling strategies by adding isotope-labeled precursors and by manipulating components of the biosynthetic pathways. We anticipate many more new labeling schemes to be introduced in the near future that would fundamentally overcome some of the limitations in NMR spectroscopy of large proteins.

Another continuing development that would provide more labeling options is the cell-free expression of proteins. The open nature of the translation reactions allows the addition of many different compounds, such as protease and RNase inhibitors, ligands, or chaperones, directly into the reaction. Owing to the lack of metabolic scrambling, amino acid type-specific labeling is possible in almost any combination (87). These methods are still very costly at present, but are expected to be increasingly affordable.

The above-described labeling schemes all selectively but uniformly label certain types of chemical groups in a polypeptide

chain. There is great interest however in methods that selectively label only one or more segments of a protein. The basic approach involves splitting a polypeptide chain, expressing and labeling it separately, and splicing. Two intein-based approaches, Expressed Protein Ligation (EPL) and Protein Trans-Splicing (PTS), have been employed to produce segmental labeled protein (88). EPL is based on a reaction involving two protein fragments containing an α -thioester at the C-terminus of the first fragment and an α -cysteine at the N-terminus of the second fragment; a cysteine is required at the ligation site. EPL has been frequently used for segmental isotope labeling of proteins (89–92). In the PTS method, a particular intein is split into two fragments, which have no activity on their own. After mixing them together in solution, they become active and can perform a splicing reaction that results in the fused protein (93, 94).

4.2. NMR Instrument

As in the case of X-ray and electron crystallography, advance in instrumentation has been one of the main driving forces in enhancing the capability of solution NMR spectroscopy. High field instruments such as the 800 and 900 MHz NMR spectrometers are now common around the world, while construction of commercial 1.3 GHz spectrometers is under way. It is important to note however that higher field magnet does not necessarily yield better NMR data. In general, spectra recorded at higher magnetic field have better sensitivity and resolution if the protein is rigid. But this is often not the case because many interesting proteins currently being studied undergo conformational exchanges in solution, in particular those having multiple physiological states. For these molecular systems, high field magnets could further amplify resonance broadening due to chemical shift exchange. What has fundamentally boosted NMR capability in the past decade is the use of cryogenic probes. These “cold” probes, in which the receiver coils are kept at ~ 30 K, can provide up to fourfold gains in the signal-to-noise ratio (S/N) under low ionic strength condition. The “cold” probe technology fundamentally changed the protein concentration requirement for NMR structure determination from > 1 mM to 0.5 mM. Since a large portion of the probe electronics in the current generation of the cryoprobe is still at room temperature, we are hopeful that another two- to threefold gains in S/N may be achieved in the near future. Such improvement would dramatically increase the applicability of solution NMR in studying membrane proteins because many membrane proteins can be made soluble to about 0.2 mM.

4.3. Structure Calculation Program

The powerful combination of rMD and SA has been the dominant tool for generating new NMR structures in the past two decades. Despite its great success, we should not neglect the fact that the MD/SA protocol is still subject to the local minima problem for

complex systems and thus potentially structure calculation artifacts. For example, the rmsd of an NMR ensemble can be affected by various parameters of the MD/SA procedure such as temperature, size of annealing steps, mass of atoms, and even mathematical properties of the target functions. As the computer processor speed is increasing rapidly, it is time for the NMR community to envision a new generation of structure calculation tools that exhaustively search for conformations that are consistent with experimental and/or knowledge-based restraints. The completeness of the search would ensure that structure calculation does not miss any native conformations, and assess the structural precision from data alone (complete data-driven structure determination). A comparison between a complete search method and MD/SA has been made for the homo-pentameric phospholamban protein, showing that rmsd of the structural ensemble obtained from a complete grid search is substantially larger (95). The complete search method is, however, not yet practical due to the unrealistically long computation time required.

An alternate, more feasible approach is to exhaustively search the protein database for structural fragments that best fit experimental data, commonly known as the Molecular Fragment Replacement (MFR) method. The MFR method was first used in crystallography for building molecular fragments into low-resolution crystallographic density (96), and was subsequently applied in NMR to fragments that agree with RDCs (97) and other NMR data such as chemical shifts (98, 99). Although in principle the MFR approach limits conformational space to what has already been observed, in practice this approach demonstrated to be very effective because (1) the Protein Data Bank is so large and diverse that its content is an excellent representation of the conformational space of proteins and nucleic acids in nature and (2) the method greatly reduces the search space such that it is computationally affordable. Furthermore, the database of crystal structures is expanding at a rapid pace and thus the MFR approach would only become more powerful with time. Our expanding knowledge of structure will increasingly facilitate *de novo* NMR structure determination. The same is true for cryo-EM, where computer algorithms that effectively utilize the knowledge from the existing structure database have been largely responsible for dramatically increasing the resolution of cryo-EM structures (100). As solution NMR gradually collects structural restraints for larger proteins, there could be new opportunities for combining the RDC/chemical shift MFR method with low-resolution EM density, that is, to fit NMR-derived fragments into EM density. The EM density, although at low resolution, solves a big problem for NMR—determining the global fold or shape of a molecule by NOE restraints.

References

- Bloch F, Hansen WW, Packard M (1946) The nuclear induction experiment. *Phys Rev* 70:474–485
- Purcell EM, Torrey HC, Pound RV (1946) Resonance absorption by nuclear magnetic moments in a solid. *Phys Rev* 69:37–38
- Pervushin K, Riek R, Wider G, Wuthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94(23):12366–12371
- Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced ^1H - ^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* 125:10420–10428
- Meyer B, Peters T (2003) NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl* 42(8):864–890
- Battiste JL, Pestova TV, Hellen CU, Wagner G (2000) The eIF1A solution structure reveals a large RNA-binding surface important for scanning function. *Mol Cell* 5(1):109–119
- Gross JD et al (2003) Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell* 115(6):739–750
- Marintchev A, Frueh D, Wagner G (2007) NMR methods for studying protein-protein interactions involved in translation initiation. *Methods Enzymol* 430:283–331
- Babu YS, Bugg CE, Cook WJ (1988) Structure of calmodulin refined at 2.2 Å resolution. *J Mol Biol* 204(1):191–204
- Ikura M et al (1992) Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* 256(5057):632–638
- Barbato G, Ikura M, Kay LE, Pastor RW, Bax A (1992) Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry* 31(23):5269–5278
- Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445:618–622
- Loria JP, Rance M, Palmer AG 3rd (1999) A relaxation-compensated carr-purcell-meiboom-gill sequence for characterizing chemical exchange by NMR spectroscopy. *J Am Chem Soc* 121(10):2331–2332
- Palmer AG 3rd, Kroenke CD, Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* 339:204–238
- Eisenmesser EZ, Bosco DA, Akke M, Kern D (2002) Enzyme dynamics during catalysis. *Science* 295(5559):1520–1523
- Eisenmesser EZ et al (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature* 438(7064):117–121
- Schnell JR, Chou JJ (2008) Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 451(7178):591–595
- Overhauser AW (1953) Polarization of nuclei in metals. *Phys Rev* 91(2):476
- Carver TR, Slitcher CP (1953) Polarization of nuclear spins in metal. *Phys Rev* 92:212
- Becerra LR, Gerfen GJ, Temkin RJ, Singel DJ, Griffin RG (1993) Dynamic nuclear polarization with a cyclotron resonance maser at 5 T. *Phys Rev Lett* 71(21):3561–3564
- Hall DA et al (1997) Polarization-enhanced NMR spectroscopy of biomolecules in frozen solution. *Science* 276(5314):930–932
- Sounier R, Blanchard L, Wu Z, Boisbouvier J (2007) High-accuracy distance measurement between remote methyls in specifically protonated proteins. *J Am Chem Soc* 129(3):472–473
- Zuiderweg ER, Fesik SW (1989) Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* 28(6):2387–2391
- Kay LE, Clore GM, Bax A, Gronenborn AM (1990) 4-Dimensional heteronuclear triple-resonance NMR-spectroscopy of interleukin-1-beta in solution. *Science* 249(4967):411–414
- Tugarinov V, Choy WY, Orekhov VY, Kay LE (2005) Solution NMR-derived global folded of a monomeric 82-kDA enzyme. *Proc Natl Acad Sci USA* 102(3):622–627
- Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13(3):289–302
- Bowers PM, Strauss CE, Baker D (2000) De novo protein structure determination using sparse NMR data. *J Biomol NMR* 18(4):311–318
- Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 104(23):9615–9620

29. Gong H, Shen Y, Rose GD (2007) Building native protein conformation from NMR backbone chemical shifts using Monte Carlo fragment assembly. *Protein Sci* 16(8):1515–1521
30. Spera S, Bax A (1991) An empirical correlation between protein backbone conformation and Ca and C β chemical shifts. *J Am Chem Soc* 113:5490–5492
31. Wishart DS, Sykes BD (1994) The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* 4(2):171–180
32. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44(4):213–223
33. Cheung MS, Maguire ML, Stevens TJ, Broadhurst RW (2010) DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J Magn Reson* 202(2):223–233
34. Bax A et al (1994) Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol* 239:79–105
35. Deber CM, Torchia DA, Blout ER (1971) Cyclic peptides. I. Cyclo(tri-l-prolyl) and derivatives. Synthesis and molecular conformation from nuclear magnetic resonance. *J Am Chem Soc* 93(19):4893–4897
36. Demarco A, Llinas M, Wuthrich K (1978) Analysis of H-1-NMR Spectra of ferrichrome peptides. 2. Amide resonances. *Biopolymers* 17(3):637–650
37. Vuister GW, Wang AC, Bax A (1993) Measurement of three-bond nitrogen-carbon J couplings in proteins uniformly enriched in nitrogen-15 and carbon-13. *J Am Chem Soc* 115(12):5334–5335
38. Grzesiek S, Vuister GW, Bax A (1993) A simple and sensitive experiment for measurement of J_{CC} couplings between backbone carbonyl and methyl carbons in isotopically enriched proteins. *J Biomol NMR* 3(4):487–493
39. MacKenzie KR, Prestegard JH, Engelman DM (1996) Leucine side-chain rotamers in a glycoporphin A transmembrane peptide as revealed by three-bond carbon-carbon couplings and ¹³C chemical shifts. *J Biomol NMR* 7(3):256–260
40. Hu J-S, Grzesiek S, Bax A (1997) Chi1 angle information from a simple two-dimensional NMR experiment which identifies trans ³J_{NCg} couplings in isotopically enriched proteins. *J Biomol NMR* 9:323–328
41. Hu JS, Bax A (1997) Determination of phi and chi(1) angles in proteins from C-13-C-13 three-bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond? *J Am Chem Soc* 119(27):6360–6368
42. Abragham A (1961) The principles of nuclear magnetism. Clarendon Press, Oxford, England
43. Saupe A, Englert G (1963) High-resolution nuclear magnetic resonance spectra of orientated molecules. *Phys Rev Lett* 11:462–464
44. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1995) Nuclear magnetic dipole interactions in field-oriented proteins – information for structure determination in solution. *Proc Natl Acad Sci USA* 92(20):9279–9283
45. Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278(5340):1111–1114
46. Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* 5(12):1065–1074
47. Clore GM, Starich MR, Gronenborn AM (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J Am Chem Soc* 120(40):10571–10572
48. Prosser RS, Losonczi JA, Shiyonovskaya IV (1998) Use of a novel aqueous liquid crystalline medium for high-resolution NMR of macromolecules in solution. *J Am Chem Soc* 120(42):11010–11011
49. Barrientos LG, Dolan C, Gronenborn AM (2000) Characterization of surfactant liquid crystal phases suitable for molecular alignment and measurement of dipolar couplings. *J Biomol NMR* 16(4):329–337
50. Ruckert M, Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc* 122(32):7793–7797
51. Tycko R, Blanco FJ, Ishii Y (2000) Alignment of biopolymers in strained gels: a new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *J Am Chem Soc* 122(38):9340–9341
52. Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S (2000) Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J Biomol NMR* 18(4):303–309
53. Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 293(5529):487–489

54. Chou JJ, Gaemers S, Howder B, Louis JM, Bax A (2001) A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *J Biomol NMR* 21(4):377–382
55. Chou JJ, Kaufman JD, Stahl SJ, Wingfield PT, Bax A (2002) Micelle-induced curvature in a water-insoluble HIV-1 Env peptide revealed by NMR dipolar coupling measurement in stretched polyacrylamide gel. *J Am Chem Soc* 124(11):2450–2451
56. Douglas SM, Chou JJ, Shih WM (2007) DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proc Natl Acad Sci USA* 104(16):6644–6648
57. Lorieau J, Yao L, Bax A (2008) Liquid crystalline phase of G-tetrad DNA for NMR study of detergent-solubilized proteins. *J Am Chem Soc* 130:7536–7537
58. Kontaxis G, Clore G, Bax A (2000) Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. *J Magn Reson* 143(1):184–196
59. Chou JJ, Delaglio F, Bax A (2000) Measurement of one-bond ^{15}N - $^{13}\text{C}'$ dipolar couplings in medium sized proteins. *J Biomol NMR* 18(2):101–105
60. Jaroniec CP, Ulmer TS, Bax A (2004) Quantitative J correlation methods for the accurate measurement of $^{13}\text{C}'$ - ^{13}C dipolar couplings in proteins. *J Biomol NMR* 30(2):181–194
61. Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* 39(18):5355–5365
62. Qin PZ et al (2007) Measuring nanometer distances in nucleic acids using a sequence-independent nitroxide probe. *Nat Protoc* 2(10):2354–2365
63. Williamson MP, Havel TF, Wüthrich K (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^1H nuclear magnetic resonance and distance geometry. *J Mol Biol* 182:295–315
64. Kaptein R, Zuiderweg ERP, Scheek RM, Boelens R, Vangunsteren WF (1985) A protein-structure from nuclear magnetic-resonance data – LAC repressor headpiece. *J Mol Biol* 182(1):179–182
65. Clore GM et al (1986) The 3-dimensional structure of alpha-1-purothionin in solution – combined use of nuclear-magnetic-resonance, distance geometry and restrained molecular-dynamics. *EMBO J* 5(10):2729–2735
66. Clore GM, Brunger AT, Karplus M, Gronenborn AM (1986) Application of molecular-dynamics with interproton distance restraints to 3-dimensional protein-structure determination – a model study of crambin. *J Mol Biol* 191(3):523–551
67. Nilges M, Clore GM, Gronenborn AM (1988) Determination of 3-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett* 229(2):317–324
68. Liu Y, Engelman DM, Gerstein M (2002) Genomic analysis of membrane protein families: abundance and conserved motifs. *Genome Biol* 3(10):research0054
69. Ulmschneider MB, Sansom MS (2001) Amino acid distributions in integral membrane protein structures. *Biochim Biophys Acta* 1512(1):1–14
70. Van Horn WD et al (2009) Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase. *Science* 324(5935):1726–1729
71. Liang B, Bushweller JH, Tamm LK (2006) Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. *J Am Chem Soc* 128(13):4389–4397
72. Vold RR, Prosser RS (1996) Magnetically oriented phospholipid bilayered micelles for structural studies of polypeptides. Does the ideal bicelle exist? *J Magn Reson B* 113(3):267–271
73. Bayburt TH, Sligar SG (2003) Self-assembly of single integral membrane proteins into soluble nanoscale phospholipid bilayers. *Protein Sci* 12(11):2476–2481
74. Raschle T et al (2009) Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. *J Am Chem Soc* 131(49):17777–17779
75. Gardner KH, Kay LE (1998) The use of ^2H , ^{13}C , ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* 27:357–406
76. Tugarinov V, Kay LE (2004) An isotope labeling strategy for methyl TROSY spectroscopy. *J Biomol NMR* 28:165–172
77. LeMaster DM, Richards FM (1988) NMR sequential assignment of *E. coli* thioredoxin utilizing random fractional deuteration. *Biochemistry* 27:142–150

78. Tugarinov V, Kanelis V, Kay LE (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* 1(2): 749–754
79. Tugarinov V, Muhandiram DR, Ayed A, Kay LE (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase G. *J Am Chem Soc* 124:10025–10035
80. Gans P et al (2010) Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew Chem Int Ed Engl* 49(11):1958–1962
81. Fischer M et al (2007) Synthesis of a C-13-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. *Chembiochem* 8(6):610–612
82. Gelis I et al (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* 131(4):756–769
83. Isaacson RL et al (2007) A new labelling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of Alanine residues. *J Am Chem Soc* 129(50):15428–15429
84. Ayala I, Sounier R, Use N, Gans P, Boisbouvier J (2009) An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J Biomol NMR* 43(2):111–119
85. LeMaster DM, Kushlan DM (1996) Dynamical mapping of E. coli thioredoxin via C-13 NMR relaxation analysis. *J Am Chem Soc* 118(39):9255–9264
86. Takeuchi K, Sun ZYJ, Wagner G (2008) Alternate C-13-C-12 labeling for complete mainchain resonance assignments using C alpha direct-detection with applicability toward fast relaxing protein systems. *J Am Chem Soc* 130(51):17210–17211
87. Kainosho M et al (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* 440:52–57
88. Skrisovska L, Schubert M, Allain FHT (2010) Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. *J Biomol NMR* 46(1):51–65
89. Camarero JA et al (2002) Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc Natl Acad Sci USA* 99(13):8536–8541
90. Skrisovska L, Allain FHT (2008) Improved segmental isotope labeling methods for the NMR study of multidomain or large proteins: application to the RRM of Npl3p and hnRNP L. *J Mol Biol* 375(1):151–164
91. Vitali F et al (2006) Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO J* 25(1):150–162
92. Xu R, Ayers B, Cowburn D, Muir TW (1999) Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc Natl Acad Sci USA* 96(2):388–393
93. Muralidharan V, Muir TW (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat Methods* 3(6):429–438
94. Xu MQ, Evans TC (2005) Recent advances in protein splicing: manipulating proteins in vitro and in vivo. *Curr Opin Biotechnol* 16(4):440–446
95. Oxenoid K, Chou JJ (2005) The structure of phospholamban pentamer reveals a channel-like architecture in membranes. *Proc Natl Acad Sci USA* 102(31):10870–10875
96. Jones TA, Thirup S (1986) Using known substructures in protein model-building and crystallography. *EMBO J* 5(4):819–822
97. Delaglio F, Kontaxis G, Bax A (2000) Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J Am Chem Soc* 122(9):2142–2143
98. Shen Y et al (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105(12):4685–4690
99. Raman S et al (2010) NMR structure determination for larger proteins using backbone-only data. *Science* 327(5968):1014–1018
100. Gonen T et al (2005) Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* 438(7068):633–638