Protein Structure in Anisotropic Environments: Unique Structural Fold from Orientational Constraints

J. R. QUINE,1 T. A. CROSS2

1Department of Mathematics, Institute of Molecular Biophysics, Florida State University, Tallahassee FL 32306-4510 E-mail: quine@math.fsu.edu
2Department of Chemistry, National High Magnetic Field Laboratory, Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32310-4005; E-mail: cross@magnet.fsu.edu

ABSTRACT: The mathematical foundation of the determination of protein structure from orientational constraints is described. The tools used are vector algebra, gram matrices, and determinants. The discussion begins in the general abstract setting and proceeds to a discussion of how the methods can be applied to the determination of protein structure using solid state nuclear magnetic resonance. Examples are given relating to the structure of the peptide gramicidin A. © 2000 John Wiley & Sons, Inc. Concepts Magn Reson 12: 71–82, 2000.

KEY WORDS: solid-state NMR; protein structure; orientational constraints; anisotropic environment; gram matrix; gramian; gramicidin A

1. INTRODUCTION

The development of a unique three-dimensional structure from orientational constraints is possible. Here the ambiguities are discussed and the mathematics for achieving a unique three-dimensional structure is described. The ambiguities are of two types, chiral and tensor. The chiral ambiguities arise because of the sign of a vector cross-product which cannot be determined from orientational constraints alone. While some of these can be found from the chirality of the amino acid, others related to the positions of peptide planes must be resolved in refinement. The tensor ambiguities arise because there are multiple solutions for bond orientation cosines from tensor values. With enough tensor values, essentially an overdetermined system, these ambiguities can be resolved. While the covalent geometry of the amino acids and their sequence is assumed,
assumptions about secondary or tertiary structure are not made.

Although it is beyond the scope of this report, the utilization of distance constraints or torsional constraints in conjunction with orientational constraints would be very beneficial, especially for defining specific contact regions between structural domains or resolving a chirality ambiguity. The mathematical development in this paper uses gram matrices which are closely related to the distance matrices used in the study of distance constraints (1), and so this treatment can efficiently be combined with distance matrix methods. Furthermore, there is much to be gained by combining the constraints with other structural results whether it be quantitative circular dichroism (CD) determinations of helical percentage or crystallographic data that provides modest resolution structural information.

2. STRUCTURE

The concept of structure arises from an attempt to isolate the properties of a molecule that result from its geometric configuration in space. The molecule is thought of as frozen in time and given by a set of space coordinates giving the position of each atom. A structure will be thought of as a sequence of points or atoms $P_0, \ldots, P_n$ in three-dimensional space, but we need to remember that this structure is both a time-averaged structure and a dynamic structure.

Central to the study of structure is the group $E$ of rigid, orientation preserving motions of Euclidean three-dimensional space. An element of $E$ is either a translation or a screw rotation about some fixed line. Two sequences of points give equivalent structures if they can be made to coincide by an element of $E$, that is, if there is a motion in $E$ which maps one sequence of coordinates onto the other.

An oriented structure is a structure together with a fixed direction in the laboratory frame denoted $\mathbf{B}$. Oriented structures arise from solid state nuclear magnetic resonance (NMR)–derived orientational constraints, and $\mathbf{B}$ is thought of as a direction for the magnetic field. There are two possible choices for the direction $\mathbf{B}$, two unit vectors pointing in opposite directions, and the particular one chosen is arbitrary. Experimental data does not distinguish between $\mathbf{B}$ and $-\mathbf{B}$, and this causes ambiguities in the interpretation of the data as discussed in Section 8.

Central to the study of oriented structures is $E_{\mathbf{B}}$, the subgroup of $E$ which leaves $\mathbf{B}$ fixed. Elements of this group consist of translations or screw rotations about axes parallel to $\mathbf{B}$. Two sequences of points give equivalent oriented structures if they can be made to coincide by element of $E_{\mathbf{B}}$.

Also useful is the study of structural elements, or substructures, structures for a subset of the given set of atoms. Given structures for disjoint sets of atoms, the question naturally arises as to how to piece them together into a complete structure. An element of $E$ must be found which brings a representative of the first structural element into proper alignment with the second. To give an oriented structure for all the atoms, oriented structural elements must be brought into alignment using an element of $E_{\mathbf{B}}$. These oriented structural units are more rigid since the group $E_{\mathbf{B}}$ has less parameters than the whole group $E$. The vector $\mathbf{B}$ keeps all the substructures in proper alignment. These ideas are given a more explicit mathematical formulation in the sections below.

3. ORIENTATIONAL CONSTRAINTS

Both distance constraints and orientational constraints are obtainable by solid-state nuclear magnetic resonance (NMR). A distance constraint gives information on $|\mathbf{P}_i \cdot \mathbf{P}_j|$, the distance between a pair of atoms. An orientational constraint gives information on the angle between a vector $\mathbf{P}_i \cdot \mathbf{P}_j$ joining bonded atoms and a magnetic field direction $\mathbf{B}$. When using distance constraints, squared distances $|\mathbf{P}_i \cdot \mathbf{P}_j|^2$ are put into a matrix called the distance matrix. Complete knowledge of this matrix determines the structure (1, 2). In the study of orientational constraints it is useful to form matrices of dot products, or gram matrices. Given vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, the dot product $\mathbf{v}_1 \cdot \mathbf{v}_2$ is defined geometrically as $|\mathbf{v}_1| \cdot |\mathbf{v}_2| \cdot \cos \theta$ where $\theta$ is the angle between the two vectors. This definition is independent of the coordinate frame in which the vectors are represented, but the dot product can also be computed easily from the coordinates of the vectors in a frame. If the vectors are given in some orthonormal coordinate frame $(i, j, k)$ by $\mathbf{v}_1 = a_i i + b_j j + c_k k$ and $\mathbf{v}_2 = a_i i + b_j j + c_k k$ then $\mathbf{v}_1 \cdot \mathbf{v}_2 = a_i a_2 + b_i b_2 + c_i c_2$. The vectors can also be converted to column matrices

$$
\mathbf{v}_1 = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} \quad \mathbf{v}_2 = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}
$$
and the dot product is given as the transpose of \(v_1\) times \(v_2\),

\[v_1 \cdot v_2 = v_1^T v_2\]

The angle between a vector \(P_i P_{i+1}\) and \(B\) will be called the bond orientation angle and its cosine the bond orientation cosine. In practice orientation cosines are obtained only for a pair of covalently bonded atoms and it is convenient to order the sequence \(P_i\) of atoms so that each atom is bonded to the ones adjacent in the sequence, i.e., there is information only about the orientation of \(P_i P_{i+1}\) for each \(i\). The sequence or atoms may be, for example, \(N-C-n-C_1-N\) along the protein backbone, or alternatively the sequence may follow the bonds along a side chain.

To set up the notation, let \(u_i\) be the unit vector in the direction of the bond \(P_i P_{i+1}\),

\[u_i = \frac{P_i P_{i+1}^\times}{|P_i P_{i+1}|}\]

and let \(\kappa_i\) be the bond orientation cosine defined by

\[B \cdot u_i = \kappa_i\]

for \(i = 1, \ldots, n - 1\). Information about the values of \(\kappa_i\) obtained from solid-state NMR experiments are called orientational constraints.

Determining structure of proteins using orientational constraints requires knowledge of the covalent geometry, the angles between covalent bond directions from the same atom, and the distances between covalently bonded atoms. These bond angles and bond lengths are well documented (3, 4). For notation let \(\beta_i\) be the bond angle cosine defined by

\[u_{i-1} \cdot u_i = \beta_i\]

for \(i = 2, \ldots, n - 1\), and let \(l_i\) be the length of the \(i\)th bond defined by

\[|P_i P_{i+1}^\times| = l_i\]

for \(i = 1, \ldots, n - 1\). In the following discussions the sequences \(\beta_i\) and \(l_i\) will be assumed as given, fixed values, with the structure to be solved as a function of the orientational constraints. Typically the geometry about the protein backbone heteroatoms is either approximately trigonal (\(\beta_i \approx \frac{1}{2} = \cos 60^\circ\)) or approximately tetrahedral (\(\beta_i \approx 1/3\)). In units of Angstroms \(l_i\) is between 1 and 1.5 depending on the type of bond.

Our goal is to determine the structure, a set of coordinates for \(P_1, \ldots, P_n\), from a set of orientational constraints. With orientational constraints for each structural element, a finite set of possibilities for the sequence of orientation cosines \(\kappa_i\) is obtained. With enough orientational constraints it is possible to obtain a unique sequence of orientation cosines (except for multiplication of all by \(-1\) which gives a rotated set of coordinates) despite several ambiguities. First, experiments do not distinguish between \(B\) and \(-B\), and so information is obtained only for \(|B \cdot u_i|\). Second, as discussed in Section 8, certain tensor values are determined only in their absolute value. These are referred to as tensor ambiguities and can be resolved with enough tensor observations.

Even with a given sequence of bond orientation cosines there are still many possible structures. This is a result of a chirality, a sign of a vector triple product, which cannot be determined experimentally from the backbone data which are planar. The mathematical formulation for these ambiguities is found in the following sections, and a description is given of strategies for resolving them in an effort to obtain a unique structural solution.

It should be mentioned that the mathematical formulation presumes that exact values are available for the orientation constraints. In fact, there is always a certain amount of experimental error. The effect of this error on the final solution is not discussed here, but this mathematical formulation is a starting point for an analysis of the error.

4. GRAM MATRICES

A useful tool in working with orientational constraints in the gram matrix, also called the metric matrix or metric tensor, since it gives complete information about the angles between pairs of vectors in a list. Some properties of gram matrices for three vectors are summarized in this section; more detailed information can be found in the section of crystallographic computing in (5).

Given vectors \(v_1, v_2\) and \(v_3\), the gram matrix is the symmetric matrix containing all possible inner products of these,

\[\text{gram}(v_1, v_2, v_3) = \begin{pmatrix}
  v_1 \cdot v_1 & v_1 \cdot v_2 & v_1 \cdot v_3 \\
  v_2 \cdot v_1 & v_2 \cdot v_2 & v_2 \cdot v_3 \\
  v_3 \cdot v_1 & v_3 \cdot v_2 & v_3 \cdot v_3
\end{pmatrix}\]
If \( \mathbf{v} = a\mathbf{v}_1 + b\mathbf{v}_2 + c\mathbf{v}_3 \) is a vector given as a linear combination of \( \mathbf{v}_1, \mathbf{v}_2 \) and \( \mathbf{v}_3 \), then its squared length is given in terms of its coordinates \( a, b, \) and \( c \) as

\[
|a\mathbf{v}_1 + b\mathbf{v}_2 + c\mathbf{v}_3|^2 = (a, b, c)\text{gram}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) \begin{pmatrix} a \\ b \\ c \end{pmatrix}
\]

and for this reason the matrix is sometimes called the metric matrix.

If the vectors are written as column matrices in some fixed frame, or lab frame \( \text{LAB} \), and if a \( 3 \times 3 \) matrix \( (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) \) is constructed from these, then

\[
\text{gram}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)
\]

Since

\[
\det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3) \quad [1]
\]

it follows that

\[
\det \text{gram}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = |\mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3)|^2 \quad [2]
\]

and so this determinant is nonnegative. It is zero exactly when the three vectors are coplanar. A geometric interpretation of \( \det \text{gram}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) \) is the squared volume of the rectangular parallelepiped formed by the three vectors \( \mathbf{v}_1, \mathbf{v}_2, \) and \( \mathbf{v}_3 \) (Fig. 1). If the vectors are in a plane, then this volume is 0. The determinant \( \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) \) is the scalar triple product which gives the signed volume of the parallelepiped. For unit vectors it also has an interpretation in terms of spherical trigonometry (see (6)).

If \( \mathbf{u}_1, \mathbf{u}_2, \) and \( \mathbf{u}_3 \) are unit vectors then

\[
\det \text{gram}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = 1 - x^2 - y^2 - z^2 + 2xyz \quad [3]
\]

Figure 1  The scalar triple product \( \mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3) \) is the signed volume of the parallelepiped formed by the three vectors. If \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) are in the plane of the page and \( \mathbf{v}_2 \) behind the page, then the volume is positive.

Figure 2  Let \( \theta_1 = \arccos x, \theta_2 = \arccos y \) and \( \theta_3 = \arccos z \) be the side lengths of a spherical triangle. If the vertices are considered as unit vectors, the non-negativity of the gram determinant, \( 1 - x^2 - y^2 - z^2 + 2xyz \geq 0 \), implies that if two of the angles are small, then the third one must also be small.

where

\[
\begin{align*}
x &= \mathbf{u}_1 \cdot \mathbf{u}_2 \\
y &= \mathbf{u}_1 \cdot \mathbf{u}_3 \\
z &= \mathbf{u}_2 \cdot \mathbf{u}_3
\end{align*}
\]

Note that the right-hand side of [3] is a symmetric function of \( x, y, \) and \( z \) which is nonnegative. This observation is important in determining which values of \( x, y, \) and \( z \) can be entries in the gram matrix

\[
\begin{pmatrix}
1 & x & y \\
x & 1 & z \\
y & z & 1
\end{pmatrix}
\]

of three unit vectors. This may then be applied to the situation when one of the vectors is a unit direction \( \mathbf{B} \) for the magnetic field and the other two are directions for bonds, helping to determine which orientational constraints lead to possible structures (7). Letting \( \mathbf{B} = \mathbf{u}_3, \) for example, and \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \) bonds at 60\(^\circ\) with \( x = \mathbf{u}_1 \cdot \mathbf{u}_3 = 0.5, \) then \( y = \mathbf{u}_1 \cdot \mathbf{u}_3 = 0.95, \) and \( z = \mathbf{u}_2 \cdot \mathbf{u}_3 = 0.95 \) cannot be the cosines of angles between the bond vectors and \( \mathbf{B} \) since substitution of these values in [3] gives a negative value, \(-0.1525\). The geometric explanation is that \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \) are very close to the direction of \( \mathbf{B} = \mathbf{u}_3 \) so the angle between them cannot be as large as 60\(^\circ\) (See Fig. 2.).

5. FRAMES

A frame is defined as an ordered set \((\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)\) of three linearly independent vectors. The vectors are linearly independent if and only if \( \det \text{gram}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) \neq 0. \) A frame is said to be orthonormal or an orthogonal coordinate system if the gram matrix is the identity. If \( \mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3) > 0, \) the frame is said to be right handed. Frames are often chosen to be orthonormal and right handed.
6. Algorithm for Structure from Orientational Constraints

The problem is to determine a set of coordinates for $P_1, \ldots, P_n$ from given sequences $\kappa_i$, $\beta_i$, and $l_i$ of bond orientation and bond angle cosines and lengths as described in Section 3. By definition $|\kappa_i| \leq 1$, but for the following analysis the assumption $|\kappa_i| < 1$ is needed, so that no bond is parallel to $B$. Otherwise a 0 appears in the denominator in Eq. [9] and others below. The structure determined is an oriented structure with $B$ taken for convenience as $k$ in the laboratory frame $(i, j, k)$. It is sufficient to find the unit vectors $u_i$ since the structure can then be constructed recursively from the bond lengths by setting

$$P_{i+1} = P_i + l_i u_i$$  \[5\]

So an algorithm is needed to find unit vectors $u_1, \ldots, u_{n-1}$ such that $B \cdot u_i = \kappa_i$ for $i = 1, \ldots, n-1$ and $u_{i-1} \cdot u_i = \beta_i$ for $i = 2, \ldots, n-1$. The two point structure is easy, let $u_1 = \sqrt{1 - \kappa_1^2}i + \kappa_1k$ in the frame LAB. The structure is determined except for a rotation about $B$.

6.1. Basic Computation for a Three Point Structure

For a three point structure, assume that $u_1$ is given as above. What are the possibilities for $u_2$? In other words, how many three point oriented structures for $P_1, \ldots, P_3$ are possible by piecing together oriented structures for $P_1, P_2$ and $P_2, P_3$? The answer is 0, 1, or 2 as the following computation shows (see Fig. 3).
To find an expression for \( \mathbf{u}_2 \) in terms of \( \mathbf{u}_1 \), use the frame \( (\mathbf{u}_1, \mathbf{B}, \mathbf{u}_1 \times \mathbf{B}) \) and write \( \mathbf{u}_2 \) in terms of this frame,

\[
\mathbf{u}_2 = a\mathbf{u}_1 + b\mathbf{B} + c\mathbf{u}_1 \times \mathbf{B}.
\]  

[6]

To find \( a, b, \) and \( c \) take the dot product of [6] with \( \mathbf{u}_1, \mathbf{B}, \text{ and } \mathbf{u}_1 \times \mathbf{B} \), respectively, to get

\[
\begin{pmatrix}
\beta_2 \\
\kappa_2 \\
(\mathbf{u}_1 \times \mathbf{B}) \cdot \mathbf{u}_2
\end{pmatrix} = \text{gram}(\mathbf{u}_1, \mathbf{B}, \mathbf{u}_1 \times \mathbf{B}) \begin{pmatrix}
a \\
b \\
c
\end{pmatrix}
\]  

[7]

Using [4] and inverting, get

\[
\begin{pmatrix}
a \\
b \\
c
\end{pmatrix} = \begin{pmatrix}
1 & \kappa_1 & 0 \\
\kappa_1 & 1 & 0 \\
0 & 0 & 1 - \kappa_1
\end{pmatrix}^{-1} \times \begin{pmatrix}
\beta_2 \\
\kappa_2 \\
(\mathbf{u}_1 \times \mathbf{B}) \cdot \mathbf{u}_2
\end{pmatrix}
\]  

[8]

Now \( \mathbf{B} \cdot (\mathbf{u}_1 \times \mathbf{u}_2) \) can be found up to sign from \( \kappa_1, \kappa_2, \) and \( \beta_2 \) since by [2] and [3],

\[
|\mathbf{B} \cdot (\mathbf{u}_1 \times \mathbf{u}_2)| = \sqrt{g_2}
\]

where

\[
g_2 = 1 - \kappa_1^2 - \kappa_2^2 - \beta_2^2 + 2\kappa_1\kappa_2\beta_2
\]


\[
\mathbf{u}_2 = \frac{1}{(1 - \kappa_1^2)} \left( (\beta_2 - \kappa_2 \kappa_1) \mathbf{u}_1 + (\kappa_2 - \beta_2 \kappa_1) \mathbf{B} + \varepsilon \sqrt{g_2} \mathbf{u}_1 \times \mathbf{B} \right)
\]  

[9]

where \( \varepsilon = -\text{sign} \mathbf{B} \cdot (\mathbf{u}_1 \times \mathbf{u}_2) = \pm 1 \). Since \( \varepsilon \) is not determined by the data, there are two possible values of \( \mathbf{u}_2 \) if \( g_2 \neq 0 \), one value if \( g_2 = 0 \). There is no solution if \( g_2 < 0 \), and so the data does not correspond to any structure.

**Remark.** As \( |\kappa_1| \) gets close to 1, the solution becomes unstable. In fact, there are an infinite number of solutions \( \mathbf{u}_2 \) if \( |\kappa_1| = 1 \) and \( \beta = \kappa_2 \).

### 6.2. The General Algorithm

For three atoms and \( \kappa_1, \kappa_2, \beta_2, l_1, \) and \( l_2 \) given then zero, one, or two oriented structures can be determined depending on whether \( g_2 \) is \( < 0, = 0, \) or \( > 0 \). The structure is solved by finding \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \) in terms of \( \kappa_1, \kappa_2, \) and \( \beta_2 \).

In the general situation the sequences \( \kappa_i, \ i = 1, \ldots, n - 1 \) and \( \beta_i, \ i = 2, \ldots, n - 1 \) are given and unit vectors \( \mathbf{u}_1, \ldots, \mathbf{u}_{n-1} \) are sought such that

\[
\mathbf{u}_{i-1} \cdot \mathbf{u}_i = \beta_i,
\]

\[
\mathbf{u}_i \cdot \mathbf{B} = \kappa_i.
\]

To find the sequence \( \mathbf{u}_i \), the equation [9] is used \( n - 2 \) times. Let

\[
g_i = \det \text{gram}(\mathbf{B}, \mathbf{u}_i, \mathbf{u}_{i-1})
\]

\[
= 1 - \kappa_i^2 - \kappa_{i-1}^2 - \beta_i^2 + 2\kappa_i\kappa_{i-1}\beta_i
\]  

[10]

To solve the structure \( g_i \) must be nonnegative for \( i = 2, \ldots, n \). This is a way to eliminate values \( \kappa_i \) which are solutions to the orientational constraints but which do not correspond to actual structures. This will be referred to as the gramin test.

Supposing, however, that \( g_i \geq 0 \) for all \( i \) then the \( \mathbf{u}_i \) can be found recursively, i.e., \( \mathbf{u}_i \) is found and then each subsequent \( \mathbf{u}_i \) found in terms of \( \mathbf{u}_{i-1} \). First define \( \mathbf{u}_1 \) and \( \mathbf{B} \) by

\[
\mathbf{u}_i = \sqrt{(1 - \kappa_i^2)} \mathbf{i} + \kappa_i \mathbf{k} \quad \mathbf{B} = \mathbf{k}
\]  

[11]

in the LAB frame. Next choose a sequence \( \varepsilon_i, \ i = 2, \ldots, n - 1 \) of values +1 or −1. Then define

\[
\mathbf{u}_i = \frac{1}{(1 - \kappa_{i-1}^2)} \left( (\beta_i - \kappa_i \kappa_{i-1}) \mathbf{u}_{i-1} 
\right.
\]

\[
+ \left. (\kappa_i - \beta_i \kappa_{i-1}) \mathbf{B} + \varepsilon_i \sqrt{g_{i-2}} \mathbf{u}_{i-1} \times \mathbf{k} \right)
\]  

[12]

This gives a structure compatible with the bond angle and bond orientation data and with \( \mathbf{B} \cdot (\mathbf{u}_{k-1} \times \mathbf{u}_k) = -\varepsilon_k \sqrt{g_k} \). Because there is a structure for each choice of sequence \( \varepsilon_i \), there are generally \( 2^{n-2} \) possible structures unless some of these chiralities are known.

**Remark.** For fixed sequences \( \beta_i \) and \( l_i \), the sequences of bond orientation cosines \( \kappa_i \) and chiralities \( \varepsilon_i \) give the same structure as the sequences \( -\kappa_i \) and \( \varepsilon_i \). The two structures differ by a 180° rotation around some axis perpendicular to \( \mathbf{B} \). 


6.3. Equations in the LAB Frame

It is useful to formulate the algorithm giving coordinates of the unit vectors \( u_i \) in the LAB frame \((i, j, k)\) where \( k = B \). Taking dots products of \([11], [12]\) with \( i, j, \) and \( k \) gives the following version of the algorithm

\[
u_1 \cdot i = \sqrt{1 - \kappa_i^2} \quad u_1 \cdot j = 0 \quad u_1 \cdot k = \kappa_1\]

\[
u_k \cdot i = A_k (u_{k-1} \cdot i) - B_k (u_{k-1} \cdot j)\]

\[
u_k \cdot j = B_k (u_{k-1} \cdot i) + A_k (u_{k-1} \cdot j)
\]

\[
u_k \cdot k = \kappa_k\]

where

\[A_k = \beta_k - \kappa_k \kappa_{k-1}\]

\[B_k = -\epsilon_k \sqrt{B_k} \frac{1}{1 - \kappa_k^2} - \kappa_{k-1}\]

and \( g_k \) is given by \([10]\).

Note that only the \( i \) and \( j \) coordinates depend on the chirality sequence. The \( k \) coordinate depends only on the bond orientation cosines \( \kappa_i \).

7. TORSION ANGLE

To study the structures obtained, it is convenient to have a formula for the torsion angle which can be computed from the sequences of bond angle and bond orientation cosines, \( \beta \), and \( \kappa_i \). For oriented structures there is the possibility of computing torsion angles using the direction \( B \) in place of a bond direction. The torsion angles involving the vector \( B \) are referred to as absolute torsion angles because the torsion is measured with one of the three directions the fixed direction \( B \). The torsion angles \( \phi, \psi \) in proteins are relative torsion angles because they depend only on the structure and not fixed \( B \) direction. By writing the relative torsion angle as a sum of absolute torsion angles, the \( \phi \) and \( \psi \) angles can be computed directly from the data \([6]\). In the remainder of this section the relationship between relative and absolute torsion angles is described.

Recall the definition of torsion angle. Given three vectors \( v_1, v_2 \) and \( v_3 \), usually bond vectors, the torsion angle \( \text{Tor} \ (v_1, v_2, v_3) \) is defined as the angle \( \tau \) such that rotation about the axis \( v_2 \) an angle of \( \tau \) sends

\[
\frac{v_1 \times v_2}{|v_1 \times v_2|} \quad \text{to} \quad \frac{v_2 \times v_3}{|v_2 \times v_3|}
\]

There is a useful formula for computing the torsion angle. Say that \( \theta \) is the argument of a vector \((x, y) \neq (0, 0)\) written \( \theta = \text{arg}(x, y) \) if \(-180^\circ < \theta \leq 180^\circ\)

\[
\cos \theta = \frac{x}{\sqrt{x^2 + y^2}} \quad \sin \theta = \frac{y}{\sqrt{x^2 + y^2}}
\]

(See Fig. 4.) Note that

\[\text{arg}(x, -y) = -\text{arg}(x, y)\]

and

\[\text{arg}(-x, -y) = \text{arg}(x, y) \pm 180^\circ\]

The following gives a formula for the torsion angle in terms of the argument,

\[
\text{Tor} \ (v_1, v_2, v_3) = \text{arg} \left( -|v_2|^2 v_1 \cdot v_3 + (v_2 \cdot v_1)(v_2 \cdot v_3), \frac{|v_2| v_1 \cdot (v_2 \times v_3)}{|v_2|} \right).
\]

The vectors \( v_i \) can be replaced by unit vectors \( u_i = v_i/|v_i| \) in the same direction and the torsion remains unchanged. The torsion angle has the following properties

\[
\text{Tor} \ (u_1, u_2, u_3)
\]

\[+ \text{Tor} \ (-u_3, u_2, u_4) = \text{Tor} \ (u_1, u_2, u_4)\]

\[\text{Tor} \ (-u_1, u_2, u_3) = \text{Tor} \ (u_1, u_2, u_3) \pm 180^\circ\]

Suppose that \( C_1—N—C_\alpha—C'-N' \) is a sequence of atoms along a protein backbone. The

![Figure 4](image-url)
\( \phi \) and \( \psi \) torsion angles are defined (according to IUPAC standards) as

\[
\phi = \text{Tor} \left( \overrightarrow{C_1N}, \overrightarrow{NC_\alpha}, \overrightarrow{C_\alpha C_1} \right) \\
\psi = \text{Tor} \left( \overrightarrow{NC_\alpha}, \overrightarrow{C_\alpha C_1}, \overrightarrow{C_1N_1} \right)
\]

For each sequence of three adjacent bond vectors \( \mathbf{u}_{i-1}, \mathbf{u}_i \), and \( \mathbf{u}_{i+1} \), two absolute torsion angles \( \text{Tor} (\mathbf{u}_{i-1}, \mathbf{u}_i, \mathbf{B}) \) and \( \text{Tor} (\mathbf{B}, \mathbf{u}_i, \mathbf{u}_{i+1}) \) can be defined. By \([17]\) torsion angle at \( \mathbf{u}_i \) is then the sum of these \( \pm 180^\circ \). The formula \([16]\) shows that \( \text{Tor} (\mathbf{u}_{i-1}, \mathbf{u}_i, \mathbf{B}) \) depends on \( \epsilon_i \) and \( \text{Tor} (\mathbf{B}, \mathbf{u}_i, \mathbf{u}_{i+1}) \) depends on \( \epsilon_{i+1} \). It is convenient to let \( \tau_i \) and \( \hat{\tau}_i \) be the values of these two angles, respectively, when \( \epsilon_i = \epsilon_{i+1} = 1 \).

\[
\tau_i = \arg(-\kappa_{i-1} + \beta_i \kappa_i, -\sqrt{3}) \\
\hat{\tau}_i = \arg(-\kappa_{i+1} + \beta_{i+1} \kappa_i, -\sqrt{3})
\]

Then using \([17]\), the relative torsion angle is

\[
\text{Tor} (\mathbf{u}_{i-1}, \mathbf{u}_i, \mathbf{u}_{i+1}) = \epsilon_i \tau_i + \epsilon_{i+1} \hat{\tau}_i \pm 180^\circ
\]

This formula is also discussed in \((8)\).

8. TENSORS

The sequence \( \kappa_i \) for unit vectors obtained from bond vectors along the backbone is determined experimentally from various solid state NMR observations. These experiments measure the values of certain quadratic tensors. Each tensor has an orthonormal set of principal axes. The one corresponding to the largest principal value will be called the major or unique axis. If two principal values are the same, the tensor is said to be symmetric.

As discussed in a companion paper \((9)\), the most commonly used tensors in determining orientational constraints are the chemical shift, dipolar, and quadrupolar. Values of each of these tensors are in terms of one or two bond orientation cosines. The strategy in computing structures is first to solve the equations to determine a unique sequence \( \kappa_k \) and then use the algorithm described above. However, since the tensor values do not distinguish between \( \mathbf{B} \) and \( -\mathbf{B} \), there is another set of \( \pm 1 \) ambiguities that arises in this stage of the analysis.

The dipolar and quadrupolar static tensors are approximated as symmetric tensors. Values of these tensors determined experimentally give the value of \( h = |3(\mathbf{B} \cdot \mathbf{u})^2 - 1| \), where \( \mathbf{u} \) is a unit vector in the direction of a covalent bond, which can be solved \( |\mathbf{B} \cdot \mathbf{u}|^2 = (1 \pm h)/3 \). This gives a unique value for \( |\mathbf{B} \cdot \mathbf{u}| \) if the tensor value is larger than half maximum, \( 1 < h \leq 2 \), otherwise two values are obtained. In any case, at most four values \( \pm \sqrt{(1 \pm h)/3} \) are obtained for the bond orientation cosine \( \mathbf{B} \cdot \mathbf{u} \).

The chemical shift tensor does not give the value of a bond orientation cosine, but rather an equation relating two bond orientation cosines. The principal axes of the chemical shift tensor are denoted \( \sigma_{11}, \sigma_{22}, \) and \( \sigma_{33} \), the principal values by \( \sigma_{11}, \sigma_{22}, \) and \( \sigma_{33} \), and the frame is called the principal axis frame \((\text{PAS})\). The tensor value, or observed chemical shift, for a given orientation is

\[
\sigma_{\text{obs}} = \sigma_{11} (\mathbf{B} \cdot \mathbf{\sigma}_{11})^2 + \sigma_{22} (\mathbf{B} \cdot \mathbf{\sigma}_{22})^2 + \sigma_{33} (\mathbf{B} \cdot \mathbf{\sigma}_{33})^2
\]

Since \( \mathbf{B} \) is a unit vector,

\[
(\mathbf{B} \cdot \mathbf{\sigma}_{11})^2 + (\mathbf{B} \cdot \mathbf{\sigma}_{22})^2 + (\mathbf{B} \cdot \mathbf{\sigma}_{33})^2 = 1
\]

The change of frame from the molecular frame to \((\text{PAS})\) has been discussed in detail in a companion paper \((9)\) and can be taken as given. Suppose that two bonds at the atom where the chemical shift is measured are \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \). When two principal axes are in the plane of the two bonds, as typically happens with \(^{15}\)N tensors in a peptide plane,

\[
\sigma_{11} = A \mathbf{u}_1 + B \mathbf{u}_2 \\
\sigma_{33} = C \mathbf{u}_1 + D \mathbf{u}_2
\]

for known constants \( A, B, C, \) and \( D \) determined from peptide plane geometry \((\text{see Section 9})\). By combining \([20]\)–\([22]\) the equation constraining the two bond orientation cosines \( \mathbf{u}_1 \cdot \mathbf{B} \) and \( \mathbf{u}_2 \cdot \mathbf{B} \) is obtained

\[
\sigma_{\text{obs}} - \sigma_{22} = (\sigma_{11} - \sigma_{22})(A \mathbf{u}_1 \cdot \mathbf{B} + B \mathbf{u}_2 \cdot \mathbf{B})^2 + (\sigma_{33} - \sigma_{22})(C \mathbf{u}_1 \cdot \mathbf{B} + D \mathbf{u}_2 \cdot \mathbf{B})^2
\]

Suppose also from dipolar or quadrupolar tensors the values

\[
h_1 = |3(\mathbf{B} \cdot \mathbf{u}_1)^2 - 1| \quad h_2 = |3(\mathbf{B} \cdot \mathbf{u}_2)^2 - 1|
\]
are known. Possibilities for the pair of bond angle cosines \((\mathbf{B} \cdot \mathbf{u}_1, \mathbf{B} \cdot \mathbf{u}_2)\) are then found by solving for them from values of \(\sigma_{\text{obs}}, h_1,\) and \(h_2,\) if at least two of these three values are known, and the number of solutions depends on which ones are known. In the worst case there are 16 possibilities, but in the best case, when all three of \(\sigma_{\text{obs}}, h_1,\) and \(h_2\) are known and one of \(h_1,\) and \(h_2\) is greater than half maximal, there are only two solutions for the pair \((\mathbf{B} \cdot \mathbf{u}_1, \mathbf{B} \cdot \mathbf{u}_2).\)

Similar techniques can be used along the backbone using chemical shift, dipolar, quadrupolar data in conjunction with peptide plane geometry. The sequences \(\kappa_i\) can also be limited to ones for which all \(g_i\) are nonnegative. Thus with enough tensor observations a unique sequence \(\kappa = \mathbf{B} \cdot \mathbf{u}\) can be obtained. This has been achieved for the gramicidin A structure discussed in a companion paper (9) and below.

In summary, the experimental values of quadratic tensors in combination with the known covalent geometry gives a unique or near unique sequence \(\kappa_i\) of bond orientation cosines. However, each one of these sequences gives generally \(2^{n-2}\) possible structures from the chirality ambiguity. Despite this very large solution set (two orientations for each peptide plane), all of these solutions lead to the same fold for the gramicidin structure, the same hydrogen bonding pattern, the same helical sense, and the same residues per turn. The reason for this is that the two orientations result in virtually identical positions for the \(\text{C}_\alpha\) carbons relative to the helix axis and hence the secondary and tertiary structure is essentially unaffected by these chirality ambiguities.

9. PEPTIDE PLANE GEOMETRY

It has been mentioned that peptide plane geometry is an important element in obtaining structure from orientational constraints. This section describes the geometry and how it is used.

Tables 1 and 2 and Fig. 5 define unit bond vectors for the peptide plane and the tetrahedral \(\alpha\) carbons and give approximate angles between pairs of these bond vectors as estimated from (3, 4) and other sources. Values are given to three significant figures but this degree of accuracy is not implied. For comparison, values of the cosines are also given for perfect trigonal and tetrahedral geometry, i.e., 120° interior angles in the peptide plane, and \(\arccos(-1/3)\) interior angles at a tetrahedral \(\alpha\) carbon.

### Table 1. Unit Vectors Corresponding to Bond Directions Along a Segment of a Protein Backbone

<table>
<thead>
<tr>
<th>Unit v.</th>
<th>Bond dir.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(u_1)</td>
<td>(\text{C}_\alpha - \text{C}_1)</td>
</tr>
<tr>
<td>(u_2)</td>
<td>(\text{C}_1 - \text{N})</td>
</tr>
<tr>
<td>(u_3)</td>
<td>(\text{N} - \text{C}_\alpha)</td>
</tr>
<tr>
<td>(u_4)</td>
<td>(\text{N} - \text{H})</td>
</tr>
<tr>
<td>(u_5)</td>
<td>(\text{C}_\alpha - \text{C}_1)</td>
</tr>
<tr>
<td>(u_6)</td>
<td>(\text{C}_1 - \text{H})</td>
</tr>
</tbody>
</table>

### Table 2. Peptide Plane Geometry Given for Perfect Trigonal and Tetrahedral Geometry, and for Experimentally Determined Geometry

<table>
<thead>
<tr>
<th>B.a. cosine</th>
<th>Trig./tetr.</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>(u_1 \cdot u_2)</td>
<td>1/2</td>
<td>(\cos 64° \approx 0.438)</td>
</tr>
<tr>
<td>(u_2 \cdot u_3)</td>
<td>1/2</td>
<td>(\cos 58° \approx 0.530)</td>
</tr>
<tr>
<td>(u_3 \cdot u_4)</td>
<td>1/2</td>
<td>(\cos 59° \approx 0.515)</td>
</tr>
<tr>
<td>(u_1 \cdot u_3)</td>
<td>(-1/2)</td>
<td>(-\cos 63° \approx -0.454)</td>
</tr>
<tr>
<td>(u_2 \cdot u_3)</td>
<td>1/3</td>
<td>(\cos 69° \approx 0.358)</td>
</tr>
<tr>
<td>(u_3 \cdot u_4)</td>
<td>1/3</td>
<td>(\cos 71° \approx 0.325)</td>
</tr>
<tr>
<td>(u_1 \cdot u_4)</td>
<td>(-1/3)</td>
<td>(-\cos 71° \approx -0.325)</td>
</tr>
</tbody>
</table>

These dot products can be used to write all vectors in the peptide plane as a linear combination of \(u_2\) and \(u_3\) and this is useful in computing the relationship between various tensor values as indicated in Section 8. For perfect trigonal geometry,

\[ u_1 = u_2 - u_3 = u_3 \]

and for the experimental values given,

\[ u_1 = 0.98u_2 - 1.0u_3 \]
\[ u_3 = 1.0u_2 - 0.99u_3 \]

[25]

Using these equations, the values of \(\mathbf{B} \cdot \mathbf{u}_1,\) and \(\mathbf{B} \cdot \mathbf{u}_3\) can be determined from the values of \(\mathbf{B} \cdot \mathbf{u}_2\) and \(\mathbf{B} \cdot \mathbf{u}_4,\)

Figure 5 Notation for unit bonds along a protein backbone.
Similarly, the vector \( \mathbf{u}_4 \) at the \( \alpha \) carbon can be written as a linear combination of \( \mathbf{u}_3 \), \( \mathbf{u}_4 \), and \( \mathbf{u}_3 \times \mathbf{u}_4 \). For perfect tetrahedral geometry,

\[
\mathbf{u}_4' = \frac{1}{2} (\mathbf{u}_3 - \mathbf{u}_4) + \frac{\sqrt{3}}{2} \mathbf{u}_3 \times \mathbf{u}_4
\]

\[
\approx 0.5 (\mathbf{u}_3 - \mathbf{u}_4) + 0.866 \mathbf{u}_3 \times \mathbf{u}_4
\]

where the “plus” sign occurs at an \( L \) residue and the “minus” sign at a \( D \) residue. For the experimental values,

\[
\mathbf{u}_4' = 0.51 (\mathbf{u}_3 - \mathbf{u}_4) + 0.88 \mathbf{u}_3 \times \mathbf{u}_4 \quad [26]
\]

Using this equation, the value of \( \mathbf{B} \cdot \mathbf{u}_4' \) can be determined from the values of \( \mathbf{B} \cdot \mathbf{u}_3 \) and \( \mathbf{B} \cdot \mathbf{u}_4 \). Since \( \mathbf{u}_4' \) is not in the plane of \( \mathbf{u}_3 \) and \( \mathbf{u}_4 \), \( \mathbf{B} \cdot \mathbf{u}_4' \) is referred to as an out-of-plane constraint.

Bond lengths are also required to determine the structure from orientational constraints. The only approximate distances (in Angstroms) used here will be

\[
l_1 = \overline{C_aC_1} = 1.53
\]

\[
l_2 = \overline{C_1N} = 1.34
\]

\[
l_3 = \overline{NC_a} = 1.45
\quad [27]
\]

From these distances, the virtual bond vector \( \mathbf{v} \) from \( C_a \) to subsequent \( C_a \) can be found in terms of \( \mathbf{u}_2 \) and \( \mathbf{u}_3' \),

\[
\mathbf{v} = l_1 \mathbf{u}_1 + l_2 \mathbf{u}_2 + l_3 \mathbf{u}_3 = 4.3 \mathbf{u}_2 - 3.0 \mathbf{u}_3' \quad [28]
\]

For comparison, with all \( l_i \) equal to 1.5 and with perfect trigonal and tetrahedral geometry, \( \mathbf{v} = 4.5 \mathbf{u}_2 - 3 \mathbf{u}_3' \).

### 10. EXAMPLES

#### 10.1. A Diplane

The following example based on (10) illustrates how the above ideas can be used. It shows how a \( C_a^{1-2}H \) quadrupolar observation can be used to determine a chirality and reduce the number of possible bond orientation cosines.

Consider a dipeptide considered as a five point structure consisting of atoms \( \text{C}_1, \text{N}, \text{C}_a, \text{C}_1' \), and \( \text{N}' \) along the backbone of a protein. Given \( \text{NC}_1 \) and NH dipolar observations and \( ^{15}\text{N} \) chemical shift data at each \(^{15}\text{N} \), suppose that using methods described in Section 8 the orientation cosines of \( \text{C}_1\text{N} \) and \( \overline{\text{NC}_a} \) in the first peptide plane can be reduced to the possibilities \( (\kappa_1, \kappa_2) = (0.4, -0.55) \) and the orientations cosines of \( \text{C}_a\text{C}_1 \) and \( \overline{\text{C}_1'\text{N}'} \) in the second peptide plane can be reduced to the possibilities \( (\kappa_3, \kappa_4) = (-0.15, 0.75) \). Since the two sequences of orientation cosines \( \kappa_i \) and \( -\kappa_i \) give the same set of structures, there are only two sequences of orientation cosines to consider, \( g_1 \) and \( g_2 \) as listed in Table 3. The values \( g_i \) are all non-negative so neither set of orientations can be rejected using the gramian test. Each sequence \( \kappa_i \) corresponds to eight different structures given by eight sequences \( \varepsilon_2, \varepsilon_3, \varepsilon_4 \) of chiralities. Thus there are 16 possible structures which can be derived from these orientational constraints.

The number of possibilities can be reduced to a unique sequence \( \kappa_i \) and just two chiralities \( \varepsilon_2, \varepsilon_4 \) for four structures, by the use of an out-of-plane constraint at the \( \text{C}_a \). Let \( \mathbf{u}_4 \) be a unit vector in the direction of \( \overline{\text{C}_a^{1-2}H} \) and suppose the residue is an \( L \) residue and that measurement of a quadrupolar splitting gives

\[
|3(\mathbf{B} \cdot \mathbf{u}_4')^2 - 1| = 1.8 \quad [29]
\]

By Eq. [26] for tetrahedral geometry at an \( L \) residue,

\[
\mathbf{B} \cdot \mathbf{u}_4' = 0.51 (\kappa_2 - \kappa_3) + 0.88 \varepsilon_3 \sqrt{g_3} \quad [30]
\]

Among the four possible choices of \( g_1 \) or \( g_2 \) and \( \varepsilon_3 = \pm 1 \), the only one satisfying [30] and [29] is the sequence of orientations \( g_2 \) with \( \varepsilon_3 = -1 \). Now having eliminated \( g_1 \) and found \( \varepsilon_3 \) the only possible choices are the chiralities \( \varepsilon_2 \) and \( \varepsilon_4 \) for a total of four possible structures for the diplane. Since, as can be seen from Table 3, the determinants \( g_i \) are small for \( i = 2 \) and \( i = 4 \), the basic recursion formula [12] shows that these structures are fairly close to each other.

To learn more about the structures, use [18] to compute the absolute torsion angles

| Table 3. Two Sets of Orientational Constraints for a Five Point Structure Consisting of Atoms \( \text{C}_1—\text{N}—\text{C}_a—\text{C}_1'—\text{N}' \) Along a Protein Backbone |
|---|---|---|---|
| \( i \) | \( \beta_i \) | \( \kappa_i \) | \( g_i \) | \( \kappa_i \) | \( g_i \) |
| 1 | 0.550 | -0.550 | 0.023 | -0.550 | 0.023 |
| 2 | 0.358 | -0.150 | 0.606 | 0.150 | 0.488 |
| 3 | 0.438 | 0.750 | 0.125 | -0.750 | 0.125 |
\[ \tau_i = \text{Tor}(u_{i-1}, u_i, B) \] and \[ \hat{\tau}_i = \text{Tor}(B, u_i, u_{i+1}) \] computed for \( \varepsilon_i = +1 \) for \( i = 2, 3, 4 \). Get

\[ \tau_2 = -167^\circ \quad \hat{\tau}_2 = -116^\circ \]
\[ \tau_3 = -49.1^\circ \quad \hat{\tau}_3 = -23.4^\circ. \]  

Now from [19] the (\( \phi, \psi \)) torsion angles for these structures are found as indicated in Table 4.

For the gramicidin structure, the above technique applied to each diplane was used to determine the chirality \( \varepsilon_i \) at each \( C_\alpha \), and to help determine a unique sequence of orientations \( \kappa_i \) for the backbone. The chiralities for adjacent bonds in each peptide plane were the remaining unknowns and all choices gave the same fold for the backbone. These chiralities were finally determined in the refinement.

10.2. Length of the Gramicidin Backbone

Although the initial structures of gramicidin A depend on the choice of the sequence \( \varepsilon_i \), the length of the projection of the backbone onto the \( B = k \) direction does not. This follows from [14] since the \( k \) coordinate is independent of the chiralities. In this section the length of a 14 peptide plane segment of the backbone is computed and shown to be about 9.5, consistent with the head-to-head dimer structure in which a single helix should span only half of the lipid bilayer (Fig. 6).

In Table 5 the values of \( h(B \cdot nc) \) and \( h(B \cdot nh) \) are listed for 14 residues, where \( nh \) is the unit vector in the direction \( N-H \) and \( nc \) is the unit vector in the direction \( N-C_1 \) and \( h(z) = |3z^2 - 1| \) is the absolute value of the spherical harmonic. These values are computed directly from dipolar splitting observations. For given \( h_1 \), there are four possible solutions \( z \) for \( h_1 = h(z) \), \( z = \delta_1 \sqrt{(1 + \delta_2 h_1)/3} \) where \( \delta_1 \) and \( \delta_2 \) can be \( \pm 1 \). For the NH splitting \( \delta_2 = 1 \) at all residues since \( h \) is greater than half maximal (see Section 8). The values \( \delta_1 \) are found to alternate between 1 and \(-1 \). For the NC splittings, both \( \delta_1 \) and \( \delta_2 \) alternate.

### Table 4. \( \phi \) and \( \psi \) Angles for the Example Diplane Structure Given as a Function of the Two Unknown Chiralities

<table>
<thead>
<tr>
<th>( \varepsilon_2 )</th>
<th>( \varepsilon_4 )</th>
<th>( \phi )</th>
<th>( \psi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>129</td>
<td>-154</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>129</td>
<td>-107</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>104</td>
<td>-154</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>104</td>
<td>-107</td>
</tr>
</tbody>
</table>

The length of the projection of 14 peptide plane segment of the backbone of gramicidin A can be computed directly from orientational constraints.

The length of the projection onto \( k \) of this 14 peptide plane segment is the sum of the projections of the 14 virtual bond vectors. By [28] this is computed to be \( l = -3.0 s_{nh} - 4.3 s_{nc} \) where \( s_{nh} \) is the sum of the values in the third column of Table 5 and \( s_{nc} \) is the sum of the values in the fifth column. Summing gives \( l = 10.4 \) The maximum value for this, within experimental error, is found to be about 13.

### Table 5. NH and NC Dipolar Splitting Data for 14 Peptide Planes in the Gramicidin A Backbone. The Function \( h(z) = |3z^2 - 1| \) is the Absolute Value of the Spherical Harmonic

<table>
<thead>
<tr>
<th>Res. n.</th>
<th>( h(B \cdot nh) )</th>
<th>( B \cdot nh )</th>
<th>( h(B \cdot nc) )</th>
<th>( B \cdot nc )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.56</td>
<td>0.924</td>
<td>0.717</td>
<td>0.756</td>
</tr>
<tr>
<td>2</td>
<td>1.93</td>
<td>-0.988</td>
<td>0.528</td>
<td>-0.713</td>
</tr>
<tr>
<td>3</td>
<td>1.52</td>
<td>0.917</td>
<td>0.646</td>
<td>0.740</td>
</tr>
<tr>
<td>4</td>
<td>1.83</td>
<td>-0.971</td>
<td>0.450</td>
<td>-0.695</td>
</tr>
<tr>
<td>5</td>
<td>1.61</td>
<td>0.933</td>
<td>0.493</td>
<td>0.705</td>
</tr>
<tr>
<td>6</td>
<td>1.95</td>
<td>-0.991</td>
<td>0.409</td>
<td>-0.685</td>
</tr>
<tr>
<td>7</td>
<td>1.58</td>
<td>0.927</td>
<td>0.553</td>
<td>0.719</td>
</tr>
<tr>
<td>8</td>
<td>1.83</td>
<td>-0.971</td>
<td>0.383</td>
<td>-0.679</td>
</tr>
<tr>
<td>9</td>
<td>1.29</td>
<td>0.873</td>
<td>0.551</td>
<td>0.719</td>
</tr>
<tr>
<td>10</td>
<td>1.85</td>
<td>-0.975</td>
<td>0.287</td>
<td>-0.655</td>
</tr>
<tr>
<td>11</td>
<td>1.43</td>
<td>0.900</td>
<td>0.613</td>
<td>0.733</td>
</tr>
<tr>
<td>12</td>
<td>1.87</td>
<td>-0.978</td>
<td>0.357</td>
<td>-0.672</td>
</tr>
<tr>
<td>13</td>
<td>1.27</td>
<td>0.869</td>
<td>0.517</td>
<td>0.711</td>
</tr>
<tr>
<td>14</td>
<td>1.90</td>
<td>-0.983</td>
<td>0.399</td>
<td>-0.683</td>
</tr>
</tbody>
</table>
11. CONCLUSION

A new, powerful approach for structure determination has been demonstrated. The precision and accuracy of these constraints have been shown by the initial structural characterization of the gramicidin structure in which the hydrogen bond distances between turns of the helix were within 0.5 Å rmsd despite having no distance constraints between turns of the helix. In other words, the small angular errors for each orientational constraint summed over seven residues (6.5 residues per turn) leads to just a 0.5 Å error. This is the advantage of using absolute constraints, those that constrain the molecule to the laboratory frame. However, this is not to say that distance or torsional constraints are not important. Indeed, the monomers–monomer junction of gramicidin could not be described with orientational constraints, but required distance constraints. Furthermore, torsional constraints may prove to be fundamentally important to sort out the occasional structural ambiguity, such as a chirality. However, it should be noted that the gramicidin structural refinement which included constraints for the sidechains as well as the backbone and the Chemistry at Harvard Macromolecular Mechanics (CHARMm) potential energy resulted in a unique solution for most of the backbone chiralities.

ACKNOWLEDGMENTS

The authors thank Jeff Denny for help creating the figures for this paper.

REFERENCES