
Protein Structure Determination from NMR Data

7.1. INTRODUCTION AND HISTORICAL OVERVIEW

Although NMR was discovered in 1946, its application to biological systems only started in the late 1960s and early 1970s. The application was very limited due to the poor sensitivity and very low resolution offered by the one-dimensional techniques used at that time. Two major breakthroughs in the 1970s revolutionized the field: Fourier transformation (FT) NMR that allowed rapid recording of NMR signals and two-dimensional NMR spectroscopy that dramatically increased the spectral resolution. These advances in combination with the development of stable magnets at higher fields led to explosive investigations using NMR in the late 1970s and early 1980s, which centered on exploring its potential in determining the three-dimensional structures of macromolecules. Even though X-ray crystallography was already a method of choice for structure determination during that period, it was believed that NMR may provide complementary structural information in a more physiologically relevant solution environment. Moreover, since some biomolecules are difficult to crystallize, NMR could be used as an alternative method for obtaining three-dimensional structures.

In the mid-1980s, several groups reported the first generation of solution structures of proteins and oligonucleotides using 2D NMR methods. The protocols used in these NMR structure calculations proved to be valid when the same structure of the α -amylase inhibitor Tendamistat was determined in 1986 independently by NMR and crystallographic groups. After that, the field witnessed an exponential growth with the excitement over NMR being another powerful method for macromolecule structure determination. However, it was soon realized that signal degeneracy and the intrinsic relaxation behavior of macromolecules limited 2D NMR application within the range of small proteins and nucleic acids (<10 kDa).

In the late 1980s and early 1990s, another quantum jump came when multidimensional heteronuclear NMR methods were developed thereby pushing the molecular size limit of NMR structures up to ~35 kDa. Advances in molecular biology that led to the overexpression and isotope labeling of proteins also played an important role. Hence multidimensional heteronuclear NMR has opened the door to studying a wide variety of proteins and protein domains. Two recent developments including TROSY and residual dipolar coupling (RDC) now promise to allow NMR to study even larger proteins and protein complexes. Hence, although NMR is still in its developmental stage and lags behind macromolecular crystallography by almost 30 years (the first crystal structure of a protein was published in 1957 whereas the first NMR structure

came to the world in the mid-1980s), it has certainly become one of the most powerful players in molecular/structural biology. Today, one fifth of the macromolecular structures deposited in the PDB (Protein Databank) were derived from NMR spectroscopy. Despite its size limitation for macromolecular structure determination, NMR has the following unique features: (a) it allows structural studies in a physiologically relevant solution environment, which avoids experimental artifacts such as crystal packing seen in crystal structures; (b) it allows structural studies of some molecules that are difficult to crystallize such as flexible protein domains, weakly bound protein complexes, etc; (c) it can provide information about protein dynamics, flexibility, folding/unfolding transitions, etc (see Chapter 8). With the completion of the human genome, NMR will also play a major role in the post-genome era in areas such as structural genomics and proteomics.

The outline for NMR-based structure determination (Wüthrich, 1986) shown in Figure 7.1 includes three stages: (a) sample preparation, NMR experiments, data processing; (b) sequence specific assignment, NOESY assignment, assignments of other conformational restraints such as J coupling, hydrogen bonding, dipolar coupling; and (c) structure calculation and structure refinement. One starts with a well-behaved sample (protein, nucleic acid, etc.) and performs a suite of NMR experiments designed for resonance assignment and structural analysis. Four important parameters are generated for structure calculations: (a) chemical shifts that provide mostly secondary structural information for proteins; (b) J coupling constants that provide geometric angles within molecules; (c) nuclear Overhauser effects (NOEs) that provide ^1H - ^1H distances within 5 Å. The NOE data are considered to be the most important and are rich in providing especially tertiary structural information; the more recently introduced RDCs can provide valuable structural restraints. This fourth parameter is complementary to NOE data since it provides long-range distance information (>5 Å) whereas the NOE data is restricted to <5 Å. Each of the parameters will first briefly be described and then a protocol will be used to

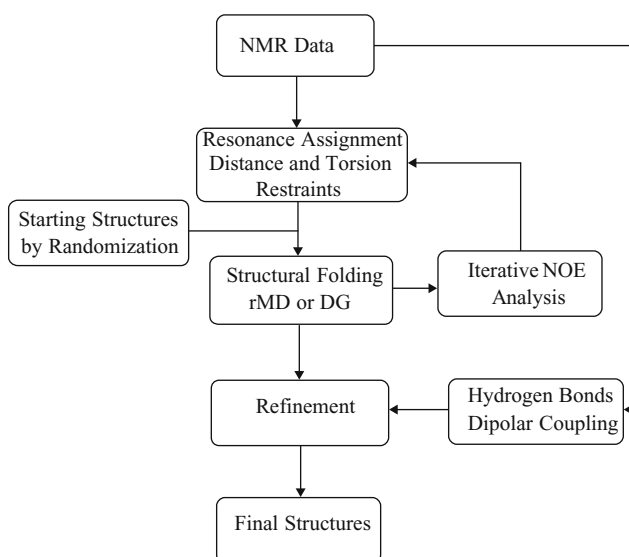


Figure 7.1. Strategy for NMR-based structure determination.

describe in detail how a structure is calculated and how each parameter is implemented during the structure calculation.

Key questions to be addressed in the current chapter include:

1. What types of structural information can be obtained based on the results of the resonance assignments?
2. What are the methods currently used in the structure calculation from NMR data?
3. How are J coupling constants used in the structure determination of a protein?
4. What other nuclear interactions can be utilized for structure determination?
5. How are the NMR parameters utilized in structure calculation?
6. What are the strategies to carry out the calculation?
7. How is the structural quality analyzed?
8. How are the precision and accuracy of the calculated structures determined?
9. What is the role of iterative NOE analysis during the structure calculation?
10. Step-by-step illustration of structural calculation using a typical XPLOR protocol.

7.2. NMR STRUCTURE CALCULATION METHODS

To date, the majority of structures characterized by NMR spectroscopy are obtained using distance and orientational restraints derived from NOE, the J coupling constant, RDC, and chemical shift measurements. The calculation of a three-dimensional structure is usually formulated as a minimization problem for a target function that measures the agreement between a conformation and the given set of restraints. There are several algorithms developed over the past two decades, of which four are widely used (Table 7.1): (a) metric matrix distance geometry, represented by the DG-II protocol; (b) variable target function method, represented by the DIANA protocol; (c) Cartesian space or torsion angle space restrained molecular dynamics (rMD), represented by the XPLOR (or CNS) protocol; (d) torsion angle dynamics, represented by the DYANA protocol. These methods generate and refine biomolecular structures by searching globally to get an ensemble of molecular structures that fit with the experimentally measured restraints within the range of experimental error. Distance geometry and rMD (simulated annealing) calculations are discussed in this section.

TABLE 7.1
Structure Calculation Methods and Programs

| Method | Program ^a | References |
|--|----------------------|------------------------------------|
| Metric matrix distance geometry | DIG-II | Havel (1991) |
| Variable target function method | DIANA | Güntert <i>et al.</i> (1991) |
| Cartesian space or torsion angle space rMD | AMBER | Pearlman <i>et al.</i> (1991) |
| | CHARM | Brooks <i>et al.</i> (1983) |
| | GROMOS | van Gunsteren <i>et al.</i> (1996) |
| | XPLOR | Brünger <i>et al.</i> (1992) |
| | CNS | Brünger <i>et al.</i> (1998) |
| Torsion angle dynamics | DYANA | Güntert <i>et al.</i> (1997) |

^aThese are the programs currently used for structure calculation.

7.2.1. Distance Geometry

One approach is to generate structures with the distance and orientational restraints derived from NOEs and J coupling constants using a metric matrix or dihedral angle space distance algorithm. The metric matrix method utilizes all possible interatomic distances as restraints, including the known distances from covalent bonds and experimentally estimated distances from NOE data to generate an n -dimensional matrix for a molecule with n atoms (Crippen, 1981; Braun, 1987). For the remaining distances, upper and lower bounds are chosen and altered until no further alterations can be made according to the triangle inequalities:

$$\begin{aligned} u_{ij} &\leq u_{ik} + u_{kj} \\ l_{ij} &\leq l_{ik} + l_{kj} \end{aligned} \quad (7.1)$$

in which i, j, k are the three atoms defining a triangle, and u and l are the upper and lower distance bounds between any two given points of the triangle. An ensemble of structures is generated from randomly selected distances within the boundary conditions.

The initial structures obtained by the randomization are further refined by minimizing a penalty function V (or potential function) such as:

$$V = \sum_{i>j} \begin{cases} k(r_{ij}^2 - l_{ij}^2)^2 & \text{if } r_{ij} < l_{ij} \\ k(r_{ij}^2 - u_{ij}^2)^2 & \text{if } r_{ij} < u_{ij} \\ 0 & \text{if } l_{ij} \leq r_{ij} \leq u_{ij} \end{cases} \quad (7.2)$$

in which r_{ij} is the distance between atoms i and j , k is a weighting factor, u, l are defined as in Equation (7.1). The true global minimum is found only if $V = 0$, meaning that all restraints are satisfied. However, in practice, V is always greater than zero because of the insufficient number of restraints available. The function V is also called a target function. The minimization is achieved by comparing the interproton distances of the calculated structures with the distances within the chosen boundaries. The improved distance geometry uses dihedral angles rather than Cartesian coordinates to fold protein structures based on the short-range restraints and then expands the calculation to eventually include all restraints. These distance geometry methods (e.g., the DIANA program) have played an important role in the determination of protein structures by solution NMR.

7.2.2. Restrained Molecular Dynamics

Restrained molecular dynamics methods, calculate the structures using NMR experimental restraints and energy minimization with potential energy functions similar to the above restrained potential energy function. The potential energy (or target function) is calculated for an array of initial atomic coordinates based on a series of potential energy functions (van Gunsteren, 1993):

$$V_{\text{tot}} = V_{\text{classic}} + V_{\text{NOE}} + V_J \text{ coupling} + V_{\text{Hbond}} + V_{\text{dipolar}} + V_{\text{cs}} + V_{\text{other}} \quad (7.3)$$

in which V_{classic} is the potential function from the classic energy of the molecule, which contains $V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}} + V_{\text{van der waals}} + V_{\text{electrostatic}}$. The rest of Equation (7.3)

takes the NMR data in terms of distances derived from NOE, torsion angles from J coupling constants, molecular bond orientation restraints from dipolar couplings, chemical shifts and other restraints such as disulfide bridges, hydrogen bonding, and planarity. Although torsion angles and dipolar coupling restraints are sometimes not used, NOE distance restraints are always used in the rMD calculations. Among several different functions used to characterize the potential energy, a flat-well potential is frequently used, which consists of the energy contributions of NOE distance violations relative to the lower and upper distance bounds (Clore *et al.*, 1986):

$$\begin{aligned}
 V_{\text{NOE}} &= \sum_i^{\text{all NOEs}} V_{\text{NOE}i} \\
 &= \sum_i^{\text{all NOEs}} \begin{cases} k_{\text{NOE}}(r_i - r_u)^2 & \text{if } r_i > r_u \\ k_{\text{NOE}}(r_i - r_l)^2 & \text{if } r_i < r_l \\ 0 & \text{if } r_l < r_i < r_u \end{cases} \quad (7.4)
 \end{aligned}$$

in which k_{NOE} is the force constant of the NOE potential function (also called the target function), r_l and r_u are the lower and upper distance bounds for individual NOE intensities, respectively, and r_i is the interproton distance for each proton pair from the generated structure. If the precise interproton distances (instead of a range of distances) are obtained from NOESY spectra with different mixing times, the NOE potential function can be described using a biharmonic potential:

$$V_{\text{NOE}} = \sum_i^{\text{all NOEs}} \begin{cases} C_1(r_i - r_0)^2 & \text{if } r_i > r_0 \\ C_2(r_i - r_l)^2 & \text{if } r_i < r_0 \end{cases} \quad (7.5)$$

in which C_1 and C_2 are the force constants which are temperature dependent and r_i and r_0 are the calculated and experimental distance, respectively.

An alternative method to calculate the NOE potential energy is to compare the calculated NOE intensities from the structure to the experimental NOE intensities at any given step during an rMD calculation using the NOE potential function (Brünger, 1992):

$$V_{\text{NOE}} = k(\mathbf{a}_{\text{exp}} - \mathbf{a}_{\text{cal}})^2 \quad (7.6)$$

in which \mathbf{a}_{exp} and \mathbf{a}_{cal} are, respectively, matrices of experimental NOE cross-peak intensities used for the calculation and calculated intensities from the structure obtained by the MD simulation.

The energy barriers between local minima are more easily overcome in rMD because molecular dynamics is used in the energy minimization, which makes the method less sensitive to the initial structures (Allen and Tildesley, 1987). An MD simulation is performed by solving Newton's equations of motion using the forces generated by varying the potential energies of the macromolecular structures. A minimum energy structure is obtained by solving the first derivative of the potential energy with respect to the coordinates of each atom using the condition that the derivative is zero. From Newton's law, the force for an individual atom

can be written as:

$$F = ma = -\frac{dV}{dr} \quad (7.7)$$

or

$$-\frac{dV}{dr} = m \frac{d^2r}{dt^2} \quad (7.8)$$

in which m is the mass of the atom, a is the acceleration, V is the potential energy, t is time, and r the coordinates of the atom. The equation of motion is solved by numerical integration algorithms, and the trajectory for each atom as a function of time is calculated.

In order to maintain an accurate and stable simulation, the time step should be kept sufficiently smaller than the fastest local motion of the molecule. Typically, the time step size chosen in the simulation is \sim fs (10^{-15} s) for simulations in a ps (10^{-12} s) timescale. During the simulation, energy barriers of the system (whose amplitude approximately equals kT) are passed by raising the temperature of the system high enough to increase the kinetic energy during the simulation so that a global energy minimum is located by balancing both the classical energy terms and those fit with the experimental restraints. In the first stage of the calculation, an ensemble of initial structures is selected by randomization. The initial structures must gain kinetic energy, which is commonly provided by increasing the temperature of the simulated system, to move away from their local energy minima and then pass over higher energy barriers. With the higher energy, the system contains a greater range of structural space. The system is then slowly cooled down to room temperature. During the cooling process the system energy is minimized over the surface of potential energy to search for stable structures at low temperature. The cycles of heating and cooling are repeated until an ensemble of stable structures with an acceptable penalty (or violations) is eventually determined.

A typical procedure (Güntert, 1998, 2003; de Alba and Tjandra, 2002; Lipstitz and Tjandra, 2004) for structure calculation includes (a) a stage of randomization in which a set of initial structures is generated with the idealized covalent geometry restraints such as bond length, bond angles, dihedral angles, and improper torsions; (b) global folding in which a variety of energy terms with both the geometry restraints and experimentally obtained distance and torsion restraints are used to obtain folded structures; and (c) refinement which utilizes the same energy terms as in the previous stage but in a smaller step size (typical \sim fs) for ps MD processes to refine the structures generated in the previous stage. The refinement can also involve RDCs in which the structures are refined using observed RDCs by increasing dipolar force constants slowly and simultaneously refining the principal components of the alignment tensor.

7.3. NMR PARAMETERS FOR STRUCTURE CALCULATION

7.3.1. Chemical Shifts

In principle, chemical shifts of NMR-active nuclei such as ^1H , ^{15}N , and ^{13}C are dictated by the structural and chemical environment of the atoms (see Chapter 1). Vigorous efforts have

been made to deduce protein structure from chemical shifts. Chemical shift-based secondary structural prediction has met with some success. In particular, the deviations of $^{13}\text{C}^\alpha$ (and, to some extent, $^{13}\text{C}^\beta$) chemical shifts from their random coil values can be well correlated with the α -helix or β -sheet conformations: $^{13}\text{C}^\alpha$ chemical shifts larger than the random coil values tend to occur for helical residues whereas the opposite is observed for β -sheet residues (Spera and Bax, 1991; de Dios *et al.*, 1993, Wishart *et al.*, 1991, 1992). A good correlation was also observed for proton shifts with secondary structures: $^1\text{H}^\alpha$ shifts smaller than the random coil values tend to occur for helical residues whereas the opposite is observed for β -sheet residues. Although the information is useful for tertiary structure calculations (Luginbühl *et al.*, 1995; Kuszewski *et al.*, 1995a, b), it is much more valuable for the initial secondary structural analysis in combination with NOE data (see below).

Methods have been developed to predict the dihedral angles using backbone chemical shifts, such as the programs TALOS (torsion angle likelihood obtained from the shift and sequence similarity, Cornilescu *et al.*, 1999) and SHIFTOR (Zhang, 2001). The prediction is based on the observation that similar amino acid sequences with similar backbone chemical shifts have similar backbone torsion angles. First, TALOS breaks the sequence of a target protein into overlapping amino acid triplets. Then, for each triplet, the program searches its database which contains proteins with known chemical shifts ($^1\text{H}^\alpha$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}'$, and ^{15}N) and high-resolution X-ray crystal structures to compare the chemical shift and sequence similarity. The torsion angles for the central residue from the best 10 matches are chosen as the predicted torsion angles for the residue, which are used as backbone dihedral angles in the structure calculation. Typically, TALOS can predict the dihedral angles for $\sim 70\%$ of the residues within $\pm 20^\circ$. The incorrect predictions can be removed by the inconsistency with other types of constraints during structure calculation.

7.3.2. J Coupling Constants

J coupling constants are derived from the scalar interactions between atoms. They provide geometric information between atoms in a molecule. The most useful and obtainable coupling constants are vicinal scalar coupling constants, 3J , between atoms separated from each other by three covalent bonds. Its relation with dihedral angle θ is defined as follows (Karplus

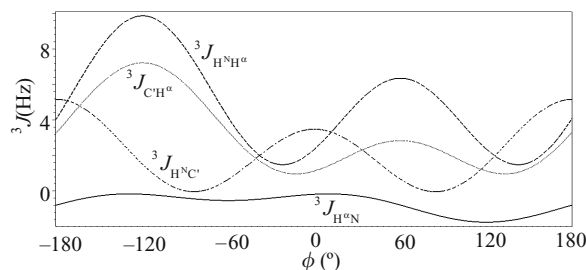


Figure 7.2. Karplus curves describing the relationships of vicinal J coupling constants and torsion angle ϕ using the constants listed in Table 7.2. The solid line is for the coupling between protons, dotted lines for the heteronuclear couplings. The angle $\theta = \phi + \text{offset}$.

TABLE 7.2
Karplus Constants

| Coupling constant | Torsion angle | Dihedral | A | B | C | Offset (°) | References |
|--|---------------|---|-------|-------|-------|------------|--------------------------------|
| $^3J_{\text{H}^{\text{N}}\text{H}^{\alpha}}$ | ϕ | $\text{H}^{\text{N}}-\text{N}-\text{C}^{\alpha}-\text{H}^{\alpha}$ | 6.51 | -1.76 | 1.6 | -60 | Vuister and Bax (1993) |
| | | | 6.98 | -1.38 | 1.72 | -60 | Wang and Bax (1996) |
| $^3J_{\text{H}^{\text{N}}\text{C}'}$ | | $\text{H}^{\text{N}}-\text{N}-\text{C}^{\alpha}-\text{C}'$ | 4.32 | 0.84 | 0.00 | 180 | Wang and Bax (1996) |
| $^3J_{\text{C}'_{i-1}\text{H}^{\alpha}}$ | | $\text{C}'-\text{N}-\text{C}^{\alpha}-\text{H}^{\alpha}$ | 3.75 | 2.19 | 1.28 | 120 | Wang and Bax (1996) |
| $^3J_{\text{H}^{\alpha}\text{N}_{i+1}}$ | ψ | $\text{H}^{\alpha}-\text{C}^{\alpha}-\text{C}'-\text{N}$ | -0.88 | -0.61 | -0.27 | -120 | Wang and Bax (1995) |
| $^3J_{\text{H}^{\alpha}\text{H}^{\beta}}$ | χ^1 | $\text{H}^{\alpha}-\text{C}^{\alpha}-\text{C}^{\beta}-\text{H}^{\beta}$ | 9.50 | -1.60 | 1.80 | -120/0 | de Marco <i>et al.</i> (1978a) |
| $^3J_{\text{NH}^{\beta}}$ | | $\text{N}-\text{C}^{\alpha}-\text{C}^{\beta}-\text{H}^{\beta}$ | -4.40 | 1.20 | 0.10 | 120/-120 | de Marco <i>et al.</i> (1978b) |
| $^3J_{\text{C}'\text{H}^{\beta}}$ | | $\text{C}'-\text{C}^{\alpha}-\text{C}^{\beta}-\text{H}^{\beta}$ | 7.20 | -2.04 | 0.60 | 0/120 | Fischman <i>et al.</i> (1980) |

equation, Karplus, 1959, 1963):

$$^3J(\theta) = A \cos^2 \theta + B \cos \theta + C \quad (7.9)$$

in which A , B , C are coefficients for various types of couplings, and θ is the dihedral angle. Using the Karplus relationship, one can convert the J coupling constants into the dihedral angles, commonly, ϕ , ψ , and χ^1 . The dihedral angles can be determined by best fitting the measured J values to the corresponding values calculated with the above equation for known protein structures (see Figure 7.2). These dihedral angles can be used as structural restraints later during calculations (see below). Table 7.2 lists 3J commonly used for deducing various dihedral angles for proteins.

7.3.3. Nuclear Overhauser Effect (NOE)

NOEs are the most important NMR parameters for structure determination because they provide short-range as well as long-range distance information between pairs of hydrogen atoms separated by less than 5 Å. Whereas the short-range NOEs are valuable for defining secondary structure elements such as α -helix or β -sheet, the long-range NOEs provide crucial tertiary structural information (Wüthrich, 1986). The intensity of an NOE (I) is related to the distance r between the proton pair, $I = f(\tau_c) < r^{-6} >$ in which $f(\tau_c)$ is a function of the rotational correlation time τ_c of the molecule. Because of many technical factors such as highly variable τ_c for different molecules at different temperatures and solvent conditions, it is common to use intensity I (or cross-peak volume) to obtain qualitative distance information. The information is usually grouped into three different distance categories: 1.8–2.5 Å (strong), 1.8–3.5 Å (medium), and 1.8–5.0 Å (weak). Note the lower bound for all three categories is 1.8 Å, corresponding to the van der Waals repulsion range. This treatment is due to the consideration that weak NOEs may not be related to longer distances such as >4 Å. Instead, they may be related to the chemical exchange or protein motions that diminish the NOE intensities.

When performing a NOESY experiment to obtain NOE information, it is important to choose a proper mixing time that is in principle proportional to the distance (NOE intensity). Short mixing times may lead to the loss of weak NOEs that may contain important tertiary structural information. However, long mixing times may induce so-called “spin-diffusion,”

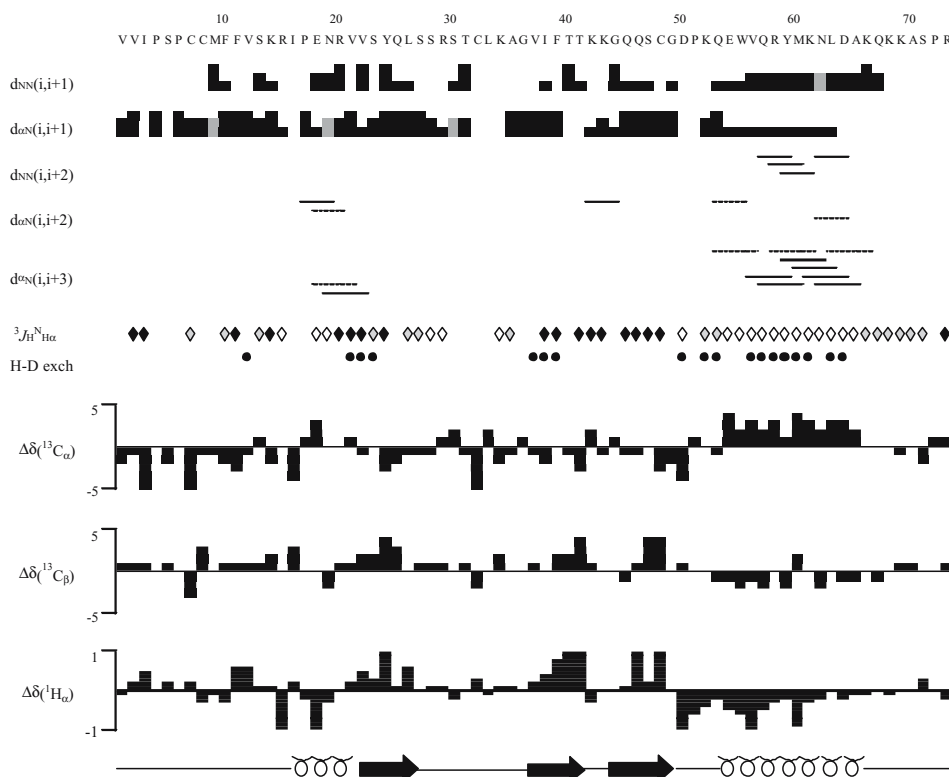


Figure 7.3. Experimental restraints for eotaxin-2, including NH exchange $^3J_{\text{HNH}\alpha}$ coupling constants, sequential, short- and medium-range NOEs and H^α , C^α , and C^β secondary shifts, along with the secondary structure deduced from the data (reproduced with permission from Mayer and Stone, *Biochemistry*, 39, 8328, 2000. Copyright © 2000 American Chemical Society). The amino acid sequence and numbering are shown at top. Sequential N–N and α –N NOEs are indicated by black bars; the thickness of the bar represents the strength of the observed NOE. The presence of medium-range N–N and α –N NOEs is indicated by solid lines. Gray bars and dashed lines represent ambiguous assignments. $^3J_{\text{HNH}\alpha}$ coupling constants are represented by diamonds corresponding to values of <6 Hz (open), 6–8 Hz (gray), and >8 Hz (black). Residues whose amide protons show protection from exchange with solvent are indicated with filled circles. The chemical shift indices shown for C^α , C^β , and H^α were calculated according to the method developed by Wishart *et al.* (1992). The locations of the secondary structure elements identified in the calculated family of structures are shown at the bottom.

that is, NOEs indirectly generated by spins in the vicinity $>5 \text{ \AA}$. The mixing time can be accurately determined by analyzing an NOE build-up curve (Neuhaus and Williamson, 1989). However, the build-up curves vary considerably among different spins in the same molecule such as between methylene protons and methyl protons. A compromise is usually given to suppress spin diffusion and to maintain sufficient NOE intensities based on NOE build-up curves. As a rule of thumb, 80–120 ms is usually used for small–medium sized proteins. Sometimes, NOESY experiments with several different mixing times are performed to make sure the “spin-diffusion” peaks are not picked.

Several programs are available for NOE analyses such as nmrvue and PIPP. These programs can store all the assigned NOEs in a table that can be converted into distance format

for structure calculations (see Appendix B1). The assigned NOEs can also be plotted as a function of protein sequence to gain information about the protein secondary structural information and topology of the tertiary fold in conjunction with chemical shifts and J coupling information (Figure 7.3).

7.3.4. Residual Dipolar Couplings

Between the early 1980s and late 1990s, all NMR structures were determined based primarily on NOE data supplemented by J coupling constants and chemical shifts. In the late 1990s, a new class of conformational restraints emerged, which originate from internuclear residual dipolar coupling in weakly aligned media such as bicelles (Tjandra and Bax, 1997; Prestegard *et al.*, 2000; de Alba and Tjandra, 2002; Lipstitz and Tjandra, 2004). The RDC gives information on angles between covalent bonds and on long-range order. The addition of this structural parameter has proven to greatly improve the precision as well as the accuracy of NMR structures.

Although internuclear DD (dipole–dipole) couplings are typically averaged out due to molecular tumbling, RDC occurs when there is a small degree of molecular alignment with the external magnetic field (see Chapter 1). The RDCs are manifested as small, field-dependent changes of the splitting normally caused by one-bond J couplings between directly bound nuclei. With the assumption of an axially symmetric magnetic susceptibility tensor and neglecting the contribution from “dynamic frequency shifts,” the frequency difference $\Delta\nu^{\text{obs}}$ between the apparent J -values at two different magnetic field strengths, B_0^1 and B_0^2 , is given by (Tjandra *et al.*, 1997):

$$\Delta\nu^{\text{obs}} = \frac{\hbar\gamma_a\gamma_b\chi_a S}{30\pi r k T} (B_0^2 - B_0^1) (3\cos^2\theta - 1) \quad (7.10)$$

in which \hbar is Planck’s constant divided by 2π , k is the Boltzmann constant, T is the temperature in Kelvin, χ_a is the axial component of the magnetic susceptibility tensor, S is the order parameter for internal motion, r is the distance between coupled nuclei a and b, and γ_a and γ_b are the gyromagnetic ratios of a and b, respectively. The structural information is contained in the angle θ between the covalent bond formed by two scalar coupled atoms a and b and the main axis of the magnetic susceptibility tensor. It is then straightforward to add an orientational restraint term to the target function of a structure calculation program that measures the deviation between the experimental and calculated values of θ .

An alternative way to obtain RDC orientational restraints without the assumption of an axially symmetric magnetic susceptibility tensor is to obtain the magnitude and relative orientation of the alignment tensor. The value for the RDC ($\Delta\nu_D$) is extracted from the difference between the splittings observed in alignment medium ($\Delta\nu_A$) and in isotropic solution ($\Delta\nu_J$):

$$\Delta\nu_D = \Delta\nu_A - \Delta\nu_J \quad (7.11)$$

The dipolar couplings can be determined using IPAP type experiments (see Chapter 5). As discussed in Chapter 1, the RDC provides orientational information according to Equation (1.68). In order to use RDCs as structural restraints, the magnitudes of the axial and rhombic components of the alignment tensor and their relative orientation with respect to the magnetic field

must be determined. The magnitude and rhombicity of the alignment tensor can be obtained by examining the powder pattern distribution of all normalized observed dipolar couplings for the molecule. When structures are calculated, all of the variables can be obtained by fitting the equation with a large number of RDCs. If the structures are accurate, the calculated dipolar couplings of the structures will be in good agreement with the observed RDCs within the experimental error range. Such orientational restraints have been shown to improve the quality of the structures. They are also extremely valuable when calculating protein complex structures.

7.4. PRELIMINARY SECONDARY STRUCTURAL ANALYSIS

Prior to structure calculations, it is useful to determine the secondary structure using a combination of chemical shifts, J coupling constants, and NOE data. As mentioned above, ^{13}C chemical shifts are particularly indicative of α -helices and β -sheets. Three-bond $J_{\text{NH-H}\alpha}$ coupling constants are often small for helical residues (<5 Hz) but large for β -sheet residues (>8 Hz). Regular secondary structure elements can also be easily identified from sequential NOEs, as each type of secondary structure element is characterized by a particular pattern of short range NOEs ($|r_i - r_j| < 5\text{\AA}$). For instance, α -helices are characterized by a stretch of strong and medium $\text{NH}_i\text{--NH}_{i+1}$ NOEs, and medium or weak $\text{C}^\alpha\text{H}_i\text{--NH}_{i+3}$, $\text{C}^\alpha\text{H}_i\text{--C}^\beta\text{H}_{i+3}$ NOEs, and $\text{C}^\alpha\text{H}_i\text{--NH}_{i+1}$ NOEs, sometimes supplemented by $\text{NH}_i\text{--NH}_{i+2}$ and $\text{C}^\alpha\text{H}_i\text{--NH}_{i+4}$ NOEs. β -strands, on the other hand, are characterized by very strong $\text{C}^\alpha\text{H}_i\text{--NH}_{i+1}$ NOEs and by the absence of other short-range NOEs involving the NH and C^αH protons. β -sheets can be identified and aligned from interstrand NOEs involving the NH, C^αH , and C^βH protons. Hydrogen-exchange experiments are also often performed to extract information for slowly exchanging amides that are normally involved in helices or β -sheets. This adds great confidence later when dealing with H-bonds of the backbone amides involved in helices or β -sheets.

A computer program written in a shell script can convert all the NOE, J coupling, and exchange data, and chemical shifts into a figure for analyzing the secondary structures of proteins. This is illustrated in Figure 7.3. Note that this approach tends to perform poorly in ill-defined secondary structural regions such as loops. In addition, the exact start and end of helices tends to be less accurate since the pattern of these parameters is similar to that present in turns. Thus, a turn at the end of a helix could be misinterpreted as being part of the helix. In the case of β -sheets, the definition of the start and end is more accurate as alignment is accomplished from interstrand NOEs involving NH and C^αH protons.

Another preliminary structural analysis is the stereospecific assignment of diastereotopic protons. There are two major types of diastereotopic protons in amino acids: (a) methylene protons in Lys, Arg, etc.; and (b) methyl groups of Val, Leu. If the signals are well-resolved, stereospecific assignments of β -methylene protons can be assigned by a combination of ^{15}N -edited TOCSY and ^{15}N -edited NOESY. Some methylene protons can also be stereospecifically assigned during the course of structure calculations. Stereospecific assignments of Val and Leu methyls can be made experimentally by the partial ^{13}C labeling or fractional deuteration method (Neri *et al.*, 1989; Senn *et al.*, 1989) provided the signals are well resolved. If signals are degenerate or weak, which prevents the stereospecific assignments, diastereotopic protons

have to be referred to as pseudoatoms, which may result in less-well-defined structures. Stereospecific assignments not only provide more accurate distance information but also provide dihedral angle information including χ^1 and χ^2 (Powers, *et al.*, 1991). Hence, it is important to have as many stereospecific assignments as possible in order to obtain a high quality structure.

7.5. TERTIARY STRUCTURE DETERMINATION

7.5.1. Computational Strategies

Because proteins typically consist of more than a thousand atoms that are restrained by thousands of experimentally determined NOE restraints in conjunction with stereochemical and steric conditions, it is in general neither feasible to do an exhaustive search of allowed conformations nor to find solutions by interactive model building. In practice, as mentioned in the previous section, the calculation of a molecular structure is performed by minimizing the target function that represents the agreement between a conformation and a set of experimentally derived restraints. In the following section, a step-by-step description of structure calculation is provided using the most widely used XPLOR protocol.

7.5.2. Illustration of Step-by-Step Structure Calculations Using a Typical XPLOR Protocol

General guidance for the rMD protocol is given in Table 7.3. A complete protocol for protein structural calculations using a simulated annealing XPLOR program (sa.inp) is provided in Appendix B1. The file names in bold require modifications for specific protein structure determination, and generally include the different input files such as distance restraints, PDB coordinates, etc. In the protocol, readers are referred to specific remarks on important lines such as “read the PSF file and initial structure,” which will help in understanding the protocol.

7.5.2.1. Preparation of Input Files

1. *Example of NOE table.* All the assigned NOEs in a peak-pick table generated by programs such as PIPP or nmrPipe can be converted into a distance restraint table using a shell script. An example of an XPLOR distance restraint table can be found in Appendix B2.
2. *Example of dihedral angle restraint table.* Dihedral angles derived from J coupling constants can be assembled into the format for the XPLOR program (Appendix B3).
3. *Example of chemical shift restraint table.* The carbon shifts for C^α and C^β can be formatted for the XPLOR program (Appendix B4).
4. *Example of H-bond table.* Although NMR experiments have been developed to directly measure the H-bonds, most of the H-bond restraints are still derived indirectly from amide exchange experiments. These H-bond restraints are normally used for structure refinement after the initial structure is calculated. The H-bond input table is as shown in Appendix B5.

TABLE 7.3
Structure Calculation Protocol Using rMD

Randomization

10 ps restrained molecular dynamics

Energy terms: bonds, angles, improper torsions

Temperature: 1,000 K

Number of calculated structures: 100

500 cycles of Powell energy minimization

Global Folding

5 ps restrained molecular dynamics

Energy terms: bonds, angles, improper torsions, NOE (Soft-square NOE potential), van der Waals (Lennard-Jones), van der Waals radii are scaled by 0.9

$k_{\text{NOE}} = 30 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, $k_{\text{dih}} = 10 \text{ kcal mol}^{-1} \text{ rad}^{-2}$

Step size: 5 fs

Number of steps: 1,000

Temperature: 2,000 K

15 ps restrained molecular dynamics while cooling to 300 K

34 cycles of 0.44 ps each

Energy terms: bonds, angles, improper torsions, NOE (Soft-square NOE potential), van der Waals (Lennard-Jones), van der Waals radii are gradually reduced by factor 0.9 to 0.8

k_{NOE} is gradually increased from 2 to 30 $\text{kcal mol}^{-1} \text{ \AA}^{-2}$,

$k_{\text{dih}} = 200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$

Step size: 5 fs

Total number of steps: 3,000

Temperature decrement 50 K per step

500 cycles of Powell energy minimization

Refinement

500 cycles of energy minimization

Energy terms: bonds, angles, improper torsions, NOE, van der Waals (Lennard-Jones)

Soft-square NOE potential

$k_{\text{NOE}} = 50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, $k_{\text{dih}} = 5 \text{ kcal mol}^{-1} \text{ rad}^{-2}$

2.5 ps restrained molecular dynamics while increasing force constants of all torsion angles

Energy terms: bonds, angles, improper torsions, soft-square NOE potential

20 cycles of 0.125 ps each

k_{NOE} : from 2.2 to 30 $\text{kcal mol}^{-1} \text{ \AA}^{-2}$, increased by a factor of 1.14 per cycle

k_{dih} : from 1 to 200 $\text{kcal mol}^{-1} \text{ rad}^{-2}$, increased by a factor of 1.304 per cycle

Step size: 0.5 fs

Temperature: 2,000 K

200 cycles of energy minimization

Energy terms: bonds, angles, improper torsions, NOE, van der Waals (Lennard-Jones)

Restrained molecular dynamics while cooling to 300 K

34 cycles of 0.44 ps each

Step size: 0.5 fs

Temperature decrement 50 K per step

500 cycles of Powell energy minimization

Refinement with dipolar couplings

Energy terms: bonds, angles, improper torsions, NOE, van der Waals (Lennard-Jones), electrostatic, dipolar

500 cycles of energy minimization

$k_{\text{dipo}} = 0.001 \text{ cal mol}^{-1} \text{ \AA}^{-2}$

15 ps restrained molecular dynamics

50 cycles of 0.3 ps each

k_{NOE} : from 0.001 to 0.2 $\text{kcal mol}^{-1} \text{ \AA}^{-2}$, increased by a factor of 1.11 per cycle

Step size: 0.5 fs

Temperature: 300 K

500 cycles of energy minimization

7.5.2.2. Preparation of Initial Random-Coil Coordinates and Geometric File

1. Input file to generate random-coil coordinates based on the protein sequence (Appendix B6).
2. Input file to generate geometric PSF file (Appendix B7). This file contains information on the molecular bonds, angles, peptide planes, etc. present in the structure (i.e., how the atoms are connected together).

7.5.2.3. Randomization

In the initial stage of the calculation, an array of random (or semi-random) initial structures is generated based on covalent geometry restraints including bond length, bond angles, dihedral angles, and improper torsions. After 10 ps of rMD dynamics is carried out at a temperature of 1,000 K, a total of 50 to 100 initial structures are obtained, which will be used for the structure calculation using experimental restraints in the next step of the calculation. The energy of the randomized structures is minimized by 500 cycles of Powell minimization (Brooks *et al.*, 1983) against the force of bond length, bond angles, dihedral angles, and improper torsions.

7.5.2.4. First-Round Structure Calculation—Global Folding

After the starting structures are obtained and all other files in bold in sa.inp are prepared, one can start the first-round of structure calculations. On a UNIX-based SGI workstation on which XPLOR or CNS is installed, simply type “XPLOR <sa.inp> sa.out” to initiate the calculation process. The detailed process and output parameters are all contained in the sa.out file. Calculation is often terminated in the beginning due to errors in the input files, nomenclature, metal coordination, etc. These errors are usually reflected in the sa.out file and the readers are referred to the XPLOR or CNS manual for instructions on specific file format. PDB coordinates of a set of calculated structures are stored in the directory during the calculation for visualization and analysis.

The starting structures are first calculated by 5 ps of rMD at 2,000 K with a step size of 5 fs and forces of the covalent geometry restraints such as bond length, bond angles, dihedral angles, improper torsions, and van der Waals (Lennard–Jones bonds), and experimental restraints of NOE and J coupling. The NOE restraints are used with a force constant of $k_{\text{NOE}} = 30\text{--}50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, whereas the torsion angle restraints are applied usually with a relatively weak constant, $k_{\text{dih}} = 5\text{--}10 \text{ kcal mol}^{-1} \text{ rad}^{-2}$. The soft repulsive van der Waals radii (Lennard–Jones) are scaled by a factor of 0.9. In the next step, the temperature of the system is decreased by 50 K per cycle during the slow cooling down to 300 K by 34 cycles of rMD calculation with a step size of 5 fs and 0.44 ps for each cycle. During the cooling, the van der Waals radii are reduced from 90% to 80% of their true values, k_{NOE} is gradually increased from 2 to 30 $\text{kcal mol}^{-1} \text{ \AA}^{-2}$, and $k_{\text{dih}} = 200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$. The last step of the first-round of calculations consists of 500 cycles of Powell energy minimizations. The above procedure is looped for 100 cycles.

Although one will use as many NOEs as possible for the structure calculation, the inter-residual distance restraints play a more important role in calculation. In order to obtain a high quality structure, more than 10 distance restraints should be used for each residue.

7.5.2.5. NOE Violations and Removal of Incorrect Distance Restraints

Once the first-round structures are calculated using the experimental restraints, mainly NOE data, it is necessary to analyze and validate the derived structures. Because of

experimental errors and imperfect restraints, the calculated structures always contain violations of distance and torsion angles. A distance violation is the difference between the interproton distance in the structure and the closest distance bound (upper or lower) defined by the observed NOE intensity. NOE violations are output into a file after the calculation. At this stage, one should carefully examine the violations which appear in a large number of the structures. These consistent violations are likely caused by misassignment of NOEs or incorrect NOE volume integration. Frequently, finding the consistent violations is not always straightforward because the structure calculation is done by minimizing the potential function over all restraints. As a result, the large violations caused by the incorrect restraints are spread to neighboring regions, leading to a region being consistently violated with lesser scale, or sometimes to violations too small to be recognized after being distributed over a large number of minor violated restraints. Removal of the incorrect restraints after the first-round calculation will improve the quality of structures.

7.5.2.6. Iterative Steps for NOE Analysis and Structure Calculations

Because it is not possible to assign all NOESY cross-peaks after sequence specific assignment due to chemical shift degeneracy and inconsistency in some extent of NOESY cross-peak positions compared to those obtained by resonance assignment, only a fraction of NOESY cross-peaks are assigned unambiguously and used for the structure calculation at the initial stage. Even with only 30% of the final number of NOEs, the generated structures are usually well defined although the resolution is relatively low. These structures are then used to resolve the ambiguous NOESY cross-peaks based on the spatial information of the first-round structures. In order to utilize the NOEs, criteria must be set such as chemical shift tolerance range (usually <0.02 ppm) and corresponding distance between the proton pair in the structures. The newly assigned NOEs are then used as restraints for the next round of structure calculations. It is necessary to carry out several rounds of NOE assignment and structure calculation to assign a majority of ambiguous NOE cross-peaks.

The above process of assignment/calculation can also be performed automatically. First, the NOEs are listed with possible assignments for a given chemical shift tolerance. After the first-round calculation, a program such as ARIA (Nilges *et al.*, 1997; Nilges and O'Donoghue, 1998; Linge *et al.*, 2001, 2003) uses output structures to reduce the assignment possibilities by comparing the interproton distance from the structures to that from the ambiguous NOEs for all assignment possibilities. The new distance restraints are tested during the next round of calculations. Usually, multiple restraints are given to the reassigned ambiguous NOEs, of which only one will be correct. Therefore, these restraints are taken to be more flexible during the structure calculation.

An alternative approach for iterative NOE analysis is back-calculation of NOESY cross-peaks based on the generated structure. Once the folded structures are generated by rMD, the intensities (in terms of volume) of NOESY diagonal and cross-peaks can be calculated for the structures using a relaxation matrix approach with the consideration of spin diffusion. During the back-calculation, the relaxation matrix is first defined with the assumption of isotropic motion in absence of the cross-correlation contribution to the relaxation. The relaxation matrix is then used to calculate theoretical NOESY spectra from the calculated structures. The theoretical NOESY spectrum is compared with the experimental data either manually or automatically to assign additional NOESY cross-peaks which are used as distance restraints for further structure calculations as described above.

7.5.3. Criteria of Structural Quality

The quality of structures is usually reported in terms of several statistical characters including rmsd (root mean square deviation) of the distance and dihedral restraints, rmsd of idealized covalent geometry, rmsd of backbone atoms, rmsd of heavy atoms, the total number of distance violations, and dihedral restraint violations. These criteria provide insights about how consistent the structures are with the experimental restraints. However, these statistical criteria are imperfect when describing the accuracy of the structural calculations. The quality factor is a preferable parameter to describe the consistency of the derived structures with the experimentally determined restraints. The quality factor or Q factor is defined for a type of restraint A as follows (Yip and Case 1989; Nilges *et al.*, 1991; Baleja *et al.*, 1990; Gochin and James 1990):

$$Q = \frac{\text{rms}(A^{\text{obs}} - A^{\text{calc}})}{\text{rms}(A^{\text{obs}})} = \left[\frac{\sum_i (a_i^{\text{obs}} - a_i^{\text{calc}})^2}{\sum_i (a_i^{\text{obs}})^2} \right]^{1/2} \quad (7.12)$$

in which $\text{rms}(A^{\text{obs}} - A^{\text{calc}})$ is the root mean square of the difference between the observed values and calculated value of restraint, and $\text{rms}(A^{\text{obs}})$ is a normalization factor. The restraint can be the NOE intensity (defined by peak volume), J coupling constant, or RDC. As the equation depicts, the Q factor can be used as an indicator for how close the calculated structures are to the actual one for a given set of NMR data.

7.5.4. Second-Round Structure Calculation—Structure Refinement

The second-round calculation involves structural refinement by optimizing the calculated structures after the folding stage with small time step rMD processes while simultaneously minimizing the Q factor of the generated structures. Usually, hydrogen bonding restraints observed for slowly exchangeable amide protons of secondary structural elements are included in the target function during the final stage of the refinement since the hydrogen bonding distance cannot be longer than 2.4 Å and the bonding angle must be within $\pm 35^\circ$ of linearity. If RDC data are available, the refinement is also carried out using an empirical energy term containing dipolar couplings for the target function. Since dipolar couplings provide long-range restraints, they can be used to correct misassigned NOEs, and hence reduce the number of violations and increase the accuracy of the structures.

7.5.5. Presentation of the NMR Structure

Once the structures are determined from NMR data, it is necessary to display them. In a ribbon representation of the average structure, the secondary structural elements are easily recognized, whereas the deviation of the determined structures is visualized by the backbone superposition of a set of final structures. The flexible and rigid regions of the structures are clearly indicated by the superposition representation. The structures can also be represented by molecular surface or electrostatic potential, which is helpful in studying the binding sites of a complex or the overall shape of a molecule. The detailed molecular structures can also be displayed.

There are many software packages available for displaying molecular structures in both schematic and detailed representations, of which MOLMOL (MOleculE analysis and MOleculE display) and MOLSCRIPT are widely used. The software takes the coordinates of the atoms in a structure file to generate three-dimensional structures in the above representations. In addition to displaying superimposed structures, MOLMOL (www.mol.biol.ethz.ch/wuthrich/software/molmol) can be used to display hydrogen bonds, electrostatic surfaces, and Ramachandran plots. A unique feature of MOLSCRIPT is that the output image can be saved in various formats such as Postscript, JPEG, GIF, and many other image formats (<http://www.avatar.se/molscript>). Midasplus is another frequently used program for structural display, which also calculates molecular surfaces, electrostatic potentials and draw distances between protons (<http://www.cgl.ucsf.edu/outreach/midasplus>). Figure 7.4 (see color insert after Chapter 8) shows a sample display of a set of structures in ribbon and superimposition representations using MOLMOL.

7.5.6. Precision of NMR Structures

The precision of NMR structures is related to the precision of the experimental data. Errors in measurements of NOE, J coupling, and dipolar coupling will affect the precision in the estimation of distance and orientational restraints derived from the data. The precision of the calculated structures is usually presented in terms of the atomic rmsd, such as the rmsd of backbone atoms and the rmsd of all heavy atoms. A low rmsd value of the structures indicates that the calculated structures are close to the average structure, which represents a high precision of the structure calculation. A smaller range of the errors in the restraints will produce structures with improved rmsd values. Several factors will contribute errors in the measurements such as low digital resolution in multidimensional experiments, noise level, resonance overlapping, etc. The RMSD of the calculated NOE intensity, J coupling, and chemical shift for the structures compared to the experimental data will also validate the quality of NMR structures as mentioned previously.

In general, an increase in the number of experimental restraints will improve the precision of the calculated structures. However, the precision of the structure determination does not guarantee the accuracy of the NMR structures. For example, if the distances derived from NOE are scaled by a factor due to incorrect NOE volume measurement, the calculated structures will be significantly different from the structures obtained with the correct distance restraints. Therefore, the accuracy of NMR structures is required to be examined with additional criteria.

7.5.7. Accuracy of NMR Structures

It is thought that an accurate structure should not have substantial violations in Ramachandran diagrams and covalent bond geometry. Programs have been developed such as PROCHECK (Laskowski *et al.*, 1996) and WHAT_CHECK (Hooft *et al.*, 1996) for checking the values of bond lengths and angles, the appearance of Ramachandran diagrams, the number and scale of violations of experimental restraints, potential energy, etc. Structures with poor scores do not necessarily indicate errors in the structure, but they require attention to locate possible misassigned experimental data. On the other hand, structures with high scores also do not assure the accuracy of the calculation.

As mentioned earlier in this chapter, the quality factor is frequently used to describe the consistency of the generated structures with the experimentally obtained restraints. Actual

NMR structures must possess a small Q factor. Consequently, the Q factor is often minimized during the refinement stage of the structural calculation. Although an ensemble of structures can be obtained with small rmsd and Q factor values, the accuracy of the structures cannot be validated using the restraints which are used to generate the structures. The accuracy of the structures requires cross-validation with other criteria, of which a free R factor has been used for this purpose (Brünger, 1992). The idea is to set aside some portion of the experimental data which will be used for validation of the accuracy of NMR structures. The prerequisite to do this is that there must be sufficient restraints to generate the structures after excluding those to be set aside. For example, NMR structures can be calculated and refined using restraints from the measurements of NOE, J coupling, chemical shift, and hydrogen bonding. RDCs can then be used for validation of the accuracy and to further refine the calculated structure. Good agreement of the validated structures with the refined structures using the dipolar couplings will confirm the accuracy of the NMR structures. NOE back-calculation is also a valuable indicator of the accuracy of structures determined using NMR data.

7.6. PROTEIN COMPLEXES

7.6.1. Protein–Protein Complexes

Protein–protein interactions play an essential role at various levels of the information flow associated with various biological processes such as gene transcription and translation, cell growth and differentiation, neurotransmission, and immune responses. The interactions frequently lead to changes in shape and dynamics as well as in the chemical or physical properties of the proteins involved. Solution NMR spectroscopy provides a powerful tool to characterize these interactions at the atomic level and at near-physiological conditions. With the use of isotopic labeling, structures of many protein complexes in the 40 kDa total molecular mass regime have been determined (Clore and Gronenborn, 1998). The development of novel NMR techniques and sample preparation has been increasing the mass size further for the structural determination of protein complexes. Furthermore, NMR has been utilized to quickly identify the binding sites of the complexes based on the results of chemical shift mapping or hydrogen bonding experiments. Because it is particularly difficult and sometimes impossible to crystallize weakly bound protein complexes ($K_d > 10^{-6}$), the chemical mapping method is uniquely suitable to characterize such complexes. The binding surfaces of small to medium proteins with molecular mass <30 kDa to large target proteins (unlabeled, up to 100 kDa) can be identified by solution NMR in combination with isotopic labeling NMR (Matsuo *et al.*, 1999; Takahashi *et al.*, 2000). As discussed in Chapter 6, structures of small ligands weakly bound to the proteins can be determined by transferred NOE experiments. The structures of peptides or small protein domains of weakly bound protein complexes can also be characterized by NMR techniques, which may be beneficial to the discovery and design of new drugs with high affinity. In addition to the structural investigation of protein complexes, NMR is a unique and powerful technique to study the molecular dynamics involved in protein–protein reorganization.

7.6.2. Protein–Peptide Complexes

The contact surface contributing to the interactions of high affinity and specificity generally involves 30 or less amino acid residues from each protein of the complex (de Vos *et al.*,

1992; Song and Ni, 1998). Frequently, this contact surface is located in a single continuous fragment of one of the proteins, which can be identified by mutation and deletion experiments. Therefore, fragments can be chemically synthesized in large amount and studied by ^1H NMR experiments owing to their small molecular size (Wüthrich, 1986). In the study of protein–peptide complexes, samples prepared according to the procedure discussed in Chapter 3 for isotopic-labeled protein and unlabeled peptide are most commonly used since the availability of labeled peptide is prohibited by the expense of chemical synthesis from labeled amino acids and the difficulty of biosynthesis due to peptide stability problems during expression and purification.

Data collection and resonance assignment for the complex can be carried out in three stages: for labeled protein, for unlabeled peptide, and for the complex. In the first two stages, the protein and peptide are treated as independent entities. Standard multidimensional heteronuclear experiments can be carried out using the complex sample for resonance assignment, including J coupling measurement, NOE analysis, and RDC measurement. Backbone H^{N} , N , C^{α} , C' resonances are assigned using HNCO, HNCA, HN(CA)CO and HN(CO)CA, and aliphatic side-chain resonances using CBCA(CO)NH, CBCANH, 3D or 4D HCCH–TOCSY, and ^{15}N edited TOCSY as described previously. The distance restraints are obtained from 3D ^{13}C or 4D ^{13}C – ^{13}C NOESY based on the resonance assignments (Qin *et al.*, 2001).

QUESTIONS

- 7.1. Why are small rmsd values of the calculated structure insufficient to describe the accuracy of the structure?
- 7.2. Why is the temperature increased to 2,000 K and then cooled down to 300 K during rMD calculation?
- 7.3. How are the chemical shift indexes used for identifying secondary structural elements?
- 7.4. Why is iterative NOE analysis important to the structure calculation?
- 7.5. What is the Q factor of the structure calculation? What does it describe?

7.1. APPENDIX B1—SA.INP—XPLOR PROTOCOL FOR PROTEIN STRUCTURE CALCULATION

REMARKS This protocol has very slow cooling with increase of vdw evaluate (\$seed = 287346589)

```
set seed $seed end
!-----
! read in the PSF file and initial structure
param @parallhdg_ILK.pro end
structure @ILK_new.psf end
coord @ILK_aves_min.pdb

coord copy end
```

```

!-----
! set the weights for the experimental energy terms
evaluate ($knoe = 25.0) !noes
evaluate ($asym = 0.1) !slope of NOE potential
evaluate ($kcdi = 10.0) !torsion angles
!-----
! The next statement makes sure the experimental energies are used in the
! calculation, and switches off the unwanted energy terms.
! note that the NMR torsions are only switched on in the cooling stage
! we include the noncrystallographic symmetry right from the start
!-----
! Read experimental restraints
noe
  reset
  nrestraints = 6000      ! allocate space for NOEs
  ceiling 100

  set echo off message off end
  class all
  set message off echo off end
    @ILK_mod1.tbl
    @hbond.tbl
  set echo on message on end

  averaging all center
  potential all square
  scale all $knoe

  sqconstant all 1.0
  sqexponent all 2
  ! soexponent all 1
  ! rswitch all 1.0
  ! sqoffset all 0.0
  ! asymptote all 2.0
end

couplings
potential harmonic
  class phi
  force 1.0
  nres 300
  degeneracy 1
  coefficients 6.98 -1.38 1.72 -60.0
  @dihed_talos.tbl
end

carbon
  nres = 200
  class all
  force 0.5
  potential harmonic
  @rcoil_c13.tbl
  @expected_edited_c13.tbl
  @shift_qm.tbl
end
evaluate ($rcon = 0.003)

parameters
nbonds
  atom
  nbxmod 3
  wmin = 0.01 ! warning off

```

```

cutnb = 4.5      ! nonbonded cutoff
tolerance 0.5
repel = 0.9      ! scale factor for vdW radii = 1 ( L-J radii)
rexp = 2         ! exponents in (rirex - R0irex)rexp
irex = 2
rcon = $rcon     ! actually set the vdW weight
end
end

set message off echo off end
restraints dihed
  scale $kcdi
  @dihed_talos.tbl
end
set message on echo on end

flags
exclude *include bonds angle impr vdw noe cdih coup carb end
evaluate ($cool_steps = 3000)
evaluate ($init_t = 2000.01)
vector do (mass = 100.0)(all)      ! uniform mass for all atoms
vector do (fbeta = 10.0)(all)     ! coupling to heat bath
eval ($endcount = 100)
coor copy end

eval ($count = 0)
while ($count < $endcount) loop main
evaluate ($count = $count + 1)

coor swap end
coor copy end

vector do (x=xcomp)(all)
vector do (y=ycomp)(all)
vector do (z=zcomp)(all)

evaluate ($ini_rad = 0.9)          evaluate ($fin_rad = 0.80)
evaluate ($ini_con = 0.004)        evaluate ($fin_con = 4.0)
evaluate ($ini_ang = 0.4)          evaluate ($fin_ang = 1.0)
evaluate ($ini_imp = 0.1)          evaluate ($fin_imp = 1.0)
evaluate ($ini_noe = 2.0)          evaluate ($fin_noe = 30.0)
evaluate ($knoe = $ini_noe)        ! slope of NOE potential
evaluate ($kcdi = 10.0)            ! torsion angles

noe
  averaging all center
  potential all square
  scale all $knoe
  sqconstant all 1.0
  sqexponent all 2
end

restraints dihed
  scale $kcdi
end
evaluate ($rcon = 1.0)

parameters
nbonds
atom
nbxmod 3

```

```

wmin = 0.01      ! warning off
cutnb = 100      ! nonbonded cutoff
tolerance 45
repel = 1.2      ! scale factor for vdW radii = 1 ( L-J radii)
rexp = 2         ! exponents in (r^irex - R0^irex)^rexp
irex = 2
rcon = $rcon     ! actually set the vdW weight
end
end

constraints

    interaction (not name ca) (all)
    weights * 1 angl 0.4 impr 0.1 vdw 0 elec 0 end
    interaction (name ca) (name ca)
    weights * 1 angl 0.4 impr 0.1 vdw 1.0 end
end

dynamics verlet
    nstep = 1000      !
    timestep = 0.005  !
    iasvel = maxwell   firsttemp = $init_t
    tcoupling = true
    tbath = $init_t
    nprint = 50
    iprfrq = 0
    ntrfr = 99999999
end

parameters
nbonds
atom
nbxmod 3
wmin = 0.01      ! warning off
cutnb = 4.5      ! nonbonded cutoff
tolerance 0.5
repel = 0.9      ! scale factor for vdW radii = 1 ( L-J radii)
rexp = 2         ! exponents in (r^irex - R0^irex)^rexp
irex = 2
rcon = 1.0       ! actually set the vdW weight
end
end

evaluate ($kcdi = 200)
restraints dihed
scale $kcdi
end

evaluate ($final_t = 100)      { K }
evaluate ($tempstep = 50)     { K }
evaluate ($ncycle = ($init_t-$final_t)/$tempstep)
evaluate ($nstep = int($cool_steps/$ncycle))
evaluate ($bath = $init_t)
evaluate ($k_vdw = $ini_con)
evaluate ($k_vdwfact = ($fin_con/$ini_con)^(1/$ncycle))
evaluate ($radius = $ini_rad)
evaluate ($radfact = ($fin_rad/$ini_rad)^(1/$ncycle))
evaluate ($k_ang = $ini_ang)
evaluate ($ang_fac = ($fin_ang/$ini_ang)^(1/$ncycle))
evaluate ($k_imp = $ini_imp)
evaluate ($imp_fac = ($fin_imp/$ini_imp)^(1/$ncycle))
evaluate ($noe_fac = ($fin_noe/$ini_noe)^(1/$ncycle))

```

```

evaluate ($knoe = $ini_noe)

vector do (vx = maxwell($bath)) (all)
vector do (vy = maxwell($bath)) (all)
vector do (vz = maxwell($bath)) (all)

evaluate ($i_cool = 0)
while ($i_cool < $ncycle) loop cool
  evaluate ($i_cool = $i_cool+1)
  evaluate ($bath = $bath - $tempstep)
  evaluate ($k_vdw = min($fin_con,$k_vdw*$k_vdwfact))
  evaluate ($radius = max($fin_rad,$radius*$radfact))
  evaluate ($k_ang = $k_ang*$ang_fac)
  evaluate ($k_imp = $k_imp*$imp_fac)
  evaluate ($knoe = $knoe*$noe_fac)

  constraints interaction (all) (all) weights
    * 1 angles $k_ang improper $k_imp
  end end
  parameter nbonds
    cutnb = 4.5 rcon = $k_vdw nbxmod = 3 repel = $radius
  end end
  noe scale all $knoe end

dynamics verlet
  nstep = $nstep timestep = 0.005 iasvel = current firsttemp = $bath
  tcoupling = true tbath = $bath nprint = $nstep iprfreq = 0
  ntrfr = 99999999
end
end loop cool

mini powell nstep = 500 nprint = 50 end

{* NOE Data Analysis *}
print threshold = 0.5 noe
evaluate ($noe5 = $violations)
print threshold = 0.0 noe
evaluate ($noe0 = $violations)
evaluate ($rms_noe = $result)

{* CDIH Data Analysis *}
print threshold = 5.0 cdih
evaluate ($cdih5 = $violations)
print threshold = 0.0 cdih
evaluate ($cdih0 = $violations)
evaluate ($rms_cdih = $result)

{* BOND Data Analysis *}
print thres = 0.05 bond
evaluate ($bond5 = $violations)
evaluate ($rms_bond = $result)

{* ANGLE Data Analysis *}
print thres = 5.0 angle
evaluate ($angle5 = $violations)
evaluate ($rms_angle = $result)

{* IMPROPER Data Analysis *}
print thres = 5.0 improper
evaluate ($improper5 = $violations)
evaluate ($rms_improper = $result)

```

```

{* ENERGY Data Analysis *}
energy end

{* J-coupling constant analysis *}
couplings print threshold 1.0 all end
evaluate ($rms_coup = $result)
evaluate ($viol_coup = $violations)

{* Carbon chemical shift analysis *}
carbon print threshold = 1.0 end
evaluate ($rms_carbshift = $rmsca)
evaluate ($rms_carbbshift = $rmscb)
evaluate ($viol_carb = $violations)

remarks =====
remarks          noe, cdih, bonds, angles, improp
remarks violations.: $noe5[I5], $cdih5[I5], $bond5[I5], $angle5[I5], $improper5[I5]
remarks RMSD      .: $rms_noe[F6.3], $rms_cdih[F6.3], $rms_bond[F6.3], $rms_angle[F6.3],
  $rms_improper[F6.3]
remarks 0-viol    .: $noe0[I5], $cdih0[I5]
remarks coup, carb-a, carb-b
remarks violations: $viol_coup[I5], $viol_carb[I5], \
remarks RMSD:      $rms_coup[F6.3], $rms_carbshift[F6.3], \
  $rms_carbbshift[F6.3], \

remarks =====
remarks overall = $ener
remarks noe = $NOE
remarks dih = $CDIH
remarks vdW = $VDW
remarks bon = $BOND
remarks ang = $ANGL
remarks imp = $IMPR
remarks coup = $COUP
remarks carb = $CARB
remarks prot = $PROT

remarks =====
evaluate ($file = "ILK_tal_" + encode($count) + ".pdb")
write coordinates output = $file end
end loop main
write coordinates output = $filename end
stop

```

7.2. APPENDIX B2—EXAMPLE OF NOE TABLE

```

!K1
assign (resid 1 and name HG#) (resid 1 and name HD#)2.5 0.7 0.2    !#A 526  9.18e+05
assign (resid 1 and name HG#) (resid 1 and name HB#)3.0 1.2 0.3    !#A 521  2.31e+05
assign (resid 1 and name HD#) (resid 1 and name HE#)2.5 0.7 0.2    !#A 518  5.72e+05
assign (resid 1 and name HG#) (resid 1 and name HE#)3.0 1.2 0.3    !m#A 516  4.30e+05
assign (resid 1 and name HG#) (resid 1 and name HA)4.0 2.2 2.0    !#A 510  2.25e+05
assign (resid 1 and name HD#) (resid 1 and name HA)4.0 2.2 1.0    !#A 509  1.45e+05
assign (resid 1 and name HB#) (resid 1 and name HA)3.0 1.2 0.3    !#A 508  3.20e+05
assign (resid 1 and name HE#) (resid 1 and name HA)4.0 2.2 1.0    !#A 500  8.04e+04
assign (resid 1 and name HA) (resid 2 and name HB)4.0 2.2 1.0    !#A 512  1.24e+05
assign (resid 1 and name HB#) (resid 2 and name HA)4.0 2.2 2.0    !added
!assign (resid 1 and name HG#) (resid 2 and name HA)4.0 2.2 2.0    !added
!assign (resid 1 and name HG#) (resid 3 and name HA#)4.0 2.2 1.0    !#A 513  9.34e+04
!assign (resid 1 and name HG#) (resid 3 and name HN)4.0 2.2 1.0    !added

```



```
assign (resid 1 and name HB#) (resid 3 and name HN)4.0 2.2 1.0      !added
assign (resid 1 and name HB#) (resid 3 and name HA#)4.0 2.2 2.0      !added
!assign (resid 1 and name HD#) (resid 3 and name HA#)4.0 2.2 2.0      !added
assign (resid 1 and name HA) (resid 4 and name HB#)4.0 2.2 1.0      !jun
assign (resid 1 and name HB#) (resid 4 and name HN)4.0 2.2 2.0      !m#A 498 8.90e+04

!V2
assign (resid 2 and name HG1#) (resid 2 and name HB)2.5 0.7 0.2      !#A 314 6.73e+05
assign (resid 2 and name HG1#) (resid 2 and name HA)3.0 1.2 0.3      !m#A 267 5.15e+05
assign (resid 2 and name HG1#) (resid 3 and name HN)4.0 2.2 1.0      !#A 601 1.48e+05
assign (resid 2 and name HB) (resid 3 and name HA#)4.0 2.2 1.0      !#A 293 1.42e+05
assign (resid 2 and name HG1#) (resid 3 and name HA#)4.0 2.2 1.0      !m#A 269 2.73e+05
assign (resid 2 and name HA) (resid 3 and name HA#)4.0 2.2 1.0      !m#A 254 2.15e+05
assign (resid 2 and name HG1#) (resid 3 and name HN)4.0 2.2 2.0      !#A 173 2.73e+05
assign (resid 2 and name HB) (resid 3 and name HN)4.0 2.2 1.0      !#A 166 1.49e+05
assign (resid 2 and name HA) (resid 3 and name HN)3.0 1.2 0.3      !m#A 57 5.58e+05
assign (resid 2 and name HB) (resid 4 and name HN)4.0 2.2 1.0      !#A 230 5.67e+04
assign (resid 2 and name HG1#) (resid 4 and name HN)4.0 2.2 2.0      !#A 229 9.28e+04
assign (resid 2 and name HG2#) (resid 4 and name HN)4.0 2.2 2.0      !#A 603 4.72e+04
!assign (resid 2 and name HG2#) (resid 4 and name HE#)4.0 2.2 2.0      !#A 611 1.30e+05
!assign (resid 2 and name HB) (resid 4 and name HD#)4.0 2.2 2.0      !#A 224 7.79e+04
assign (resid 2 and name HA) (resid 4 and name HD#)4.0 2.2 2.0      !#A 187 5.78e+04
assign (resid 2 and name HA) (resid 4 and name HN)4.0 2.2 1.0      !#A 99 1.15e+05
assign (resid 2 and name HB) (resid 5 and name HE#)4.0 2.2 2.0      !#A 440 1.13e+05
assign (resid 2 and name HG1#) (resid 5 and name HD#)4.0 2.2 2.0      !#A 212 1.60e+05
assign (resid 2 and name HG1#) (resid 5 and name HE#)4.0 2.2 2.0      !#A 433 2.33e+05
assign (resid 2 and name HA) (resid 5 and name HD#)4.0 2.2 2.0      !#A 406 9.65e+04
assign (resid 2 and name HB) (resid 5 and name HD#)4.0 2.2 2.0      !#A 220 1.00e+05
assign (resid 2 and name HA) (resid 5 and name HE#)4.0 2.2 2.0      !#A 186 8.51e+04
assign (resid 2 and name HA) (resid 5 and name HB#)4.0 2.2 1.0      !#Jun

!G3
assign (resid 3 and name HA#) (resid 3 and name HN)2.5 0.7 0.2      !#A 58 4.42e+05
assign (resid 3 and name HA#) (resid 4 and name HA)4.0 2.2 1.0      !m#A 257 1.69e+05
assign (resid 3 and name HA#) (resid 4 and name HN)3.0 1.2 0.3      !#A 59 6.34e+05
assign (resid 3 and name HN) (resid 4 and name HE#)4.0 2.2 1.0      !#A 28 3.96e+04
assign (resid 3 and name HN) (resid 4 and name HN)3.0 1.2 0.3      !#MA 4 6.69e+04
assign (resid 3 and name HA#) (resid 5 and name HB1)4.0 2.2 2.0      !m#A 503 6.23e+04
assign (resid 3 and name HA#) (resid 5 and name HE#)4.0 2.2 2.0      !olga
assign (resid 3 and name HA#) (resid 5 and name HN)4.0 2.2 1.0      !#A 78 1.39e+05
assign (resid 3 and name HN) (resid 5 and name HD#)4.0 2.2 2.0      !#A 50 3.75e+04
assign (resid 3 and name HN) (resid 6 and name HN)4.0 2.2 1.0      !#A 20 4.57e+04
assign (resid 3 and name HA#) (resid 6 and name HN)4.0 2.2 1.0      !#A 486 7.65e+04
assign (resid 3 and name HA#) (resid 6 and name HB#)4.0 2.2 2.0      !added
assign (resid 3 and name HA#) (resid 6 and name HG#)4.0 2.2 2.0      !added

!F4
assign (resid 4 and name HA) (resid 4 and name HB2)2.5 0.7 0.2      !#A 242 3.84e+05
assign (resid 4 and name HA) (resid 4 and name HB1)3.0 1.2 0.3      !#A 241 3.46e+05
assign (resid 4 and name HB1) (resid 4 and name HD#)3.0 1.2 0.3      !m#A 204 4.66e+05
assign (resid 4 and name HB2) (resid 4 and name HD#)3.0 1.2 0.3      !m#A 203 4.42e+05
assign (resid 4 and name HA) (resid 4 and name HD#)4.0 2.2 1.0      !m#A 175 4.05e+05
assign (resid 4 and name HB1) (resid 4 and name HN)3.0 1.2 0.3      !#A 102 2.49e+05
assign (resid 4 and name HB2) (resid 4 and name HN)3.0 1.2 0.3      !#A 101 2.48e+05
assign (resid 4 and name HA) (resid 4 and name HN)3.0 1.2 0.3      !#A 60 2.23e+05
assign (resid 4 and name HD#) (resid 4 and name HE#)2.5 0.7 0.2      !#A 46 2.48e+06
assign (resid 4 and name HE#) (resid 4 and name HN)4.0 2.2 1.0      !#A 40 1.05e+05
assign (resid 4 and name HD#) (resid 4 and name HN)4.0 2.2 1.0      !m#A 38 1.77e+05
assign (resid 4 and name HN) (resid 5 and name HB1)4.0 2.2 1.0      !#A 428 6.11e+04
assign (resid 4 and name HD#) (resid 5 and name HB1)4.0 2.2 1.0      !#A 206 9.60e+04
assign (resid 4 and name HA) (resid 5 and name HN)3.0 1.2 0.3      !m#A 61 3.98e+05
assign (resid 4 and name HN) (resid 5 and name HD#)4.0 2.2 2.0      !#A 39 7.34e+04
```

```

assign (resid 4 and name HE#) (resid 5 and name HN)4.0 2.2 1.0      !#A 36  1.29e+05
assign (resid 4 and name HD#) (resid 5 and name HN)4.0 2.2 1.0      !#A 35  1.07e+05
assign (resid 4 and name HN) (resid 5 and name HN)3.0 1.2 0.3        !#MA 8  1.34e+05
assign (resid 4 and name HA) (resid 7 and name HN)4.0 2.2 1.0        !#A 392  2.71e+05
assign (resid 4 and name HB1) (resid 7 and name HN)4.0 2.2 2.0       !#A 116  1.09e+05
!assign (resid 4 and name HD#) (resid 7 and name HD#)4.0 2.2 2.0     !added  1.09e+05
assign (resid 4 and name HD#) (resid 8 and name HB2)4.0 2.2 2.0      !added  1.09e+05
assign (resid 4 and name HD#) (resid 17 and name HG#)4.0 2.2 2.0

!F5
assign (resid 5 and name HA) (resid 5 and name HB1)3.0 1.2 0.3       !#A 240  2.24e+05
assign (resid 5 and name HB2) (resid 5 and name HB1)2.5 0.7 0.2      !#A 232  5.40e+05
assign (resid 5 and name HB2) (resid 5 and name HD#)3.0 1.2 0.3      !m#A 202  5.11e+05
assign (resid 5 and name HB1) (resid 5 and name HD#)2.5 0.7 0.2      !#A 201  4.51e+05
assign (resid 5 and name HA) (resid 5 and name HD#)3.0 1.2 0.3       !#A 176  4.58e+05
assign (resid 5 and name HB1) (resid 5 and name HN)3.0 1.2 0.3       !#A 104  1.68e+05
assign (resid 5 and name HB2) (resid 5 and name HN)4.0 2.2 1.0       !m#A 103  2.72e+05
assign (resid 5 and name HA) (resid 5 and name HN)3.0 1.2 0.3       !#A 64  3.98e+05
assign (resid 5 and name HD#) (resid 5 and name HE#)2.5 0.7 0.2      !#A 45  3.13e+06
assign (resid 5 and name HD#) (resid 5 and name HN)3.0 1.2 0.3       !#A 34  1.97e+05
assign (resid 5 and name HE#) (resid 6 and name HB#)4.0 2.2 1.0       !#A 447  1.20e+05
assign (resid 5 and name HD#) (resid 6 and name HB#)4.0 2.2 1.0       !#A 446  1.23e+05
assign (resid 5 and name HE#) (resid 6 and name HD#)4.0 2.2 2.0       !#A 441  2.27e+05
!assign (resid 5 and name HD#) (resid 6 and name HG#)4.0 2.2 1.0      !#A 221  1.05e+05
assign (resid 5 and name HE#) (resid 6 and name HG#)4.0 2.2 1.0       !#A 214  1.02e+05
assign (resid 5 and name HD#) (resid 6 and name HA)4.0 2.2 1.0        !#A 183  1.27e+05
assign (resid 5 and name HB2) (resid 6 and name HN)4.0 2.2 1.0       !m#A 112  1.86e+05
assign (resid 5 and name HB1) (resid 6 and name HN)4.0 2.2 1.0       !#A 111  1.39e+05
assign (resid 5 and name HA) (resid 6 and name HN)3.0 1.2 0.3        !#A 62  3.22e+05
assign (resid 5 and name HD#) (resid 6 and name HN)4.0 2.2 1.0       !#A 37  1.27e+05
assign (resid 5 and name HE#) (resid 6 and name HN)4.0 2.2 1.0       !#A 33  8.23e+04

INCOMPLETED

```

7.3. APPENDIX B3—EXAMPLE OF DIHEDRAL ANGLE RESTRAINT TABLE

```

!remark phi angle constraints
!! r22
assign (resid 21 and name c and segid b) (resid 22 and name n and segid b)
      (resid 22 and name ca and segid b) (resid 22 and name c and segid b) 1.0 -64.0 20.0 2
!! r23
assign (resid 22 and name c and segid b) (resid 23 and name n and segid b)
      (resid 23 and name ca and segid b) (resid 23 and name c and segid b) 1.0 -67.0 20.0 2
!! r24
assign (resid 23 and name c and segid b) (resid 24 and name n and segid b)
      (resid 24 and name ca and segid b) (resid 24 and name c and segid b) 1.0 -73.0 20.0 2
!! r25
assign (resid 24 and name c and segid b) (resid 25 and name n and segid b)
      (resid 25 and name ca and segid b) (resid 25 and name c and segid b) 1.0 -79.0 20.0 2
!! r26
assign (resid 25 and name c and segid b) (resid 26 and name n and segid b)
      (resid 26 and name ca and segid b) (resid 26 and name c and segid b) 1.0 -68.0 20.0 2
!! r27
assign (resid 26 and name c and segid b) (resid 27 and name n and segid b)
      (resid 27 and name ca and segid b) (resid 27 and name c and segid b) 1.0 -66.0 20.0 2
!! r28
assign (resid 27 and name c and segid b) (resid 28 and name n and segid b)
      (resid 28 and name ca and segid b) (resid 28 and name c and segid b) 1.0 -94.0 20.0 2
!! r29
assign (resid 28 and name c and segid b) (resid 29 and name n and segid b)

```

```
(resid 29 and name ca and segid b) (resid 29 and name c and segid b) 1.0 -62.0 20.0 2
!! r30
assign (resid 29 and name c and segid b) (resid 30 and name n and segid b)
(resid 30 and name ca and segid b) (resid 30 and name c and segid b) 1.0 -65.0 20.0 2
!! r31
assign (resid 30 and name c and segid b) (resid 31 and name n and segid b)
(resid 31 and name ca and segid b) (resid 31 and name c and segid b) 1.0 -63.0 20.0 2
!! r32
assign (resid 31 and name c and segid b) (resid 32 and name n and segid b)
(resid 32 and name ca and segid b) (resid 32 and name c and segid b) 1.0 -63.0 20.0 2
!! r33
assign (resid 32 and name c and segid b) (resid 33 and name n and segid b)
(resid 33 and name ca and segid b) (resid 33 and name c and segid b) 1.0 -63.0 20.0 2
!! r34
assign (resid 33 and name c and segid b) (resid 34 and name n and segid b)
(resid 34 and name ca and segid b) (resid 34 and name c and segid b) 1.0 -64.0 20.0 2
!! r35
assign (resid 34 and name c and segid b) (resid 35 and name n and segid b)
(resid 35 and name ca and segid b) (resid 35 and name c and segid b) 1.0 -67.0 20.0 2
!! r36
assign (resid 35 and name c and segid b) (resid 36 and name n and segid b)
(resid 36 and name ca and segid b) (resid 36 and name c and segid b) 1.0 -63.0 20.0 2
!! r37
assign (resid 36 and name c and segid b) (resid 37 and name n and segid b)
(resid 37 and name ca and segid b) (resid 37 and name c and segid b) 1.0 -64.0 20.0 2
!! r38
assign (resid 37 and name c and segid b) (resid 38 and name n and segid b)
(resid 38 and name ca and segid b) (resid 38 and name c and segid b) 1.0 -64.0 20.0 2
!! r39
assign (resid 38 and name c and segid b) (resid 39 and name n and segid b)
(resid 39 and name ca and segid b) (resid 39 and name c and segid b) 1.0 -63.0 20.0 2
!! r40
INCOMPLETED
!remark psi angles constraints
!! r22
assign (resid 22 and name n and segid b) (resid 22 and name ca and segid b)
(resid 22 and name c and segid b) (resid 23 and name n and segid b) 1.0 -41.0 20.0 2
!! r23
assign (resid 23 and name n and segid b) (resid 23 and name ca and segid b)
(resid 23 and name c and segid b) (resid 24 and name n and segid b) 1.0 -39.0 20.0 2
!! r24
assign (resid 24 and name n and segid b) (resid 24 and name ca and segid b)
(resid 24 and name c and segid b) (resid 25 and name n and segid b) 1.0 -30.0 20.0 2
!! r25
assign (resid 25 and name n and segid b) (resid 25 and name ca and segid b)
(resid 25 and name c and segid b) (resid 26 and name n and segid b) 1.0 -33.0 20.0 2
!! r26
assign (resid 26 and name n and segid b) (resid 26 and name ca and segid b)
(resid 26 and name c and segid b) (resid 27 and name n and segid b) 1.0 -36.0 20.0 2
!! r27
assign (resid 27 and name n and segid b) (resid 27 and name ca and segid b)
(resid 27 and name c and segid b) (resid 28 and name n and segid b) 1.0 -34.0 20.0 2
!! r28
assign (resid 28 and name n and segid b) (resid 28 and name ca and segid b)
(resid 28 and name c and segid b) (resid 29 and name n and segid b) 1.0 -9.0 20.0 2
!! r29
assign (resid 29 and name n and segid b) (resid 29 and name ca and segid b)
(resid 29 and name c and segid b) (resid 30 and name n and segid b) 1.0 -36.0 20.0 2
!! r30
assign (resid 30 and name n and segid b) (resid 30 and name ca and segid b)
(resid 30 and name c and segid b) (resid 31 and name n and segid b) 1.0 -40.0 20.0 2
```

```

!! r31
assign (resid 31 and name n and segid b) (resid 31 and name ca and segid b)
      (resid 31 and name c and segid b) (resid 32 and name n and segid b) 1.0 -42.0 20.0 2
!! r32
assign (resid 32 and name n and segid b) (resid 32 and name ca and segid b)
      (resid 32 and name c and segid b) (resid 33 and name n and segid b) 1.0 -40.0 20.0 2
!! r33
assign (resid 33 and name n and segid b) (resid 33 and name ca and segid b)
      (resid 33 and name c and segid b) (resid 34 and name n and segid b) 1.0 -44.0 20.0 2
!! r34
assign (resid 34 and name n and segid b) (resid 34 and name ca and segid b)
      (resid 34 and name c and segid b) (resid 35 and name n and segid b) 1.0 -42.0 20.0 2
!! r35
assign (resid 35 and name n and segid b) (resid 35 and name ca and segid b)
      (resid 35 and name c and segid b) (resid 36 and name n and segid b) 1.0 -35.0 20.0 2
!! r36
assign (resid 36 and name n and segid b) (resid 36 and name ca and segid b)
      (resid 36 and name c and segid b) (resid 37 and name n and segid b) 1.0 -42.0 20.0 2
!! r37
assign (resid 37 and name n and segid b) (resid 37 and name ca and segid b)
      (resid 37 and name c and segid b) (resid 38 and name n and segid b) 1.0 -40.0 20.0 2
!! r38

INCOMPLETED

```

7.4. APPENDIX B4—EXAMPLE OF CHEMICAL SHIFT TABLE FOR TALOS

REMARK AlfaIIb fused to MBP in complex with beta3, input for TALOS

DATA SEQUENCE KVGFFKRNRP PLEEDDEEGE

VARS RESID RESNAME ATOMNAME SHIFT
 FORMAT %4d %1s %4s %8.3f

| | | | |
|---|---|----|--------|
| 1 | K | N | 120.93 |
| 1 | K | HA | 4.08 |
| 1 | K | C | 176.52 |
| 1 | K | CA | 56.26 |
| 1 | K | CB | 32.81 |
| 2 | V | N | 117.94 |
| 2 | V | HA | 4.15 |
| 2 | V | C | 176.47 |
| 2 | V | CA | 62.48 |
| 2 | V | CB | 32.55 |
| 3 | G | N | 109.71 |
| 3 | G | HA | 3.88 |
| 3 | G | C | 173.63 |
| 3 | G | CA | 45.15 |
| 4 | F | N | 118.02 |
| 4 | F | HA | 4.56 |
| 4 | F | C | 175.36 |
| 4 | F | CA | 57.93 |
| 4 | F | CB | 39.55 |
| 5 | F | N | 118.99 |
| 5 | F | HA | 4.58 |
| 5 | F | C | 175.12 |
| 5 | F | CA | 57.51 |
| 5 | F | CB | 39.60 |
| 6 | K | N | 120.89 |
| 6 | K | HA | 4.21 |
| 6 | K | C | 175.94 |
| 6 | K | CA | 56.20 |

| | | | |
|----|---|----|--------|
| 6 | K | CB | 33.00 |
| 7 | R | N | 119.97 |
| 7 | R | HA | 4.28 |
| 7 | R | C | 175.84 |
| 7 | R | CA | 55.98 |
| 7 | R | CB | 30.90 |
| 8 | N | N | 118.01 |
| 8 | N | HA | 4.66 |
| 8 | N | C | 175.17 |
| 8 | N | CA | 53.14 |
| 8 | N | CB | 38.78 |
| 9 | R | N | 119.81 |
| 9 | R | HA | 4.62 |
| 9 | R | CA | 55.37 |
| 9 | R | CB | 30.19 |
| 10 | P | HA | 4.67 |
| 11 | P | HA | 4.39 |
| 11 | P | C | 176.81 |
| 11 | P | CA | 63.10 |
| 11 | P | CB | 31.84 |
| 12 | L | N | 119.06 |
| 12 | L | HA | 4.33 |
| 12 | L | C | 174.45 |
| 12 | L | CA | 55.18 |
| 12 | L | CB | 42.19 |
| 13 | E | N | 119.81 |
| 13 | E | HA | 4.29 |
| 13 | E | C | 176.23 |
| 13 | E | CA | 56.30 |
| 13 | E | CB | 29.93 |
| 14 | E | N | 119.17 |
| 14 | E | HA | 4.30 |
| 14 | E | C | 177.34 |
| 14 | E | CA | 56.30 |
| 14 | E | CB | 29.93 |
| 15 | D | N | 119.13 |
| 15 | D | HA | 4.63 |
| 15 | D | C | 175.70 |
| 15 | D | CA | 53.91 |
| 15 | D | CB | 40.64 |
| 16 | D | N | 119.14 |
| 16 | D | HA | 4.61 |
| 16 | D | C | 175.99 |
| 16 | D | CA | 53.92 |
| 16 | D | CB | 40.67 |
| 17 | E | N | 119.02 |
| 17 | E | HA | 4.30 |
| 17 | E | C | 176.37 |
| 17 | E | CA | 56.29 |
| 17 | E | CB | 29.86 |
| 18 | E | N | 119.40 |
| 18 | E | HA | 4.30 |
| 18 | E | C | 176.81 |
| 18 | E | CA | 56.38 |
| 18 | E | CB | 29.80 |
| 19 | G | N | 108.49 |
| 19 | G | HA | 3.96 |
| 19 | G | C | 173.15 |
| 19 | G | CA | 45.20 |
| 20 | E | N | 123.55 |
| 20 | E | HA | 4.16 |
| 20 | E | CA | 57.45 |
| 20 | E | CB | 30.62 |

7.5. APPENDIX B5—EXAMPLE OF HYDROGEN BOND TABLE

```

assign (resid 2 and name o) (resid 6 and name n) 3.0 0.7 0.5
assign (resid 2 and name o) (resid 6 and name hn) 2.5 0.7 0.5
assign (resid 3 and name o) (resid 7 and name n) 3.0 0.7 0.5
assign (resid 3 and name o) (resid 7 and name hn) 2.5 0.7 0.5
assign (resid 4 and name o) (resid 8 and name n) 3.0 0.7 0.5
assign (resid 4 and name o) (resid 8 and name hn) 2.5 0.7 0.5
assign (resid 5 and name o) (resid 9 and name n) 3.0 0.7 0.5
assign (resid 5 and name o) (resid 9 and name hn) 2.5 0.7 0.5

```

7.6. APPENDIX B6—EXAMPLE OF INPUT FILE TO GENERATE A RANDOM-COIL COORDINATES

```

remarks file nmr/generate_template.inp
remarks Generates a "template" coordinate set. This produces
remarks an arbitrary extended conformation with ideal geometry.
remarks
remarks Author: Axel T. Brunger

topology reset @topallhdg_new.pro end
parameter reset @parallhdg_new.pro end

{====>}
structure @alfa_RQ.psf end          {*Read structure file.*}

vector ident (x) (all)
vector do (x = x/10.) (all)
vector do (y = random(0.5)) (all)
vector do (z = random(0.5)) (all)

vector do (fbeta = 50) (all)      {*Friction coefficient, in 1/ps.*}
vector do (mass = 100) (all)      {*Heavy masses, in amus.*}

parameter
  nbonds
    cutnb = 5.5 rcon = 20. nbxmod = -2 repel = 0.9 wmin = 0.1 tolerance = 1.
    rexp = 2 irexp = 2 inhibit = 0.25
  end
end

flags exclude * include bond angle vdw end

minimize powell nstep = 50 nprint = 10 end

flags include impr end

minimize powell nstep = 50 nprint = 10 end

dynamics verlet
  nstep = 50 timestep = 0.001 iasvel = maxwell firsttemp = 300.
  tcoupling = true tbath = 300. nprint = 50 iprfreq = 0
end

parameter
  nbonds
    rcon = 2. nbxmod = -3 repel = 0.75
  end
end

```

```

minimize powell nstep = 100 nprint = 25 end

dynamics verlet
  nstep = 500 timestep = 0.005 iasvel = maxwell firsttemp = 300.
  tcoupling = true tbath = 300. nprint = 100 iprfrq = 0
end

flags exclude vdw elec end
vector do (mass = 1.) ( name h* )
hbuild selection = ( name h* ) phistep = 360 end
flags include vdw elec end

minimize powell nstep = 200 nprint = 50 end
      (*Write coordinates.*)

remarks produced by nmr/generate_template.inp
write coordinates output = alfa_RQ_00.pdb end

stop

```

7.7. APPENDIX B7—EXAMPLE OF INPUT FILE TO GENERATE A GEOMETRIC PSF FILE

```

remarks file nmr/generate.inp
remarks Generate structure file for a protein
remarks using the SA parameter and topology files.

topology
  @../topallhdg_new.pro
end      (*Read topology file      *)

segment      (*Generate protein      *)
  name = " "      (*This name has to match the *)
                (*four characters in columns 73 *)
                (*through 76 in the coordinate *)
                (*file, in XPLOR this name is *)
                (*name is referred to as SEGId. *)

chain
  @TOPPAR:toph19.pep      (*Read peptide bond file      *)
  sequence LYS VAL GLY PHE PHE LYS GLN ASN ARG PRO
            PRO LEU GLU GLU ASP ASP GLU GLU GLY GLU
  end
end
end

write structure output = alfa_RQ.psf end
stop

```

REFERENCES

- Allen, M.P. and D.J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford (1987).
 Baleja, J.D., P.T. Pon, and B.D. Sykes, *Biochemistry* **29**, 4828 (1990).
 Braun, W., *Quart. Rev. Biophys.* **19**, 115 (1987).
 Brooks, B.R., R. Brucoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).

- Brünger, A., *Nature (London)* **355**, 472 (1992).
- Brünger, A.T., *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR*, Yale University Press, New Haven (1992).
- Brünger, A.T., P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve *et al.*, *Acta Cryst.* **D54**, 905 (1998). <http://cns.csb.yale.edu/>.
- Clore, G.M. and A.M. Gronenborn, *Trends Biotech.* **16**, 22 (1998).
- Clore, G.M., A.T. Brünger, M. Karplus, and A.M. Gronenborn, *J. Mol. Biol.* **191**, 523 (1986).
- Crippen, G.M., *Distance Geometry and Conformational Calculations*, Research Studies Press (1981).
- Cormilescu, G., F. Delaglio, and A. Bax, *J. Biomol. NMR*, **13**, 289 (1999). <http://spin.niddk.nih.gov/bax/software/TOLAS/info.html>.
- de Alba, E. and N. Tjandra, *Prog. Nucl. Magn. Reson. Spectrosc.* **40**, 175 (2002).
- de Dios, A.C., J.G. Pearson, and E. Oldfield, *Science* **260**, 1491 (1993).
- de Marco, A., M. Llinás, and K. Wüthrich, *Biopolymers* **17**, 2727 (1978b).
- de Marco, A., M. Llinás, and K. Wüthrich, *Biopolymers* **17**, 617 (1978a).
- de Vos, A.M., M. Ultsch, and A.A. Kossiakoff, *Science*, **255**, 306 (1992).
- Fischman, A.J., D.H. Live, H.R. Wyssbrod, W.C. Agosta, and D. Cowburn, *J. Am. Chem. Soc.* **102**, 2533–2539. (1980).
- Gochin, M. and T.L. James, *Biochemistry* **29**, 11172 (1990).
- Güntert, P., *Prog. Nucl. Magn. Reson. Spectrosc.* **43**, 105 (2003).
- Güntert, P., *Quart. Rev. Biophys.* **31**, 2 (1998).
- Güntert, P., C. Mumenthaler, and K. Wüthrich, *J. Mol. Biol.* **273**, 283 (1997).
- Güntert, P., W. Braun, and K. Wüthrich, *J. Mol. Biol.* **217**, 517 (1991). <http://www.mol.biol.ethz.ch/wuthrich/software/>.
- Havel, T.F., *Prog. Biophys. Mol. Biol.* **56**, 43 (1991).
- Hooft, R.W., G. Vriend, C. Sander, and E.E. Abola, *Nature (London)*, **381**, 272 (1996).
- Karplus, K., *J. Am. Chem. Soc.* **85**, 2870 (1963).
- Karplus, K., *J. Phys. Chem.* **30**, 11 (1959).
- Kuszewski, J., A.M. Gronenborn, and G.M. Clore, *J. Magn. Reson.* **B107**, 293 (1995a).
- Kuszewski, J., J. Qin, A.M. Gronenborn, and G.M. Clore, *J. Magn. Reson.* **B106**, 92 (1995b).
- Laskowski, R.A., J.A.C. Rullmann, M.W. MacArthur, R. Kaptein, and J.M. Thornton, *J. Biomol. NMR* **8**, 477 (1996).
- Linge, J.P., M. Habeck, W. Rieping, and M. Nilges, *Bioinformatics* **19**, 315 (2003).
- Linge, J.P., S.I. O'Donoghue, and M. Nilges, *Meth. Enzymol.* **339**, 71 (2001).
- Lipstiz, R.S. and N. Tjandra, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 387 (2004).
- Luginbühl, P., T. Szyperski, and K. Wüthrich, *J. Magn. Reson.* **B109**, 229 (1995).
- Mastsuo, H., K.J. Walters, K. Teruya, T. Tanaka, G.T. Gassner, S.J. Lippard *et al.*, *J. Am. Chem. Soc.* **121**, 9903 (1999).
- Mayer, K.L. and M.J. Stone, *Biochemistry* **39**, 8382 (2000).
- Neri, D., T. Szyperski, G. Otting, H. Senn, and K. Wüthrich, *Biochemistry* **28**, 7510 (1989).
- Neuhaus, D. and M.P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH Publishers, New York (1989).
- Nilges, M. and S.I. O'Donoghue, *Prog. NMR Spectrosc.* **32**, 107 (1998).
- Nilges, M., J. Habazettl, A. Brünger, and T.A. Holak, *J. Mol. Biol.* **219**, 499 (1991).
- Nilges, M., M.J. Macias, S.I. O'Donoghue, and H. Oschkinat, *J. Mol. Biol.* **269**, 408 (1997).
- Pearlman, D.A., D.A. Case, J.C. Caldwell, G.L. Seibel, U.C. Singh, P. Weiner *et al.*, *Amber 4.0*, University of California, San Francisco (1991). <http://www.amber.ucsf.edu/amber/amber.html>.
- Powers, R., G.M., Clore, A.M. Gronenborn, and A. Bax, *J. Magn. Reson.* **94**, 209 (1991).
- Prestegard, J.H., H.M. Al-Hashimi, and J.R. Tolman, *Quart. Rev. Biophys.* **33**, 371 (2000).
- Qin, J., O. Vinogradova, and A.M. Gronenborn, *Meth. Enzymol.* **339**, 377 (2001).
- Senn, H., B. Werner, B.A. Messerle, C. Weber, R. Traber, and K. Wüthrich, *FEBS Lett.* **249**, 113 (1989).
- Song, J. and F. Ni, *Biochem. Cell Biol.* **76**, 177 (1998).

- Spera, S. and A. Bax, *J. Am. Chem. Soc.* **113**, 5490 (1991).
- Takahashi, H., T. Nahanishi, K. Kami, Y. Arata, and I. Shimada, *Nat. Struct. Biol.* **7**, 220 (2000).
- Tjandra, T. and A. Bax, *Science*, **278**, 1111 (1997).
- Tjandra, N., J.G. Omichinski, A.M. Gronenborn, G.M. Clore, and A. Bax, *Nature Struct. Biol.* **4**, 732 (1997).
- van Gunsteren, W.F., S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Krüger, A.E. Mark *et al.*, *Biomolecular Simulation: the GROMOS96 Manual and User Guide*, vdf Hochschulverlag, Zürich (1996).
- van Gunsteren, W.F., *Curr. Opin. Struct. Biol.* **3**, 277 (1993).
- Vinogradova, O., A. Velyvis, A. Velyvience, B. Hu, A.T. Haas, E.F. Plow, and J. Qin, *Cell*, **110**, 587–597 (2002).
- Vuister, G.W. and A. Bax, *J. Am. Chem. Soc.* **115**, 7772 (1993).
- Wang, A.C. and A. Bax, *J. Am. Chem. Soc.* **117**, 1810 (1995).
- Wang, A.C. and A. Bax, *J. Am. Chem. Soc.* **118**, 2483 (1996).
- Wishart, D.S., F.M. Richards, and B.D. Sykes, *Biochemistry* **31**, 1647 (1992).
- Wishart, D.S., F.M. Richards, and B.D. Sykes, *J. Mol. Biol.* **222**, 311 (1991).
- Wüthrich, K., *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York (1986).
- Yip, P. and D.A. Case, *J. Magn. Reson.* **83**, 643 (1989).
- Zhang, H., M.Sc. Thesis, University of Alberta (2001). <http://redpoll.pharmacy.ualberta.ca>.