



## Mathematical Aspects of Protein Structure Determination with NMR Orientational Restraints

J. R. QUINE\*<sup>†</sup>

Department of Mathematics,  
Florida State University,  
Tallahassee, FL 32306,  
USA  
National High Magnetic Field Laboratory,  
Florida State University,  
Tallahassee, FL 32306,  
USA

TIMOTHY A. CROSS<sup>†</sup>

Department of Chemistry,  
Florida State University,  
Tallahassee, FL 32306,  
USA  
National High Magnetic Field Laboratory,  
Florida State University,  
Tallahassee, FL 32306,  
USA

MICHAEL S. CHAPMAN<sup>†</sup>

Department of Chemistry,  
Florida State University,  
Tallahassee, FL 32306,  
USA

RICHARD BERTRAM<sup>†</sup>

Department of Mathematics,  
Florida State University,  
Tallahassee, FL 32306,  
USA

The field of structural biology is becoming increasingly important as new technological developments facilitate the collection of data on the atomic structures of proteins and nucleic acids. The solid-state NMR method is a relatively new biophysical technique that holds particular promise for determining the structures of peptides and proteins that are located within the cell membrane. This method

---

\*Corresponding address: Department of Mathematics, Florida State University, Tallahassee, FL 32306, USA. *E-mail*: quine@math.fsu.edu

<sup>†</sup>Also at: Kasha Laboratory of Biophysics, Florida State University, Tallahassee, FL 32306, USA.

provides information on the orientation of the peptide planes relative to an external magnetic field. In this article, we discuss some of the mathematical methods and tools that are useful in deriving the atomic structure from these orientational data. We first discuss how the data are viewed as tensors, and how these tensors can be used to construct an initial atomic model, assuming ideal stereochemistry. We then discuss methods for refining the models using global optimization, with stereochemistry constraints treated as penalty functions. These two processes, initial model building followed by refinement, are the two crucial steps between data collection and the final atomic model.

© 2004 Society for Mathematical Biology. Published by Elsevier Ltd. All rights reserved.

## 1. INTRODUCTION

This article discusses some recent developments in protein structure determination from nuclear magnetic resonance (NMR) data, particularly with regard to solid-state NMR of aligned samples and the observation of anisotropic NMR observables such as dipolar couplings and chemical shifts. These methods give orientational restraints rather than the distance or torsional restraints familiar from solution NMR.

Solid-state NMR is particularly promising as an aid for solving membrane protein structures using techniques such as uniformly aligned membrane protein samples in lipid bilayers (Cross and Quine, 2000). Membrane proteins form about 30% of many genomes (Wallin and Von Heijne, 1998), but only 0.5% of known structures. This paucity of structural information is due largely to the difficulty in obtaining membrane protein crystals for x-ray crystallography and in obtaining isotropic solutions for solution NMR. The techniques most appropriate for membrane proteins do not have the same long history of methods development, being newer techniques than crystallography.

This article has two parts. We first describe the mathematical tools useful in the interpretation of anisotropic NMR data. We also describe an algorithm that can help in the construction of an initial molecular model from this data. Most of these mathematical tools are useful in the analysis of both solution and solid-state NMR data. In the second part of the article we discuss computational methods that can be used to improve on or refine the initial model. Atomic refinement is commonly applied by x-ray crystallographers as a means of incorporating the x-ray diffraction data into the model building process. This computational methodology can be adapted to solid-state NMR. The computations involve nonlinear restrained optimization in which a structural model is derived that is consistent with both the experimental data and with *a priori* understanding of the molecular geometry of the components of macromolecules, based on studies of simpler model systems.

## 2. STRUCTURE OF PROTEINS AND NMR RESTRAINTS

In this section we describe an algorithm for obtaining a structure of a long polymer or protein from the orientations of a sequence of bonds. This discussion is useful for establishing the mathematical framework for using orientational restraints.

**2.1. Protein structure and discrete curves.** A protein structure can be represented as a collection of atoms together with their coordinates in three-dimensional space,  $\mathbb{R}^3$ . This may be a list of all atoms in the protein, or a list of the ones that can be observed. Thus, if  $N$  atoms are listed with coordinates we have a vector in  $(\mathbb{R}^3)^N$ . Two structures are the same if one can be transformed into the other by a sense-preserving Euclidean motion. If  $\mathbf{E}$  is the group generated by rotations and translations, it can be thought of as acting on all the coordinates listed, and a structure is an element of  $(\mathbb{R}^3)^N/\mathbf{E}$ .

The protein molecule is a sequence of amino acids of 20 different kinds. The peptide bond links into a polymeric backbone individual amino acids with 20 types of side-chains. It is convenient to think of a protein as a collection of discrete curves. This is useful both in understanding the torsion angle description of protein structures and the method of using orientational restraints and dynamics (see Section 3.3.2) to determine protein structures.

A *discrete curve* is a sequence of points  $\mathbf{p}_0, \dots, \mathbf{p}_n$  in three-dimensional space. These points can be thought of as atoms and the line segments joining atoms in the sequence can be thought of as covalent bonds. The backbone of a protein is a discrete curve consisting of points representing the atoms  $-C'-N-C_\alpha-C'-$  proceeding from the N-terminus to the C-terminus. By putting the atoms in sequential order, side-chains can also be made into a discrete curve. Thinking of a protein as a curve allows us to abstract some ideas from differential geometry to study the structure.

2.1.1. *Frenet frames.* A Frenet formalism for discrete curves will be described briefly. The idea of a Frenet frame, or moving frame, for differentiable space curves can be modified for use with discrete space curves. Let

$$s_j = |\mathbf{p}_{j+1} - \mathbf{p}_j|$$

and define a unit tangent vector at  $\mathbf{p}_j$ ,  $j = 0, \dots, n - 1$ , by

$$\mathbf{t}_j = \frac{\mathbf{p}_{j+1} - \mathbf{p}_j}{s_j}. \quad (1)$$

The points of the curve can be reconstructed up to translation from the sequences  $\{\mathbf{t}_j\}$  and  $\{s_j\}$  by

$$\mathbf{p}_k - \mathbf{p}_0 = \sum_{j=0}^{k-1} s_j \mathbf{t}_j. \quad (2)$$

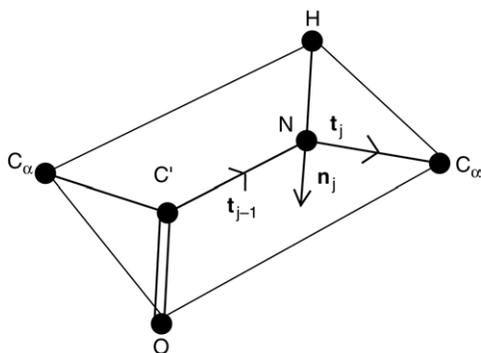


Figure 1. Atoms forming the peptide plane, and the vectors  $\mathbf{t}_j$  and  $\mathbf{n}_j$  of the Frenet frame that span the plane.

If  $\mathbf{t}_{j-1}$  and  $\mathbf{t}_j$  are not parallel, binormal and normal vectors can be given by

$$\mathbf{b}_j = \frac{\mathbf{t}_{j-1} \times \mathbf{t}_j}{|\mathbf{t}_{j-1} \times \mathbf{t}_j|}, \quad \mathbf{n}_j = \mathbf{b}_j \times \mathbf{t}_j \quad (3)$$

and a right-handed orthogonal frame by

$$\mathbf{F}_j = (\mathbf{t}_j, \mathbf{n}_j, \mathbf{b}_j). \quad (4)$$

This will be referred to as the Frenet frame at  $\mathbf{p}_j$ . (Vectors are column vectors, and a frame is a sequence of three linearly independent vectors considered as columns of a nonsingular  $3 \times 3$  matrix. Orthogonal frames correspond to orthogonal matrices and right-handed orthogonal frames to rotation matrices.)

The Frenet frames can be thought of as molecular frames along the molecule. The plane formed by the tangent and the normal vector at a point contains the point together with the previous and the subsequent points. At a nitrogen atom in the protein backbone, the vectors  $\mathbf{t}$  and  $\mathbf{n}$  span the peptide plane (Fig. 1) and  $\mathbf{b}$  is a peptide plane normal.

The Frenet frames are also related to the torsion angles used in the study of molecular structure. The relationship of one Frenet frame to the next is given by

$$\mathbf{F}_{j+1} = \mathbf{F}_j \mathbf{R}_1(\tau_j) \mathbf{R}_3(\theta_{j+1}) \quad (5)$$

where  $\theta_j = \arccos(\mathbf{t}_{j-1} \cdot \mathbf{t}_j)$  is the exterior bond angle at  $\mathbf{p}_j$ , and  $\tau_j$  is the angle of torsion about the bond direction  $\mathbf{t}_j$  (Fig. 2). Here

$$\mathbf{R}_1(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \quad \mathbf{R}_3(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (6)$$

are rotation matrices. Thus the discrete curve can be reconstructed up to a Euclidean motion from the sequences  $\{s_j\}$  of bond lengths,  $\{\mathbf{t}_{j-1} \cdot \mathbf{t}_j\}$  of bond angle cosines, and  $\{\tau_j\}$  of torsion angles.

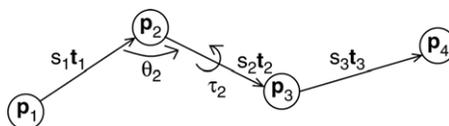


Figure 2. The covalent bond angle  $\theta_2$  at  $\mathbf{p}_2$  and the angle of torsion  $\tau_2$  about the bond direction  $\mathbf{t}_2$ .

2.1.2. *Standard protein geometry.* For finding and describing protein structures with limited structural data it is often assumed that bond lengths and angles have standard (or ideal) values depending on the type of bond (see Section 3.1.1). With bond angles and bond lengths given, the above discussion shows how the structure of the protein backbone can be determined given a torsion angle for each bond.

2.2. *Restraints from NMR.* Since about 1957, the technique of nuclear magnetic resonance (NMR) has been used to find the structure of peptides and, more recently, of proteins. Today, there are two different methods, solution NMR and solid-state NMR. In the solution NMR method the molecules are tumbling in the solvent during the experiment at a rate that is fast on the NMR timescale (Evans, 1995); in the solid-state NMR method the molecules are rigid or they are restricted to anisotropic motion. This dichotomy between isotropic and anisotropic motion is not absolute; there is always some motion of the molecule and the motion may be more or less restricted. An intermediate case is exemplified by the situation where residual dipolar couplings are measured by means of solution NMR, as discussed later.

The difference between the NMR techniques of structure determination is in the type of geometric information that can be obtained from the experiment. The solution NMR method predominantly measures distance restraints and the solid-state NMR method predominantly measures orientational restraints. A distance restraint is an equation or inequality involving the distance between two atoms in a molecule. An orientational restraint is an equation or inequality involving an angle between the external magnetic field direction and the vector between two atoms in the molecule. Often the atoms are covalently bonded and this angle is referred to as a bond orientation angle. A hybrid distance/orientational restraint is given by residual dipolar couplings in the solution NMR method.

The difference between orientational restraints and distance restraints can be thought of in terms of groups of Euclidean motions. Distance information is the only type of information invariant under the full group  $\mathbf{E}$  of rigid body motions. Let  $\mathbf{B}$  be the *unit direction* of the magnetic field (usually  $\mathbf{B}$  has magnitude equal to the strength of the field), and let  $\mathbf{E}_{\mathbf{B}}$  be the subgroup leaving  $\mathbf{B}$  fixed. The group  $\mathbf{E}_{\mathbf{B}}$  can also be thought of as the group generated by translations and rotations about the axis  $\mathbf{B}$ . Orientational restraints are invariant under the group  $\mathbf{E}_{\mathbf{B}}$ , but not under the full group  $\mathbf{E}$ .

2.2.1. *Distance restraints.* The basic principles of distance geometry can be expressed in terms of a distance matrix. For a sequence of points  $\mathbf{p}_0, \dots, \mathbf{p}_n$ , the distance matrix  $\mathbf{D}$  is defined to be the matrix with entries  $|\mathbf{p}_i - \mathbf{p}_j|^2$ . For the purpose of finding the structure, we may assume that  $\mathbf{p}_0 = \mathbf{0}$ . We also define the gram matrix  $\mathbf{G}$  for the sequence of vectors  $\mathbf{p}_1, \dots, \mathbf{p}_n$  to be the matrix with entries  $\mathbf{p}_i \cdot \mathbf{p}_j$ . The identity  $|\mathbf{p}_i - \mathbf{p}_j|^2 = |\mathbf{p}_i|^2 - 2\mathbf{p}_i \cdot \mathbf{p}_j + |\mathbf{p}_j|^2$  shows that the gram matrix can be computed from the distance matrix using row and column operations. A set of coordinates for the points can be obtained by using the eigenvalues and eigenvectors of the symmetric matrix  $\mathbf{G}$  to write

$$\mathbf{G} = \mathbf{M}^t \mathbf{M} \quad (7)$$

where  $\mathbf{M}$  is a  $3 \times n$  matrix. The columns of  $\mathbf{M}$  are then coordinates for the points  $\mathbf{p}_1, \dots, \mathbf{p}_n$ . To write  $\mathbf{G}$  as the ‘square’ of a  $3 \times n$  matrix  $\mathbf{M}$  as in (7),  $\mathbf{G}$  must be positive definite and rank 3 and this restricts which matrices can be distance matrices (Havel and Dress, 1993). So without perfect data, the gram matrix computed from the data might have rank greater than 3, which would result in the coordinates of the atoms being in some higher dimensional space.

2.2.2. *Orientalional restraints.* While in principle it is possible to determine the structure using complete distance information, in practice such information is rarely, if ever, available. The data is supplemented with tables of average bond angles and lengths in peptides.

Mathematically, the method for obtaining coordinates from orientational restraints is similar to that for obtaining distance restraints (Brenneman and Cross, 1990; Ramamoorthy *et al.*, 1995) in that matrices of dot products are used. A complete gram matrix is not available because dot products are available only for selected vectors. Average bond angles and lengths supplement the information in the matrices.

Orientalional restraints are easiest to obtain in the solid-state NMR method for vectors joining covalently bonded atoms. To compute a structure from these restraints, consider a discrete curve and suppose that the values  $s_j$ , representing bond lengths, and  $\mathbf{t}_{j-1} \cdot \mathbf{t}_j$ , representing the negatives of the cosines of bond angles, are known. If orientations  $\mathbf{B} \cdot \mathbf{t}_j$  (cosines of angles of bonds with the chosen direction of the magnetic field) are known and none of these values are  $\pm 1$ , then there are only a finite number of structures possible for the curve. It is sufficient to find coordinates for the unit vectors  $\mathbf{t}_j$ . These can be found recursively (Quine, 1999). Suppose, for example, that  $\mathbf{B}$ ,  $\mathbf{t}_{j-1}$ ,  $\mathbf{B} \cdot \mathbf{t}_j$ , and  $\mathbf{t}_{j-1} \cdot \mathbf{t}_j$  are given; then

$$\mathbf{t}_j = \frac{1}{(1 - \kappa_j^2)} ((\beta_j - \kappa_j \kappa_{j-1}) \mathbf{t}_{j-1} + (\kappa_j - \beta_j \kappa_{j-1}) \mathbf{B} + \varepsilon_j \sqrt{g_j} \mathbf{t}_{j-1} \times \mathbf{B}) \quad (8)$$

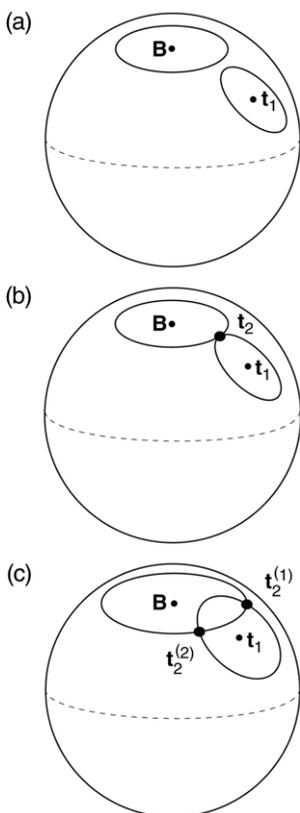


Figure 3. Illustration of (8). The sphere represents the set of unit vectors. The equations  $\kappa_2 = \mathbf{B} \cdot \mathbf{t}_2$  and  $\beta_2 = \mathbf{t}_1 \cdot \mathbf{t}_2$  indicate that given  $\mathbf{t}_1$ , the vector  $\mathbf{t}_2$  lies on both of two circles about  $\mathbf{B}$  and  $\mathbf{t}_1$ . (a) If  $g_2 < 0$  the circles do not intersect and there is no solution. (b) If  $g_2 = 0$  then the circles intersect at one point and there is only one possibility for  $\mathbf{t}_2$ . (c) If  $g_2 > 0$  then there are two possibilities for  $\mathbf{t}_2$  distinguished by chirality  $\varepsilon_2$ . The vector  $\mathbf{t}_2^{(1)}$  corresponds to  $\varepsilon_2 = -1$  and the vector  $\mathbf{t}_2^{(2)}$  corresponds to  $\varepsilon_2 = 1$ .

where

$$\kappa_j = \mathbf{B} \cdot \mathbf{t}_j \quad \beta_j = \mathbf{t}_{j-1} \cdot \mathbf{t}_j \quad \varepsilon_j = \pm 1$$

and where

$$g_j = \det \begin{pmatrix} 1 & \kappa_{j-1} & \kappa_j \\ \kappa_{j-1} & 1 & \beta_j \\ \kappa_j & \beta_j & 1 \end{pmatrix}.$$

Equation (8) determines the vector  $\mathbf{t}_j$  up to two possibilities depending on  $\varepsilon_j$  (Fig. 3).

Since there is a choice of  $\varepsilon = \pm 1$  at each stage, and since these values cannot be determined from the values of  $\mathbf{B} \cdot \mathbf{t}_j$ , there are  $2^{n-1}$  structures for this curve compatible with the constraints. Since

$$\varepsilon_j = -\text{sign } \mathbf{B} \cdot (\mathbf{t}_{j-1} \times \mathbf{t}_j), \quad (9)$$

the structural elucidation requires the determination of chiralities. Knowledge of the values  $\mathbf{B} \cdot \mathbf{t}_j$  and  $\varepsilon_j$  is equivalent to the knowledge of the coordinates of  $\mathbf{B}$  in each Frenet frame and this determines the structure using (2), (9), and (8).

Sometimes the coordinates of  $\mathbf{B}$  in a Frenet frame can be determined by obtaining orientations of other vectors whose coordinates are known in the Frenet frame. At an alpha carbon, for example, the four bond directions are known to have approximate tetrahedral geometry. Choosing the alpha carbon and any three neighboring atoms, i.e., any three of the four bonds, the sign of the scalar triple product of three bond directions is known from the chirality (usually  $L$ ) of the amino acid. In this case any of the four vectors can be written uniquely in the Frenet frame and so the coordinates of  $\mathbf{B}$  in the Frenet frame are known from the dot products of  $\mathbf{B}$  with any three of the four alpha carbon bond vectors.

The problem of obtaining chiralities is analogous to the phase problem in x-ray crystallography, where a set of phases must be chosen and used with the diffraction intensity data to obtain a structure.

Note that  $g_j$  is the determinant of the gram matrix of the sequence of vectors  $\mathbf{B}$ ,  $\mathbf{t}_{j-1}$ ,  $\mathbf{t}_j$ , and as such should be non-negative. Problems in solving for the structure occur if inconsistencies in the data cause this determinant to be negative. This is similar to the situation in distance geometry when imperfect distance constraints give gram matrices which are not of rank 3.

The expression  $1 - (\mathbf{B} \cdot \mathbf{t}_{j-1})^2$  in the denominator of (8) also causes problems in this method of solving structures from orientational restraints. The recursive solution works only if none of the tangent vectors of the discrete curve are parallel to  $\mathbf{B}$ , in which case there can be an infinite number of structures consistent with the data. This is because changing the single torsion angle around a bond vector parallel to  $\mathbf{B}$  will give a curve with the same sequence of values  $\{s_j\}$ ,  $\{\mathbf{B} \cdot \mathbf{t}_j\}$ , and  $\{\mathbf{t}_{j-1} \cdot \mathbf{t}_j\}$ . The algorithm is also numerically unstable if any of the values  $\mathbf{B} \cdot \mathbf{t}_j$  is close to  $\pm 1$ , that is, if any of the vectors  $\mathbf{t}_j$  are nearly parallel to  $\mathbf{B}$ .

**2.3. Obtaining orientational restraints.** Orientational restraints are obtained from the solid-state NMR method and can also be obtained from the solution NMR method from residual dipolar couplings. In general, an NMR experiment detects the radio-frequency precession of nuclear spins in a molecule (Fig. 4). In quantum mechanics, this frequency represents a discrete difference in energy levels and the energy levels are eigenvalues of a Hamiltonian matrix. The strongest interaction affecting the energy is given by the Zeeman Hamiltonian,  $H = B_0\gamma I_z$ , where  $B_0$  is the intensity of the magnetic field,  $\gamma$  is a gyromagnetic ratio depending only on the type of the atom, and  $I_z$  is a spin matrix with the  $z$  axis being in the direction of the magnetic field. The Zeeman Hamiltonian represents the effect of the magnetic field on the nuclear spin. Using Schrödinger's equation with the Zeeman Hamiltonian shows that in the absence of other interactions the spin of a nucleus precesses at a frequency  $B_0\gamma$  radians per second, the Larmor frequency.

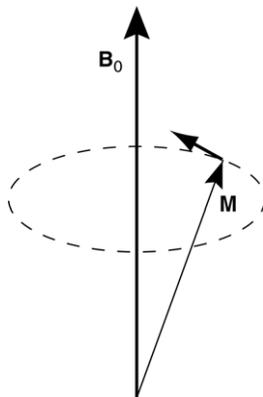


Figure 4. NMR experiments detect the radio-frequency precession about a magnetic field  $\mathbf{B}_0$  of the bulk magnetization  $\mathbf{M}$  of nuclear spins of atoms in a molecule.

What makes structure determination possible is that the spins in the molecule interact, so the observed frequency differs from the Larmor frequency. In quantum mechanics the interaction arises from other Hamiltonians added to the Zeeman Hamiltonian. These Hamiltonians are functions of the orientation of the molecule with respect to the magnetic field direction  $\mathbf{B}$ . The dependence of these interactions on the orientation can be detected only if the molecules maintain some average orientation with respect to  $\mathbf{B}$  as in the solid-state NMR method. In the solution NMR method the orientational dependence is typically averaged out due to isotropic tumbling of the molecules.

The orientational restraints discussed can be thought of as tensors because they are given as a quadratic expression in the coordinates of the unit magnetic field direction. Any quadratic expression  $a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{23}yz + 2a_{13}xz$  can be written as a symmetric matrix with entries  $a_{ij}$ . As a tensor, it can be discussed in terms of its principal values and principal axis frame, and the latter can be expressed in terms of the molecular (Frenet) frames we have discussed.

2.3.1. *The dipolar interaction.* The most common interaction used to obtain orientational restraints is the dipolar interaction. The effect of a spin 1/2 atom on another, for example, results in splitting the Larmor frequency into two frequencies, a single peak becoming a doublet (Fig. 5). The difference of these two frequencies is a function of the angle between  $\mathbf{B}$  and the vector joining the two atoms. This difference, representing a change in energy  $\nu$ , is given by

$$\nu = \frac{\nu_{\parallel}}{2}(3(\mathbf{u} \cdot \mathbf{B})^2 - 1) \quad (10)$$

where  $\mathbf{u} = \mathbf{r}/r$  is the unit vector in the direction of a vector  $\mathbf{r}$  joining the centers of the two atoms,  $\nu_{\parallel} = C\gamma_1\gamma_2/r^3$ ,  $\gamma_1$  and  $\gamma_2$  are the gyromagnetic ratios of the two atoms, and  $C$  is a constant. Only the absolute value  $|\nu|$  of the dipolar coupling is

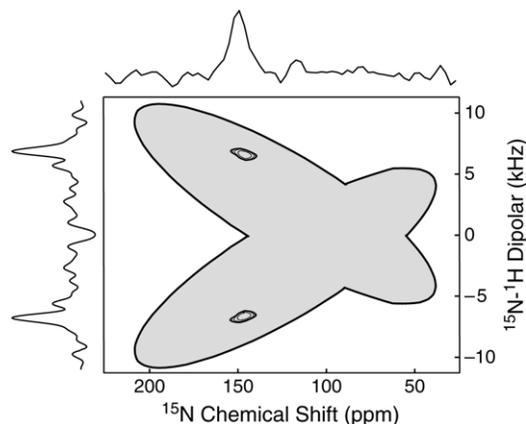


Figure 5. A 2D PISEMA spectrum (superimposed on a powder pattern). The vertical 1D spectrum shows the dipolar splitting. The separation between the peaks is  $2\nu$ . The position of the peak on the horizontal spectrum indicates the chemical shift  $\sigma$ .

typically observed in solid state experiments. Problems in determining the sign of  $\nu$  are discussed in Section 3.4.1.

The experimentally measured value of  $|\nu|$  is the simplest example of an orientational restraint. In many instances dipolar interactions between bonded atoms are measured where the distance  $r$  between them is known. Thus  $\nu_{\parallel}$  in equation (10) is typically assumed constant, and the equation is solved up to a finite number of possibilities for  $\mathbf{u} \cdot \mathbf{B}$ , and the orientation of one bond direction with respect to the magnetic field is constrained to a finite number of possible values.

The dipolar restraint can be expressed more generally in terms of a tensor. Let  $\mathbf{F}_P$  be a principal axis frame for the tensor, an orthonormal frame with the third vector, the  $z$  direction, in the direction  $\mathbf{u}$  of a covalent bond along which a dipolar interaction is measured. If  $(x, y, z)$  are the coordinates of  $\mathbf{B}$  in this frame, then the splitting (10) is equal to  $\frac{\nu_{\parallel}}{2}(3z^2 - 1)$ . Since  $\mathbf{B}$  is a unit vector, the splitting is equal to  $\frac{\nu_{\parallel}}{2}(2z^2 - x^2 - y^2)$  and this can be thought of as the quadratic tensor given in its principal axis frame  $\mathbf{F}_P$  by the diagonal matrix  $\frac{\nu_{\parallel}}{2}\text{diag}(2, -1, -1)$ . This is a traceless tensor and the corresponding function is harmonic, the zonal spherical harmonic.

**2.3.2. The chemical shift.** The deviation from the Larmor frequency due to the spins of neighboring orbiting electrons is called the chemical shift. The chemical shift can also be expressed as a tensor, a quadratic expression in  $\mathbf{B}$ . In contrast to the dipolar tensor it is directly proportional to the intensity of the magnetic field. The principal values are not easily computed from quantum mechanics and must be experimentally determined (Oas *et al.*, 1987; Brender *et al.*, 2001; Lee *et al.*, 2001; Teng *et al.*, 1992).

**2.3.3. NMR tensors and the Frenet frame.** The structural significance of the observed values of the NMR tensors depends on knowing, in addition to the

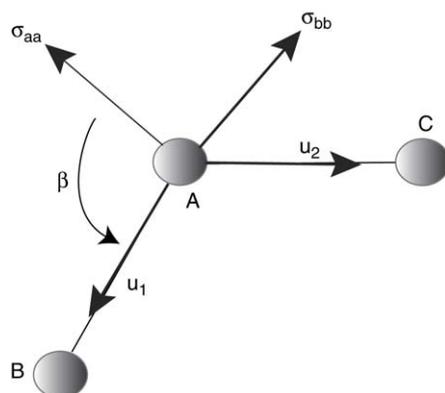


Figure 6. The principal major axis  $\sigma_{aa}$  for the  $^{15}\text{N}$  chemical shift tensor at atom A is often in the plane of the three bonded atoms and is determined by an angle  $\beta$  from one of the unit bond vectors. By experimental characterization, one other major axis  $\sigma_{bb}$  is generally in the plane with the third major axis parallel to  $\mathbf{u}_1 \times \mathbf{u}_2$  so  $\beta$  characterizes the  $\mathbf{F}_P$  with respect to the Frenet frame.

principal values of the tensors, the relationship of the principal axis frames to the Frenet or molecular frame.

For the dipolar tensor, the unique principal axis is in the direction of a bond vector, and this bond vector can be expressed in the Frenet frame using the local geometry of the molecule.

The chemical shift tensor can also be expressed as a tensor in terms of a Frenet frame  $\mathbf{F}$ . The tensor is given by specifying the principal values and writing the principal axis frame  $\mathbf{F}_P$  as  $\mathbf{F}_P = \mathbf{F}\mathbf{R}$  for some fixed rotation matrix  $\mathbf{R}$ . Due to the spatial geometry of the orbiting electrons in a peptide plane, this rotation matrix for backbone atoms can often be given as a matrix of the form  $\mathbf{R}_3(\beta)$  for some angle  $\beta$  (Fig. 6). Both  $\mathbf{R}$  and the principal values are measured by powder experiments where a sample is studied with the molecules in random orientations. Often from rotation patterns using a single crystal it can be deduced that  $\beta$  and the principal values are constant for a given type of atom, e.g., a backbone nitrogen in a protein structure (Mai *et al.*, 1993).

2.3.4. *2D NMR, PISEMA.* Two-dimensional methods using NMR such as PISEMA allow the measurement of both the anisotropic chemical shift and the dipolar splitting from a single signal (Ramamoorthy and Opella, 1995; Tian *et al.*, 1998; Ramamoorthy *et al.*, 1999). This also provides added information in determining whether  $\nu = |\nu|$  or  $\nu = -|\nu|$ . This sign indeterminacy is often a problem in getting full structural information from orientation restraints. From (10) it follows that  $\nu$  is between  $-\frac{\nu_{\parallel}}{2}$  and  $\nu_{\parallel}$  and so if  $\nu$  is greater than  $\frac{\nu_{\parallel}}{2}$  then  $\nu = |\nu|$ . Methods such as PISEMA give added information on the sign indeterminacy. The possible values  $(\sigma, \nu)$  of the chemical shift and dipolar splitting are given as an

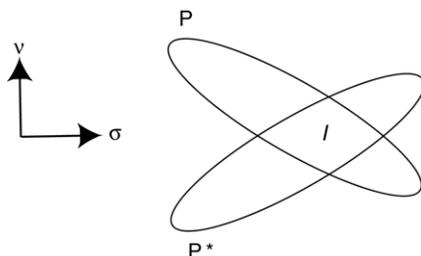


Figure 7. The powder pattern for the PISEMA experiment is sketched as the union of two ellipse-shaped regions  $P$  and  $P^*$ . The chemical shift, dipolar splitting pair  $(\sigma, \nu)$  must be in  $P$ . If  $(\sigma, \pm|\nu|)$  are not in the intersection  $I$  of  $P$  and  $P^*$  then the sign of  $\nu$  is determined by this fact.

ellipse-shaped set,  $P$ , related to the powder pattern (Fig. 7). We also consider the reflection  $P^*$  of  $P$  consisting of all possible values of  $(\sigma, -\nu)$ . If the resonance does not lie in the intersection of these sets, then the sign of  $\nu$  can be determined.

Another useful technique for resolving degeneracies with PISEMA is using the characteristic two-dimensional patterns made by protein helices. This is related to the concepts of *PISA wheels* (Marassi and Opella, 2000; Denny *et al.*, 2001) and *dipolar waves* (Mascioni and Veglia, 2003; Mesleh and Opella, 2003).

**2.4. Tensor averaging.** All samples experience some form of motion. At one extreme, isotropic motion, the observed tensors are averaged over all possible orientations and the observed value is

$$\text{Av } \mathbf{T} = \frac{1}{4\pi} \int_{\mathbf{B}} \mathbf{B}' \mathbf{T} \mathbf{B} dA = \frac{1}{3} \text{Trace}(\mathbf{T}),$$

where  $\mathbf{B}$  is considered as a point on the sphere of unit vectors,  $dA$  is area measure on the sphere, and  $\mathbf{T}$  is the observed tensor. (Although the tensor is taking a random orientation in the lab frame, it is easier to think of  $\mathbf{T}$  as fixed and  $\mathbf{B}$  at a random position in the principal axis frame of  $\mathbf{T}$ .)

2.4.1. *Residual dipolar couplings.* In the solution NMR method the observed value of the dipolar coupling tensor is zero since it is a traceless tensor. Suppose, however, that the motion is not perfectly isotropic, possibly due to diamagnetic susceptibility of the molecules or to other large molecules hindering the motion, and that the preference of certain orientations can be given by a weighting with a symmetric quadratic tensor  $\mathbf{S}$ , so that the observed value of the tensor  $\mathbf{T}$  is

$$\text{Av}_{\mathbf{S}} \mathbf{T} = \frac{1}{4\pi} \int_{\mathbf{B}} (\mathbf{B}' \mathbf{T} \mathbf{B}) (\mathbf{B}' \mathbf{S} \mathbf{B}) dA,$$

(see Prestegard *et al.*, 1999). Computing in the principal axis frame of  $\mathbf{S}$ , suppose  $\mathbf{S} = \text{diag}(\chi_1, \chi_2, \chi_3)$  and  $\mathbf{T} = (t_{ij})$ ; then

$$\text{Av}_{\mathbf{S}} \mathbf{T} = \frac{1}{5} \left( t_{11} \left( \chi_1 + \frac{\chi_2 + \chi_3}{3} \right) + t_{22} \left( \chi_2 + \frac{\chi_3 + \chi_1}{3} \right) + t_{33} \left( \chi_3 + \frac{\chi_1 + \chi_2}{3} \right) \right). \quad (11)$$

If the tensor  $\mathbf{T}$  is traceless ( $t_{11} + t_{22} + t_{33} = 0$ ), then from (11) any scalar tensor can be added to  $\mathbf{S}$  without changing the average. So  $\mathbf{S}$  can be assumed to be traceless and (11) becomes

$$\text{Av}_{\mathbf{S}} \mathbf{T} = \frac{2}{15} (t_{11}\chi_1 + t_{22}\chi_2 + t_{33}\chi_3). \quad (12)$$

Applying this to the dipolar tensor (10) it follows that

$$\text{Av}_{\mathbf{S}} v = \frac{v_{\parallel}}{5} (u_1^2 \chi_1 + u_2^2 \chi_2 + u_3^2 \chi_3) \quad (13)$$

where  $(u_1, u_2, u_3)$  are the coordinates of the bond direction vector  $\mathbf{u}$  in the principal axis frame of  $\mathbf{S}$ . Writing in spherical coordinates,

$$(u_1, u_2, u_3) = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^t$$

the average tensor can also be written as

$$\text{Av}_{\mathbf{S}} v = \frac{v_{\parallel}}{10} ((\chi_1 - \chi_2) \sin^2 \theta \cos 2\phi + \chi_3 (3 \cos^2 \theta - 1)). \quad (14)$$

The result of the above discussion is that the observed dipolar tensor under this form of non-isotropic motion is not zero as is typical of solution NMR, but is given by the absolute value of a traceless tensor  $\text{Av}_{\mathbf{S}} v$  with the principal axis frame the same as that of  $\mathbf{S}$ . The observed absolute value of  $\text{Av}_{\mathbf{S}} v$  is referred to as a *residual dipolar coupling*.

Residual dipolar couplings give important orientational restraints and information about the direction vector  $\mathbf{u}$  of a bond, but the orientation is with respect to the principal axis frame of  $\mathbf{S}$ . This frame is not known *a priori* but must be determined from the data, as must the principal values  $\chi_1$ ,  $\chi_2$ , and  $\chi_3$  (Clore *et al.*, 1998; Tjandra *et al.*, 2000). So, in addition to the structure, the traceless tensor  $\mathbf{S}$  must be found, which adds five new parameters to the ones that determine structure.

2.4.2. *Magic angle spinning.* In another tensor averaging procedure, called magic angle spinning, the sample is spun at a high frequency about an axis making an angle  $\theta$  with the magnetic field. To compute the average let  $\mathbf{T} = (t_{ij})$  be the coordinates of a symmetric tensor in a frame where the  $z$  axis is along the axis of the spinning sample. Let  $\mathbf{B} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^t$ , then

$$\text{Av}_{\phi} \mathbf{T} = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{B}' \mathbf{T} \mathbf{B} d\phi = \frac{\text{Trace}(\mathbf{T})}{2} \sin^2 \theta + \frac{t_{33}}{2} (3 \cos^2 \theta - 1). \quad (15)$$

If the tensor is traceless and the sample is spun at the magic angle where  $3 \cos^2 \theta - 1 = 0$ , then the observed value of the tensor is 0. In the case of the dipolar splitting, since the tensor is traceless, no splitting is observed in magic angle spinning, and in this way other interactions can be more readily observed.

### 3. FORCE FIELDS AND ATOMIC REFINEMENT

The method described above for using orientational information to obtain atom locations is useful in the determination of an initial protein structure. To improve on this it is appropriate to refine the structure using all available data and all stereochemical information as *restraints* rather than *constraints*. That is, rather than strictly enforcing agreement of the model with the experimental data, one constructs a potential energy function consisting of a sum of stereochemical force fields and penalty functions for deviation from experimental data. Refinement then consists of minimizing the energy function using a combination of global and local minimization algorithms. In this section we first describe the components of the energy function, and then the optimization strategies commonly applied to locate the global minimum, which corresponds to the native conformation of the model protein (Anfinsen, 1973). We then discuss how orientational information from solid-state NMR can be used in refinement, and some of the mathematical properties of the orientational penalty functions.

**3.1. Stereochemical force fields.** Atoms within a protein are affected by two types of interatomic forces: those due to covalent bonding and those due to nonbonded interactions. Although these forces can be described with quantum mechanical formulations, the vast majority of protein structure determinations are made using simpler classical formulations. A number of different force field sets have been developed for proteins and nucleic acids, parametrized for specific sets of amino acids or nucleotides. We will focus on the CHARMM force field set (Brooks *et al.*, 1983), since this is used in the popular atomic refinement software packages X-Polar and CNS (Brünger, 1992a,b; Brünger *et al.*, 1998). This was also used in the computer software TORC, developed for refinement using orientational data (Ketchem *et al.*, 1997).

3.1.1. *Bonded interactions.* Covalent bond lengths and angles are quite rigid and predictable. Lengths range from  $\approx 1 \text{ \AA}$  for bonds involving H to  $\approx 1.5 \text{ \AA}$  for all other bonds. Angles depend on the hybridization and thus the valence of the atoms, with angles of  $\approx 109^\circ$  for tetrahedral arrangements (e.g., bonds at  $C_\alpha$ ) and angles of  $\approx 120^\circ$  for trigonal-planar arrangements (e.g., bonds at carbonyl C and amide N) (MacKerell *et al.*, 1998). The deviation from mean values is small, with a standard deviation of less than  $0.02 \text{ \AA}$  for bond lengths and  $\approx 2^\circ$  for bond angles (Hendrickson, 1985). To restrain bond lengths and angles harmonic force fields are

used, penalizing deviation from mean ‘ideal’ values:

$$E_{\text{bonds}} = \sum_{\text{bonds}} K_b (b - b_0)^2 \quad (16)$$

$$E_{\text{angles}} = \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2, \quad (17)$$

where summation is over all covalent bonds,  $b_0$ ,  $\theta_0$  are the ideal bond lengths and angles, and  $K_b$ ,  $K_\theta$  are force constants. Values of these parameters depend on the types of atoms and their locations in the protein main chain or side-chains. The harmonic restraints are analogous to the restoring energy for two masses coupled by a spring.

Torsion angles describe the ‘twisting’ of two bonds about a third. Consider four atoms  $\mathbf{p}_1, \dots, \mathbf{p}_4$  joined sequentially through covalent bonds (Fig. 2). Let  $\mathbf{a} = \mathbf{p}_2 - \mathbf{p}_1$ ,  $\mathbf{b} = \mathbf{p}_3 - \mathbf{p}_2$ ,  $\mathbf{c} = \mathbf{p}_4 - \mathbf{p}_3$ . Then the torsion angle about  $\mathbf{b}$  is the dihedral angle between the plane spanned by  $\mathbf{a}$ ,  $\mathbf{b}$  and the plane spanned by  $\mathbf{b}$ ,  $\mathbf{c}$ . Thus, it is the angle between vectors normal to these planes,  $\mathbf{a} \times \mathbf{b}$  and  $\mathbf{b} \times \mathbf{c}$ , respectively. A convenient formula for computing dihedral angles uses the argument of a 2D vector  $(x, y)$ , written as  $\arg(x, y)$  with  $-180^\circ < \arg(x, y) < 180^\circ$ . This is the angle made by the vector with the positive  $x$  axis. The torsion angle about  $\mathbf{b}$  is then

$$\tau = \arg(-|\mathbf{b}|^2 \mathbf{a} \cdot \mathbf{c} + (\mathbf{b} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{c}), |\mathbf{b}| \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})). \quad (18)$$

The torsion angle energy term is periodic, reflecting different hybridizations about the bonded atoms:

$$E_{\text{torsion}} = \sum_{\tau} K_\tau \cos(n\tau - \delta), \quad (19)$$

where  $n$  is an integer, and  $K_\tau$ ,  $n$ ,  $\delta$  depend on the atoms forming the torsion angle. Torsion angles formed by atoms in aromatic groups are restrained by a harmonic energy function to maintain planarity, and a harmonic energy function is used to enforce chirality about the peptide bond. These are sometimes called *improper angles*:

$$E_{\text{improper}} = \sum_{\text{improper}} K_{\text{imp}} (\tau - \tau_0)^2, \quad (20)$$

where  $\tau_0$  is the ideal torsion angle value.

3.1.2. *Nonbonded interactions.* In addition to covalent interactions, protein atoms interact through van der Waals and electrostatic forces. The van der Waals interaction between two atoms is repulsive for short distances due to overlap of their electron clouds, and attractive for larger distances due to mutual induction of

electrostatic dipoles formed from local fluctuations of electron density. The force between the atoms is zero at the van der Waals contact distance, which has been measured or determined from quantum calculations for different combinations of atoms. The van der Waals interaction is often described by a Lennard-Jones potential:

$$E_{\text{vdW}} = \sum_{\text{vdW}} \left( \frac{A}{r^{12}} - \frac{B}{r^6} \right), \quad (21)$$

where  $r$  is the interatomic distance and  $A$ ,  $B$  are determined by the types of atoms involved in the interaction.

Charged atoms produce electrostatic interactions with neighboring atoms, described by

$$E_{\text{elec}} = \sum_{i,j} \frac{q_i q_j}{\epsilon r_{ij}}, \quad (22)$$

where  $q_i$ ,  $q_j$  are charges on two neighboring atoms and  $r_{ij}$  is the distance between. The dielectric constant  $\epsilon$  is typically assumed to be constant, but in fact it varies throughout the protein due to the nonuniform chemical environment. Partly for this reason, and partly due to difficulties in assessing atomic charges, the electrostatic force field is often omitted, and the van der Waals interaction alone is used to describe nonbonded interactions. The omission of an explicit electrostatic force field may have serious consequences in a low dielectric constant environment, such as within a lipid layer. In such cases, charged atoms are thought to play a particularly important role in the tertiary structure.

**3.2. Penalty functions.** In principle, the sum of the stereochemical force fields

$$E_{\text{chem}} = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{torsion}} + E_{\text{improper}} + E_{\text{vdW}} + E_{\text{elec}} \quad (23)$$

is sufficient to describe the interatomic interactions, and minimization of this potential energy function should give the native conformation of the protein. However, our knowledge of interatomic forces within a protein is incomplete, and experience has shown that the stereochemical force fields alone are insufficient to describe the protein. Instead, accurate atomic structures are determined using data from x-ray crystallography or an NMR method. These data are used to restrain atom locations by introducing the penalty function

$$E_{\text{data}} = \sum_{\text{data}} (s_c - s_o)^2, \quad (24)$$

where  $s_o$  is an experimental observable and  $s_c$  is the corresponding quantity calculated from the model. Summation is over all data. This penalty term is weighted by a factor  $w$  and added to  $E_{\text{chem}}$ , completing the potential energy function:

$$E = E_{\text{chem}} + w E_{\text{data}}. \quad (25)$$

The process of minimizing this function is called *atomic refinement*.

The choice of the weight  $w$  applied to the experimental data is problematic. Large  $w$  puts more emphasis on the data, which is incomplete and subject to error, often leading to bad stereochemistry. Small  $w$  emphasizes the stereochemical force fields, which are only approximations to the interatomic forces. A good refinement requires an appropriate balancing of stereochemical and experimental restraints. The approach taken in recent years is to remove a small fraction of the data from the refinement process, and to use this as a test set (Brünger, 1992a,b). Multiple refinements are then performed, using a range of values for  $w$ . The weight that gives the best agreement between the model and the test set is then used for the final refinement; one discards other refinements.

### 3.3. Optimization techniques.

3.3.1. *Cartesian refinement.* A major obstacle to atomic refinement is global minimization of (25), since the landscape defined by this function is studded with local minima. When local minimization algorithms such as steepest descent or conjugate gradient are used, the structure often gets trapped in local minima far from the global minimum, so these may be of limited use in atomic refinement. The method used most often is molecular dynamics with simulated annealing. The dynamics is described by Newton's second law:

$$m_i \frac{d^2 \mathbf{x}_i}{dt^2} = -\nabla_i E \quad (26)$$

where  $m_i$  and  $\mathbf{x}_i$  are the mass and location of atom  $i$ , respectively. Simulated annealing (Kirkpatrick *et al.*, 1983) introduces a computational temperature, which is a measure of kinetic energy. At high temperature there is a great deal of kinetic energy, allowing the system to escape local potential energy minima. At low temperature the conformational search is more restricted by potential energy barriers. There are several approaches to temperature control (Brünger and Rice, 1997), one of which will be described here.

The *computational temperature* is defined as

$$T = \frac{2E_{\text{kin}}}{3nk_B} \quad (27)$$

where  $n$  is the number of atoms,  $k_B$  is Boltzmann's constant, and  $E_{\text{kin}}$  is the kinetic energy of the system:

$$E_{\text{kin}} = \sum_{i=1}^n \frac{1}{2} m_i \left( \frac{d\mathbf{x}_i}{dt} \right)^2. \quad (28)$$

With the *velocity scaling* approach to temperature control, the atom velocities are periodically and uniformly scaled so that the computational temperature of the

system matches a target temperature ( $T_{\text{tgt}}$ ):

$$\mathbf{v}_i = \frac{d\mathbf{x}_i}{dt} \sqrt{T_{\text{tgt}}/T}, \quad (29)$$

where  $\mathbf{v}_i$  is the new or scaled velocity of atom  $i$ . After scaling, the integration of (26) is restarted with the current atom positions and the scaled velocities. In simulated annealing, the target temperature is started at a high value (hundreds or thousands of kelvins) and slowly lowered. The annealing schedule describes the starting temperature and the rate at which temperature is lowered. The success of annealing depends largely on the schedule, and different annealing schedules seem to work best for different proteins. Following annealing, one typically applies a local minimization method such as the conjugate gradient one to move the system into the nearest local minimum.

3.3.2. *Torsion angle refinement.* The chief disadvantage of Cartesian molecular dynamics is the large number of independent variables, three times the number of atoms. Since proteins typically contain more than 1000 and often more than 10,000 atoms, this leads to two problems. First the large system is computationally expensive to integrate. Second, and more importantly, the ratio of data to variables can be small. This can lead to overfitting of the model to the available data, analogous to the overfitting of a high-degree polynomial to a small number of data points. The key to overcoming these problems is the observation that covalent bond lengths and angles are relatively inflexible, unlike the torsion angles that define the secondary and tertiary structures of the protein. This observation led to the development of *torsion angle dynamics*, where the equations of motion are written in terms of torsion angles rather than Cartesian coordinates (Diamond, 1971; Mazur and Abagyan, 1989; Rice and Brünger, 1994). The equations of motion become (Vaidehi and Goddard, 2001)

$$\mathbf{M}(\boldsymbol{\tau})\ddot{\boldsymbol{\tau}} + \mathbf{C}(\boldsymbol{\tau}, \dot{\boldsymbol{\tau}}) = \mathbf{F}(\boldsymbol{\tau}) \quad (30)$$

for the  $q \times 1$  vector of torsion angles  $\boldsymbol{\tau}$ , typically about 1/10 of the total number of Cartesian degrees of freedom (Rice and Brünger, 1994). Here  $\mathbf{M}$  is the  $q \times q$  mass matrix,  $\mathbf{C}$  is the  $q \times 1$  Coriolis force vector, and  $\mathbf{F}$  is the  $q \times 1$  vector of interatomic forces. Although  $q$  is small compared to the Cartesian degrees of freedom, (30) is still a large system with a dense mass matrix. Thus, even with the simplification introduced with torsion angle dynamics, atomic refinement with molecular dynamics is a computationally expensive procedure. However, another advantage of torsion angle refinement is that the radius of convergence to the global minimum appears to be larger than that for Cartesian refinement (Brünger and Rice, 1997).

**3.4. Refinement with orientational data.** The orientational data provided by solid-state NMR (ssNMR) measurements can be used as restraints in the potential energy function (25), allowing for refinement with this type of data alone or in combination with other data types. One software package, TORC, uses a Monte Carlo approach with simulated annealing for the refinement (Ketchem *et al.*, 1997). This program has the advantage that chirality moves are built in, which may be desirable when refining with orientational data. More recently, software was developed (Bertram *et al.*, 2000) in the form of a module for the CNS refinement package (Brünger *et al.*, 1998), allowing Cartesian or torsion angle refinement implemented with molecular dynamics. This software is faster than TORC, and has the extra advantage that it can be used to simultaneously refine against ssNMR data in conjunction with other data types.

For the one-dimensional ssNMR method, the dipolar coupling and chemical shift restraints are treated independently. More recently, correlated chemical shift and dipolar data has been used in the ssNMR method (Ramamoorthy and Opella, 1995; Tian *et al.*, 1998; Denny *et al.*, 2001). Recently developed software uses this correlated data in refinement (Bertram *et al.*, 2003). One benefit of the two-dimensional ssNMR method is that many of the dipolar sign degeneracies can be resolved by correlation with the chemical shift. We will first discuss the 1-D ssNMR restraints, then discuss how 2-D ssNMR could be used to construct better restraints.

3.4.1. *One-dimensional solid-state NMR restraints.* Anisotropic chemical shift measurements (denoted here by  $\sigma_o$ ) can be used to restrain the refinement by adding the harmonic penalty function

$$E_{cs} = \sum_{cs} (\sigma_c - \sigma_o)^2, \quad (31)$$

where  $\sigma_c$  is the chemical shift computed directly from the model, and summation is over all  $^{13}\text{C}$  and  $^{15}\text{N}$  anisotropic chemical shift measurements. A harmonic penalty function can also be used to restrain against dipolar coupling measurements. Summation is over all measurements, and the  $i$ th term in the penalty function is

$$E_{dp,i} = \begin{cases} (|v_c| - v_o)^2 & \text{if } v_o \leq \frac{v_{\parallel}}{2} \\ (v_c - v_o)^2 & \text{if } \frac{v_{\parallel}}{2} < v_o \leq v_{\parallel}, \end{cases} \quad (32)$$

where  $v_c, v_o$  are the  $i$ th calculated and observed dipolar couplings, defined by

$$v_c = \frac{v_{\parallel}}{2}(3 \cos^2 \theta_c - 1) \quad (33)$$

$$v_o = \left| \frac{v_{\parallel}}{2}(3 \cos^2 \theta_o - 1) \right|, \quad (34)$$

and where  $\theta$  is the angle between the magnetic field vector and the appropriate covalent bond vector. This angle can be computed directly from the model to

give  $\theta_c$ , but the angle  $\theta_o$  can only be inferred from the dipolar coupling measurement  $\nu_o$ . Thus, there are two degeneracies involved in the determination of  $\theta_o$ : a *sign degeneracy* due to the absolute value in (34), and a *quadratic degeneracy* due to the squaring of the cosine. The absolute value in (32) is a reflection of the sign degeneracy, since observed dipolar couplings are always positive (34), while the spherical harmonic  $\nu_c$  ranges from  $-\frac{\nu_{\parallel}}{2}$  to  $\nu_{\parallel}$ . If  $\nu_o \leq \frac{\nu_{\parallel}}{2}$ , then the model agrees with the data if  $|\nu_c| = \nu_o$ , or if  $\nu_c = \pm\nu_o$ . If  $\frac{\nu_{\parallel}}{2} < \nu_o \leq \nu_{\parallel}$ , then for agreement with the data  $\nu_c = \nu_o$  and there is no degeneracy.

For nuclei with spin greater than 1/2, such as deuterium, the distribution of charged particles generates an electric quadrupole moment. Quadrupolar interactions produce quadrupolar splittings of NMR peaks, similar to those produced by dipolar interactions. The restraining function is often similar to that for dipolar coupling, with the quadrupolar coupling constant ( $\frac{3}{4}$ QCC) replacing  $\nu_{\parallel}$ :

$$E_{qd,i} = \begin{cases} (|\nu_c| - \nu_o)^2 & \text{if } \nu_o \leq \frac{3}{8}\text{QCC} \\ (\nu_c - \nu_o)^2 & \text{if } \frac{3}{8}\text{QCC} < \nu_o \leq \frac{3}{4}\text{QCC}. \end{cases} \quad (35)$$

3.4.2. *Correlated orientational restraints.* As described in Section 2.3.4, if the anisotropic chemical shift and dipolar coupling measurements are correlated, many of the sign degeneracies in dipolar coupling can be resolved. One way to make this correlation is to obtain both measurements from the same signal, the method known as PISEMA. Alternatively, one can obtain separate measurements of  $\sigma$  and  $\nu$  corresponding to the same nitrogen atom, and then form the ordered pair  $(\sigma, \nu)$ . Regardless of how the ordered pair is obtained, one can plot the pair as a point in the  $\sigma\nu$  plane. If the point falls in the primary ellipse  $P$ , but not the reflected ellipse  $P^*$  (see Fig. 7), then  $\nu_o = |\nu_o|$ . If it falls in the reflected ellipse, but not the primary ellipse, then  $\nu_o = -|\nu_o|$ . Only if it falls in the intersection of the two ellipses does the sign degeneracy remain unresolved. Thus, an improved dipolar coupling restraint is

$$E_{dp,i} = \begin{cases} (|\nu_c| - \nu_o)^2 & \text{if unresolved} \\ (\nu_c - \nu_o)^2 & \text{if resolved.} \end{cases} \quad (36)$$

We will see that resolving the dipolar degeneracy greatly simplifies the dipolar energy landscape.

3.4.3. *Dipolar energy landscape.* The energy landscape for a protein is extremely complex, and direct visualization is not possible. However, since the dipolar coupling energy has a simple angle dependency, one can gain insight into how this type of restraint affects refinement by performing a simple graphical examination of the dipolar energy function or landscape. The quadrupolar energy landscape is similar. It is convenient to interpret  $E_{dp}$  as a family of curves

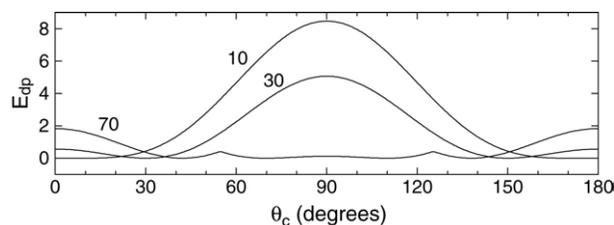


Figure 8. Three members of the family of dipolar coupling energy curves [equation (36)], corresponding to  $\theta_o = 10^\circ$ ,  $30^\circ$ , and  $70^\circ$ . In each case  $\nu_{\parallel} = 2$ .

parametrized by  $\theta_o$ , the *actual* angle between the magnetic field vector and the appropriate covalent bond vector. For a given  $\theta_o$  one can compute  $\nu_o$  using (34), and from this the dipolar coupling energy [using (32)] as a function of  $\nu_c$  or  $\theta_c$ . Energy also depends on  $\nu_{\parallel}$ . Thus,  $E_{dp} = E_{dp}(\theta_c; \theta_o, \nu_{\parallel})$ , where  $\theta_o$  and  $\nu_{\parallel}$  are parameters. In what follows we assume that the dipolar and chemical shift data have not been correlated, and use the restraint (32).

Three members of the  $E_{dp}$  family of curves are shown in Fig. 8, corresponding to  $\theta_o = 10^\circ$ ,  $30^\circ$ , and  $70^\circ$ , each with  $\nu_{\parallel} = 2$ . The curve  $E_{dp}(\theta_c; 10, 2)$  has two zeros, at  $\theta_c = 10^\circ$  and  $\theta_c = 170^\circ$ .

The curve  $E_{dp}(\theta_c; 30, 2)$  also has two zeros ( $30^\circ$  and  $150^\circ$ ), but now there is more concavity near each. This will lead to stronger restraint on the angle during refinement, since the angle is now penalized more severely for small deviations from the energy minima. Finally, the curve  $E_{dp}(\theta_c; 70, 2)$  has four zeros ( $\approx 42^\circ$ ,  $70^\circ$ ,  $110^\circ$ , and  $\approx 138^\circ$ ). Thus, even though the actual angle made with the magnetic field is  $\theta_o = 70^\circ$ , the model can make angles of  $\approx 42^\circ$ ,  $70^\circ$ ,  $110^\circ$ , or  $\approx 138^\circ$  and satisfy the dipolar data equally well, due to the degeneracies inherent in the uncorrelated dipolar coupling measurement.

As shown in Fig. 8, with small  $\theta_o$  angles there are two minima of  $E_{dp}$ . For these angles  $\nu_o > \nu_{\parallel}$ . There is a bifurcation at  $\theta_o \approx 35^\circ$ , where  $\nu_o = \nu_{\parallel}$ . Since this bifurcation reflects the sign degeneracy, we denote this angle as  $\theta_{\text{sign}}$ :

$$\theta_{\text{sign}} = \arccos(\sqrt{2/3}). \quad (37)$$

Fig. 9(a) shows two members of the  $E_{dp}$  family for  $\theta_o$  on either side of  $\theta_{\text{sign}}$ . At  $\theta_{\text{sign}}$  a new minimum emerges at  $\theta_c = 90^\circ$ , and splits into two minima symmetric about  $90^\circ$  for  $\theta_o > \theta_{\text{sign}}$ . Thus, for  $\theta_o = 36^\circ$  there are four minima.

Another bifurcation occurs when  $\nu_o = 0$ , the *magic angle*,  $\theta_{\text{magic}}$ :

$$\theta_{\text{magic}} = \arccos(\sqrt{1/3}) \approx 54.7^\circ. \quad (38)$$

At the magic angle the leftmost and rightmost pairs of minima coalesce, only to split again for larger values of  $\theta_o$  [Fig. 9(b)]. The qualitative change in the energy

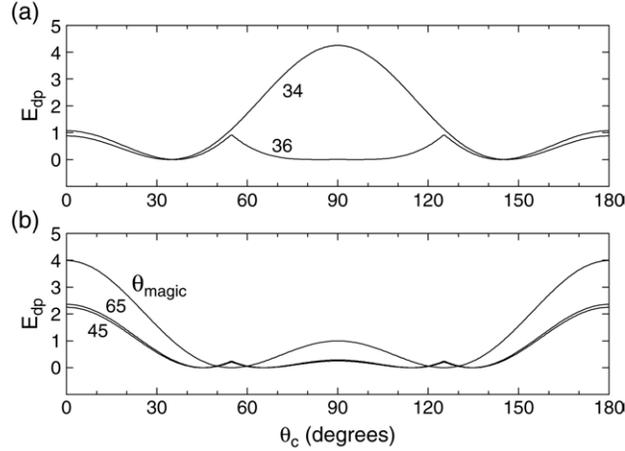


Figure 9. (a) Two members of the  $E_{dp}$  family of curves, on either side of the bifurcation at  $\theta_{\text{sign}} \approx 35^\circ$ . (b) The  $E_{dp}$  curve at the magic angle bifurcation,  $\theta_{\text{magic}} \approx 54.7^\circ$ , and two other members of the  $E_{dp}$  on either side of the bifurcation.

function is rather mild at the magic angle bifurcation, in contrast to the bifurcation at  $\theta_{\text{sign}}$ .

The structure of minima for the  $E_{dp}$  family of curves and the bifurcation points are best illustrated with circle diagrams (Fig. 10). For  $\theta_o = 10^\circ$  there are two minima, represented by filled circles in the top left circle diagram.  $\theta$  is the angle between the vertical line (magnetic field direction) and a point on the unit circle. The minimum connected by a line to the center is at the actual  $\theta$  angle,  $\theta_o$ . For larger  $\theta_o$ , the minima move along the circle towards  $90^\circ$ . At  $\theta_{\text{sign}}$  a new minimum is born, and has bifurcated by  $\theta_o = 36^\circ$ . The new and old minima approach one another for larger  $\theta_o$ , and at the magic angle they coalesce. For larger  $\theta_o$  the minima move past one another and at  $\theta = 90^\circ$  there is yet another coalescence.

Taken together, Figs. 8–10 illustrate that the dipolar coupling energy landscape is quite complex when the sign degeneracy in the dipolar coupling is not resolved. However, when the degeneracy can be resolved, perhaps by correlating data, the landscape becomes much simpler. This is illustrated in Fig. 11, where now the equation for the energy function is

$$E_{dp} = (v_c - v_o)^2. \quad (39)$$

As in Fig. 8 curves are plotted for  $\theta_o = 34^\circ, 36^\circ, 45^\circ, 55^\circ$ , and  $65^\circ$ . While the curves for  $\theta_o = 34^\circ$  and  $55^\circ$  are identical to those in Fig. 8, all other curves differ. When the sign of the dipolar coupling is resolved there are no derivative discontinuities as in Fig. 8. While the  $E_{dp}(\theta_c; 65, 2)$  curve is relatively flat for  $\theta_c \in (30^\circ, 150^\circ)$  when the degeneracy has not been resolved, it is flat over the much smaller interval  $(60^\circ, 120^\circ)$  when the degeneracy has been resolved. In terms

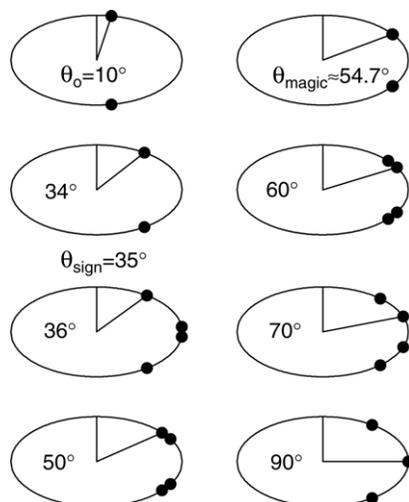


Figure 10. Unit circle diagrams illustrating the location of minima for several members of the  $E_{dp}$  family of curves.

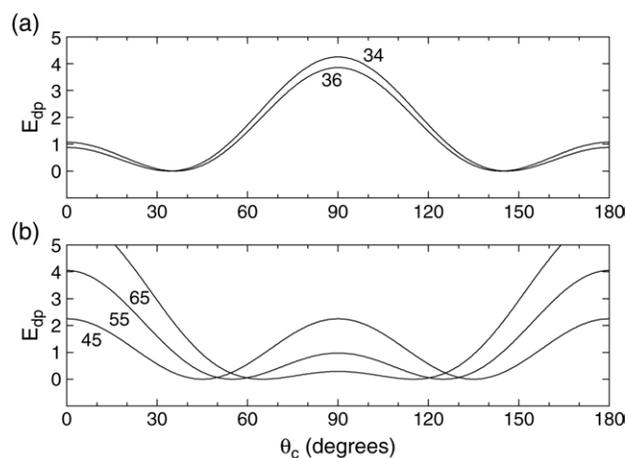


Figure 11. Dipolar coupling energy curves when sign ambiguities have been resolved. The curves deform continuously as  $\theta_o$  is increased through (a)  $\theta_{\text{sign}} \approx 35^\circ$  and through (b)  $\theta_{\text{magic}} \approx 54.7^\circ$ . Compared to Fig. 9, the curves are more suitable for restraining the  $\theta$  angle during refinement.

of Fig. 10, the extra pair of minima born at  $\theta_{\text{sign}}$  will not occur if the sign degeneracy is resolved.

In summary, there are two reasons that it is important to resolve sign degeneracies in the dipolar coupling. First, there are fewer local minima in the energy function, directly reflecting resolution of degenerate solutions. Second, when degeneracies are resolved the energy landscape is less complex and is better suited to restraining the bond angles during refinement. Correlating the chemical shift and dipolar measurements is an effective way to resolve many of the degeneracies.

#### 4. CONCLUSIONS

The use of orientational information from solid-state NMR has great potential for the structure determination of membrane proteins. This technique is still quite new, but already great experimental and theoretical progress has been made (Cross and Quine, 2000; Quine and Cross, 2000). We have discussed many of the mathematical issues involved in the interpretation of NMR data, and how this orientational data can be used to describe the atomic structure of the protein backbone. An algorithm was discussed for creating an initial atomic model, assuming standard stereochemistry of covalently bonded atoms, and the process of atomic refinement using orientational data was described. Further development of these two procedures, initial model building and subsequent atomic refinement, will help assure the best use of solid-state NMR data as it becomes available for the many membrane proteins whose structure has not yet been determined. The development of two-dimensional NMR is an extremely important recent step for the solid-state NMR method, and correlating orientational measurements is, as we show here, important from a computational perspective.

#### ACKNOWLEDGEMENTS

JRQ and TAC were partially supported by NSF grant DMS 9986036. MSC was partially supported by NSF grant DBI 9808098. All authors were partially supported by NIH grant 1P01 GM64676.

#### REFERENCES

- Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223–230.
- Bertram, R., T. Asbury, F. Fabiola, J. R. Quine, T. A. Cross and M. S. Chapman (2003). Atomic refinement with correlated solid-state NMR restraints. *J. Magn. Reson.* **163**, 300–309.
- Bertram, R., J. R. Quine, M. S. Chapman and T. A. Cross (2000). Atomic refinement using orientational restraints from solid-state NMR. *J. Magn. Reson.* **147**, 9–16.
- Brender, J. R., D. M. Taylor and A. Ramamoorthy (2001). Orientation of amide-nitrogen-15 chemical shift tensors in peptides: a quantum mechanical study. *J. Am. Chem. Soc.* **123**, 914–922.
- Brenneman, M. T. and T. A. Cross (1990). A method for the analytic determination of polypeptide structure using solid-state nuclear magnetic resonance: the metric method. *J. Chem. Phys.* **92**, 1483–1494.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
- Brünger, A. T. (1992a). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475.

- Brünger, A. T. (1992b). *X-Plor Version 3.1, A System for Crystallography and NMR*, New Haven, CT: Yale University Press.
- Brünger, A. T. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921.
- Brünger, A. T. and L. M. Rice (1997). Crystallographic refinement by simulated annealing: methods and applications. *Methods Enzymol.* **277**, 243–269.
- Clore, G. M., A. M. Gronenborn and N. Tjandra (1998). Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J. Magn. Reson.* **131**, 159–162.
- Cross, T. A. and J. R. Quine (2000). Protein structure in anisotropic environments: development of orientational constraints. *Concepts Magn. Reson.* **12**, 55–70.
- Denny, J. K., J. Wang, T. A. Cross and J. R. Quine (2001). PISEMA powder patterns and PISA wheels. *J. Magn. Reson.* **152**, 217–226.
- Diamond, R. (1971). A real-space refinement procedure for proteins. *Acta Crystallogr. A* **27**, 436–452.
- Evans, J. N. S. (1995). *Biomolecular NMR Spectroscopy*, New York: Oxford University Press.
- Havel, T. F. and A. W. M. Dress (1993). Distance geometry and geometric algebra. *Found. Phys.* **23**, 1357–1374.
- Hendrickson, W. A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115**, 252–270.
- Ketchum, R. R., B. Roux and T. A. Cross (1997). High-resolution polypeptide structure in a lamellar phase lipid environment from solid state NMR derived orientational constraints. *Structure* **5**, 1655–1669.
- Kirkpatrick, S., C. D. Gelatt Jr and M. P. Vecchi (1983). Optimization by simulated annealing. *Science* **220**, 671–680.
- Lee, D.-K., Y. Wei and A. Ramamoorthy (2001). A two-dimensional magic-angle decoupling and magic-angle turning solid-state NMR method: an application to study chemical shift tensors from peptides that are nonselectively labeled with  $^{15}\text{N}$  isotope. *J. Phys. Chem. B* **105**, 4752–4762.
- MacKerell, A. D. Jr *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem.* **102**, 3586–3616.
- Mai, W., W. Hu, C. Wang and T. A. Cross (1993). Orientational constraints as three-dimensional structural constraints from chemical shift anisotropy: the polypeptide backbone of gramicidin A in a lipid bilayer. *Protein Sci.* **2**, 532–542.
- Marassi, F. M. and S. J. Opella (2000). A solid-state NMR index of helical membrane protein structure and topology. *J. Magn. Reson.* **144**, 150–155.
- Mascioni, A. and G. Veglia (2003). Theoretical analysis of residual dipolar coupling patterns in regular secondary structures of proteins. *J. Am. Chem. Soc.* **125**, 12520–12526.
- Mazur, A. K. and R. A. Abagyan (1989). New methodology for computer-aided modelling of biomolecular structure and dynamics. (I) Non-cyclic structures. *J. Biomol. Struct. Dynam.* **4**, 815–832.
- Mesleh, M. F. and S. J. Opella (2003). Dipolar waves as NMR maps of helices in proteins. *J. Magn. Reson.* **163**, 288–299.
- Oas, T. G., C. J. Hartzell, F. W. Dahlquist and G. P. Drobny (1987). The amide  $^{15}\text{N}$  chemical shift tensors of four peptides determined from  $^{13}\text{C}$  dipole-coupled chemical shift powder patterns. *J. Am. Chem. Soc.* **109**, 5962–5966.

- Prestegard, J. H., J. R. Tolman, H. M. Al-Hashimi and M. Andrec (1999). Protein structure and dynamics from field-induced residual dipolar couplings, in *Biological Magnetic Resonance, Volume 17: Structure and Dynamics in Protein NMR*, Krishna and Berliner (Eds), New York: Plenum Publishers, pp. 311–355.
- Quine, J. R. (1999). Helix parameters and protein structures using quaternions. *Theochem* **460**, 53–66.
- Quine, J. R. and T. A. Cross (2000). Protein structure in anisotropic environments: unique structural fold from orientational constraints. *Concepts Magn. Reson.* **12**, 71–82.
- Ramamoorthy, A., F. Marassi, M. Zasloff and S. J. Opella (1995). Three-dimensional solid-state NMR spectroscopy of a peptide oriented in membrane bilayers. *J. Biol. NMR* **6**, 329–334.
- Ramamoorthy, A. and S. J. Opella (1995). Two-dimensional chemical shift/heteronuclear dipolar coupling spectra obtained with polarization inversion spin exchange at the magic angle and magic-angle sample spinning (PISEMAMAS). *Solid State NMR* **4**, 387–392.
- Ramamoorthy, A., C. H. Wu and S. J. Opella (1999). Experimental aspects of multidimensional solid-state NMR correlation spectroscopy. *J. Magn. Reson.* **140**, 131–140.
- Rice, L. M. and A. T. Brünger (1994). Torsion angle dynamics: Reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins: Struct. Funct. Genetics* **19**, 277–290.
- Teng, Q., M. Iqbal and T. A. Cross (1992). Determination of the  $^{13}\text{C}$  chemical shift and  $^{14}\text{N}$  electric field gradient tensor orientations with respect to the molecular frame in a polypeptide. *J. Am. Chem. Soc.* **114**, 5312–5321.
- Tian, F., Z. Song and T. A. Cross (1998). Orientational constraints derived from hydrated powder samples by two-dimensional PISEMA. *J. Magn. Reson.* **135**, 227–231.
- Tjandra, N., J. Marquardt and G. M. Clore (2000). Direct refinement against proton–proton dipolar couplings in NMR structure determination of macro-molecules. *J. Magn. Reson.* **142**, 393–396.
- Vaidehi, N. and W. A. Goddard III (2001). Atomic-level simulation and modeling of biomacromolecules, in *Computational Modeling of Genetic and Biochemical Networks*, J. M. Bower and H. Bolouri (Eds), Cambridge, MA: MIT Press, pp. 161–188.
- Wallin, P. J. and G. Von Heijne (1998). Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7**, 1029–1038.

Received 26 August 2003 and accepted 26 March 2004