PROTEIN STRUCTURAL ANALYSIS FROM SOLID STATE NMR DERIVED ORIENTATIONAL CONSTRAINTS

J. R. QUINE, MATT BRENNEMAN AND T. A. CROSS

Quine: Department of Mathematics, Florida State University, Tallahassee, FL 32306-3027

Brenneman: Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306

Cross: Center for Interdisciplinary Magnetic Resonance in the National High Magnetic Field Laboratory, Institute of Molecular Biophysica and Department of Chemistry, Florida State University, Tallahassee, FL 32306

ABSTRACT

High resolution orientational constraints from solid state NMR spectroscopy of uniformly aligned biological macromolecules provides 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 very 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. As a result significant advantage is demonstrated over the previous analysis methods.

1. Introduction

The advent of multi-dimensional techniques (Kumar et al., 1991, Wu et al., 1994) and the recent determination of a peptide's complete three dimensional structure (Ketchem et al., 1993; Ketchem et al., 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 is presented here to determine the initial structure of a polypeptide's backbone from experimental SS NMR derived orientational constraints which is free of these limitations.

Determining the structure of a molecule from experimental data generally involves the use of some mathematical method, 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 which 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 figure 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 this data.

Using vector analysis, analytical equations for the $\{\phi, \psi\}$ torsion angles of the protein backbone are derived. The goal is not to just develop an algorithm which 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 minimal 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 which 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 the methods and materials, some basic terminology and notation is 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 a model peptide, gramicidin A, whose structure has already been determined by SS NMR.

2. 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 $\hat{\mathbf{B}} \cdot \hat{\mathbf{a}} = \cos\theta_{\mathbf{a}}, \hat{\mathbf{B}} \cdot \hat{\mathbf{b}} = \cos\theta_{\mathbf{b}}, \hat{\mathbf{B}} \cdot \hat{\mathbf{c}} = \cos\theta_{\mathbf{c}}$ and referred to as the bond orientation cosines. The dot products $\hat{\mathbf{a}} \cdot \hat{\mathbf{b}} = \cos\xi_{\mathbf{ab}}$ and $\hat{\mathbf{b}} \cdot \hat{\mathbf{c}} = \cos\xi_{\mathbf{bc}}$ are referred to as the bond angle cosines. The torsion angle, $\mathrm{Torsion}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$, is a function, τ, of seven variables:

$$\mathrm{Torsion}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) = \tau(\cos\theta_{\mathbf{a}}, \cos\theta_{\mathbf{b}}, \cos\theta_{\mathbf{c}}, \cos\xi_{\mathbf{ab}}, \cos\xi_{\mathbf{bc}}, \epsilon_{\mathbf{ab}}, \epsilon_{\mathbf{bc}}),$$

where

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

are chiralities which 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 figures 1,2 and 3.

2.1. **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 with the first vector 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

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

(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 $\hat{\mathbf{y}} = \hat{\mathbf{z}} = \mathbf{0}$.

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 $(\hat{\mathbf{B}} \cdot \hat{\mathbf{u}}, 0, (1 - (\hat{\mathbf{B}} \cdot \hat{\mathbf{u}})^2)^{1/2})$.

2.2. **Gramians, triple products, 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

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

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

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

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 $\epsilon = \operatorname{sign}(\hat{\mathbf{a}} \cdot (\hat{\mathbf{b}} \times \hat{\mathbf{c}}))$ 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 non-negativity of the gramian is an important observation in determining bond orientation cosines from second order tensor data (Opella et al., 1987 fig. 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 non-negativity of γ distinguishes triples of dot products that can occur for three vectors in Euclidean space.

2.3. **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

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

(This angle has an equivalent description as a dihedral angle between two planes.)

Suppose that bond angle cosines $\cos\theta_{\mathbf{a}}, \cos\theta_{\mathbf{b}}$ and $\cos\theta_{\mathbf{c}}$ are known and the expression for torsion angles is given in terms of bond orientation cosines and two chiralities. The bond orientation sines $\sin\theta_{\mathbf{a}}, \sin\theta_{\mathbf{b}}$ and $\sin\theta_{\mathbf{c}}$ can be computed from the bond orientation cosines using $\sin\theta = (1-\cos^2\theta)^{1/2}$. Similarly the bond angle sines $\sin\xi_{\mathbf{ab}}$ and $\sin\xi_{\mathbf{bc}}$ can be computed. Now the triple products are computed from the gramians:

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

Define $\alpha_{\mathbf{a}}$,$\beta_{\mathbf{a}}$, $\alpha_{\mathbf{c}}$ and $\beta_{\mathbf{c}}$ by

$$(\alpha_{\mathbf{a}}, \beta_{\mathbf{a}}) = \big(\hat{\mathbf{B}} \cdot (\hat{\mathbf{a}} \times \hat{\mathbf{b}}), -\cos\theta_{\mathbf{a}} + \cos\theta_{\mathbf{b}}\cos\xi_{\mathbf{ab}}\big)$$
$$(\alpha_{\mathbf{c}}, \beta_{\mathbf{c}}) = \big(\hat{\mathbf{B}} \cdot (\hat{\mathbf{b}} \times \hat{\mathbf{c}}), \cos\theta_{\mathbf{c}} - \cos\theta_{\mathbf{b}}\cos\xi_{\mathbf{bc}}\big)$$

The torsion angle formula is

$$\begin{aligned} \phi &= \mathrm{Torsion}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) \\ &= \mathrm{sign}\,(\alpha_{\mathbf{c}}\beta_{\mathbf{a}} - \alpha_{\mathbf{a}}\beta_{\mathbf{c}}) \arccos\left(\frac{\alpha_{\mathbf{a}}\alpha_{\mathbf{c}} + \beta_{\mathbf{a}}\beta_{\mathbf{c}}}{\sin^2\theta_{\mathbf{b}} \sin\xi_{\mathbf{ab}} \sin\xi_{\mathbf{bc}}}\right) \end{aligned} \tag{3}$$

Note that all quantities in the formula except for $\epsilon_{\mathbf{ab}}$ and $\epsilon_{\mathbf{bc}}$ 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

$$\phi = \tau(\cos\theta_{\mathbf{a}}, \cos\theta_{\mathbf{b}}, \cos\theta_{\mathbf{c}}, \cos\xi_{\mathbf{ab}}, \cos\xi_{\mathbf{bc}}, \epsilon_{\mathbf{ab}}, \epsilon_{\mathbf{bc}}).$$

To prove the formula the coordinate system at $\hat{\mathbf{b}}$ of vectors

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

is used. In terms of these vectors, the projections of $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}$ on the plane perpendicular to $\hat{\mathbf{b}}$ are easily computed as

$$P\hat{\mathbf{a}} = \frac{-1}{\sin\theta_{\mathbf{b}}}(\alpha_{\mathbf{a}}\,\hat{\mathbf{y}} + \beta_{\mathbf{a}}\,\hat{\mathbf{z}})$$
$$P\hat{\mathbf{c}} = \frac{1}{\sin\theta_{\mathbf{b}}}(\alpha_{\mathbf{c}}\,\hat{\mathbf{y}} + \beta_{\mathbf{c}}\,\hat{\mathbf{z}}).$$

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

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

Noting that $|P\hat{\mathbf{a}}| = \sin\xi_{\mathbf{ab}}$ and $|P\hat{\mathbf{c}}| = \sin\xi_{\mathbf{bc}}$, the formula follows.

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,

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

2.4. **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 formula (3) by the vanishing of $\sin\theta_{\mathbf{b}}$ 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 circle in figure 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 since $d\sqrt{g} = dg/2\sqrt{g}$ 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 figure 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 is to tip the samples in the spectrometer so that $\hat{\mathbf{B}}$ will no longer be in the plane.

3. Results

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

3.1. **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 figure 4). The geometry of the peptide plane bonds and the bonds to the C_α carbons are determined by the following dot products (Engh and Huber, 1991):

$$\begin{array}{ll} \hat{\mathbf{c}} \cdot \hat{\mathbf{a}} = \cos 6^\circ & \hat{\mathbf{c}} \cdot \hat{\mathbf{n}} = \cos 75^\circ \\ \hat{\mathbf{c}} \cdot \hat{\mathbf{h}} = \cos 123^\circ & \hat{\mathbf{a}} \cdot \hat{\mathbf{c}}' = \cos 70^\circ \\ \hat{\mathbf{n}} \cdot \hat{\mathbf{a}} = \cos 59^\circ & \hat{\mathbf{h}} \cdot \hat{\mathbf{a}} = \cos 117^\circ \\ \hat{\mathbf{n}} \cdot \hat{\mathbf{h}} = \cos 58^\circ & \hat{\mathbf{d}} \cdot \hat{\mathbf{a}} = 1/3 \\ \hat{\mathbf{d}} \cdot \hat{\mathbf{c}}' = -1/3 & \end{array}$$

Using the notation for bond angle, for example, $\xi_{\mathbf{ch}} = 123^\circ$. 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 is 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.

3.2. **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:

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

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

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

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

Using equation (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.

4. Discussion

4.1. **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 in section 3.1. The assumption is made that the ω torsion angles are 180°. The equations are $\hat{\mathbf{a}} = -1.01\hat{\mathbf{h}} + 1.05\hat{\mathbf{n}}$ and $\hat{\mathbf{c}} = -1.06\hat{\mathbf{h}} + .99\hat{\mathbf{n}}$. 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.

A sequence of torsion angles can be generated from the data using equation (5) and a chosen chirality for each peptide and C_α plane. The chiralities of the C_α planes can be determined from the experimental C_α-D data. Recall that the direction of these is denoted by $\hat{\mathbf{d}}$. The formula

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

follows from writing the left hand side as the product of determinants and using the formula $\det AB = \det A \det B$. Now the alternating stereochemistry of amino acid residues in gramicidin shows that the triple product $\hat{\mathbf{d}} \cdot (\hat{\mathbf{c}}' \times \hat{\mathbf{a}})$ alternates in sign -,+,-, ... 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 it is reasonable to assume that $(\hat{\mathbf{B}} \cdot (\hat{\mathbf{c}}' \times \hat{\mathbf{a}}))$ can be taken to be negative at all C_α sites and the C_α plane chiralities are all -1. Now there are 2^{16} possible sequences of chiralities for the peptide planes to consider. The folding motif (i. e. hydrogen bonding pattern) is the same for each of these possibilities.

Taking alternate signs for the peptide plane chiralities, -1,1,-1, . . . , torsion angles in table 3 result and if the peptide plane chiralities are all +1 the the torsion angles in table 4 result from the formula in section 2.3, Additional natural possibilities are

1,-1,1, ... and all peptide plane chiralities -1. 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).

4.2. **Conclusions.** 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 resulting ambiguities, termed chiralities, are described in such a way that a very concise and simple view of the structural possibilities is developed. In general, ambiguities may suggest structural error and a poor definition of the structure, however the chirality ambiguities described here represent specific analytical solutions for the experimental data. In other words, the data is not consistent with the full range of torsion angles between the structural possibilities, such as those in table 3. Furthermore, for gramicidin all possible structures have the same hydrogen bonding pattern, helical parameters, and hence polypeptide fold. With the high resolution solid-state NMR data it is necessary to have a structural analysis method such as described here to take full advantage of the unique data.

Acknowledgement. This work has been supported by the National Science Foundation, MEB 9317111 to T. A. Cross and by a Florida State University COFRS grant to J. R. Quine. We would like to thank Timothy Havel for his helpful comments on an early draft of this paper.

References

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: Wiley, Chemometrics series 15, New York.

Cross, T. A. 1994. Structural Biology of Peptides and Proteins in Synthetic Membrane Environments by Solid-state NMR Spectroscopy. *Annual Reports on NMR Spectroscopy.* 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 Cryst.* A47:392-400.

Glusker 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. PhD Thesis Florida State University.

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. Biomolecular NMR.* in press.

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. Mag. 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, *Quarterly Review of Biophysics.* 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. Mag. Res.* 105:270-272.

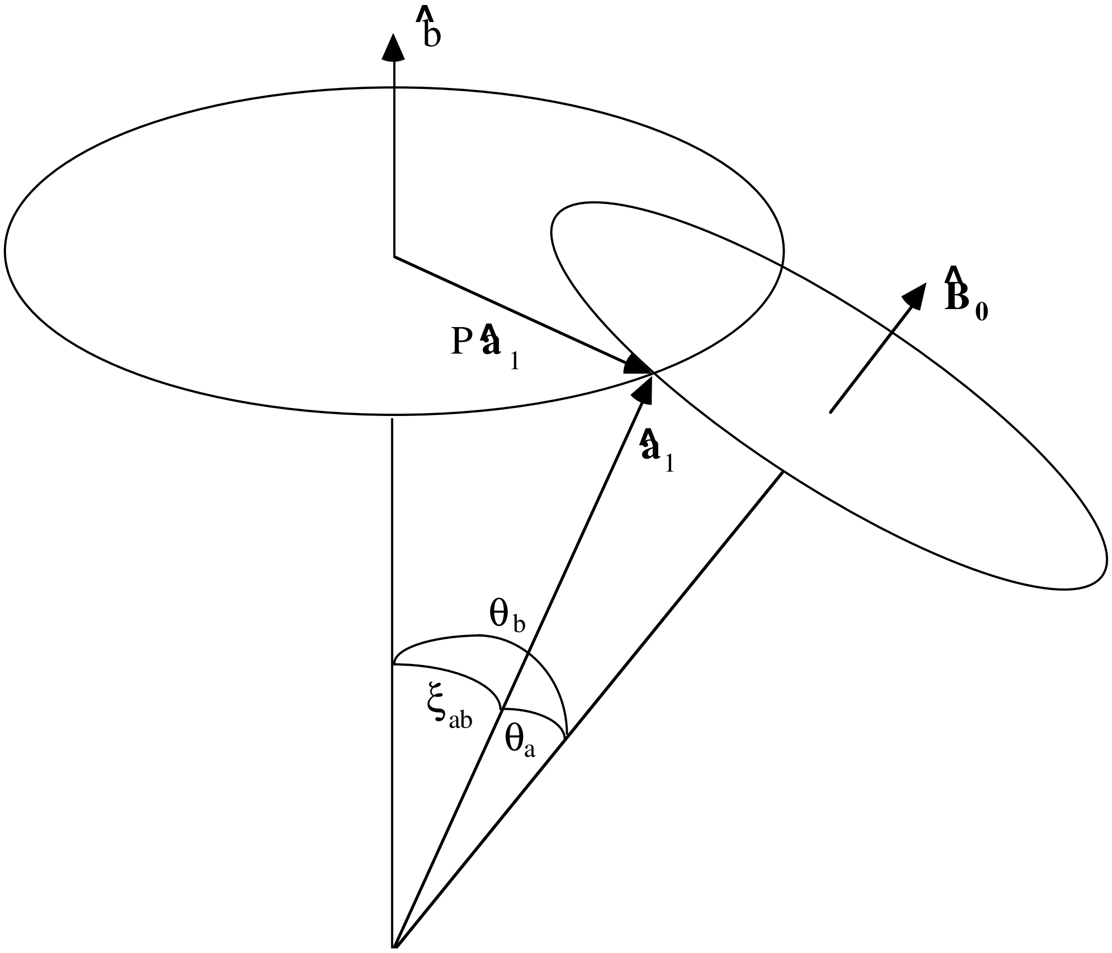

FIGURE 1. **Bond orientation and bond angle constraints.** $\hat{\mathbf{B}}_0$ and $\hat{\mathbf{b}}$ are given unit vectors and $\hat{\mathbf{a}} \cdot \hat{\mathbf{b}} = \cos \xi_{\mathbf{ab}}$, $\hat{\mathbf{a}} \cdot \hat{\mathbf{B}}_0 = \cos \theta_{\mathbf{a}}$. How does this determine the unit 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 $\hat{\mathbf{b}}$.

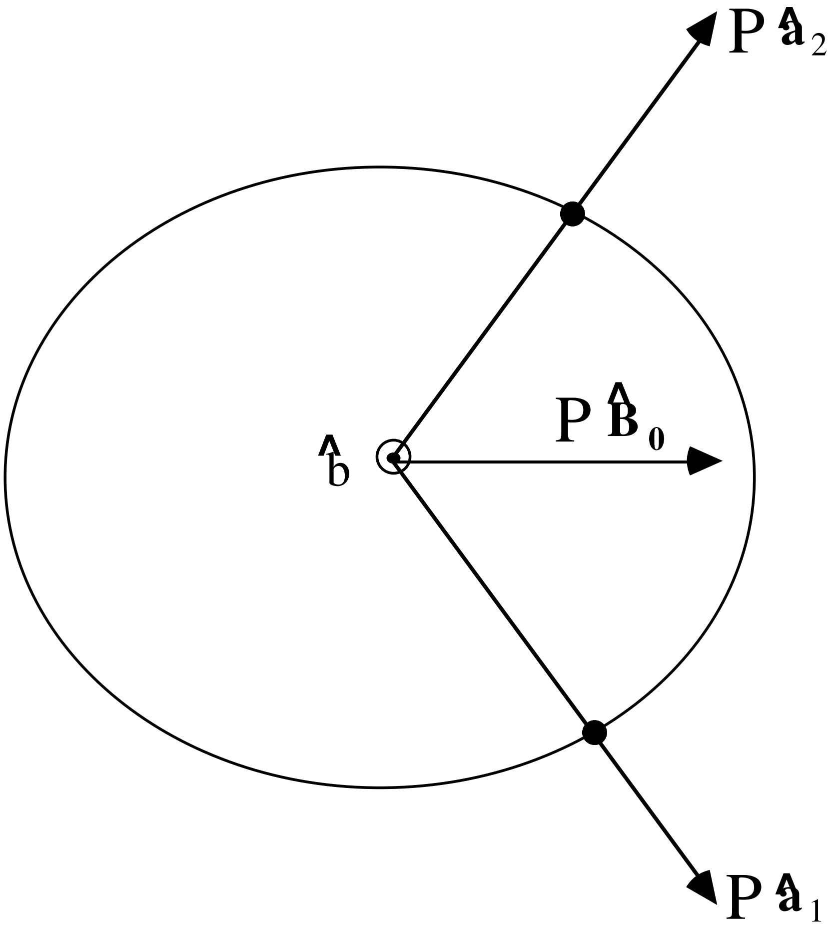

FIGURE 2. **Projection and chirality.** Projecting figure 1 onto the plane perpendicular to $\hat{\mathbf{b}}$ two possibilities $P\hat{\mathbf{a}}_1$ and $P\hat{\mathbf{a}}_2$ result. The projections are symmetric with respect to $P\hat{\mathbf{B}}_0$ and the two possibilities are distinguished by the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

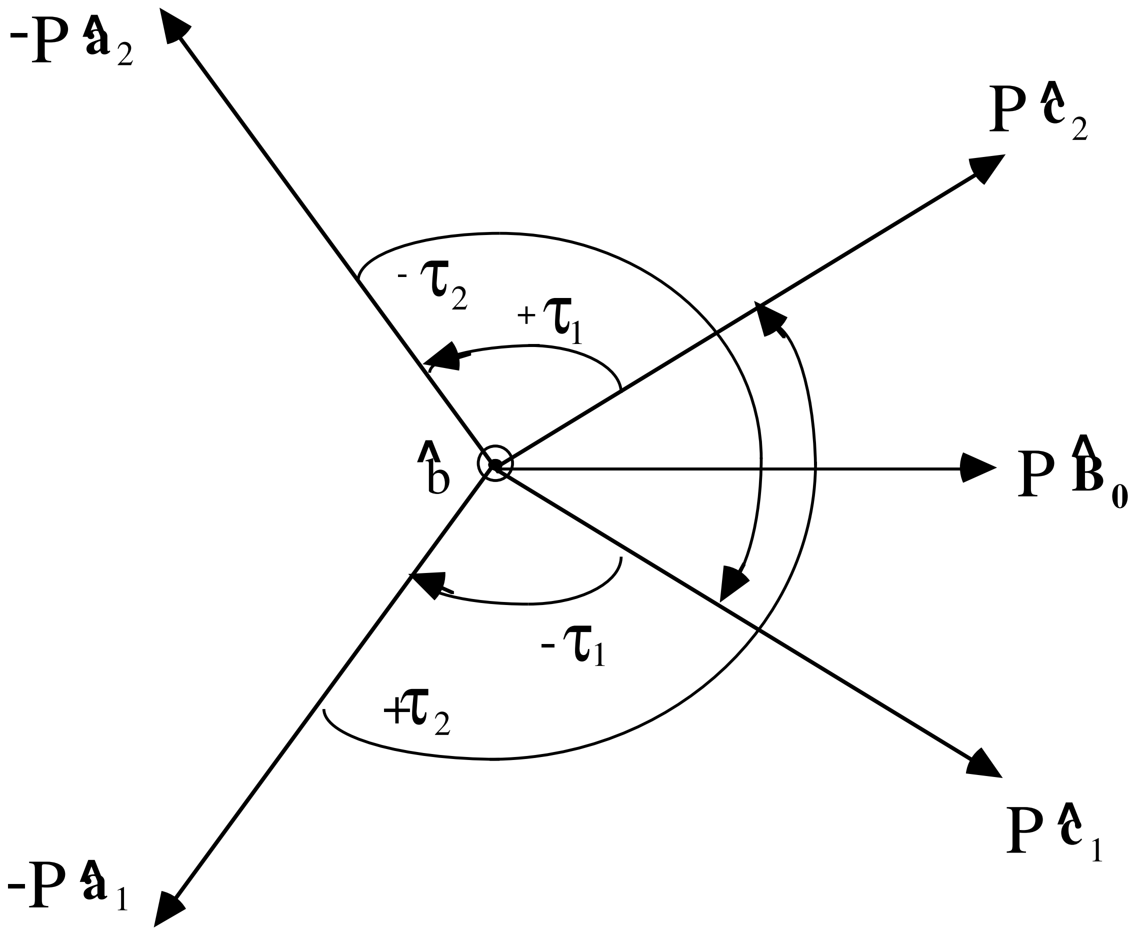

FIGURE 3. **Chirality and torsion.** The torsion angle, Torsion($\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}$), is the angle from $-P\hat{\mathbf{a}}$ to $P\hat{\mathbf{c}}$ where P is the projection onto the plane perpendicular to $\hat{\mathbf{b}}$ as in figure 2. With two possible values 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.

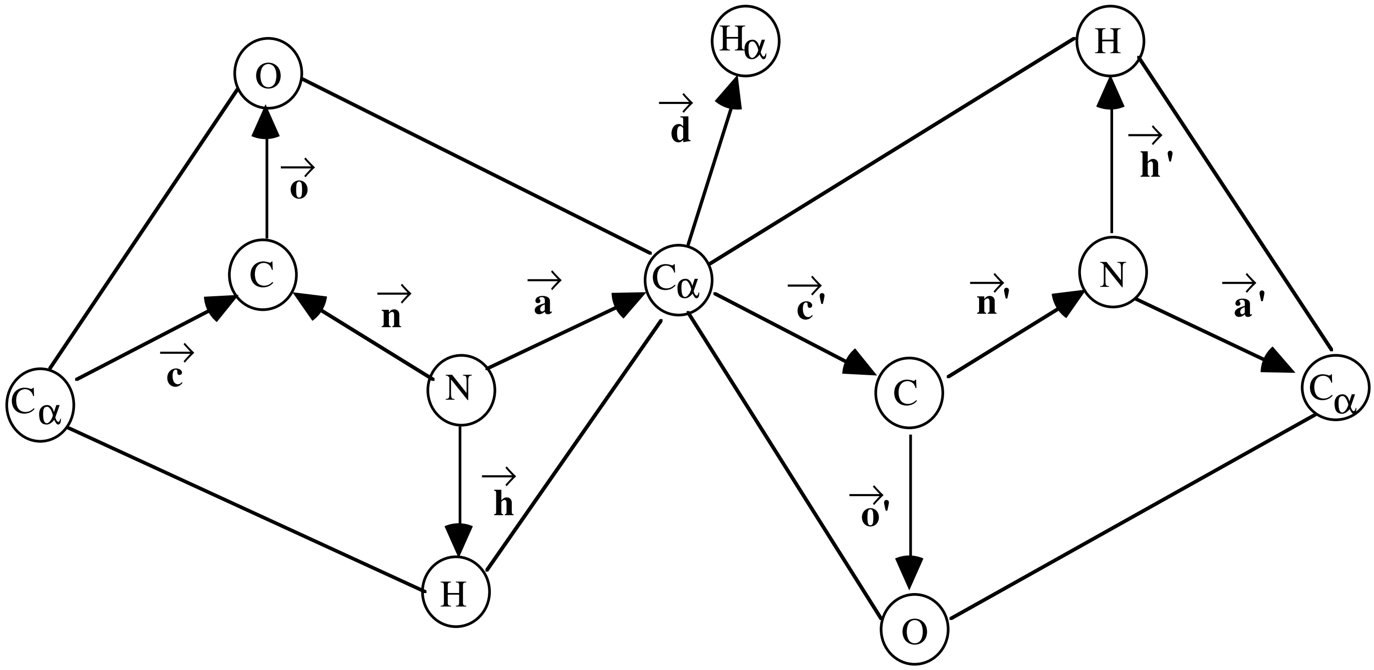

FIGURE 4. **Unit bond vectors in adjacent peptide planes.**

unit vector	chemical bond
$\hat{\mathbf{c}}$	$\mathrm{C}_\alpha - \mathrm{C}'$
$\hat{\mathbf{n}}$	$\mathrm{C}' - \mathrm{N}$
$\hat{\mathbf{a}}$	$\mathrm{N} - \mathrm{C}_\alpha$
$\hat{\mathbf{h}}$	$\mathrm{N} - \mathrm{H}$
$\hat{\mathbf{d}}$	$\mathrm{C}_\alpha - \mathrm{D}$

TABLE 1. Unit bond vectors in a peptide plane

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	

TABLE 2. Bond orientation cosines for gramicidin A obtained from SS NMR data. 'Eam' is the ethanolamine blocking group at the carboxyl terminus.

residue	ϕ	ψ
1	-139	141
2	125	-109
3	-134	153
4	119	-102
5	-142	152
6	122	-111
7	-136	150
8	118	-103
9	-134	153
10	107	-97
11	-135	146
12	114	-100
13	-135	153
14	110	-100
15	-137	154

TABLE 3. Torsion angles for gramicidin A for peptide chiralities $-1, 1, -1, \ldots$.

residue	ϕ	ψ
1	-139	114
2	151	-109
3	-134	119
4	152	-102
5	-142	118
6	154	-111
7	-136	116
8	151	-103
9	-134	106
10	153	-97
11	-135	108
12	151	-100
13	-135	105
14	157	-100
15	-137	107

TABLE 4. Torsion angles for gramicidin A for peptide chiralities $1, 1, 1, \ldots$.