Biophysical Journal Volume 72 May 1997 2342–2348

Protein Structural Analysis from Solid-State NMR-Derived Orientational Constraints

J. R. Quine,* M. T. Brenneman,# and T. A. Cross§

*Department of Mathematics; #Institute of Molecular Biophysics; and §Center for Interdisciplinary Magnetic Resonance in the National High Magnetic Field Laboratory, Institute of Molecular Biophysics, and Department of Chemistry, Florida State University, Tallahassee, Florida 32306 USA

ABSTRACT High-resolution orientational constraints from solid-state NMR spectroscopy of uniformly aligned biological macromolecules provide a great structural analysis problem. Several approaches to this problem have been made in the past. Here a vector algebra method is developed that provides analytical solutions for the torsion angles and a concise and simple view of the structural possibilities. Numerical instabilities in this approach are easily predicted. Insight into how the structural ambiguities arise in the first place and how they can be reduced in number is demonstrated with this new approach.

INTRODUCTION

The advent of multidimensional techniques (Kumar and Opella, 1991; Wu et al., 1994) and the recent determination of a peptide's complete three-dimensional structure (Ketchem et al., 1993, 1996a) using solid-state NMR (SS NMR) clearly demonstrate its potential to join the ranks of x-ray diffraction and solution NMR as another independent method for determining protein structure. The effort to further develop SS NMR for use on a routine basis is important because it is naturally suited for determining high-resolution structures of membrane proteins (Cross, 1994; Cross and Opella, 1994) and fibrous proteins (Nicholson et al., 1990; Simmons et al., 1996) in their native environment. In this respect, one particularly underdeveloped area is structural analysis: although several methods (Opella et al., 1987; Brenneman and Cross, 1990; Teng et al., 1991) have been used to determine the three-dimensional structure of peptides, all suffer from weaknesses that would affect their application to proteins. A method free of these limitations is presented here to determine the initial structure of a polypeptide's backbone from experimental SS NMR-derived orientational constraints.

Determining the structure of a molecule from experimental data generally involves the use of mathematical methods, such as the Fourier transform in x-ray crystallography or distance geometry in solution NMR. The purpose here is to develop a mathematical method to determine the complete structure of a protein's backbone from SS NMR data and to define the structural ambiguities. The input data are measurements from dipolar, quadrupolar, and chemical shift interactions (Opella et al., 1987) obtained from samples that are aligned with the direction of $\hat{\mathbf{B}}_0$, the fixed magnetic field of the spectrometer. The magnitude of these interactions allows one to solve for the angle each covalent bond makes with $\hat{\mathbf{B}}_0$, the bond's azimuthal angle. It has been shown that by measuring roughly three interactions per peptide plane, a unique solution for each bond's azimuthal angle can generally be obtained (Brenneman, Quine and Cross, unpublished; Ketchem et al., 1996b). As Fig. 1 shows, the fixed azimuthal angles, along with the covalent bond angles, act as constraints on the protein's structure. Just as in solution NMR, where the structure of a molecule is solved using distance constraints, in solid-state NMR the structure is solved using these angular or orientational constraints. Here a mathematical solution is developed for the structural analysis of these data.

Using vector analysis, analytical equations for the $\{\phi, \psi\}$ torsion angles of the protein backbone are derived. The goal is not just to develop an algorithm that computes a numerical result for the structure from the experimental data (three solutions of this type have already been published), but to use mathematics to also solve a second problem: how the structural information from the experimental data defines the structure of a molecule. A thorough understanding of this problem answers questions of significant practical importance to the experimentalist, such as the minimum data set needed to determine the molecule's structure, whether a particular data set will yield a numerically stable result, how effective a particular measurement will be in reducing the number of possible structures, and the total number of conformations consistent with the experimental data. It is the ability to solve this problem that marks the passage from simply computing numerical results to performing structural analysis. It is also the point at which this method diverges from those in the past.

In Materials and Methods, some basic terminology and notation are established, an important concept called "chirality" (following the terminology of Crippen and Havel, 1988) is introduced, and the basic equation for any general dihedral angle in terms of bond azimuthal angles and chiralities is derived. In the Results section the method is refined for solving the structure of a protein backbone. Finally, in the Discussion section, the method is applied to

Received for publication 17 September 1996 and in final form 6 February 1997.

Address reprint requests to Dr. Timothy A. Cross, National High Magnetic Field Laboratory, 1800 E. Paul Diroc Dr., Tallahassee, FL 32306-4005. Tel.: 904-644-0917; Fax: 904-644-1366; E-mail: cross@magnet.fsu.edu.

© 1997 by the Biophysical Society

0006-3495/97/05/2342/07 $2.00

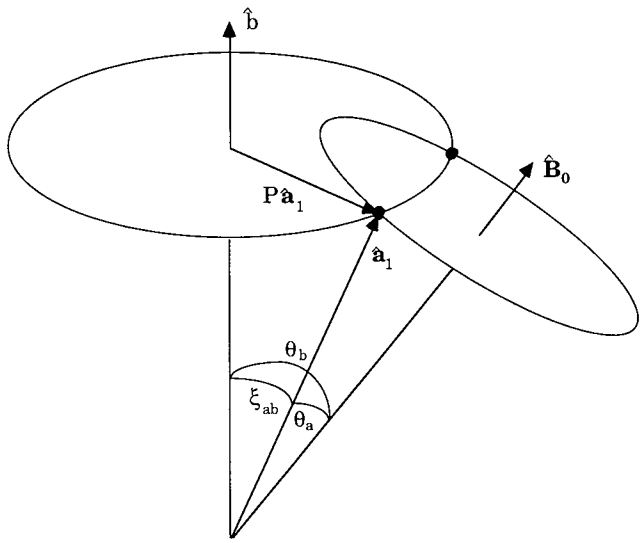

FIGURE 1 Bond orientation and bond angle constraints. $\hat{\mathbf{B}}_0$ and $\hat{\mathbf{b}}$ are given unit vectors, and $\hat{\mathbf{a}} \cdot \hat{\mathbf{b}} = \cos \xi_{ab}$, $\hat{\mathbf{a}} \cdot \hat{\mathbf{B}}_0 = \cos \theta_a$. How does this determine the unit vector $\hat{\mathbf{a}}$? There are two possibilities indicated by the intersection of the two circles. One possibility ($\hat{\mathbf{a}}_1$) is shown above, along with its projection ($P\hat{\mathbf{a}}_1$) onto the plane perpendicular to $\hat{\mathbf{b}}$.

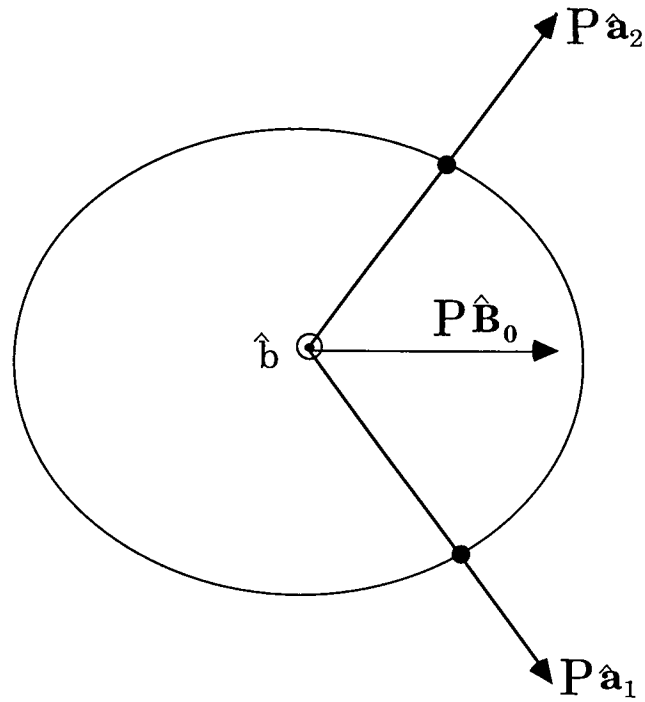

FIGURE 2 Projection and chirality. When Fig. 1 is projected onto the plane perpendicular to $\hat{\mathbf{b}}$, two possibilities ($P\hat{\mathbf{a}}_1$ and $P\hat{\mathbf{a}}_2$) result. The projections are symmetrical with respect to $P\hat{\mathbf{B}}_0$, and the two possibilities are distinguished by the sign of a triple product (a chirality): $\hat{\mathbf{b}} \cdot (\hat{\mathbf{B}}_0 \times \hat{\mathbf{a}}_2) > 0$, $\hat{\mathbf{b}} \cdot (\hat{\mathbf{B}}_0 \times \hat{\mathbf{a}}_1) < 0$.

a model peptide, gramicidin A, the structure of which has already been determined by SS NMR.

MATERIALS AND METHODS

The basic tool is a formula for the torsion angle in terms of the bond orientation cosines, the bond angle cosines, and the chiralities. Three unit vectors, $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, $\hat{\mathbf{c}}$, and a unit vector, $\hat{\mathbf{B}} = \hat{\mathbf{B}}_0$, in the direction of the field of the spectrometer are used. The dot products of the vectors with $\hat{\mathbf{B}}$ are given by $\hat{\mathbf{B}} \cdot \hat{\mathbf{a}} = \cos \theta_a$, $\hat{\mathbf{B}} \cdot \hat{\mathbf{b}} = \cos \theta_b$, $\hat{\mathbf{B}} \cdot \hat{\mathbf{c}} = \cos \theta_c$, and are referred to as the bond orientation cosines. The dot products $\hat{\mathbf{a}} \cdot \hat{\mathbf{b}} = \cos \xi_{ab}$ and $\hat{\mathbf{b}} \cdot \hat{\mathbf{c}} = \cos \xi_{bc}$ are referred to as the bond angle cosines. The torsion angle, Torsion($\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, $\hat{\mathbf{c}}$), is a function, τ, of seven variables:

$$\text{Torsion}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) = \tau(\cos \theta_a, \cos \theta_b, \cos \theta_c, \cos \xi_{ab}, \cos \xi_{bc}, \epsilon_{ab}, \epsilon_{bc}),$$

where

$$\epsilon_{ab} = \text{sign}\, \hat{\mathbf{B}} \cdot (\hat{\mathbf{a}} \times \hat{\mathbf{b}}) \text{ and } \epsilon_{bc} = \text{sign}\, \hat{\mathbf{B}} \cdot (\hat{\mathbf{b}} \times \hat{\mathbf{c}})$$

are chiralities that cannot be determined from the bond orientation and bond angle cosines.

In the subsections below, the framework for understanding and using this formula is described. The role of chiralities in determining the ambiguities in this torsion angle is illustrated in Figs. 1, 2, and 3.

Coordinate systems

A right-handed coordinate system is defined as an ordered triple $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$ of unit vectors such that $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are orthogonal and $\hat{\mathbf{z}} = \hat{\mathbf{x}} \times \hat{\mathbf{y}}$. The coordinates of any vector $\mathbf{v}$ in this system are $(\mathbf{v} \cdot \hat{\mathbf{x}}, \mathbf{v} \cdot \hat{\mathbf{y}}, \mathbf{v} \cdot \hat{\mathbf{z}})$.

Computations can be made easier by choosing the appropriate coordinate system for the problem. A coordinate system is usually chosen in which the first vector is a unit bond vector $\hat{\mathbf{u}}$. There are many possible ways of choosing the other two vectors. Here a "$\hat{\mathbf{u}}$ system based on $\hat{\mathbf{B}}$" is a convenient coordinate system for computations using bond orientation cosines. It consists of the vectors $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$, where

$$\hat{\mathbf{x}} = \hat{\mathbf{u}} \qquad \hat{\mathbf{y}} = \frac{\hat{\mathbf{B}} \times \hat{\mathbf{u}}}{(1 - (\hat{\mathbf{B}} \cdot \hat{\mathbf{u}})^2)^{1/2}} \tag{1}$$

$$\hat{\mathbf{z}} = \hat{\mathbf{x}} \times \hat{\mathbf{y}} = \frac{\hat{\mathbf{B}} - (\hat{\mathbf{B}} \cdot \hat{\mathbf{u}})\hat{\mathbf{u}}}{(1 - (\hat{\mathbf{B}} \cdot \hat{\mathbf{u}})^2)^{1/2}}.$$

(By the formula for the length of a cross product, the vector $\hat{\mathbf{y}}$ is a unit vector.) This system can be used only if $\hat{\mathbf{u}}$ and $\hat{\mathbf{B}}$ are linearly independent, that is, if $\hat{\mathbf{u}}$ is not equal to $\hat{\mathbf{B}}$ or $-\hat{\mathbf{B}}$. If $\hat{\mathbf{u}}$ is $\hat{\mathbf{B}}$ or $-\hat{\mathbf{B}}$, then $\hat{\mathbf{y}} = \hat{\mathbf{z}} = 0$.

In the $\hat{\mathbf{u}}$ system based on $\hat{\mathbf{B}}$, the coordinates of $\hat{\mathbf{B}}$ are uniquely determined from the bond orientation cosines. The coordinates are $(\hat{\mathbf{B}} \cdot \hat{\mathbf{u}}, 0, (1 - (\hat{\mathbf{B}} \cdot \hat{\mathbf{u}})^2)^{1/2})$.

Gramians, triple products, and chirality

We review here some basic vector geometry needed in the following sections. A good reference in general is the *CRC Handbook of Mathematical Sciences* (Beyer, 1987). Also see Crippen and Havel (1988) for a discussion of gramians and chirality.

If $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, and $\hat{\mathbf{c}}$ are unit vectors, then $|\hat{\mathbf{a}} \cdot (\hat{\mathbf{b}} \times \hat{\mathbf{c}})|$ is given by the square root of the determinant

$$g(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) = \begin{vmatrix} 1 & \hat{\mathbf{a}} \cdot \hat{\mathbf{b}} & \hat{\mathbf{a}} \cdot \hat{\mathbf{c}} \\ \hat{\mathbf{b}} \cdot \hat{\mathbf{a}} & 1 & \hat{\mathbf{b}} \cdot \hat{\mathbf{c}} \\ \hat{\mathbf{c}} \cdot \hat{\mathbf{a}} & \hat{\mathbf{c}} \cdot \hat{\mathbf{b}} & 1 \end{vmatrix}.$$

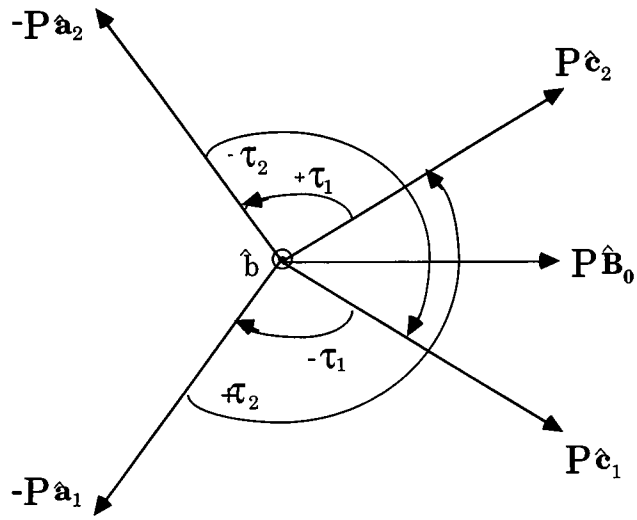

FIGURE 3 Chirality and torsion. The torsion angle, Torsion($\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, $\hat{\mathbf{c}}$), is the angle from $-P\hat{\mathbf{a}}$ to $P\hat{\mathbf{c}}$, where P is the projection onto the plane perpendicular to $\hat{\mathbf{b}}$, as in Fig. 2. With two possible values for $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}$, distinguished by chirality, there are four possible values for the torsion angle. There are two pairs, one the negative of the other.

This determinant is referred to as the gramian or the metric matrix. Because the vectors are unit vectors, each entry can be written as the cosine of the angle between vectors. The determinant is evaluated,

$$\begin{aligned} g(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) &= 1 - (\hat{\mathbf{a}}\cdot\hat{\mathbf{b}})^2 - (\hat{\mathbf{a}}\cdot\hat{\mathbf{c}})^2 - (\hat{\mathbf{b}}\cdot\hat{\mathbf{c}})^2 \\ &\quad + 2(\hat{\mathbf{a}}\cdot\hat{\mathbf{b}})(\hat{\mathbf{a}}\cdot\hat{\mathbf{c}})(\hat{\mathbf{b}}\cdot\hat{\mathbf{c}}) \\ &= \gamma(\hat{\mathbf{a}}\cdot\hat{\mathbf{b}}, \hat{\mathbf{a}}\cdot\hat{\mathbf{c}}, \hat{\mathbf{b}}\cdot\hat{\mathbf{c}}) \end{aligned}$$

where γ denotes the function in the variables x, y, and z given by

$$\gamma(x, y, z) = 1 - x^2 - y^2 - z^2 + 2xyz. \tag{2}$$

Note that this function is unchanged by permutation of the variables.

The matrix in the gramian above can be written as the product of a matrix and its transpose, where the rows of the matrix are the coordinates of $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, and $\hat{\mathbf{c}}$ in a right-handed coordinate system. The gramian is the square of the determinant of this matrix. Thus for the triple product, $|\hat{\mathbf{a}}\cdot(\hat{\mathbf{b}}\times\hat{\mathbf{c}})|^2 = \gamma(\hat{\mathbf{a}}\cdot\hat{\mathbf{b}}, \hat{\mathbf{a}}\cdot\hat{\mathbf{c}}, \hat{\mathbf{b}}\cdot\hat{\mathbf{c}})$ or $\hat{\mathbf{a}}\cdot(\hat{\mathbf{b}}\times\hat{\mathbf{c}}) = \epsilon\sqrt{\gamma(\hat{\mathbf{a}}\cdot\hat{\mathbf{b}}, \hat{\mathbf{a}}\cdot\hat{\mathbf{c}}, \hat{\mathbf{b}}\cdot\hat{\mathbf{c}})}$, where $\epsilon = \mathrm{sign}(\hat{\mathbf{a}}\cdot(\hat{\mathbf{b}}\times\hat{\mathbf{c}}))$ is the chirality of this ordered triple of vectors. The important thing to note here is that the absolute value of the gramian is a function of the three dot products (or cosines).

The gramian has a useful interpretation in terms of spherical trigonometry. If τ is the angle between the arcs $\hat{\mathbf{a}}$ to $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ to $\hat{\mathbf{c}}$ on the unit sphere, then

$$g(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) = (1 - (\hat{\mathbf{a}}\cdot\hat{\mathbf{b}})^2)(1 - (\hat{\mathbf{a}}\cdot\hat{\mathbf{c}})^2)(1 - \cos^2\tau).$$

The nonnegativity of the gramian is an important observation in determining bond orientation cosines from second-order tensor data (Opella et al., 1987, figure 9, *A*, *B*; Brenneman and Cross, 1990). In general, $\gamma(x, y, z)$ can be negative for $|x| \le 1$, $|y| \le 1$, and $|z| \le 1$, for example, $\gamma(1, 1, 0) = -1$, so the nonnegativity of γ distinguishes triples of dot products that can occur for three vectors in Euclidean space.

The torsion angle formula

In this section a derivation for the torsion angle formula as a function of bond orientation cosines and bond angle cosines is given. If $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, and $\hat{\mathbf{c}}$ are three linearly independent vectors thought of as a sequence of bond vectors, then the torsion angle can be defined as follows. Let $P\hat{\mathbf{a}} = \hat{\mathbf{a}} - (\hat{\mathbf{a}}\cdot\hat{\mathbf{b}})\hat{\mathbf{b}}$ and $P\hat{\mathbf{c}} = \hat{\mathbf{c}} - (\hat{\mathbf{b}}\cdot\hat{\mathbf{c}})\hat{\mathbf{b}}$ be the projection of $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}$, respectively, onto the plane perpendicular to $\hat{\mathbf{b}}$. The torsion angle at $\hat{\mathbf{b}}$ is the angle from $-P\hat{\mathbf{a}}$ to $P\hat{\mathbf{c}}$ measured counterclockwise around $\hat{\mathbf{b}}$. This angle is denoted by

$$\phi = \mathrm{Torsion}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}).$$

The angle is chosen in the interval from $-180°$ to $180°$. This angle has an equivalent description as a dihedral angle between two planes. In our application, the vectors will be along consecutive bonds pointing in the same direction along the backbone of a protein.

Suppose that bond angle cosines $\cos\theta_a$, $\cos\theta_b$, and $\cos\theta_c$ are known and the expression for torsion angles is given in terms of bond orientation cosines and two chiralities. The bond orientation sines $\sin\theta_a$, $\sin\theta_b$, and $\sin\theta_c$ can be computed from the bond orientation cosines using $\sin\theta = (1 - \cos^2\theta)^{1/2}$. Similarly, the bond angle sines, $\sin\xi_{ab}$ and $\sin\xi_{bc}$, can be computed. Because bond angles and bond orientation angles are defined to be in the interval from 0° to 180°, the sine of these angles is always nonnegative. Now the triple products are computed from the gramians:

$$\hat{\mathbf{B}}\cdot(\hat{\mathbf{a}}\times\hat{\mathbf{b}}) = \epsilon_{ab}\sqrt{g(\hat{\mathbf{B}}, \hat{\mathbf{a}}, \hat{\mathbf{b}})},$$

$$\hat{\mathbf{B}}\cdot(\hat{\mathbf{b}}\times\hat{\mathbf{c}}) = \epsilon_{bc}\sqrt{g(\hat{\mathbf{B}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})}.$$

Define α_a, β_a, α_c, and β_c by

$$\begin{aligned} (\alpha_a, \beta_a) &= -(\hat{\mathbf{a}}\cdot(\hat{\mathbf{B}}\times\hat{\mathbf{b}}), \cos\theta_a - \cos\theta_b\cos\xi_{ab}) \\ &= (\hat{\mathbf{B}}\cdot(\hat{\mathbf{a}}\times\hat{\mathbf{b}}), -\cos\theta_a + \cos\theta_b\cos\xi_{ab}) \\ (\alpha_c, \beta_c) &= (\hat{\mathbf{c}}\cdot(\hat{\mathbf{B}}\times\hat{\mathbf{b}}), \cos\theta_c - \cos\theta_b\cos\xi_{bc} \\ &= (\hat{\mathbf{B}}\cdot(\hat{\mathbf{b}}\times\hat{\mathbf{c}}), \cos\theta_c - \cos\theta_b\cos\xi_{bc}) \end{aligned}$$

In the coordinate system at $\hat{\mathbf{b}}$ of vectors

$$\hat{\mathbf{x}} = \hat{\mathbf{b}}, \quad \hat{\mathbf{y}} = \frac{1}{\sin\theta_b}\hat{\mathbf{B}}\times\hat{\mathbf{b}}, \quad \hat{\mathbf{z}} = \frac{1}{\sin\theta_b}(\hat{\mathbf{B}} - \cos\theta_b\hat{\mathbf{b}}),$$

the projections of $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}$ on the plane perpendicular to $\hat{\mathbf{b}}$ are easily computed as

$$P\hat{\mathbf{a}} = \frac{-1}{\sin\theta_b}(\alpha_a\hat{\mathbf{y}} + \beta_a\hat{\mathbf{z}})$$

$$P\hat{\mathbf{c}} = \frac{1}{\sin\theta_b}(\alpha_c\hat{\mathbf{y}} + \beta_c\hat{\mathbf{z}}).$$

Now the torsion angle ϕ is the angle from $-P\hat{\mathbf{a}}$ to $P\hat{\mathbf{b}}$ measured counterclockwise around $\hat{\mathbf{b}}$. Hence,

$$\cos\phi = \frac{-P\hat{\mathbf{a}}\cdot P\hat{\mathbf{c}}}{|P\hat{\mathbf{a}}||P\hat{\mathbf{c}}|}, \quad \mathrm{sign}\,\phi = \mathrm{sign}\,\hat{\mathbf{b}}\cdot(P\hat{\mathbf{c}}\times P\hat{\mathbf{a}}).$$

Noting that $|P\hat{\mathbf{a}}| = \sin\xi_{ab}$ and $|P\hat{\mathbf{c}}| = \sin\xi_{bc}$, it follows that the torsion angle is given by

$$\begin{aligned} \phi &= \mathrm{Torsion}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) \\ &= \mathrm{sign}(\alpha_c\beta_a - \alpha_a\beta_c)\arccos\left(\frac{\alpha_a\alpha_c + \beta_a\beta_c}{\sin^2\theta_b\sin\xi_{ab}\sin\xi_{bc}}\right). \end{aligned} \tag{3}$$

(Sign 0 should be defined to be 1.) Note that all quantities in the formula except for ϵ_{ab} and ϵ_{bc} can be computed in terms of the bond angle cosines and the bond orientation cosines. The formula would be too long if written

only in terms of these, so intermediate variables have been used. Thus the torsion angle formula above is a function

$$\phi = \tau(\cos\theta_a, \cos\theta_b, \cos\theta_c, \cos\xi_{ab}, \cos\xi_{bc}, \epsilon_{ab}, \epsilon_{bc}).$$

Note certain symmetries in the formula for ϕ. If the sign of both ϵ_1 and ϵ_2 is changed, then the sign of ϕ is changed. If the sequence of vectors is reversed, then the angle ϕ is unchanged. If the sign is changed on all bond orientation cosines, then the sign of ϕ is changed. Also note that by definition of the torsion angle,

$$\text{Torsion}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \text{Torsion}(\hat{\mathbf{c}}, \hat{\mathbf{b}}, \hat{\mathbf{d}}) = \text{Torsion}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{d}}). \tag{4}$$

Numerical stability

It was noted that the coordinate system at $\hat{\mathbf{b}}$ is undefined if $\hat{\mathbf{b}} = \hat{\mathbf{B}}$ or $\hat{\mathbf{b}} = -\hat{\mathbf{B}}$. This is reflected in the Eq. 3 by the vanishing of $\sin\theta_b$ in the denominator. The formula is numerically unstable, then, if $\hat{\mathbf{B}}$ and $\hat{\mathbf{b}}$ are approximately parallel. In this situation, the two circles in Fig. 1 would be close to being the same circle and the intersection points would not be well defined.

Numerical instability also occurs in the computation of the torsion angle in the following situations:

1. when $g(\hat{\mathbf{B}}, \hat{\mathbf{a}}, \hat{\mathbf{b}}) = 0$, i.e., when $\hat{\mathbf{B}}$, $\hat{\mathbf{a}}$, and $\hat{\mathbf{b}}$ are in the same plane,
2. when $g(\hat{\mathbf{B}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) = 0$, i.e., when $\hat{\mathbf{B}}$, $\hat{\mathbf{c}}$, and $\hat{\mathbf{b}}$ are in the same plane.

(Note that if $\hat{\mathbf{B}}$ and $\hat{\mathbf{b}}$ are parallel, then both of these situations occur.) In these situations there is a well-defined solution for the torsion angle, but because $d\sqrt{g} = dg/2\sqrt{g}$ is infinite, a small change in g will result in a large one for $\sqrt{g}$ and the computation is unstable. Here the circles in Fig. 1 would be tangent to one another with only one intersection point. A small change in the circles will result in a large change in the location of the intersection points. This type of situation arises for real proteins. In the case of oriented silk fibers (Nicholson et al., 1990), the $\hat{\mathbf{B}}$ field is found to lie roughly in the plane of $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}'$, and as a result any equation for the torsion angle will give unstable results. A possible solution to this problem is to use bond orientation information for an out-of-plane bond such as the C_α-H bond, $\hat{\mathbf{d}}$, and then compute a torsion angle where this bond is one of the vectors. Another possibility for molecules without axial rotation is to tip the samples in the spectrometer so that $\hat{\mathbf{B}}$ will no longer be in the plane.

RESULTS

The torsion angle formula will now be applied to the geometry of a dipeptide, two adjacent peptide planes.

Planes and unit bond vectors

Unit bond vectors are denoted as in Table 1.

The vectors $\hat{\mathbf{c}}, \hat{\mathbf{n}}, \hat{\mathbf{a}}, \hat{\mathbf{h}}$ are assumed to be in the same plane, the peptide plane. Primed vectors indicate vectors in the subsequent peptide plane (see Fig. 4). The geometry of the peptide plane bonds and the bonds to the C_α carbons are determined approximately by the following dot products (Engh and Huber, 1991):

$$\hat{\mathbf{c}}\cdot\hat{\mathbf{a}} = \cos 6^\circ \quad \hat{\mathbf{c}}\cdot\hat{\mathbf{n}} = \cos 75^\circ$$

$$\hat{\mathbf{c}}\cdot\hat{\mathbf{h}} = \cos 123^\circ \quad \hat{\mathbf{a}}\cdot\hat{\mathbf{c}}' = 1/3 = \cos 70^\circ$$

$$\hat{\mathbf{n}}\cdot\hat{\mathbf{a}} = \cos 59^\circ \quad \hat{\mathbf{h}}\cdot\hat{\mathbf{a}} = \cos 117^\circ$$

$$\hat{\mathbf{n}}\cdot\hat{\mathbf{h}} = \cos 58^\circ \quad \hat{\mathbf{d}}\cdot\hat{\mathbf{a}} = 1/3$$

$$\hat{\mathbf{d}}\cdot\hat{\mathbf{c}}' = -1/3 = \cos 110^\circ,$$

using the notation for bond angle, for example, $\xi_{ch} = 123^\circ$. Note that $\hat{\mathbf{d}}\cdot(\hat{\mathbf{c}}'\times\hat{\mathbf{a}}) < 0$ at a C_α in an L amino acid and $\hat{\mathbf{d}}\cdot(\hat{\mathbf{c}}'\times\hat{\mathbf{a}}) > 0$ at a C_α in a D amino acid.

In this paper it is assumed that bond orientation data are known, i.e., that $\hat{\mathbf{B}}\cdot\hat{\mathbf{u}}$ is known for all unit bond vectors $\hat{\mathbf{u}}$. For bonds in a peptide plane, if the dot product with $\hat{\mathbf{B}}$ for two linearly independent vectors in the plane is known, then the dot product with $\hat{\mathbf{B}}$ can be computed for all vectors in the plane.

TABLE 1 Unit bond vectors in a peptide plane

Unit vector	Chemical bond
$\hat{\mathbf{c}}$	C_α—C′
$\hat{\mathbf{n}}$	C′—N
$\hat{\mathbf{a}}$	N—C_α
$\hat{\mathbf{h}}$	N—H
$\hat{\mathbf{d}}$	C_α—D

Computation of C_α torsion angles

The ϕ and ψ angles at an α carbon bond are defined by

$$\phi = \text{Torsion}(\hat{\mathbf{n}}, \hat{\mathbf{a}}, \hat{\mathbf{c}}')$$

$$\psi = \text{Torsion}(\hat{\mathbf{a}}, \hat{\mathbf{c}}'\ \hat{\mathbf{n}}')$$

The following chiralities are also defined:

$$\epsilon_1 = \text{sign}(\hat{\mathbf{B}}\cdot(\hat{\mathbf{n}}\times\hat{\mathbf{a}}))$$

$$\epsilon_2 = \text{sign}(\hat{\mathbf{B}}\cdot(\hat{\mathbf{a}}\times\hat{\mathbf{c}}'))$$

$$\epsilon_3 = \text{sign}(\hat{\mathbf{B}}\cdot(\hat{\mathbf{c}}'\times\hat{\mathbf{n}}'))$$

$$= -\text{sign}(\hat{\mathbf{B}}\cdot(\hat{\mathbf{n}}'\times\hat{\mathbf{a}}')).$$

The chirality ϵ_1 will be defined to be the chirality of the peptide plane containing the unit bond vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{a}}$. The chirality ϵ_2 will be defined as the chirality of the α carbon containing the unit bond vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}'$. Using the formula above,

$$\begin{aligned}\phi &= \tau(\hat{\mathbf{B}}\cdot\hat{\mathbf{n}}, \hat{\mathbf{B}}\cdot\hat{\mathbf{a}}, \hat{\mathbf{B}}\cdot\hat{\mathbf{c}}', \hat{\mathbf{n}}\cdot\hat{\mathbf{a}}, \hat{\mathbf{a}}\cdot\hat{\mathbf{c}}', \epsilon_1, \epsilon_2)\\ \psi &= \tau(\hat{\mathbf{B}}\cdot\hat{\mathbf{a}}, \hat{\mathbf{B}}\cdot\hat{\mathbf{c}}', \hat{\mathbf{B}}\cdot\hat{\mathbf{n}}', \hat{\mathbf{a}}\cdot\hat{\mathbf{c}}', \hat{\mathbf{c}}'\cdot\hat{\mathbf{n}}', \epsilon_2, \epsilon_3).\end{aligned} \tag{5}$$

For the problem of two adjacent peptide planes, there are three chiralities, giving eight possible pairs of dihedral angles. For a polypeptide there are chiralities to be determined for each peptide plane (determined by vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{a}}$) and α carbon plane (determined by vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}'$). Thus the number of possible structures from a given set of bond orientation cosines, two for each peptide plane, is 2^{2n-1}, where n is the number of peptide planes. The number $2n - 1$ represents the number of peptide planes, n, plus the

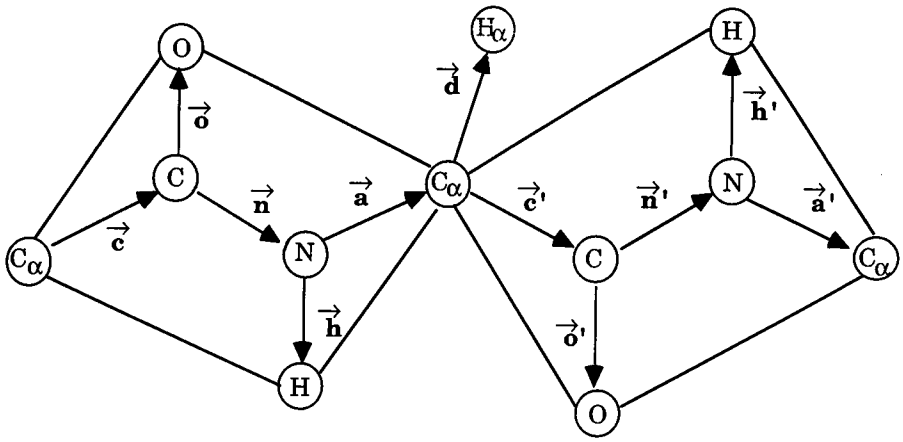

FIGURE 4 Unit bond vectors in adjacent peptide planes.

number of α carbons, $n - 1$. If data are available for bonds not in a peptide plane, such as the C_α—D bond, then some of the chiralities for α carbon planes can be determined, reducing the number of possibilities. An example of this is given in the Discussion section.

Using Eq. 4,

$$\phi = \text{Torsion}(\hat{\mathbf{n}}, \hat{\mathbf{a}}, \hat{\mathbf{B}}) + \text{Torsion}(\hat{\mathbf{B}}, \hat{\mathbf{a}}, \hat{\mathbf{c}}')$$
$$= \text{Torsion}(\hat{\mathbf{n}}, \hat{\mathbf{a}}, \hat{\mathbf{c}}'),$$

and this formula can be used to derive equation 13 in Teng et al. (1991). The sign ambiguities in the latter can now be explained in terms of chiralities, i.e., signs of triple products of vectors.

DISCUSSION

Gramicidin A

By using gramicidin A as an example, it is shown how the structure is affected by the choice of chiralities. Consider the following sample data in Table 2 (Ketchem et al., 1993). The components $\hat{\mathbf{B}} \cdot \hat{\mathbf{h}}$, $\hat{\mathbf{B}} \cdot \hat{\mathbf{n}}$, and $\hat{\mathbf{B}} \cdot \hat{\mathbf{d}}$ are obtained from experimental data, and the $\hat{\mathbf{B}} \cdot \hat{\mathbf{a}}$ and $\hat{\mathbf{B}} \cdot \hat{\mathbf{c}}$ components are obtained from these using the geometry of the peptide plane as above (Planes and unit bond vectors). The assumption is made that the ω torsion angles are 180°. The equations relating the unit bond vectors in the plane are $\hat{\mathbf{a}} = -1.01\hat{\mathbf{h}} + 1.05\hat{\mathbf{n}}$ and $\hat{\mathbf{c}} = -1.06\hat{\mathbf{h}} + 0.99\hat{\mathbf{n}}$. Vectors $\hat{\mathbf{h}}$, $\hat{\mathbf{n}}$, $\hat{\mathbf{a}}$, $\hat{\mathbf{c}}$ in Table 2 are those associated with the peptide plane containing the ^{15}N of the identified residue. The value of $\hat{\mathbf{B}} \cdot \hat{\mathbf{d}}$ at the Gly site is taken to be consistent with alternating D and L residues.

Obtaining the data in Table 2 requires elucidating some ambiguities distinct from the ones due to chirality studied here. They arise from the fact that tensors are invariant upon replacing $\hat{\mathbf{B}}$ by $-\hat{\mathbf{B}}$. Consequently, even after considering data from ^{15}N chemical shift, N—H and N—C^1 dipolar tensors, $(\hat{\mathbf{B}} \cdot \hat{\mathbf{n}}, \hat{\mathbf{B}} \cdot \hat{\mathbf{h}})$ is indistinguishable from $(-\hat{\mathbf{B}} \cdot \hat{\mathbf{n}}, -\hat{\mathbf{B}} \cdot \hat{\mathbf{h}})$. This is resolved by looking at C_α—D quadrupolar splittings (Ketchem et. al., 1996b). A table of values for

TABLE 2 Bond orientation cosines for gramicidin A obtained from SS NMR data

Residue	$\hat{\mathbf{B}} \cdot \hat{\mathbf{h}}$	$-\hat{\mathbf{B}} \cdot \hat{\mathbf{n}}$	$\hat{\mathbf{B}} \cdot \hat{\mathbf{a}}$	$\hat{\mathbf{B}} \cdot \hat{\mathbf{c}}$	$\hat{\mathbf{B}} \cdot \hat{\mathbf{d}}$
Val	−0.96	0.45	0.49	0.58	0.95
Gly	0.92	−0.75	−0.13	−0.23	−0.93
Ala	−0.98	0.4	0.58	0.66	0.94
D-Leu	0.92	−0.74	−0.15	−0.25	−0.93
Ala	−0.98	0.43	0.54	0.62	0.93
D-Val	0.93	−0.71	−0.19	−0.29	−0.92
Val	−0.98	0.43	0.54	0.62	0.94
D-Val	0.92	−0.73	−0.17	−0.27	−0.9
Trp	−0.99	0.44	0.54	0.62	
D-Leu	0.88	−0.72	−0.14	−0.23	−0.94
Trp	−0.99	0.48	0.49	0.58	
D-Leu	0.89	−0.74	−0.13	−0.23	−0.95
Trp	−0.98	0.46	0.5	0.59	
D-Leu	0.86	−0.72	−0.12	−0.21	−0.94
Trp	−0.97	0.45	0.51	0.6	
Eam	0.86	−0.72	−0.12	−0.21	

Eam, Ethanolamine blocking group at the carboxyl terminus.

$\hat{\mathbf{B}} \cdot \hat{\mathbf{n}}$ and $\hat{\mathbf{B}} \cdot \hat{\mathbf{h}}$ is obtained for each monomer. Table 2 is for the monomer with the N terminal-to-C terminal axis in the same direction as $\hat{\mathbf{B}}$.

For this monomer the C_α chiralities, sign $\hat{\mathbf{B}} \cdot (\hat{\mathbf{a}} \times \hat{\mathbf{c}}')$, are all + 1. This is apparent from Table 2 and the following calculation. Recall that the C_α—D direction is denoted $\hat{\mathbf{d}}$. The formula

$$(\hat{\mathbf{d}} \cdot (\hat{\mathbf{a}} \times \hat{\mathbf{c}}'))(\hat{\mathbf{B}} \cdot (\hat{\mathbf{a}} \times \hat{\mathbf{c}}')) = \begin{vmatrix} \hat{\mathbf{d}} \cdot \hat{\mathbf{B}} & \hat{\mathbf{d}} \cdot \hat{\mathbf{a}} & \hat{\mathbf{d}} \cdot \hat{\mathbf{c}}' \\ \hat{\mathbf{a}} \cdot \hat{\mathbf{B}} & 1 & \hat{\mathbf{a}} \cdot \hat{\mathbf{c}}' \\ \hat{\mathbf{c}}' \cdot \hat{\mathbf{B}} & \hat{\mathbf{c}}' \cdot \hat{\mathbf{a}} & 1 \end{vmatrix}$$

follows from writing the left-hand side as the product of determinants and using the formula det AB = det A det B (Beyer, 1987). Now the alternating stereochemistry of amino acid residues in gramicidin, L, D, L . . . , shows that the triple product $\hat{\mathbf{d}} \cdot (\hat{\mathbf{a}} \times \hat{\mathbf{c}}')$ alternates in sign, +, −, +, ... (Its absolute value is about $4/(3\sqrt{3})$.) Computation of the determinant on the right at the C_α sites where $\hat{\mathbf{B}} \cdot \hat{\mathbf{d}}$ is known indicates that the sign alternates, +, −, +, ... Thus $(\hat{\mathbf{B}} \cdot (\hat{\mathbf{a}} \times \hat{\mathbf{c}}'))$ is positive and the C_α plane chiralities

are all +1. This is the same as it would be for a right-handed α-helix, with the axis pointing in the direction of the magnetic field.

A sequence of torsion angles can be generated from the data by using Eq. 5 and a chosen chirality for each peptide and C_α plane. There are 2^{16} possible sequences of peptide plane chiralities to consider for gramicidin. The data are periodic, reflecting a β-sheet type of a structure with a period of two peptide planes. If it is assumed that the chiralities are likewise periodic, then the number of possible chirality sequences reduces to 4.

Taking alternate signs for the peptide plane chiralities, $-1, 1, -1, \ldots$, torsion angles ϕ_1 and ψ_1 in Table 3 result, and if the peptide plane chiralities are all -1, the torsion angles ϕ_2 and ψ_2 in Table 3 result from the formula above (The torsion angle formula). Additional possibilities are $+1, -1, +1, \ldots$, and all peptide plane chiralities $+1$. These structures are also referred to (Ketchem et al., 1996a) as g2328, g2424, g2823, and g2727, respectively. The folding motif (i.e., hydrogen-bonding pattern) is the same for each of these possibilities. This is because the direction of the helix axis, and hence the direction of **B**, is less than about 15° out of each peptide plane. Thus the virtual C_α—C_α bond vector is little affected by choosing the opposite chirality. These four torsion angle solution sets are a representative set of initial structures, and refinement against all experimental constraints and the CHARMM force field of all four of these leads to the same final structure and a unique set of chiralities independent of the chiralities used in the initial structure (Ketchem, 1995).

What pattern of chiralities should be expected from other peptides? For an ideal α-helix, with chiralities as defined here, if the direction of the magnetic field is in the direction of the helix axis, the chiralities are all -1. This is because for a regular helix the helix axis, $\hat{\mathbf{A}}$, is oriented in each peptide plane such that $\hat{\mathbf{A}} \cdot (\hat{\mathbf{n}} \times \hat{\mathbf{a}})$ is negative, with the C—O bond pointing away from the axis, or "out."

TABLE 3 Torsion angles for gramicidin A for peptide plane chiralities −1, +1, −1, . . . (ϕ_1 and ψ_1) and for chiralities −1, −1, −1, . . . , (ϕ_2 and ψ_2)

Residue	ϕ_1	ψ_1	ϕ_2	ψ_2
1	−139	141	−139	114
2	125	−109	151	−109
3	−134	153	−134	119
4	119	−102	152	−102
5	−142	152	−142	118
6	122	−111	154	−111
7	−136	150	−136	116
8	118	−103	151	−103
9	−134	153	−134	106
10	107	−97	153	−97
11	−135	146	−135	108
12	114	−100	151	−100
13	−135	153	−135	105
14	110	−100	157	−100
15	−137	154	−137	107

CONCLUSIONS

With high-resolution solid-state NMR it is necessary to have a structural analysis method such as described here to take full advantage of the unique data. Presented here is a structural analysis method based on vector algebra that provides analytical torsion angle solutions from solid-state NMR-derived orientational constraints. The assumption is made that all of the ω torsion angles are 180°. The resulting ambiguities, termed "chiralities," are described in such a way that a concise and simple view of the structural possibilities is developed. Each sequence of chiralities represents a specific analytical solution for the experimental data. In other words, the data are not consistent with the full range of torsion angles between the structural possibilities, such as those in Table 3.

The torsion angle solution is unstable in some situations, but these are anticipated to be rare throughout a helical protein structure, and therefore it should have little consequence for this structural method. The chirality ambiguities have no influence on the hydrogen-bonding pattern, helical parameters, and hence polypeptide fold in gramicidin. This is due to the fortunate situation that a change in chirality does not greatly affect the direction of the C_α—C_α axis in either a β-helix or an α-helix that is aligned parallel to the magnetic field direction. It is anticipated that defining the chiralities in molecules where the helical axis is not aligned with the field may be easier experimentally than when it is parallel. Current research in our laboratory will clarify this point.

This work has been supported by the National Science Foundation, DMB 9317111, to TAC, and by a Florida State University COFRS grant to JRQ.

We would like to thank Timothy Havel for his helpful comments on an early draft of this paper.

REFERENCES

Beyer, W. H. 1987. CRC Handbook of Mathematical Sciences. CRC Press, Boca Raton, FL.

Brenneman, M. T., and T. A. Cross. 1990. A novel method for the analytical determination of protein structure using solid state NMR: the "metric method." *J. Chem. Phys.* 90:1483–1494.

Crippen, G. M., and T. F. Havel. 1988. Distance Geometry and Molecular Conformation. Research Studies Press: Chemometrics series 15. Wiley, New York.

Cross, T. A. 1994. Structural biology of peptides and proteins in synthetic membrane environments by solid-state NMR spectroscopy. *Annu. Rep. NMR Spectrosc.* 29:124–158.

Cross, T. A., and S. J. Opella. 1994. Solid-state NMR structural studies of peptides and proteins in membranes. *Curr. Opin. Struct. Biol.* 4:574–578.

Engh, R. A., and R. Huber. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr.* A47:392–400.

Gluske, J. P., and K. N. Trueblood. 1985. Crystal Structure Analysis: A Primer. Oxford University Press, New York.

Ketchem, R. R. 1995. Structural determination and refinement of the gramicidin A transmembrane channel as studied by solid-state nuclear magnetic resonance spectroscopy. Ph.D. thesis. Florida State University, Tallahassee, FL.

Ketchem, R. R., W. Hu, and T. A. Cross. 1993. High-resolution conformation of gramicidin A in a lipid bilayer by solid-state NMR. *Science.* 261:1457–1460.

Ketchem, R. R., K. C. Lee, S. Huo, and T. A. Cross. 1996a. Macromolecular structural elucidation with solid-state NMR derived orientational constraints. *J. Biomol. NMR.* 8:1–14.

Ketchem, R. R., B. Roux, and T. A. Cross. 1996b. Computational refinement through solid-state NMR and energy constraints of a membrane bound polypeptide. *In* Biological Membranes. K. M. Merz, Jr., and B. Roux, editors. Springer-Verlag, New York. 299–322.

Kumar, B. S. A., and S. J. Opella. 1991. High-resolution heteronuclear dipolar solid-state NMR spectroscopy. *J. Magn. Res.* 95:417–420.

Nicholson, L. K., T. Asakura, M. Demura, and T. A. Cross. 1990. A method for studying the structure of uniaxially aligned biopolymers using solid state NMR ^{15}N-NMR: application to *Bombyx mori* silk fibronin fibers. *Biopolymers.* 33:847–861.

Opella, S. J., P. L. Stewart, and K. G. Valentine. 1987. Protein structure by solid state NMR spectroscopy. *Q. Rev. Biophys.* 19:7–49.

Simmons, A. H., C. A. Michal, and L. W. Jelinski. 1996. Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science.* 271:84–87.

Teng, Q., L. K. Nicholson, and T. A. Cross. 1991. Experimental determination of torsion angles in the polypeptide backbone of the gramicidin A channel by solid state nuclear magnetic resonance. *J. Mol. Biol.* 218: 601–619.

Wu, C. H., A. Ramamoorthy, and S. J. Opella. 1994. Two-dimensional ^{1}H/^{15}N heteronuclear correlation—^{15}N/^{15}N dilute spin-exchange solid-state NMR spectroscopy. *J. Magn. Res.* 105:270–272.