Protein Structure Determination by NMR Spectroscopy

Kiattawee Choowongkomon

Abstract

The Nuclear magnetic resonance (NMR) method for protein structure determiniation in solution is now firmly established as an alternative method beside X-ray crystallography. NMR can give a detailed picture of the three-dimensional structure of protein at atomic resolution. It also allows one to study conformational dynamics, real time enzyme kinetic, and very week to tight protein-protein interaction. The intention of this review is to summarize the methods for protein structure determination together with other developmental NMR techniques for studying biophysical properties of proteins by NMR. Useful implementations of these experiments based on the practical experience of the author are presented as overview for general scientists who want to understand what NMR can do with the proteins.

Key words: NMR, protein structure, determination

Department of Biochemistry, Faculty of Science, Kasetsart University, 10900.
E-mail: fsciktc@ku.ac.th

Background

Nuclear Magnetic Resonance (NMR) spectroscopy was first discovered in 1946 (Bloch,1946). It took another 40 years to develop the technique and make it possible to determine protein structures (Williamson *et al.*,1985). NMR is the only solution structural technique that is comparable to X-ray diffraction techniques. By September, 2005, more than 4400 (15%) of the biomolecular entries deposited in the Protein Data Bank (www.pdb.org) have been solved by NMR methods. Although the percentage of structures solved by NMR is low, NMR plays a crucial role in solving the structures of membrane proteins, highly flexible proteins, and many proteins that can not be crystallized.

i. NMR Signal

NMR methods for the study of protein structures depend on the sensitive variation of the resonance frequency of a nuclear spin in an external magnetic field with the different chemical structure, the conformation, and the solvent environment (1987). Three of the four most abundant elements in proteins, H, C, and N, have naturally occurring isotopes with nuclear spin ½ which are suitable for high-resolution NMR experiments in solution. The proton (^{1}H) has the highest natural abundance (99.98%) whereas both ^{13}C and ^{15}N have low natural abundance (1.11% and 0.37%, respectively). The lower abundance of ^{13}C and ^{15}N is routinely overcome by overexpression of proteins in isotope-labeled media. The nuclear spin gives rise to a magnetic dipole moment:

$$\mu = \gamma \cdot I \qquad (1)$$

where μ is magnetic dipole moment, γ is gyromagnetic ratio of the nucleus, and I is spin angular momentum.

If such an atomic nucleus is placed into external magnetic field B_0, it has energy associated with the interaction between its magnetic moment and the external field:

$$E = -\mu \cdot B_0 \qquad (2)$$

For a spin ½ nucleus, two (2I+1) orientations are possible ($I_z = \pm ½$) and the energy levels are populated according to a Boltzmann distribution. In order to induce nuclear magnetic resonance, an oscillatory magnetic field has to be applied at the frequency which corresponds to the separation of the two spin energy levels. One of two ways to do that is with a very short radiofrequency pulse which encodes the range of frequencies covering the entire frequency range (Fourier transform, FT NMR). Another method is sweeping through the radiofrequencies in the expected frequency range (continuous wave, CW NMR). FT NMR is more sensitive and a used in biomolecular NMR. The response signal obtained from a FT NMR experiment is free-induction decay (FID) which is a superposition of the frequencies of all spins in the molecule as a function of time, F(t). In order to obtain the spectrum $F(\omega)$ (intensity as a function of frequency), a Fourier transformation is performed to translate the function in the time domain into the frequency domain.

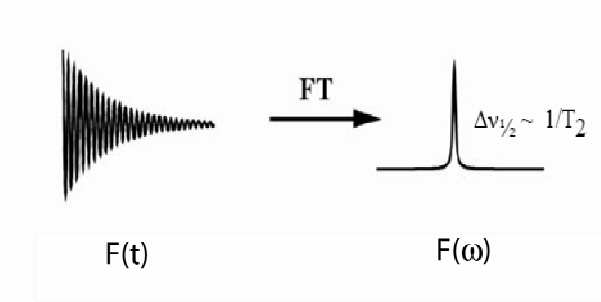

Figure 1: The Fourier transformation of the NMR signal. NMR signal is free-induction decay (FID) which is a function of time, F(t). The Fourier transformation is used to translate the function in the time domain into the frequency domain, F(ω).

ii. Relaxation

FID signal decreases with time, which implies that component of the magnetization vector in the xy-plane must be shrinking. This effect can be explained by two processes. Both reflect the fact that the nuclear spins are not in thermal equilibrium with their environments.

After thermal equilibrium the spins have a Bolzmann distribution, with more spin in the lower state than in the upper state. However, a magnetization vector in the xy-plane immediately after a 90 pulse has equal number of upper and lower spins. The populations revert to their thermal equilibrium values exponentially. As they do so, the z-component of the magnetization reverts to its equilibrium value with a time constant called the longitudinal relaxation time T_1. T_1 is caused by fluctuating local magnetic fields arising from the motion of the molecules. These fluctuations can stimulate the spins to change from upper to lower state and vice versa, and to relax towards the thermal equilibrium population.

The second aspect of spin relaxation is the fanning-out of the spin in the xy-plane if they process at different rates. The magnetization vector is large when all the spins are bunched together immediately after a 90 pulse. However, this orderly bunching of spins is not at equilibrium and, even if there were no spin-lattice relaxation, we would expect the individual spins to spread out until they were uniformly distributed with all possible angles around the z-axis. At that stage, the components of magnetization vector in the plane would be zero. The randomization of the spin directions occurs exponentially with a time constant called the transverse relaxation time T_2. The linewidth of an NMR signal is dependent on the exponential decay of the FID, which is governed by the transverse relaxation time T_2.

iii. Chemical Shifts

Each spin in a protein gives rise to a resonance. The exact resonance frequency depends on the chemical environment of each spin. The nuclear spins are shielded from the static magnetic field B_0. As a result, each spin in a protein resonates at a slightly different frequency. These frequencies are called chemical shifts and are measured in parts per million (ppm) in order to have chemical shift values independent of the static magnetic field strength.

iv. J-couplings

J-couplings or scalar couplings are mediated through chemical bonds connecting two spins. The energy levels of each spin are slightly altered depending on the spin state of a scalar coupled spin, giving a rise to a splitting of the resonances.

Couplings are well-correlated with the backbone dihedral angle via the Karplus equation (Karplus,1963):

$$^3J = A\cos 2\theta + B\cos \theta + C \quad (3)$$

where θ is the dihedral angle and A, B and C are adjustable parameters.

v. Nuclear Overhauser Effect (NOE)

The NOE is a result of cross-relaxation between dipolar coupled spin/spin interaction through space. The dipolar couplings are usually in the kHz range, and depend on the distance between two spins and the orientation of internuclear vector with respect to the static magnetic field B_0. Due to fast isotropic thumbling of proteins in solution, the dipolar couplings are averaged to zero. However, the dipolar couplings give rise to spin-spin and spin-lattice relaxation. The NOE is a result of cross-relaxation between spins and is defined by the transition rate W_0 and W_2 which involve spin flips of both spins. The NOE allows to transfer magnetization from one spin to another through space and is observable only if the distance between two spins less than about 5 Å (NOE ~ $1/r^6$).

vi. Residual Dipolar Couplings (RDC)

RDCs are a useful phenomenon to characterize structure because they depend on distance, orientation, and dynamics of a single bond vector in an overall molecular frame. This dipolar coupling is averaged to zero under isotropic conditions, but it can be only observed under anisotropic conditions. If only one of 10^3 -10^4 molecules is aligned in media, the dipolar couplings are scaled down from kHz range to few Hz. This allows utilizing the distance and the angle dependence of dipolar couplings as orientational restraints. The observable dipolar coupling in solution can be expressed as:

$$D = S\,[D_a\,(3\cos^2\theta-1) + 3/2\,D_r \sin^2\theta\cos 2\varphi]\,/r^3 \quad (4)$$

Where D is size of dipolar splitting, S is an order parameter characterizing the rigidity of the molecule, D_a and D_r, are axial and rhombic components of the alignment tensor, r is the length of the internuclear vector, θ is the angle between r and D_a and φ is an angle between r and D_r.

While NOE are local-distance restraints, RDCs provide long-range orientational information and are now widely utilized in structure calculation. The RDCs can be used for refinement of protein structure calculations. They are also used to determine the relative domain orientation of multidomain molecules (Fischer *et al.*,1999) (Mollova and Pardi,2000) and protein-ligand complexes (Weaver and Prestegard,1998) (Pan *et al.*,2002).

The measurement of residual dipolar couplings in solution has been accomplished by weakly aligning a molecule in the liquid state. There are several methods to weakly align proteins for RDC measurements. These include bicelles using different detergents (Ottiger and Bax,1998), phage particles (Pf1) (Hansen *et al.*,1998), purple membrane fragments (PMFs) (Sass *et al.*,2000), and strain-induced gels (Ishii *et al.*,2001).

Solution NMR Techniques

The molecular weights of proteins that can be characterized by NMR are limited to about 30 kDa due to relaxation and overlap of crosspeaks. The line widths in the NMR spectra are inversely proportional to the relaxation rates (T_2). Therefore the signal-to-noise in NMR spectra of larger molecules is poor. However, some larger proteins can be studied by NMR by dividing it into a smaller part, called a

domain. Most large proteins consist of a number of globular domains of 100-200 amino acids in size. Each domain often has a modular function, e.g. one domain is required to bind to the ligand whereas another domain has enzyme activity. Therefore, in order to understand a molecular interaction or reaction, it is often sufficient to determine the structure of individual domain of interest in the large protein.

A high concentration of protein in solution (at least 1 mM) is required, since NMR is a low-sensitivity technique. Normally, extracting and purifying proteins from a natural source is not enough to achieve the necessary protein concentration. Peptide synthesis is an alternative method to get a peptide or protein of interests. However, longer peptides may not fold correctly and get a low yield. The overexpression of a protein of interest in a bacterial host, normally *E. coli*, is routinely achieved. The benefit of bacterial cells is that the protein can be isotopically labeled with ^{13}C, ^{15}N and also ^{2}H isotopes. Furthermore, some eukaryotic peptides can be folded correctly by themselves or bacterial protein folding machineries.

All NMR experiments are carried out using pulse sequences. Several pulse sequences have been developed to obtain different information from the peptide chain. In 1D spectrum, every peak usually corresponds to one nucleus in the protein. The proton spectra of proteins have general characteristics which can be divided into 3 majors regions, 0-3.5 ppm for proton from , CH, CH_2, CH_3 the side-chain amino acid , 3.5-7 ppm for proton from H^{α}, and water, and 6-11 ppm for H^{N} and aromatic protons (Figure 2.2). 1D proton spectra are far too complex for interpretation as most of the signals overlap.

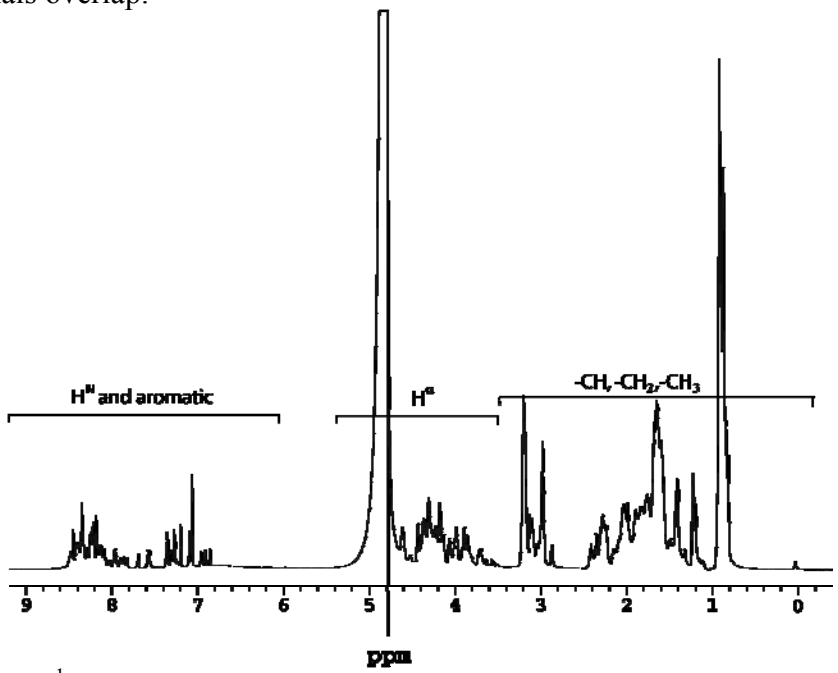

Figure 2: 1D-^{1}H- NMR spectra of the juxtamembrane domain of epidermal growth factor receptor peptide from residues Arg645 to Gly697 in water at 25 °C.

By introduction of additional spectral dimensions, these spectra can be simplified. Spectra with two or more dimensions always have off-diagonal peaks, called crosspeaks, which correspond to an interaction between two or more nuclei. There are two types of two dimension (2D) experiments; homonuclear and heteronuclear 2D experiments. The homonuclear 2D spectra are widely used for the structure determination of proteins that can give interaction information between

proton and proton. The heteronuclear 2D spectra of isotopic labeling atoms (e.g. ^{15}N or ^{13}C) can give information between isotopic atom and other protons. The most widely used heteronuclear 2D spectra are ^{1}H-^{15}N-heteronuclear single quantum correlation (HSQC) which correlates the nitrogen atom of an NH group with the directly attached proton (Figure 2.3). Each signal in a HSQC spectrum represents a proton that is bound to a nitrogen atom. This HSQC spectrum is called a "fingerprint" as each protein has different crosspeak patterns. It can be used to quick highlight whether the peptide is folded. Crosspeak patterns of the folded peptides usually spread out in larger range of ppm in both dimensions of HSQC spectra while the crosspeaks of the aggregated or unfolded peptides cluster in the center of HSQC spectra.

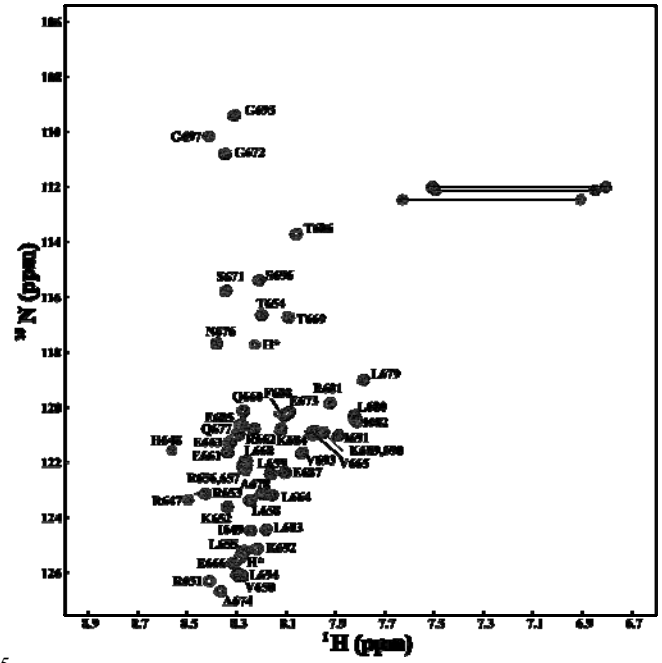

Figure 3: ^{1}H-^{15}N- HSQC spectrum of the juxtamembrane domain from residue Arg645 to Gly697 of epidermal growth factor receptor in water 25°C.

Three-dimensional experiments are performed by adding an additional spectral dimension in pulse sequences to spread out the signal in the third dimension, allowing one to observe the correlation of three spins with each other. The addition of the third dimension increases the resolution and reduces the overlap. Homonuclear three dimension (3D) experiments are not normally utilized due to long time requirement. On the other hand, the heteronuclear 3D experiments are widely used as they give more sensitive and require less time, but need isotopical labeling of proteins. Series of the heteronuclear 3D experiments are now routinely used for assignment purposes and structure determination.

Solving protein structures by NMR

i. Assignment Strategy

Assignment strategies based on homonuclear 2D experiments were developed by Wüthrich (1986). Sequence specific resonance assignments can be obtained based on three experiments: 2D-^{1}H-^{1}H-total correlation spectroscopy (TOCSY) (braunschweiler and Ernst,1983), 2D-^{1}H-^{1}H-nuclear Overhauser effect spectroscopy (NOESY) (Kumar *et al.*,1980), and 2D-double quantum filtered correlation

spectroscopy (DQF-COSY) (Rance *et al.*,1983). The TOCSY experiment correlates both backbone and sidechain proton which can then be used to identify the amino acid type. The COSY experiment permits one to trace H-H connectivities and all sidechain protons in an amino acid when combined with TOCSY data. The NOESY experiment yields through-space correlations between protons in close proximity which gives information of sequential neighbor and through space connections due to structure. All atoms of the previous residues could be mostly observed which is very useful for sequential assignment. The sequence-specific resonance assignment usually starts on backbone NOE correlations then confirms the connection to the primary sequence by sidechain NOE correlations. The chains of peptide can build up from small linking of sequential amino acids. Eventually, peptide chain can be completed. The NOE assignment process is also responsible for determining all the short-, medium- and long-ranges interresidue NOE correlations. These give information of secondary structure of peptides by analyzing the patterns of NOE correlations (1986). Furthermore, tertiary folding of peptides can also derived the network distance patterns from NOEs. However, the degeneracy and overlapping of chemical shifts sometimes make NOE assignment more difficult in longer peptides. Thus the isotopic-labeling peptide approaches have been developed.

Developments of assignment strategies using uniformly-$^{13}C/^{15}N$-labeled samples allows precise resonance assignments and extend to allow the study of the structure of larger proteins. Commonly, the assignment strategy begins with the identification of hydrogen-nitrogen (H^N-N) pairs in the backbone of amino acid residues in the sequence, and is followed by the identification of sequential connections of the H^N-N pairs with sequential experiments such as HNCA, and HN(CO)CA. The sequential connections are confirmed and expanded by incorporating in experiments such as: HNCACB (Wittekind and Mueller,1993), CBCA(CO)HN (Grzesiek and Bax,1992), HNCO (Ikura *et al.*,1990) , and HN(CA)CO (Yamazaki *et al.*,1994) experiments. During this process C^α, H^α, C^β, H^β and CO nuclei are identified. After amino acids have been connected into chains, the chains are placed into the amino acid sequence of the protein based on the chemical shifts of C^α, H^α, C^β, and H^β and CO nuclei to identify the amino acid type. Side-chain assignments are made after nearly complete backbone assignments. The sidechain nuclei are identified using different TOCSY experiments for example, HCCH-TOCSY (Kay *et al.*,1993), H(CCO)NH (Montelione *et al.*,1992), and (H)C(CO)NH (Montelione *et al.*,1992). These experiments correlate the nuclei of the sidechains with the known backbone nuclei and also help to confirm the amino acid type identification at the same time. In principle, nearly all nuclei should be identified by this protocol; however, missing resonances are possible due to motion and/or chemical shift exchange.

After all the resonance assignments are complete, they are used to assign all peaks that are observed in both ^{15}N- and ^{13}C-HSQC-NOESY experiments (pascal *et al.*,1994). Separation of the proton-proton interaction into a third dimension can help to solve spectral ambiguities which limit the analysis of conventional 2D NMR spectra. The network of backbone-backbone, sidechain-backbone, and sidechain-sidechain interactions can be determined from ^{15}N- and ^{13}C-HSQC-NOESY experiments. The ^{15}N-HSQC-NOESY allows detection of protons that are closer than 6 Å of ^{15}N bound proton whereas the ^{13}C-HSQC-NOESY allows detection of protons that are closer than 6 Å of a ^{13}C bound proton. The secondary structure elements can be identified by analyzing the chemical shifts together with NOE pattern whereas tertiary folds can be characterized by the network of observed distances.

The assignment of NOEs in NOESY experiments is a time consuming process that takes several months to complete. Several automated NOE assignment programs are developed to reduce the time for NMR structure determination. These include ARIA (Linge *et al.*,2003) AUTOSTRUCTURE (Huang *et al.*,2003), CANDID (Herrmann *et al.*,2002), KNOWNOE/AUREMOL (Gronwald *et al.*,2002), SANE (Duggan *et al.*,2001). Although the promising computer automated NOE determination will continue to grow in importance, up to date there is any program can assignment without an error. It always requires carefully inspection and manually intervening to get the correct assignments.

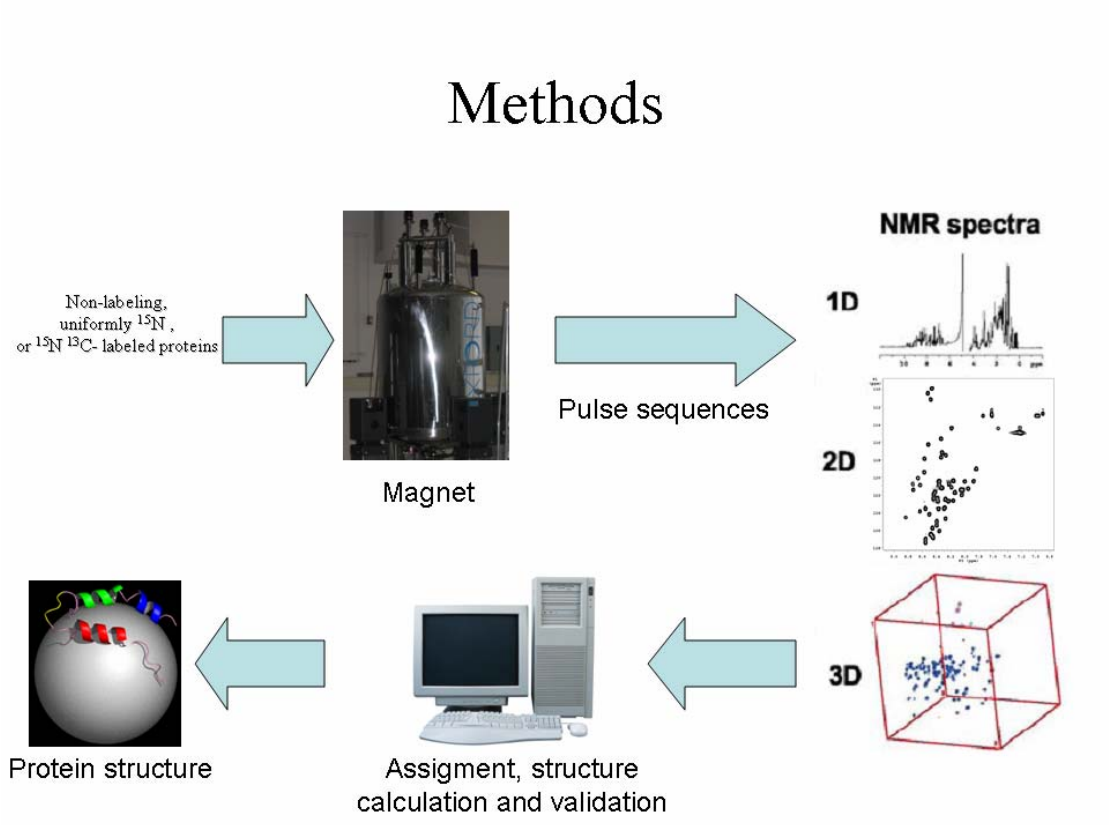

Figure 4: Summary of general strategy for solving protein structure by NMR methods

ii. Structural Restraints

There are three common types of structural restraints, distance, angle, and orientation restraints. Distance restraints are acquired from NOESY experiments. In NOESY experiment the dipole-dipole cross relaxation mechanism is utilized to extract the through space distance between nuclei since intensities of NOESY crosspeaks are proportional to the distances between nuclei to the power of minus six (with the limitation of distances less than 5 Å). The transfer of magnetization over the chemical bonds, called scalar- or J-coupling, allows one to measure the angle restraints via several experiments. The rate of magnetization transfer depends on the local conformation of atoms. The 3-bond scalar couplings between H^N and H^α ($^3J_{N\alpha}$) in the peptide backbone can be measured in an HNHA experiment (Vuister and Bax,1993), and provides information on the backbone torsion angle ϕ (CO_{i-1}-N_i-C^α_i-CO_i). Torsion angle restraints can also be obtained from chemical shift analysis using the TALOS program (Cornilescu *et al.*,1999). Furthermore, RDCs also provide long-range orientational information. RDCs are typically not used in initial structure

calculation but rather in the refinement stage of structure calculations. Including RDC in initial calculations may trap the structure in a local minimum, leading to convergence problems (Meiler *et al.*,2000).

iii. Structure Calculation Methods

Structure calculations are done by either one of three methods; distance geometry, simulated annealing, and torsion angle dynamics.

Distance geometry was the first approach used for the structure calculation of proteins by NMR data (Braun *et al.*,1981) (Havel *et al.*,1983a) (Havel *et al.*,1983b). Distance Geometry is a procedure for transforming a "distance matrix" representation of a structure to a 3-D Cartesian representation. Ideally, the accuracy and precision of the distance matrix is sufficient enough that diagonalization of the metric matrix yields an unambiguous, physically realistic structure. The geometrical relationships from all available distance constraints, bond, torsion angles, and van der Waals radii are used to obtain approximate coordinates from interatomic distances and then are further refined.

Simulated annealing is widely used today. It is a molecular dynamics simulation, which takes place directly in the Cartesian coordinate system (Nilges *et al.*,1988) (Scheek *et al.*,1989) (Brunger and Nilges,1993). Simulated annealing methods are performed with the programs CNS (Brunger *et al.*,1998), X-PLOR (1992), or X-PLOR-NIH (Schwieters *et al.*,2003). In this method, a starting structure is heated to a high temperature in a simulation. During many slow cooling steps the starting structure can evolve toward the energetically favorable final structure under the influence of a force fieled derived from experimental restraints, distances and dihedral angles.

At the beginning of the calculation, the energy of the starting structure is minimized by moving the atoms until the structure reaches an energy minimum. The obtained structure may be in a local minimum that does not satisfy all constraints. Such structures cannot reach the global energy minimum, as they cannot cross the energy barrier between local and global minima. This problem can be overcome if kinetic energy is introduced. This is achieved by heating the system to very high temp (> 1000 K). The system then has enough energy to cross the barrier and reaches the global energy minimum. After the high temperature steps, atomic velocities are slowly reduced in many steps. At each temperature, the system is left to develop under the influence of the potential field. While the temperature of the system is reduced, the force constants in the experimental constraints are simultaneously raised in order to weight them more strongly. The result of the simulation is a minimum energy of protein structure, but the possibility that this structure is in a local minimum rather than the global minimum can be excluded. Therefore, at least twenty different starting structures with random folds are used. These resulting structures are iteratively reused as starting structures for another simulated annealing with slightly changed input protocols, until no further reduction in global energy is observed and the structures converge in conformational space.

Torsion angle dynamics uses torsion angles instead of Cartesian coordinates as degrees of freedom and a fast recursive algorithm to integrate the equations of motions (Guntert *et al.*,1997). Torsion angle dynamics are implemented in the program DYANA and its successor CYANA. The principal difference of torsion angle dynamics is that it works with internal coordinates rather than Cartesian coordinates. Therefore, the degree of freedom is decreased by tenfold since the covalent structure parameters including bond lengths, bond angles, chiralities and

planarities are kept fixed at their optimal value during calculation. This is more efficient than Cartesian space calculations which allow for the use of longer time-steps and /or higher temperatures in the structure calculation.

iv. Structure Validation

Protein structures solved by NMR are validated by studying the geometry of the molecule, the potential energy in the molecule and the overall fold. The geometry of the proteins are analyzed by programs WHATIF (Vriend,1990) and PROCHECK (Laskowski *et al.*,1993), including an NMR-specific extension called PROCHECK-NMR (Laskowski *et al.*,1996). These programs try to assess the quality of a structure primarily by checking whether a number of different parameters are in agreement with values in databases that have been derived from both high-resolution X-ray and NMR structures. This is usually done by analyzing the Ramachandran plot between phi and psi angles of each residue in proteins. In addition, PROCHECK-NMR can read NMR experimental restraints in a variety of formats and provide measurements for the agreement of the experimental restraints with the calculated structure. PROCHECK-NMR can also analyze ensembles of NMR structures which usually the preferred representation of the NMR protein structures. The potential energy is calculated using force fields, parameterizing intramolecular forces. If the potential energy is high, indicating a violation of structure from experimental restraints, the calculated structure of the proteins may be wrong. The overall fold of proteins can be analyzed from the ensemble of calculated structures. This can be visually judged by superposing the backbone atoms of all ensemble structures. If the structures have not converged to a single fold, this could indicate an erroneous structure.

Other NMR experiments for studying proteins

i. Labeling and filtering experiments

Labeling and filtering techniques employ the properties of a directly attached heteronucleus to select or destroy the signal of specific resonances. This technique can use for studying the complexes between labeled and unlabeled molecules and also use to directly determine interaction details between sidechains of peptides in order to reduce the amount of cross-peaks from both molecules. This technique allows us to observe only adjacent atoms from unlabeled molecules with labeled molecules.

ii. Chemical shift perturbation mapping

Chemical shifts are sensitive probes for both the local environment and the local structure. This can be used for studying the changes in chemical shifts as a function of an interaction partner. Interaction of peptide with lipid environment e.g. micelles and responsible residues can be determined by analyzing the chemical shift perturbation. Furthermore, this mapping can be used to derive the structure restraints for calculated structures of both partner by program HADDOCK (Dominguez *et al.*,2003).

iii. Relaxation

The advantage of NMR techniques is to use for studying the dymanic behavior of biomolecules. The dynamic behavior of the N-H bond vector can be derived from the relaxation rates of the ^{15}N nucleus. Typically, three rates are measured, ^{15}N T_1 (longitudinal) relaxation rates, ^{15}N T_2 (transverse) relaxation rates, and {1H}-15N-

NOE. These data can be measured on each residue which can use to map the dynamic of molecules. One popular way to interpret the measured rate is the 'model-free' approach. Model-free was a widely method to explain the dynamic of each residue in the molecule (Palmer *et al.*,1991) (Mandel *et al.*,1995). Deriving model-free equation gives results of both overall rotational correlation time T_c, local correlation time T_e, and order parameter S^2. S^2 is a measurement for the degree of structure with S^2=1.0 indicating fully ordered and S^2=0 indicating random order.

iv. Hydrogen/ Deuterium exchange

The presence of secondary structural units in proteins can slow down the Hydrogen/ Deuterium exchange in deuterium solution. This information can use to define the H-bond restraints (Berjanskii *et al.*,2000)**.** This can also use to probe the protein folding (Englander and Mayne,1992).

v. Diffusion experiment

Pulse-field gradient (PFG) NMR spectroscopy is a well-established technique for the measurement of diffusion coefficients by using a pulsed field gradient longitudinal eddy-current delay experiment (Altieri *et al.*,1995). For a single diffusion species, the signal attenuation in the presence of a single pair of pulsed field gradients is given by

$$I = I_0 \exp [-D (\Delta - \delta/3 - \tau/2)G^2\gamma^2\delta^2] \quad (5)$$

Where I_0 is the intensity of the signal in the absence of a gradient pulse, Δ is the diffusion delay, δ is the duration of the bipolar gradient pulse pair, τ is the time between the positive and negative gradient pulses, G is the gradient amplitude. γ is the gyromagnetic ratio

The diffusion coefficient of molecule can use to determine the relative size of molecule since it is inversely proportional to the hydrodynamic radius by the Stokes-Einstein relationship. The major changes in conformation that are accompanied by a change in the hydrodynamic radius including self-association and conformational changes should be detected through the change in the molecular diffusion coefficients.

Conclusion

The recent development have been tried to extend the limitation of protein NMR techniques in many way. For example, the large size and slow tumbling of the full-length EGFR nevertheless may be possible to study by NMR spectroscopy using newly developed techniques such as TROSY- and CRINEPT-based experiments (Riek *et al.*,2000) (Weigelt,1998). However, significant optimizing conditions and tremendous efforts will be needed for this study. Furthermore, the 2 -4 weeks of time consuming in spectra acquiring process and few months for data analysis are a drawback for NMR techniques when comparing to X-ray crystallography which takes only few minutes of acquiring time and few days of data analysis. The tremendous efforts have been tried to reduce these limitations by a development of reduced-dimensional spectroscopy such as G-matrix Fourier transform (GFT) NMR spectroscopy (Atreya and Szyperski,2004) together with development of automated assignment strategy in NMR spectra which can save more than 90% of both processings. In the next decade, protein NMR techniques may overcome all the technical limitation and may use as a tool in any laboratory as molecular biology does today.

References

Altieri, A., Hinton, D. P., and Byrd, R. A. 1995. Association of biomolecular systems via plused field gradient NMR self-diffusion measurements. J. Am. Chem. Soc. 117:7566-7567.

Atreya, H. S. and Szyperski, T. 2004. G-matrix Fourier transform NMR spectroscopy for complete protein resonance assignment . Proc. Natl. Acad. Sci. U. S. A. 101:9642-7.

Berjanskii, M. V., Riley, M. I., Xie, A., Semenchenko, V., Folk, W. R., and Van Doren, S. R. 2000. NMR structure of the N-terminal J domain of murine polyomavirus T antigens. Implications for DnaJ-like domains and for mutations of T antigens. J. Biol. Chem. 275:36094-103.

Bloch, F. 1946. Nuclear induction. Phys. Rev. 70:460-474.

Braun, W., Bosch, C., Brown, L. R., Go, N., and Wüthrich, K. 1981. Combined use of proton-proton Overhauser enhancements and a distance geometry algorithm for determination of polypeptide conformations. Application to micelle-bound glucagon. Biochim. Biophys. Acta. 667:377-396.

braunschweiler, L. and Ernst, R. R. 1983. Coherence transfer by isotopic mixing: application to proton correlation spectroscopy. J. Magn. Reson. 53:521-528.

Brunger, A. T. (1992) *X-PLOR version 3.1. A system for X-ray crstallography and NMR,* Yale University Press, New Haven, CT.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta. Crystallogr. D Biol. Crystallogr. 54 (Pt 5):905-921.

Brunger, A. T. and Nilges, M. 1993. Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy. Q. Rev. Biophys. 26:49-125.

Cornilescu, G., Delaglio, F., and Bax, A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR. 13:289-302.

Dominguez, C., Boelens, R., and Bonvin, A. M. 2003. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. 125:1731-7.

Duggan, B. M., Legge, G. B., Dyson, H. J., and Wright, P. E. 2001. SANE (Structure Assisted NOE Evaluation): an automated model-based approach for NOE assignment. J. Biomol. NMR. 19:321-329.

Englander, S. W. and Mayne, L. 1992. Protein folding studied using hydrogen-

exchange labeling and two-dimensional NMR. Annu. Rev. Biophys. Biomol. Struct. 21:243-65.

Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) *The principles of nuclcear magnetic resonance in one and two dimensions,* Clarendon Press, Oxford

Fischer, M. W., Losonczi, J. A., Weaver, J. L., and Prestegard, J. H. 1999. Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. Biochemistry. 38:9013-9022 .

Gronwald, W., Moussa, S., Elsner, R., Jung, A., Ganslmeier, B., Trenner, J., Kremer, W., Neidig, K. P., and Kalbitzer, H. R. 2002. Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). J. Biomol. NMR. 23:271-287.

Grzesiek, S. and Bax, A. 1992. Improved 3D triple resonance NMR techniques applied to a 31 kDa protein. J. Magn. Reson. 96:432-440.

Guntert, P., Mumenthaler, C., and Wüthrich, K. 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. 273:283-298.

Hansen, M. R., Mueller, L., and Pardi, A. 1998. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat. Struct. Biol. 5:1065-1074.

Havel, T. F., Crippen, G. M., Kuntz, I. D., and Blaney, J. M. 1983a. The combinatorial distance geometry method for the calculation of molecular conformation. II. Sample problems and computational statistics. J. Theor. Biol. 104:383-400.

Havel, T. F., Kuntz, I. D., and Crippen, G. M. 1983b. The combinatorial distance geometry method for the calculation of molecular conformation. I. A new approach to an old problem. J. Theor. Biol. 104:359-381.

Herrmann, T., Guntert, P., and Wüthrich, K. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. 319:209-227.

Huang, Y. J., Swapna, G. V., Rajan, P. K., Ke, H., Xia, B., Shukla, K., Inouye, M., and Montelione, G. T. 2003. Solution NMR structure of ribosome-binding factor A (RbfA), a cold-shock adaptation protein from Escherichia coli. J. Mol. Biol. 327:521-536.

Ikura, M., Kay, L. E., and Bax, A. 1990. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry. 29:4659-4667.

Ishii, Y., Markus, M. A., and Tycko, R. 2001. Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel. J. Biomol. NMR. 21:141-151.

Karplus, M. 1963. Vicinal Proton Coupling in Nuclear Magnetic Resonance. J. Am. Chem. Soc. 85:2870-2871.

Kay, L. E., Xu, G., Singer, A. U., Muhandiram, D. R., and Forman-Kay, J. D. 1993. A gradient enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlation in H_2O samples of proteins. J. Magn. Reson. B. 101:333-337.

Kumar, A., Ernst, R. R., and Wüthrich, K. 1980. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem. Biophys. Res. Commun. 95:1-6.

Laskowski, R. A., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. 1993. PROCHECK: a program to check stereochemical quality of protein structure. J. Appl. Cryst. 26:283-291.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR. 8:477-486.

Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. 2003. ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics. 19:315-316.

Mandel, A. M., Akke, M., and Palmer, A. G. 3rd 1995. Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. J. Mol. Biol. 246:144-163.

Meiler, J., Blomberg, N., Nilges, M., and Griesinger, C. 2000. A new approach for applying residual dipolar couplings as restraints in structure elucidation. J. Biomol. NMR. 16:245-252.

Mollova, E. T. and Pardi, A. 2000. NMR solution structure determination of RNAs. Curr. Opin. Struct. Biol. 10:298-302.

Montelione, G. T., Lyons, B. A., Emerson, S. D., and Tashiro, M. 1992. An efficient triple resonance experiment using carbon 13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. J. Am. Chem. Soc. 114:10974-10975.

Nilges, M., Clore, G. M., and Gronenborn, A. M. 1988. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. FEBS. Lett. 229:317-324.

Ottiger, M. and Bax, A. 1998. Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. J. Biomol. NMR. 12:361-372.

Palmer, A. G., Rance, M., and Wright, P. E. 1991. Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Dectected Natural Abundance 13C Heteronuclear NMR Spectroscopy. J. Am. Chem. Soc. 113:4371-4380.

Pan, B., Li, B., Russell, S. J., Tom, J. Y., Cochran, A. G., and Fairbrother, W. J. 2002. Solution structure of a phage-derived peptide antagonist in complex with vascular endothelial growth factor. J. Mol. Biol. 316:769-787.

pascal, S. M., Muhandiram, D. R., Yamazaki, T., Formankay, J. D., and Kay, L. E. 1994. Simultaneous Acquisition of ^{15}N- and ^{13}C-Edited NOE Spectra of Proteins Dissolved in H_2O. J. Mag. Reson. B. 103:197-201.

Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. 1983. Improved spectral resolution in cosy 1H NMR spectra of proteins via double quantum filtering. Biochem. Biophys. Res. Commun. 117:479-485.

Riek, R., Pervushin, K., and Wüthrich, K. 2000. TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. Trends. Biochem. Sci. 25:462-468.

Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T., and 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. 18:303-309.

Scheek, R. M., van Gunsteren, W. F., and Kaptein, R. 1989. Molecular dynamics simulation techniques for determination of molecular structures from nuclear magnetic resonance data. Methods. Enzymol. 177:204-218.

Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. 2003. The Xplor-NIH NMR molecular structure determination package. J. Magn. Reson. 160:65-73.

Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. J. Mol. Graph. 8:52-56.

Vuister, G. W. and Bax, A. 1993. Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins . J. Am. Chem. Soc. 115:7772-7777.

Weaver, J. L. and Prestegard, J. H. 1998. Nuclear magnetic resonance structural and ligand binding studies of BLBC, a two-domain fragment of barley lectin. Biochemistry. 37:116-128.

Weigelt, J. 1998. Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments. J. Am. Chem. Soc. 120:10778-10779 .

Williamson, M. P., Havel, T. F., and 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.

Wittekind, M. and Mueller, L. 1993. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances. J. Magn. Reson. B. 101:201-205.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids,* Wil ey, John & Sons, New York

Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. 1994. A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity. J. Am. Chem. Soc. 116:11655-11666.