Journal of Magnetic Resonance **147,** 9–16 (2000)
doi:10.1006/jmre.2000.2193, available online at http://www.idealibrary.com on IDEAL®

Atomic Refinement Using Orientational Restraints from Solid-State NMR

Richard Bertram,*[,1] J. R. Quine,†[,]‡ Michael S. Chapman,*[,]§ and Timothy A. Cross†[,]§

Institute of Molecular Biophysics, †National High Magnetic Field Laboratory, ‡Department of Mathematics, and §Department of Chemistry, Florida State University, Tallahassee, Florida 32306

Received June 12, 2000; revised August 16, 2000

We describe a procedure for using orientational restraints from solid-state NMR in the atomic refinement of molecular structures. Minimization of an energy function can be performed through either (or both) least-squares minimization or molecular dynamics employing simulated annealing. The energy, or penalty, function consists of terms penalizing deviation from "ideal" parameters such as covalent bond lengths and terms penalizing deviation from orientational data. Thus, the refinement strives to produce a good fit to orientational data while maintaining good stereochemistry. The software is in the form of a module for the popular refinement package CNS and is several orders of magnitude faster than previous software for refinement with orientational data. The short computer time required for refinement removes one of the difficulties in protein structure determination with solid-state NMR.

1. INTRODUCTION

Solid-state NMR is a developing technique for the determination of three-dimensional molecular structures, complementing X-ray crystallography and solution NMR. Orientational restraints derived from solid-state NMR were used recently to determine the structure of the polypeptide ion channel gramicidin A ((*1*), PDB Accession No. 1MAG), the first complete structure determination with this technique. Two characteristics of solid-state NMR, the extreme precision of orientational measurements (*2*) and the requirement for uniform molecular alignment relative to an external magnetic field, make it ideal for the study of membrane proteins, which are difficult to study with other methods (*3*).

Orientational measurements from solid-state NMR typically consist of anisotropic dipolar and quadrupolar couplings, which describe the angles made between specific atomic bonds and the magnetic field direction, and chemical shifts, which describe the orientation of chemical shift tensors with respect to the field. Orientational ambiguities are inherent in this information, but most of these can be resolved with complementary spin interaction data (*4*) or stereochemical constraints and restraints.

In the structural determination of gramicidin A, Ketchem *et al.* developed a computer program to minimize an energy or penalty function consisting of stereochemical terms (based on CHARMM force fields (*5*)) and penalty terms reflecting the deviation between solid-state NMR observables and those calculated from the model (*1*). For small structural improvements, this program performed a series of atomic moves, each of which reduced the potential energy. A simulated annealing approach was used for larger structural changes, allowing atomic coordinate moves that increased the potential energy. As the computational "temperature" was lowered, such energetically unfavorable moves became less likely and there was convergence to a final minimized structure. This annealing approach is useful in global optimization as a means of escaping local minima (*6*) and is particularly useful in atomic refinement, where the potential energy landscape is characterized by many local minima (*7*).

Atomic refinement using data from crystallography or solution NMR is often performed with the computer software package CNS (*8*). This package allows the user to specify the types of stereochemical restraints to be applied (e.g., covalent bond lengths or angles) and the types and values of experimental observables (e.g., diffraction structure amplitudes or NOE measurements). The user then specifies the type of minimization that is to be performed: either least-squares minimization for small structural changes or simulated annealing for larger searches through conformation space. Unlike the Ketchem refinement program, the least-squares minimization is implemented using a gradient descent algorithm and the annealing is implemented through the numerical solution of molecular dynamics equations. This implementation is potentially much faster than the Monte-Carlo-like approach employed by Ketchem *et al.*

We have developed computer software that permits refinement using solid-state NMR orientational data with CNS. This software has two main components. The first is a module that is linked to CNS and uses the model coordinates and the data to compute chemical shift energy terms and their derivatives. The second component is a modification of a previously existing module that computes dipolar/quadrupolar coupling energy

[1] To whom correspondence should be addressed. Fax: (850) 561-1406.

1090-7807/00 $35.00

terms and their derivatives. Atomic refinement of a test molecule, a monomer of gramicidin A, shows that the new implementation is several orders of magnitude faster than the implementation of Ketchem *et al*. Equally important, since the new implementation is designed to work with the popular CNS package, the process of refinement using solid-state NMR orientational data is standardized with other types of data used for structure determination. Thus, not only is it simple to do refinement using solid-state NMR orientational data with the new implementation, but also this data can be supplemented by other data types.

In what follows, we construct energy terms for the orientational data typically obtained in solid-state NMR experiments. We then compute derivatives of these energies with respect to atomic coordinates, for use with the CNS minimization algorithms. Finally, we use a perturbed gramicidin A monomer structure to illustrate how refinement with the new software improves both stereochemistry and agreement with orientational data. In the Appendix, we give some details on running the refinement with CNS.

2. SOLID-STATE NMR ENERGY TERMS

The total potential energy minimized during atomic refinement can be written as

$$E = E_{\text{chem}} + E_{\text{data}}. \quad [1]$$

The stereochemical energy, E_{chem}, includes terms that penalize deviation of model parameters, such as covalent bond lengths, from "ideal" parameter values, inferred from investigations of small molecules. For nonbonded interactions, Lennard–Jones and Coulomb force fields are typically used to restrain distances between atom pairs (*5*). The second component of the potential energy, E_{data}, is often a quadratic penalty function for deviation of the model from the observed experimental data:

$$E_{\text{data}} = \sum_i w_i (s_c - s_o)^2. \quad [2]$$

Here, s_o is an experimental observation and s_c is an equivalent quantity calculated from the model. The observation s_o might be a crystallographic structure amplitude or an NMR NOE atom-pair distance, for example. The components of this energy term are scaled by the weighting factors w_i, which are typically different for the different data types employed in the refinement.

For refinement using solid-state NMR data, E_{data} is composed of terms for dipolar and quadrupolar couplings and for anisotropic chemical shifts:

$$E_{\text{data}} = \sum_i w_{\text{dq},i} E_{\text{dq},i} + \sum_j w_{\text{cs},j} E_{\text{cs},j}. \quad [3]$$

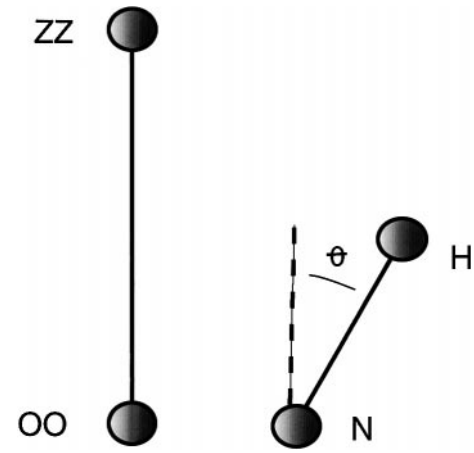

FIG. 1. Illustration of the angle θ used in the calculation of ν_c for an N–H dipolar coupling. N and H are backbone amide atoms, and OO and ZZ are pseudo-atoms used to define the external magnetic field direction.

The summations are over all couplings and chemical shifts used in the refinement. These energy terms penalize deviations of couplings and chemical shifts calculated from the model (ν_c and σ_c, respectively) from experimentally observed couplings and chemical shifts (ν_o and σ_o, respectively):

$$E_{\text{dq}} = \begin{cases} (|\nu_c| - \nu_o)^2 & \text{if } \nu_o \le K \\ (\nu_c - \nu_o)^2 & \text{if } K < \nu_o \le 2K \end{cases} \quad [4]$$

and

$$E_{\text{cs}} = (\sigma_c - \sigma_o)^2. \quad [5]$$

The positive constant K is either calculated or experimentally defined and represents $\nu_\parallel$ for the dipolar couplings and 0.75 QCC for the quadrupolar couplings, where QCC is the quadrupolar coupling constant (*2*). The N–C dipolar, N–H dipolar, quadrupolar, ^{15}N, and ^{13}C chemical shift energy terms can be weighted separately or in smaller subsets using $w_{\text{dq},i}$ and $w_{\text{cs},j}$ (Eq. [3]).

The calculated ν_c is defined as

$$\nu_c = K(3\cos^2\theta - 1), \quad [6]$$

where θ is the angle made between the magnetic field direction and appropriate covalent bonds (Fig. 1). The squaring of the cosine term introduces a twofold orientational ambiguity. A second, conditional, twofold ambiguity is evident by noting that ν_c ranges over $[-K, 2K]$, while experimental measurements of couplings (ν_o) cannot distinguish the sign. For this reason, the absolute value of ν_c is used in Eq. [4] if $\nu_o \le K$, introducing a second twofold orientational ambiguity in this case. (In the refinement of Ketchem *et al.*, the sign of ν_c was determined prior to refinement based on stereochemical con-

siderations; in this CNS implementation the sign is determined by simultaneous minimization of E_{data} and E_{chem}.) A similar energy function was used recently in refinement with solution NMR dipolar coupling data (*9*).

3. CALCULATION OF DERIVATIVES

The gradient descent and simulated annealing minimization techniques require expressions for the derivatives of all energy terms with respect to atomic coordinates. Derivatives of E_{chem} are contained within the CNS software package, as are derivatives of E_{data} for most crystallographic and solution NMR applications. Similarly, derivatives of E_{dq} and E_{cs} must be calculated and included in the modules for solid-state NMR refinement. Derivatives of dipolar coupling energy for solution NMR have been derived previously (*10*) and implemented in the CNS package. However, these require modification for solid-state NMR dipolar/quadrupolar couplings due to the absolute value in Eq. [4].

The derivative of E_{dq} with respect to x (where x is any of the nine coordinates of the three atoms) is

$$\frac{\partial E_{dq}}{\partial x} = \begin{cases} \text{sign}[3\cos^2\theta - 1]2(|\nu_c| - \nu_o)\dfrac{\partial \nu_c}{\partial x} & \text{if } \nu_o \le K \\ 2(\nu_c - \nu_o)\dfrac{\partial \nu_c}{\partial x} & \text{if } K < \nu_o \le 2K, \end{cases} \quad [7]$$

where sign$[3\cos^2\theta - 1] = \pm 1$, depending on whether $3\cos^2\theta - 1$ is positive or negative. For E_{cs},

$$\frac{\partial E_{cs}}{\partial x} = 2(\sigma_c - \sigma_o)\frac{\partial \sigma_c}{\partial x}. \quad [8]$$

3.1. Derivatives of Dipolar and Quadrupolar Couplings

The energy derivative Eq. [7] requires $\partial \nu_c/\partial x$, which by the chain rule is

$$\frac{\partial \nu_c}{\partial x} = 6K\cos\theta\frac{\partial \cos\theta}{\partial x}. \quad [9]$$

As an example of how this is computed, we consider the specific instance of an N–H dipolar coupling. In this case, θ is the angle made between the N–H bond and the magnetic field direction (Fig. 1). To define the magnetic field direction, a coordinate frame is constructed from pseudo-atoms. This frame is orthogonal, with z direction parallel to the magnetic field (Fig. 1). Then,

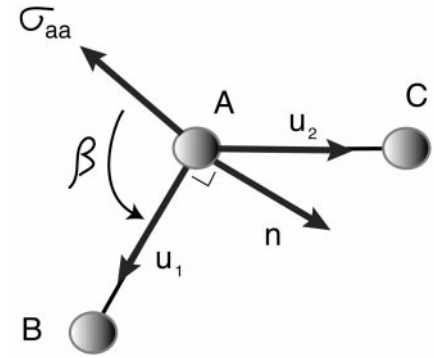

FIG. 2. Illustration of the molecular and principal axis frames. The vector $\boldsymbol{\sigma}_{aa}$ lies in the plane of the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$. The angle from $\boldsymbol{\sigma}_{aa}$ to $\mathbf{u}_1$ is β. The angle is measured counterclockwise about $\mathbf{b} = \mathbf{u}_1 \times \mathbf{n}$ so that, in the figure, β is positive.

$$\cos\theta = \frac{\mathbf{OZ}\cdot\mathbf{NH}}{|\mathbf{OZ}||\mathbf{NH}|}, \quad [10]$$

where **OZ** is the vector from the origin, OO, of the synthetic frame to the *ZZ* pseudo-atom, and **NH** is the vector from the N to the H atom. Equations for $\partial/\partial x \cos\theta$ are derived from Eq. [10] by displacing the N and H atoms in the x direction. This approach was used in (*10*) for differentiating the dipolar coupling energy term in solution NMR applications and is part of the CNS software package. The only changes required for solid-state NMR applications are the redefinition of E_{dq} (Eq. [4]) and the form of its derivative (Eq. [7]). In addition, there is no rhombicity component in ν_c (Eq. [6]).

3.2. Derivatives of Chemical Shift Tensors

3.2.1. The molecular frame. The principal axis frame of the chemical shift tensor is described in terms of a molecular frame, and the molecular frame is described in terms of two atoms bonded to a third. Suppose that the atoms are labeled A, B, and C with A bonded to B and C (Fig. 2). Suppose the atoms have coordinates given by

$$P_A = (x_A, y_A, z_A)$$
$$P_B = (x_B, y_B, z_B)$$
$$P_C = (x_C, y_C, z_C).$$

Define unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ by

$$\mathbf{u}_1 = \frac{P_B - P_A}{|P_B - P_A|}$$
$$\mathbf{u}_2 = \frac{P_C - P_A}{|P_C - P_A|}.$$

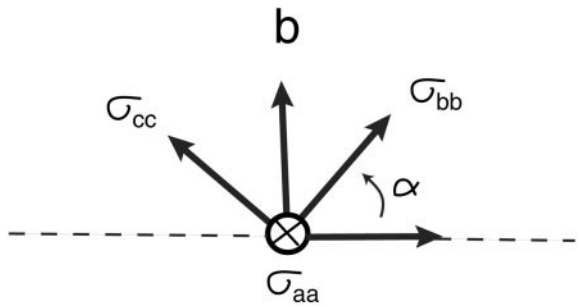

FIG. 3. The angle α gives the position $\boldsymbol{\sigma}_{bb}$ with respect to the plane of the atoms A, B, and C. In the figure, the dotted line represents the plane spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$ and $\boldsymbol{\sigma}_{aa}$ points toward the reader. The vector in the plane is $\boldsymbol{\sigma}_{bb}$ for $\alpha = 0$, i.e., $\mathbf{b} \times \boldsymbol{\sigma}_{aa}$. The angle α is measured counterclockwise about $\boldsymbol{\sigma}_{aa}$ so that, in the figure, α is positive.

Define a binormal vector **b** by

$$\mathbf{b} = \frac{\mathbf{u}_1 \times \mathbf{u}_2}{|\mathbf{u}_1 \times \mathbf{u}_2|} \qquad [11]$$

and a right-handed orthonormal molecular frame F by $F = (\mathbf{u}_1, \mathbf{n}, \mathbf{b})$, where **n** is a normal vector defined by $\mathbf{n} = \mathbf{b} \times \mathbf{u}_1$. The vectors can be thought of as column vectors and F as a 3×3 matrix.

Note that $\mathbf{u}_1$ and **n** are in the plane of the three atoms A, B, and C.

3.2.2. The principal axis frame. The principal axis frame for the chemical shift tensor is usually denoted

$$[\mathrm{PAS}] = (\boldsymbol{\sigma}_{aa}, \boldsymbol{\sigma}_{bb}, \boldsymbol{\sigma}_{cc}),$$

where $\boldsymbol{\sigma}_{aa}$, $\boldsymbol{\sigma}_{bb}$, and $\boldsymbol{\sigma}_{cc}$ are as defined in (*11*). The frame [PAS] in relation to the molecular frame is specified in terms of two angles α and β. Let

$$M_\beta = \begin{pmatrix} \cos\beta & \sin\beta & 0 \\ -\sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$N_\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix},$$

then

$$[\mathrm{PAS}] = F M_\beta N_\alpha. \qquad [12]$$

The angle α is the angle made between the plane spanned by $\mathbf{u}_1$ and **n** and that spanned by $\boldsymbol{\sigma}_{aa}$ and $\boldsymbol{\sigma}_{bb}$. The vector $\boldsymbol{\sigma}_{aa}$ is always in the $\mathbf{u}_1 - \mathbf{n}$ plane, and $\alpha > 0$ if $\boldsymbol{\sigma}_{bb}$ is on the same side of this plane as **b** (Fig. 3). In most cases considered in (*1*), $\boldsymbol{\sigma}_{aa}$ and $\boldsymbol{\sigma}_{bb}$ are both in the $\mathbf{u}_1 - \mathbf{n}$ plane and so $\alpha = 0$ and N_α is the identity. The angle β indicates the angle from the principal axis $\boldsymbol{\sigma}_{aa}$ to the A to B bond vector, measured counterclockwise around **b** (Fig. 2).

The assumption will be made that α and β remain constant as the coordinates of the atoms A, B, and C are varied. Since the chemical shift depends on the electron cloud surrounding the atoms, this is only approximately true, but the assumption will be made that the configuration of the atoms is not varied too much from standard peptide bond geometry. Similarly, the principal values of the chemical shift tensor will be assumed constant.

3.2.3. The chemical shift. The principal values of the chemical shift tensor are denoted σ_{aa}, σ_{bb}, and σ_{cc} and the chemical shift is given by

$$\sigma_c = \sigma_{aa}(\boldsymbol{\sigma}_{aa} \cdot \mathbf{B})^2 + \sigma_{bb}(\boldsymbol{\sigma}_{bb} \cdot \mathbf{B})^2 + \sigma_{cc}(\boldsymbol{\sigma}_{cc} \cdot \mathbf{B})^2,$$

where $\mathbf{B} = \mathbf{OZ}/|\mathbf{OZ}|$ is a unit direction for the magnetic field (Fig. 1). It is convenient to write this in matrix form as

$$\sigma_c = \mathbf{B}^t[\mathrm{PAS}]D[\mathrm{PAS}]^t\mathbf{B}, \qquad [13]$$

where D is the diagonal matrix

$$D = \begin{pmatrix} \sigma_{aa} & 0 & 0 \\ 0 & \sigma_{bb} & 0 \\ 0 & 0 & \sigma_{cc} \end{pmatrix}.$$

3.2.4. Derivative of the chemical shift. The goal is to find $\partial\sigma_c/\partial x$ for the atoms A, B, or C. From Eq. [13] and assuming D to be constant,

$$\frac{\partial \sigma_c}{\partial x} = 2\mathbf{B}^t \frac{\partial[\mathrm{PAS}]}{\partial x} D[\mathrm{PAS}]^t\mathbf{B}. \qquad [14]$$

By [12], assuming that M_β and N_α are constant,

$$\frac{\partial[\mathrm{PAS}]}{\partial x} = \frac{\partial F}{\partial x} M_\beta N_\alpha, \qquad [15]$$

and the problem is to compute

$$\frac{\partial F}{\partial x} = \left(\frac{\partial \mathbf{u}_1}{\partial x}, \frac{\partial \mathbf{n}}{\partial x}, \frac{\partial \mathbf{b}}{\partial x}\right).$$

3.2.5. Computation of $\partial F/\partial x$. Since F is an orthogonal matrix, $F^tF = I$, it follows by differentiating this equation that $F^t(\partial F/\partial x)$ is a skew symmetric matrix, which will be denoted by S. Thus

$$F^{\mathrm{t}}\frac{\partial F}{\partial x}=S=\begin{pmatrix}0 & p & q\\ -p & 0 & r\\ -q & -r & 0\end{pmatrix}, \qquad [16]$$

where

$$p=\mathbf{u}_1\cdot\frac{\partial \mathbf{n}}{\partial x}$$

$$q=\mathbf{u}_1\cdot\frac{\partial \mathbf{b}}{\partial x}$$

$$r=\mathbf{n}\cdot\frac{\partial \mathbf{b}}{\partial x}. \qquad [17]$$

3.2.6. Computation of p, q, and r. Since $\mathbf{n}=\mathbf{b}\times\mathbf{u}_1$, it follows from Eq. [17] that

$$p=\mathbf{u}_1\cdot\left(\mathbf{b}\times\frac{\partial \mathbf{u}_1}{\partial x}\right), \qquad [18]$$

noting that $\mathbf{u}_1\cdot(\partial\mathbf{b}/\partial x\times\mathbf{u}_1)=0$.

From Eq. [11],

$$\begin{aligned}q=\mathbf{u}_1\cdot\frac{\partial \mathbf{b}}{\partial x}&=\mathbf{u}_1\cdot\left(\frac{\partial(\mathbf{u}_1\times\mathbf{u}_2)}{\partial x}\left|\mathbf{u}_1\times\mathbf{u}_2\right|^{-1}\right.\\ &\quad\left.+(\mathbf{u}_1\times\mathbf{u}_2)\frac{\partial\left|\mathbf{u}_1\times\mathbf{u}_2\right|^{-1}}{\partial x}\right)\\ &=\mathbf{u}_1\cdot\frac{\partial(\mathbf{u}_1\times\mathbf{u}_2)}{\partial x}\left|\mathbf{u}_1\times\mathbf{u}_2\right|^{-1}\\ &=\mathbf{u}_1\cdot\left(\mathbf{u}_1\times\frac{\partial \mathbf{u}_2}{\partial x}+\frac{\partial \mathbf{u}_1}{\partial x}\times\mathbf{u}_2\right)\left|\mathbf{u}_1\times\mathbf{u}_2\right|^{-1},\end{aligned}$$

so that

$$q=\left|\mathbf{u}_1\times\mathbf{u}_2\right|^{-1}\mathbf{u}_1\cdot\left(\frac{\partial \mathbf{u}_1}{\partial x}\times\mathbf{u}_2\right). \qquad [19]$$

A similar computation shows that

$$r=\left|\mathbf{u}_1\times\mathbf{u}_2\right|^{-1}\mathbf{n}\cdot\left(\mathbf{u}_1\times\frac{\partial \mathbf{u}_2}{\partial x}+\frac{\partial \mathbf{u}_1}{\partial x}\times\mathbf{u}_2\right). \qquad [20]$$

3.2.7. Computation of $\partial\mathbf{u}_k/\partial x$, $k=1, 2$. Finally, the computation is reduced to computing $\partial\mathbf{u}_k/\partial x$ for $k=1, 2$. Write

$$\begin{aligned}\mathbf{v}_1&=P_{\mathrm{B}}-P_{\mathrm{A}}, \quad \mathbf{v}_2=P_{\mathrm{C}}-P_{\mathrm{A}}\\ \mathbf{u}_1&=\frac{\mathbf{v}_1}{|\mathbf{v}_1|}, \qquad \mathbf{u}_2=\frac{\mathbf{v}_2}{|\mathbf{v}_2|}.\end{aligned} \qquad [21]$$

Then

$$\frac{\partial \mathbf{u}_k}{\partial x}=\frac{\partial \mathbf{v}_k}{\partial x}\left|\mathbf{v}_k\right|^{-1}-\mathbf{v}_k\left|\mathbf{v}_k\right|^{-3}\mathbf{v}_k\cdot\frac{\partial \mathbf{v}_k}{\partial x} \qquad [22]$$

for $k=1, 2$ and for differentiation with respect to the atom coordinates,

$$\frac{\partial \mathbf{v}_1}{\partial x}=\begin{cases}\mathbf{i} & \text{if } x=x_{\mathrm{B}}\\ \mathbf{j} & \text{if } x=y_{\mathrm{B}}\\ \mathbf{k} & \text{if } x=z_{\mathrm{B}}\\ -\mathbf{i} & \text{if } x=x_{\mathrm{A}}\\ -\mathbf{j} & \text{if } x=y_{\mathrm{A}}\\ -\mathbf{k} & \text{if } x=z_{\mathrm{A}}\\ 0 & \text{if } x=x_{\mathrm{C}}\\ 0 & \text{if } x=y_{\mathrm{C}}\\ 0 & \text{if } x=z_{\mathrm{C}}\end{cases} \qquad [23]$$

$$\frac{\partial \mathbf{v}_2}{\partial x}=\begin{cases}\mathbf{i} & \text{if } x=x_{\mathrm{C}}\\ \mathbf{j} & \text{if } x=y_{\mathrm{C}}\\ \mathbf{k} & \text{if } x=z_{\mathrm{C}}\\ -\mathbf{i} & \text{if } x=x_{\mathrm{A}}\\ -\mathbf{j} & \text{if } x=y_{\mathrm{A}}\\ -\mathbf{k} & \text{if } x=z_{\mathrm{A}}\\ 0 & \text{if } x=x_{\mathrm{B}}\\ 0 & \text{if } x=y_{\mathrm{B}}\\ 0 & \text{if } x=z_{\mathrm{B}}.\end{cases} \qquad [24]$$

This completes the computation of S (Eq. [16]), and from this, $\partial F/\partial x=FS$. This derivative is used in Eqs. [14] and [15] to determine $\partial\sigma_c/\partial x$ and from this $\partial E_{\mathrm{cs}}/\partial x$ (Eq. [8]).

3.2.8. Typical values. This section describes some typical experimentally determined values for the chemical shift parameters β, σ_{aa}, σ_{bb}, and σ_{cc}. In the typical cases discussed in (*1*), $\alpha=0$.

For ^{15}N protein backbone chemical shift, take A = N, B = C_1 and C = C_α. Typical values are $\beta=105°$, $\sigma_{aa}=205$, $\sigma_{bb}=35$, and $\sigma_{cc}=60$. See (*12*).

For ^{13}C protein backbone chemical shift, take A = C_1, B = N, and C = C_α. Typical values are $\beta=35°$, $\sigma_{aa}=245$, $\sigma_{bb}=175$, and $\sigma_{cc}=90$. See (*11*).

For ^{15}N indole chemical shift, take A = $N_{\epsilon 1}$, B = $H_{\epsilon 1}$, and C = $C_{\epsilon 2}$. Typical values are $\beta=25°$, $\sigma_{aa}=165$, $\sigma_{bb}=115$, and $\sigma_{cc}=45$. See (*13*).

In the illustrative refinement discussed below, these typical values were applied uniformly. However, specific principal tensor values could be used for each chemical shift. As discussed in the Appendix, the principal tensor values are to be included with each chemical shift as part of the input data.

4. REFINEMENT: A TEST CASE

To illustrate how refinement with the solid-state NMR module can improve a structure, we begin with an initial structure

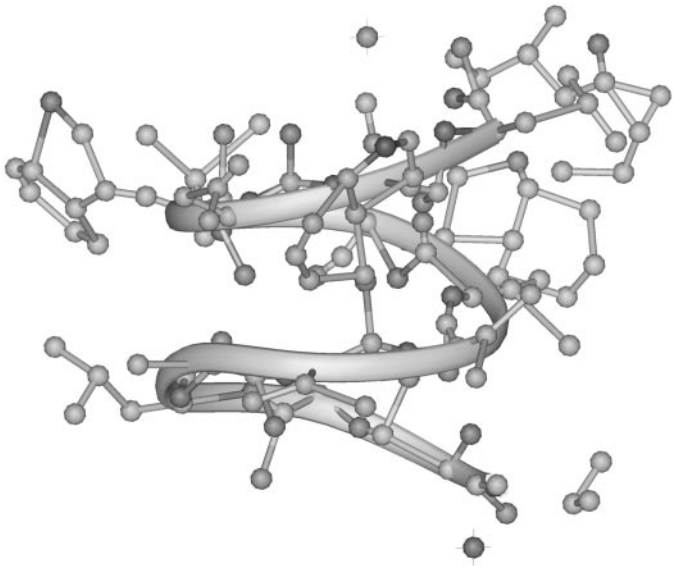

FIG. 4. Gramicidin A monomer after random perturbation of all atomic coordinates.

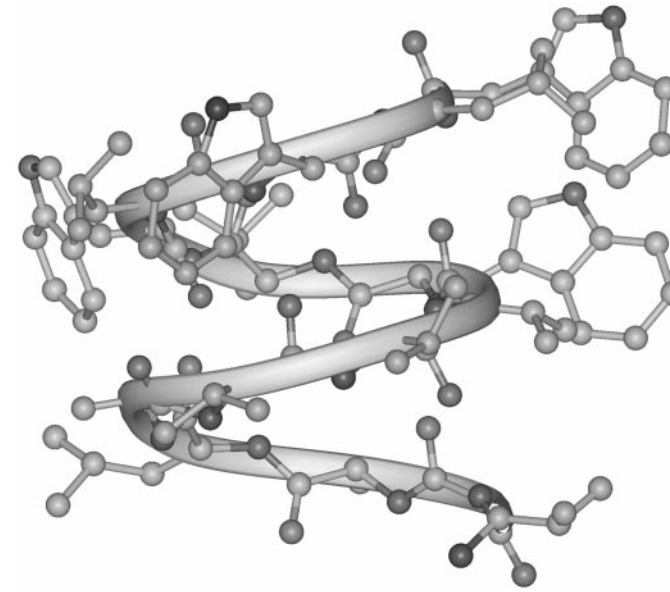

FIG. 5. Gramicidin A monomer after refinement of initial perturbed model.

produced by randomly perturbing the atom locations in a model of a gramicidin A monomer (*1*). This has the amino acid sequence Val-Gly-Ala-DLeu-Ala-DVal-Val-DVal-Trp-DLeu-Trp-DLeu-Trp-DLeu-Trp and forms a β-helix. The perturbation was achieved by adding random numbers between -0.5 and 0.5 to each coordinate of each atom, greatly distorting the backbone and side chain geometries (Fig. 4). Several rounds of refinement were performed on the perturbed structure, including both simulated annealing and conjugate gradient least-squares minimization. The stereochemical energy included terms for bonds, angles, improper angles, and van der Waals forces. Also included was an energy term for hydrogen bonds, defined as the squared difference between calculated and ideal lengths. The determination of hydrogen bond coupling was based on the model of gramicidin A from which the initial perturbation was made. Data consisted of 15 N–H dipolar couplings, 13 N–C dipolar couplings, 70 quadrupolar couplings, 14 ^{15}N chemical shifts, and 2 ^{13}C chemical shifts.

There is no *a priori* rule for choosing the weights w_{dq} and w_{cs} for the dipolar/quadrupolar coupling and chemical shift energy terms. One could refine using a range of weights, accepting the refinement that produced the best fit to the data while preserving good stereochemistry. A better approach would be to choose weights that minimize some free *R* factor, obtained through cross validation (*14, 15*). Work on this approach, using orientational restraints from solid-state NMR, is in progress. For the current illustrative example, weighting values of $w_{dq} = 300$ and $w_{cs} = 300$ were used since they yielded a good refinement, but it is unlikely that they are optimal. These uniform values were chosen for simplicity, but the program has the flexibility to assign different weights to different data sets. (In the refinement of Ketchem *et al.*, for example, the weights were scaled by the experimental error as a way of normalizing the different data types.) The starting temperature for annealing was 300 K, and Cartesian dynamics were used. Additional details are given in the Appendix.

The refined structure (Fig. 5) is greatly improved over the initial structure in terms of both the stereochemistry and the fit to the solid-state NMR data. This is illustrated in Table 1, which shows the various energy terms calculated with both models. As expected from the random coordinate perturbation, deviations from ideal covalent bond lengths and angles are severe. However, the total energy of the perturbed model is dominated by the chemical shift energy, indicating that the perturbed model disagrees greatly with this data set. After refinement, all stereochemical and solid-state NMR energy terms are greatly reduced, illustrating the simultaneous minimization of E_{chem} and E_{data}.

In this refinement, all dipolar/quadrupolar couplings were scaled by *K* (see Eq. [6]). This allows the user of the program to enter a single value ($K = 1$) in the input file, rather than

TABLE 1
Energies Before and After Refinement

Energy term	Perturbed	Refined
Bond	39,931	86
Angle	53,893	3386
Improper	48,563	1121
van der Waal	273	47
H Bond	1.5	0.6
CS	1.2×10^7	24
DQ	17,139	117

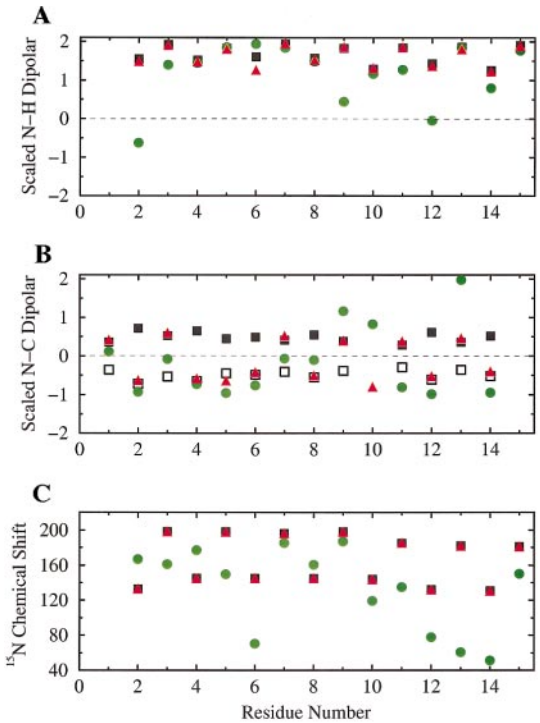

FIG. 6. (A, B) Observed and calculated dipolar couplings, scaled by K to lie between -1 and 2. (C) Observed and calculated ^{15}N chemical shifts in ppm. (circles) Perturbed model, (triangles) refined model, (solid squares) ν_o and σ_o, (open squares) $-\nu_o$. Principal values and α and β angles are the *typical values* described in the text.

entering the different $\nu_\parallel$ and QCC values. A similar approach cannot be used for chemical shift data, where each data point is associated with three parameters (the principal values) rather than one.

The fit of the calculated dipolar couplings (with $K = 1$) to the scaled observed dipolar couplings is shown in Figs. 6A and 6B (quadrupolar couplings not shown). All scaled N–H couplings are greater than unity ($\nu_o > K$) so, for each residue, the dipolar energy is minimized ($E_{dq} = 0$) only when $\nu_c = \nu_o$. In contrast, all scaled N–C dipolar couplings are less than unity, so the dipolar energy is minimized either when $\nu_c = \nu_o$ or when $\nu_c = -\nu_o$. The observed couplings ν_o are represented in Fig. 6 by filled squares and $-\nu_o$ by open squares in Fig. 6B. Despite large initial perturbations away from observed values, the refinement moves atoms so as to satisfy the dipolar coupling restraints. By satisfying these restraints, the refinement orients backbone and side chain atoms in a way that is consistent with the orientational data.

The fit to the ^{15}N chemical shift data is shown in Fig. 6C. Again, despite large initial perturbations away from observed values, the refinement moves atoms so as to satisfy the orientational restraints. In this case, the restraints orient main chain atoms, complementing the dipolar coupling restraints. As is typical for atomic refinement, fit to the data comes at the expense of higher stereochemical energy. By increasing or decreasing the weights w_{dq} and w_{cs}, more or less emphasis is put on satisfying the orientational data.

This refinement was performed on a DEC Alpha workstation and required less than 1 min of computer time. As a comparison, the Monte-Carlo-like implementation of Ketchem *et al.* required several hours to refine the gramicidin A monomer. With the new implementation, computational speed will not be a limiting factor in the structure determination of larger polypeptides or proteins using solid-state NMR.

5. DISCUSSION

We have described a new method for using orientational restraints in atomic refinement. The computer software is in the form of a module for the crystallographic and NMR refinement software package CNS, and the refinement is several orders of magnitude faster than previous refinement software using orientational restraints. This speed removes one of the difficulties in the determination of protein structures with solid-state NMR.

The refinement of molecular structures to high resolution is vitally important for assessing structure–function relationships. Gramicidin A is used as a model system in this study and provides an excellent example of the need for high resolution. While it has been known for several decades that gramicidin A forms a pore through membranes that facilitates cation transport, fundamental questions about selectivity and kinetics were not answered until the high-resolution structure became available (*1*). For instance, the rate limiting step for cation entry into the single file region of the channel was determined to be the removal of the last nonaxial water molecule in the hydration sphere. In addition, it has been clearly shown that the cation binding site is delocalized, thereby decreasing the entropic penalty for cation binding (*16*). Here, the new capabilities introduced into CNS for utilizing orientational restraints allow optimal refinements of solid-state NMR structures.

The optimal utilization of structural restraints is a major challenge in structural biology. This challenge is made more complex by a diversity of restraint and constraint types, such as stereochemical terms assessed by force fields, distances within a macromolecule associated with a known hydrogen bond, and orientations of particular atomic sites with respect to a macromolecular frame of reference. Because of the high precision of the solid-state NMR restraints, the possibilities during refinement for being trapped in local minima are great. Hence describing a high-quality initial structure is very important for the success of the refinement protocol. Fortunately, the high precision of the NMR data permits the development of such an initial structure and as shown here corrections for substantial distortions of the initial structure are well within the capabilities of the refinement protocol.

APPENDIX

The module for chemical shift derivatives and patches for standard CNS modules can be downloaded from the website http://www.sb.fsu.edu/~bertram. A sample simulated annealing CNS input file and sample data files can also be downloaded.

The sample annealing input file allows the user to enter the names of any files containing dipolar and quadrupolar coupling data and files containing chemical shift data. The dipolar/quadrupolar data files include the atoms involved in the couplings and the observed coupling values. Dipolar and quadrupolar couplings should be normalized by K and $K = 1$ entered in the CNS input file. Chemical shift data files include observed shifts and principal values for each chemical shift tensor. Values for α and β angles are entered in the CNS input file. Each data file is weighted separately, and the weights are entered in the input file. Either Cartesian or torsion angle dynamics may be used in the simulated annealing.

ACKNOWLEDGMENTS

R.B. was supported by NSF Grants DBI 9602233 and DMS 9981822. M.S.C. was partially supported by NSF Grant DBI 9808098. J.R.Q. and T.A.C. were partially supported by NSF Grant DMB 9986036. J.R.Q. was also partially supported by NSF Grant DMS 9972858.

REFERENCES

1. R. R. Ketchem, B. Roux, and T. A. Cross, High-resolution polypeptide structure in a lamellar phase lipid environment from solid state NMR derived orientational constraints, *Structure* **5,** 1655–1669 (1997).
2. R. R. Ketchem, W. Hu, and T. A. Cross, High-resolution conformation of gramicidin A in a lipid bilayer by solid-state NMR, *Science* **261,** 1457–1460 (1993).
3. R. Fu and T. A. Cross, Solid-state NMR investigation of protein and polypeptide structure, *Annu. Rev. Biophys. Biomol. Struct.* **28,** 235–268 (1999).
4. J. R. Quine and T. A. Cross, Protein structure in anisotropic environments: Unique structural fold from orientational constraints, *Concepts Magn. Reson.* **12,** 71–82 (2000).
5. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.* **4,** 187–217 (1983).
6. S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Optimization by simulated annealing, *Science* **220,** 671–680 (1983).
7. A. T. Brünger, Crystallographic refinement by simulated annealing, *J. Mol. Biol.* **203,** 803–816 (1988).
8. A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren, Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr.* **D54,** 905–921 (1998).
9. N. Tjandra, J. Marquardt, and G. M. Clore, Direct refinement against dipolar couplings in NMR structure determination of macromolecules, *J. Magn. Reson.* **142,** 393–396 (2000).
10. G. M. Clore, A. M. Gronenborn, and N. Tjandra, Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude, *J. Magn. Reson.* **131,** 159–162 (1998).
11. Q. Teng, M. Iqbal, and T. A. Cross, Determination of the ^{13}C chemical shift and ^{14}N electric field gradient tensor orientations with respect to the molecular frame in a polypeptide, *J. Am. Chem. Soc.* **114,** 5312–5321 (1992).
12. Q. Teng and T. A. Cross, The *in situ* determination of the ^{15}N chemical-shift tensor orientation in a polypeptide, *J. Magn. Reson.* **85,** 439–447 (1989).
13. W. Hu, K.-C. Lee, and T. A. Cross, Tryptophans in membrane proteins: Indole ring orientations and functional implications in the gramicidin channel, *Biochemistry* **32,** 7035–7047 (1993).
14. A. T. Brünger, Free *R* value: A novel statistical quantity for assessing the accuracy of crystal structures, *Nature* **355,** 472–475 (1992).
15. G. M. Clore and D. S. Garrett, R-factor, free R, and complete cross-validation for dipolar coupling refinement of NMR structures, *J. Am. Chem. Soc.* **121,** 9008–9012 (1999).
16. F. Tian and T. Cross, Cation transport: An example of structural based selectivity, *J. Mol. Biol.* **285,** 1993–2003 (1999).