Atomic refinement with correlated solid-state NMR restraints

Richard Bertram,[a,b,*] Tom Asbury,[b] Felcy Fabiola,[b] J.R. Quine,[a,b,c]
Timothy A. Cross,[b,c,d] and Michael S. Chapman[b,d]

[a] *Department of Mathematics, Florida State University, Tallahassee, FL 32306-4510, USA*
[b] *Kasha Institute of Biophysics, Florida State University, Tallahassee, FL 32306, USA*
[c] *National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32306, USA*
[d] *Department of Chemistry, Florida State University, Tallahassee, FL 32306, USA*

Received 24 January 2003; revised 27 March 2003

Abstract

The orientation data provided by solid-state NMR can provide a great deal of structural information about membrane proteins. The quality of the information provided is, however, somewhat degraded by sign degeneracies in measurements of the dipolar coupling tensor. This is reflected in the dipolar coupling penalty function used in atomic refinement, which is less capable of properly restraining atoms when dipolar sign degeneracies are present. In this report we generate simulated solid-state NMR data using a variety of procedures, including back-calculation from crystal structures of α-helical and β-sheet membrane proteins. We demonstrate that a large fraction of the dipolar sign degeneracies are resolved if anisotropic dipolar coupling measurements are correlated with anisotropic chemical shift measurements, and that all sign degeneracies can be resolved if three data types are correlated. The advantages of correlating data are demonstrated with atomic refinement of two test membrane proteins. When refinement is performed using correlated dipolar couplings and chemical shifts, perturbed structures converge to conformations with a larger fraction of correct dipolar signs than when data are uncorrelated. In addition, the final structures are closer to the original unperturbed structures when correlated data are used in the refinement. Thus, refinement with correlated data leads to improved atomic structures. The software used to correlate dipolar coupling and chemical shift data and to set up energy functions and their derivatives for refinement, CNS-SS02, is available at our web site.
© 2003 Elsevier Science (USA). All rights reserved.

Keywords: Refinement; Solid-state NMR; Dipolar coupling; Chemical shift

1. Introduction

Analyses of genome sequences of eubacterial, archaean, and eukaryotic organisms indicate that up to 30% of the open reading frames encode membrane proteins [1]. However, the three-dimensional structures of only a small number of membrane proteins have been solved due to practical difficulties with both solution NMR and X-ray crystallography [2]. Solid-state NMR, which makes use of samples with uniformly aligned proteins, is a technique that is well suited for membrane proteins [2,3]. Orientational data from solid-state NMR have been used to determine the atomic structure of gramicidin A [4,5] and the backbone structure of the M2 transmembrane α-helix from the acetylcholine receptor [6,7] and from the influenza A virus [8]. While there has been much progress in secondary structure determination using correlated solid-state NMR restraints [9], there has been less progress on computational methodology for tertiary structure determination. There are essentially three steps involved in atomic model development: (1) getting the data, (2) constructing an initial model, and (3) refining the initial model. In this report we focus on the third step.

Anisotropic dipolar coupling and chemical shift data from solid-state NMR provide information on the orientation of the peptide planes relative to an external magnetic field. One difficulty with the dipolar coupling is the ambiguities or degeneracies that exist when relating the measured quantity to the orientation angle of the

---

bonded atom pair [10]. One type of degeneracy, the sign degeneracy, comes about because the dipolar coupling measurement is the absolute value of the dipolar tensor value. For small angles this is not a problem, the sign is positive. However, for larger angles the degeneracy cannot be resolved using this dipolar coupling information alone.

One method for resolving sign degeneracies is to correlate anisotropic dipolar couplings with chemical shifts [11,12]. This can be done with PISEMA (polarization inversion spin exchange at magic angle) experiments, where ^{15}N–^1H dipolar couplings and ^{15}N chemical shifts from uniformly labeled samples are correlated [13]. One can also correlate separate ^{15}N chemical shift and ^{15}N–^{13}C dipolar coupling measurements corresponding to the same nitrogen atom, although this requires selective isotope labeling. Equations were derived in [11] for the powder pattern of correlated dipolar coupling and chemical shift interactions. In two regions of the (^{15}N chemical shift, ^{15}N–^1H dipolar coupling) powder pattern the sign of the dipolar coupling tensor is resolved as either positive or negative, while in the two other regions it is not. However, the degenerate region of the powder pattern is much smaller than the degenerate region that exists when the dipolar couplings are not correlated with chemical shifts. That is, when knowledge of the dipolar coupling alone is used to resolve the sign.

In the construction of an initial model from orientational constraints, the goal is to find the spherical coordinates (α, β) of the unit magnetic field direction, $\mathbf{B}_0$, in a local atomic frame. Each orientational constraint at the atom gives an equation in (α, β), and a unique solution to the system of equations is sought. Degeneracies arise because the expressions for chemical shift and dipolar coupling are quadratics, and because the measured dipolar coupling includes an absolute value. As many as 16 solutions are possible. The degeneracies can be handled algebraically, or graphically using a restriction plot [14,15]. Often the degeneracy cannot be resolved a priori. Our focus is on the sign degeneracy, which can often be resolved by correlating chemical shift and dipolar coupling data.

In this report, we simulate ^{15}N–^1H dipolar coupling and ^{15}N chemical shift data using a number of different protocols to determine the fraction of data that lie in each region of the (^{15}N chemical shift, ^{15}N–^1H dipolar coupling) powder pattern. Using random orientational sampling protocols, we find that fewer than 20% of the data lie in degenerate regions of the two-dimensional powder patterns, while more than 60% of the dipolar signs are unresolved if dipolar data are not correlated with ^{15}N chemical shifts. Data are then simulated from two membrane proteins, bacteriorhodopsin and outer membrane protein A (ompA), whose structure has been determined using X-ray crystallography. The former is

an α-protein, while the latter is a β-protein. Data simulated with these structures indicate that with bacteriorhodopsin a slightly larger fraction of dipolar signs are resolved compared with random sampling, while with ompA a significantly lower fraction is resolved. In both cases, the fraction of dipolar signs resolved using correlated data is much greater than the fraction resolved using uncorrelated data.

The second aim of this report is to demonstrate the advantages of computer software that automatically correlates dipolar coupling and chemical shift data and uses this to refine atom positions. Atomic refinement is an optimization process in which the atoms of an approximate model are shifted so as to minimize a potential energy function. This function consists of *stereochemical restraints* to enforce properties such as covalent bond lengths and angles, and *penalty restraints* to enforce compliance with experimental data. Our focus is on orientational data from solid-state NMR, both anisotropic dipolar couplings and chemical shifts. A computer program for refinement using orientational data has been developed that employs a Monte Carlo simulated annealing approach [16]. This was used in the refinement of gramicidin A [5] and the M2 monomer backbone structure from influenza A virus [8]. More recently, computer software was developed (CNS-SS) [17] that operates as a module for the refinement program CNS [18]. Here several optimization approaches can be used, including conjugate gradient minimization and simulated annealing based on molecular dynamics. One advantage of CNS-SS is the speed in which refinement can be performed [17].

As described above, when data are uncorrelated the signs of the majority of the ^{15}N–^1H dipolar couplings are typically degenerate. This results in additional local minima in the ^{15}N–^1H penalty function, and we show that the quality of the function for restraining atom positions is degraded. The new software described in this report, CNS-SS02, correlates data so that the maximal fraction of dipolar signs are resolved. This results in a reduction in the number of the local minima in the ^{15}N–^1H penalty function, and an overall improvement in the quality of the restraint. We show that refinement with correlated data is better able to correct initially incorrect dipolar signs, and produces a refined structure that is closer to that of the target structure. Thus, correlating data increases the quality of atomic refinements. Finally, we demonstrate that all dipolar signs can be resolved if three data types are correlated.

2. Mapping of the sphere to the frequency plane

The anisotropic dipolar tensor v giving the singlet splitting is axially symmetric with unique axis $\mathbf{u}$ in the direction of a covalent bond. If $v_\parallel$ is the value of v when

unit vector $\mathbf{u}$ is aligned with the magnetic field vector $\mathbf{B_0}$, then

$$v = \frac{v_\parallel}{2}(3\cos^2\theta - 1), \tag{1}$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{B_0}$. Note that the doublet splitting is $2v$ and the maximal value of the doublet splitting is $2v_\parallel$. Because of the doublet splitting of the dipolar interaction, only the absolute value of v can be experimentally determined.

The anisotropic chemical shift tensor σ is asymmetric, with principal axes σ_{11}, σ_{22}, and σ_{33}. The principal values satisfy $\sigma_{11} \leqslant \sigma_{22} \leqslant \sigma_{33}$. The value of σ is given by the quadratic function

$$\sigma = \sigma_{11}(\mathbf{B_0} \cdot \sigma_{11})^2 + \sigma_{22}(\mathbf{B_0} \cdot \sigma_{22})^2 + \sigma_{33}(\mathbf{B_0} \cdot \sigma_{33})^2. \tag{2}$$

For main chain measurements, the dipolar axis $\mathbf{u}$ may be along the amide N–H bond or the backbone N–C bond. The ^{15}N chemical shift tensor has σ_{33} in the peptide plane, making an angle β with $\mathbf{u}$ (Fig. 1), the angle being measured counterclockwise with respect to a specified peptide plane orientation. We assume that σ_{11} is also in the peptide plane (orthogonal to σ_{33}), so equations relating σ and v can be derived [11]. With this assumption, we write $\mathbf{u}$ in terms of the chemical shift principal axis frame,

$$\mathbf{u} = \sin\beta\sigma_{11} + \cos\beta\sigma_{33}. \tag{3}$$

One can also write the unit vector $\mathbf{B_0}$ in this frame, with coordinates (x, y, z) where $x = \mathbf{B_0} \cdot \sigma_{11}$, $y = \mathbf{B_0} \cdot \sigma_{22}$, $z = \mathbf{B_0} \cdot \sigma_{33}$, noting that $x^2 + y^2 + z^2 = 1$. Then, σ and v are related to $\mathbf{B_0}$ by

$$\sigma = \sigma_{11}x^2 + \sigma_{22}y^2 + \sigma_{33}z^2, \tag{4}$$

$$v = \frac{v_\parallel}{2}[3((\sin\beta)x + (\cos\beta)z)^2 - 1]. \tag{5}$$

Eqs. (4) and (5) map points (x, y, z) on the unit sphere to the frequency plane (σ, v). For $\beta > 0$ the points $(x, 0, z)$ map to an ellipse given by
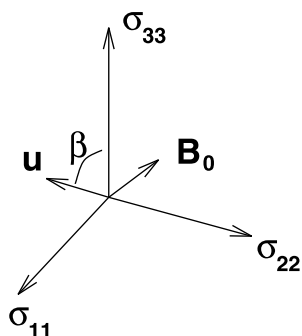
$$(\tilde{v} - \cos(2\beta)\tilde{\sigma} - \sin^2\beta)^2 = \tilde{\sigma}(1 - \tilde{\sigma})\sin^2(2\beta), \tag{6}$$

where

$$\tilde{\sigma} = \frac{\sigma - \sigma_{11}}{\sigma_{33} - \sigma_{11}}, \tag{7}$$

$$\tilde{v} = \frac{1}{3}\left(\frac{2v}{v_\parallel} + 1\right). \tag{8}$$

Eq. (6) corrects an error in [11, Eq. (4)]. Most other points map to the interior of the ellipse, while some map to a region outside the ellipse [11]. The point $(0, 1, 0)$ maps to a point $Q = (\sigma_{22}, -v_\parallel/2)$ that helps to define the extra-elliptical region. In particular, each point on the sphere maps to a point on a ray from Q to a point on the ellipse. The farther Q is from the ellipse the larger the extra-elliptical region.

Projections to the frequency plane are illustrated in Figs. 2a and b, for ^{15}N–^{1}H and ^{15}N–^{13}C dipolar cou-
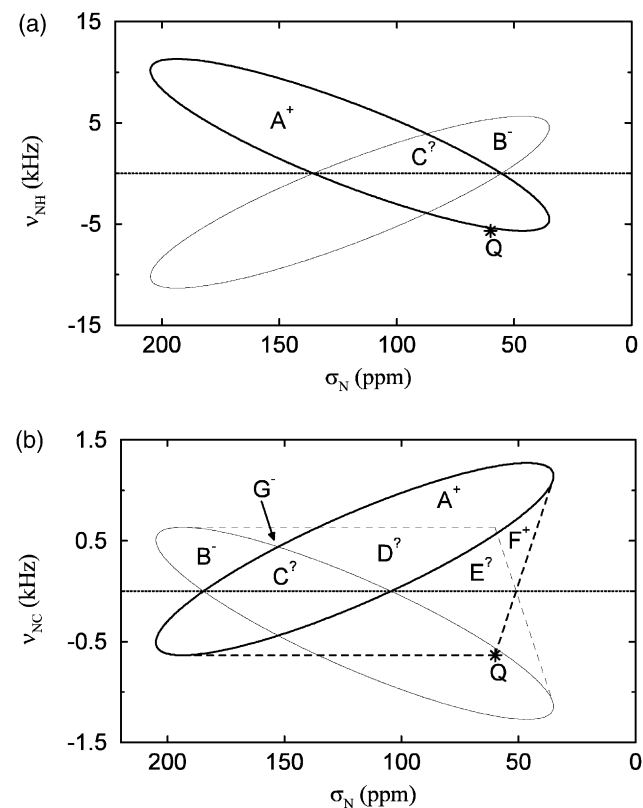

(a)

(b)

Fig. 2. Primary (heavy curve) and reflected (thin) powder pattern ellipses. Here and in all simulations, we use $\sigma_{11} = 35$ ppm, $\sigma_{22} = 60$ ppm, $\sigma_{33} = 205$ ppm, and $v_\parallel = 11.335$ kHz for ^{15}N–^{1}H and $v_\parallel = 1.271$ kHz for ^{15}N–^{13}C dipolar couplings [26]. (a) $\beta = 15°$. Three of the four regions where experimental measurements of (σ, v) may lie are labeled, along with the sign of the dipolar coupling (+, −, or degenerate (?)). The fourth region is outside and slightly above the reflected ellipse, and the sign of the coupling is negative. (b) $\beta = 105°$. There are seven distinct regions, three in extra-elliptical regions, which are now large. The boundary of the extra-elliptical region is indicated by heavy dashed lines, and the boundary of its reflection is indicated by thin dashed lines.

Fig. 1. Principal axis frame $(\sigma_{11}, \sigma_{22}, \sigma_{33})$ of the chemical shift tensor. The unique axis $\mathbf{u}$ of the symmetric dipolar coupling tensor is assumed to lie in the σ_{11}–σ_{33} plane and make an angle of β with σ_{33}. The magnetic field direction $\mathbf{B_0}$ can be written in terms of the PAF.

plings, respectively. While the image of the sphere ranges from $-v_{\parallel}/2 \leqslant v \leqslant v_{\parallel}$, experimental measurements of the dipolar coupling are positive. Hence, if a point on the sphere maps to a point with $v < 0$, then the experimental correlate of this point will lie in the reflection of the ellipse about the σ-axis of the σv-plane. Primary and reflected ellipses are shown in Fig. 2a for $\beta = 15°$, the approximate angle made between the N–H bond vector and $\boldsymbol{\sigma}_{33}$. An experimental (σ, v) measurement lying in region A is in the primary ellipse and has a positive dipolar tensor; a point in region B lies in the reflected ellipse and has a negative tensor value; a point in region C lies in both primary and reflected ellipses, so the sign of the dipolar tensor cannot be determined (it is degenerate). The point Q lies just off of the primary ellipse, so the extra-elliptical region is small and not illustrated.

The primary ellipse rotates counterclockwise as β is increased and the extra-elliptical region becomes larger. This is shown in Fig. 2b for $\beta = 105°$, the approximate angle between the main chain N–C bond vector and $\boldsymbol{\sigma}_{33}$. Rays emanating from Q traverse a large extra-elliptical region, nearly as large in area as the interior of the primary ellipse. Experimental data points in region A lie only in the primary ellipse, so the dipolar sign is positive. Points in B lie only in the reflected ellipse, so the dipolar sign is negative. Points in C lie in both and are degenerate. Points in D lie in both the primary ellipse and the reflection of the extra-elliptical region, so are degenerate. Points in E lie in both the extra-elliptical region and its reflection, so they too are degenerate. Points in F lie only in the extra-elliptical region, so the dipolar signs are positive. Finally, points in G lie in the reflected extra-elliptical region and the dipolar signs are negative. This more complicated picture applies to correlated $^{15}\text{N}-^{13}\text{C}$ dipolar couplings and ^{15}N chemical shifts, which requires selective isotope labeling.

3. Samplings of the sphere

The areas of the degenerate and non-degenerate regions in Fig. 2 do not give a true sense of the distribution of points mapped to each region since the mapping is not area preserving. One estimate can be obtained with a uniform random sampling of the unit sphere. One could fix $\mathbf{B}_0$ and rotate the principal axis frames of the chemical shift and dipolar tensors, or equivalently, fix the principal axis frames and rotate $\mathbf{B}_0$. We have chosen the latter approach, rotating $\mathbf{B}_0$ to generate a uniform random sampling of 50,000 points. Fig. 3 shows the result of this sampling, where dipolar couplings with negative sign have been reflected, as would be the case in experimental measurements.

Approximately 35% of the sampling points map to region A, where the dipolar sign is positive; 47% map to region B or the extra-elliptical region, where the sign is
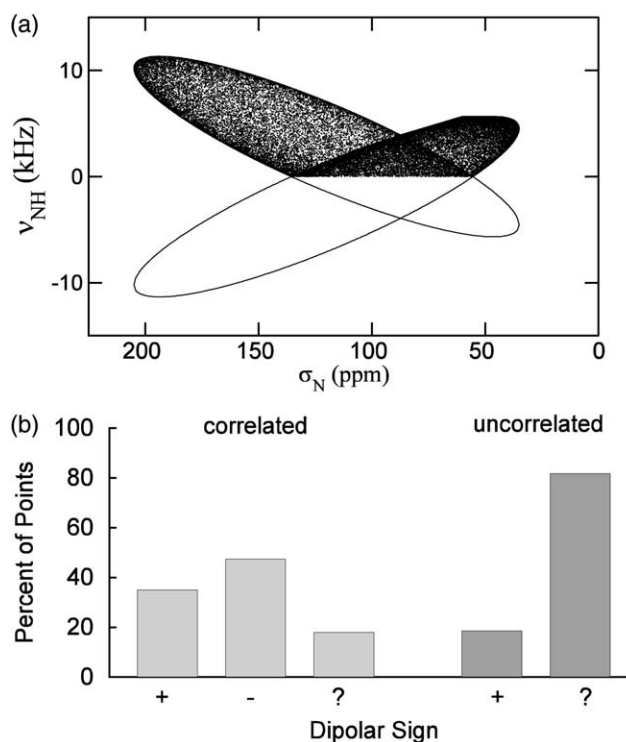

Fig. 3. (a) $(\sigma_{\text{N}}, v_{\text{NH}})$ data generated by a uniform sampling of the unit sphere (50,000 sampling points). (b) Percentage of points where the dipolar sign can be determined to be positive (+) or negative (−), or where the sign is degenerate (?). The degenerate fraction is much lower when dipolar and chemical shift data are correlated.

negative; approximately 18% map to the degenerate region C. Thus, 82% of the dipolar signs are resolved by correlating $^{15}\text{N}-^{1}\text{H}$ dipolar coupling and ^{15}N chemical shift data. In contrast, if data are not correlated, then only those points in which $v > v_{\parallel}/2$ will have the dipolar sign resolved (as positive). In the random sampling, this constitutes only 18% of the points, leaving 82% of the dipolar signs unresolved. Therefore, correlating the two sources of data leads to a great reduction in the number of degenerate dipolar signs.

Since membrane proteins often have a large helical content, a more physiological sampling of the sphere can be made using ideal α-helices. Thus, one can map points from ideal α-helices with tilt angle (relative to $\mathbf{B}_0$) randomly distributed between 0° and 90°. The distribution of points generated in this way is shown in Fig. 4 for $\beta = 15°$. With this sampling, 51% of the points map to region A with positive dipolar sign; 35% map to one of the regions with negative dipolar sign; and 14% map to a degenerate region. When data are uncorrelated, 64% of the dipolar signs are unresolved, a much larger fraction than with correlated data. Thus, correlating the two data types again leads to a great improvement in the resolution of dipolar signs, with fewer than 15% of the signs left unresolved when data are correlated.
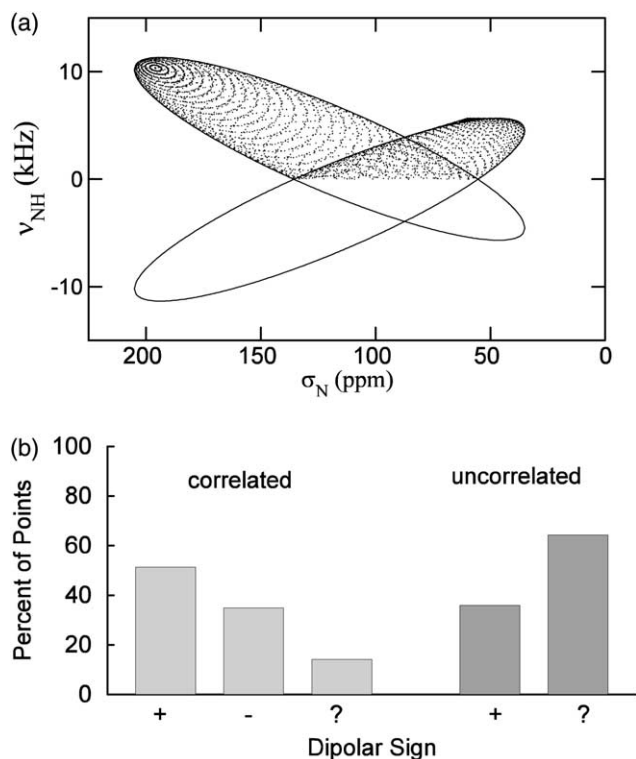
Fig. 4. (a) Mapping of data generated from an ideal α-helix ($\phi = -65°$, $\psi = -40°$) with tilt angle varied from 0° to 90°. (b) With this sampling 86% of the points map to regions where the dipolar sign is resolved. When dipolar couplings and chemical shifts are uncorrelated, fewer than 40% of the dipolar signs are resolved.
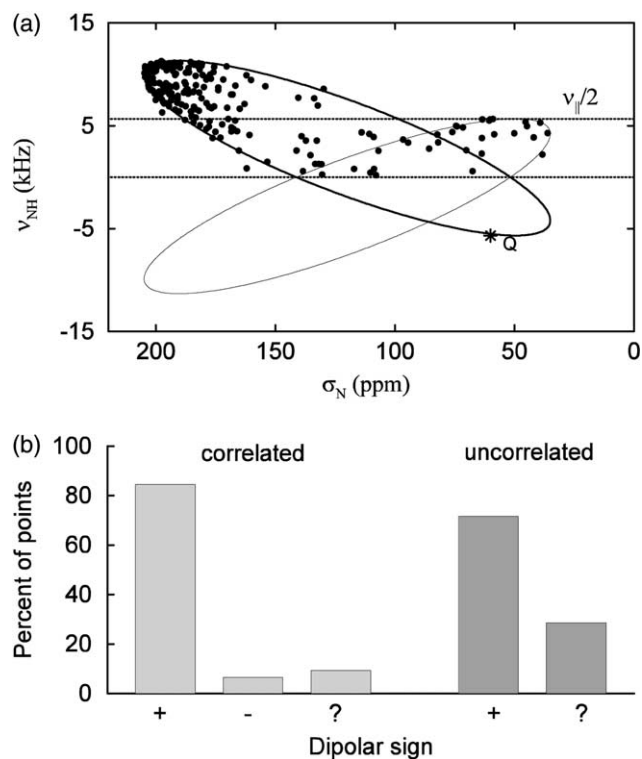

Fig. 5. (a) Two hundred and eighteen data points simulated from a 2.1 Å model of bacteriorhodopsin. (b) Approximately 91% of the dipolar signs are resolved when data are correlated, while 72% are resolved when data are uncorrelated.

As final samplings of the sphere, we generated synthetic $^{15}N-^{1}H$ dipolar coupling and ^{15}N chemical shift data from two membrane proteins whose structures have been determined by crystallography. The first structure, bacteriorhodopsin, has seven membrane-spanning α-helices. We used the 2.1 Å model from Royant et al. (PDB #1E0P) [19] to simulate the NMR data, assuming that dipolar coupling and chemical shift measurements can be made from all non-proline main chain N atoms. The mapping of these points is shown in Fig. 5a. Since the majority of the residues are contained in helices with relatively small tilt angle, most of the points map to PISA wheels in the upper left corner of the primary ellipse [20]. Thus, the majority of the dipolar signs are resolved as positive (85%) when data are correlated. A small fraction is resolved as negative (6%), and the remaining 9% are unresolved (Fig. 5b). Even when data are uncorrelated a large fraction of the dipolar signs are resolved (72%), leaving 28% of the signs unresolved. These results demonstrate that there is very little dipolar sign degeneracy for protein structures composed primarily of membrane-spanning α-helices with relatively small tilt angles. These structures include G protein coupled receptors [21] and the KcsA potassium ion channel [22].

We next simulated data from the 2.5 Å resolution model (PDB #1BXW) of the transmembrane domain of outer membrane protein A (ompA) [23]. This consists of eight β-strands forming a β-barrel. Synthetic data were generated again assuming that measurements can be made from all non-proline main chain N atoms. The mapping of these points is shown in Fig. 6a. In contrast to the mapping of bacteriorhodopsin, the points in the frequency plane are concentrated in regions B and C, so most have $v < v_{\parallel}/2$. This distribution reflects the various patterns formed by mappings of the β-strands [9]. Fig. 6b shows that when data are correlated 28% of the dipolar signs are resolved as positive, 33% as negative, and 39% of the dipolar signs are degenerate. The high percentage of degenerate signs reflects the high density of points in region C. When data are uncorrelated the situation is worse, with only 9% of the dipolar signs resolved and 91% degenerate. Thus, in this example with a structure consisting mainly of β-strands, the correlation of data is essential to provide reasonable resolution of dipolar signs.

4. Atomic refinement

Atomic refinement of a molecular structure involves the shifting of atom positions to improve agreement
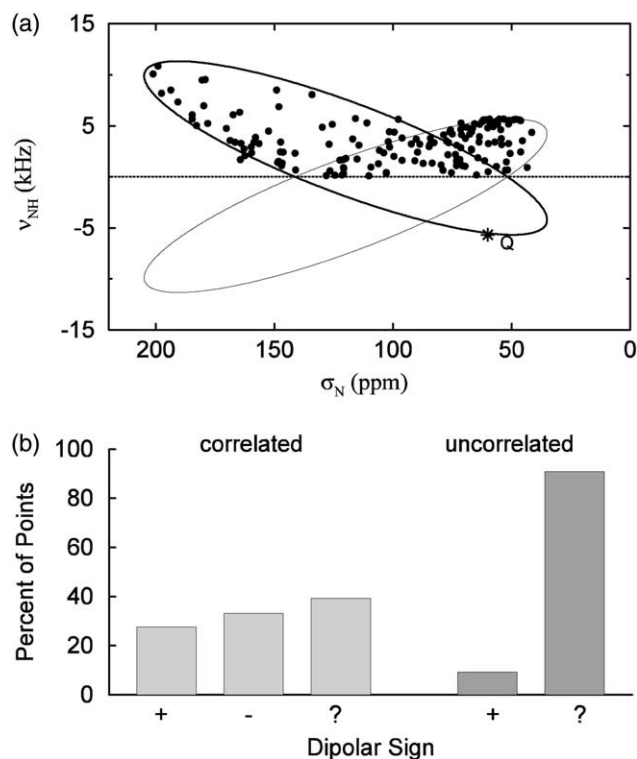
Fig. 6. (a) One hundred and sixty-three data points simulated from the 2.5 Å model of the transmembrane domain of outer membrane protein A. (b) Approximately 61% of the dipolar signs are resolved when data are correlated, compared with 9% when data are uncorrelated.

with experimental data and with stereochemical information known a priori from small molecules and quantum calculations. This is achieved by minimizing a potential pseudo-energy function E. For use with solid-state NMR, the form of the potential energy function is

$$E = E_{chem} + w_{dp}E_{dp} + w_{cs}E_{cs}, \tag{9}$$

where E_{chem} consists of stereochemical restraints on parameters such as covalent bond geometry and van der Waals interactions [24], and E_{dp}, E_{cs} are penalty functions corresponding to dipolar coupling and chemical shift data, respectively. The latter two components are weighted relative to the stereochemical terms by w_{dp} and w_{cs}. The chemical shift penalty term is the sum of simple harmonic functions,

$$E_{cs} = \sum_{cs} (\sigma_c - \sigma_o)^2, \tag{10}$$

where σ_o is the observed chemical shift and σ_c is the chemical shift calculated from the model. The dipolar coupling penalty function is more complicated due to the potential sign degeneracy, which introduces an absolute value [17]. The ith term in the dipolar penalty function is

$$E_{dp} = \begin{cases} (|v_c| - v_o)^2 & \text{if sign unresolved,} \\ (v_c - v_o)^2 & \text{if sign resolved.} \end{cases} \tag{11}$$

If the angle, θ_o, made between the appropriate co-valent bond vector (e.g., N–H vector for ^{15}N–^{1}H dipolar coupling) and $\mathbf{B}_0$ is small then $v_o > v_{\parallel}/2$ and the dipolar sign is positive. When $\theta_0 = \arccos(\sqrt{2/3}) \approx 35°$, $v_o = v_{\parallel}/2$, and for this and larger angles the dipolar sign may be positive or negative. In this case, the first expression in Eq. (11) is used in the refinement, unless the degeneracy is resolved some other way. The sign degeneracy introduces extra roots into E_{dp}, and this is reflected in the shape of the dipolar coupling energy landscape. This landscape is shown in Fig. 7a, where the dipolar energy is plotted versus θ_c, the angle between $\mathbf{B}_0$ and the N–H vector calculated from the model. For $\theta_o = 34°$ there are two minima, at $\theta_c = 34°$ and 146°, separated by a large barrier. For $\theta_o = 36°$ there are four minima: one at 36°, one at 144°, and two additional minima near and symmetric about 90°. More importantly, the energy landscape is relatively flat for all angles between 70° and 110°; the large energy barrier separating the outer minima is gone. Thus, if $\theta_o = 36°$ and the dipolar sign is not resolved, then the dipolar penalty function will not effectively restrain atoms to match the dipolar coupling measurement, which itself may be very precise. This problem persists to varying degrees for all $35° < \theta_o < 145°$.

The problem with the dipolar penalty function is alleviated when the dipolar sign is resolved by correlating with chemical shift data. In this case, the lower expression in Eq. (11) is used and no extra local minima are introduced for $\theta_o > 35°$. As shown in Fig. 7b, the dipolar energy landscape now changes smoothly as θ_o is increased from 34° to 36°, maintaining the energy barrier between the two local minima.

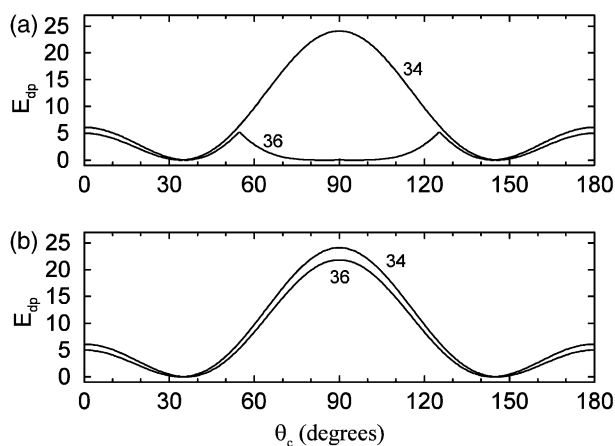From the various samplings of the sphere and the analysis of the dipolar coupling energy function above,

Fig. 7. (a) Dipolar coupling energy landscape for $\theta_o = 34°$, and 36°. There is a dipolar sign degeneracy when $\theta_o = 36°$. (b) The shape of the landscape is improved for $\theta_o = 36°$ (and larger angles) when the dipolar sign is resolved by correlating dipolar coupling and chemical shift data. In these calculations $v_{\parallel} = 11.335$.

we postulate: (1) that atomic refinement using correlated dipolar coupling and chemical shift data should be better at correcting dipolar signs that are incorrect in the initial model and (2) that the superior dipolar coupling penalty functions that result from resolving dipolar signs should lead to better refinements, even if all dipolar signs are correct in the initial model.

We have tested these hypotheses by refining the sample proteins bacteriorhodopsin and ompA, using the models to generate synthetic ^{15}N chemical shift and $^{15}N-^1H$ dipolar coupling data. Refinement was performed using the CNS-SS02 software package. This consists of two software modules that are linked to the refinement package CNS [18]. One module contains the chemical shift penalty function and its derivatives [17], while the other contains similar information for the dipolar coupling penalty function [25], as well as conditional statements to determine the dipolar signs by correlating dipolar coupling and chemical shift data. That is, conditional statements are included to determine whether a data point lies in regions A, B, or C of the powder pattern (Fig. 2a). For refinements using uncorrelated data, the conditional statements are bypassed, and the dipolar sign is degenerate unless $v_o > v_\parallel/2$. Thus, for all refinements described below both dipolar coupling and chemical shift penalty functions are included in the potential energy function, but for refinements using uncorrelated data the two data types are not correlated to provide information about the dipolar signs.

Since both of the model protein structures were solved using X-ray crystallography data, atoms will shift when these data are not included in the energy function, relieving strain in the stereochemistry. As a first step, we therefore performed a local energy minimization (using a conjugate gradient algorithm) with stereochemical restraints alone. Simulated NMR data were then calculated from this energy-minimized structure, which we denote as S_1 to distinguish from the original structure S_0. To test our first hypothesis, that refinement with correlated data should correct more dipolar signs than that performed with uncorrelated data, we applied a large perturbation to S_1 so that a large fraction of the dipolar signs are changed. This perturbation was accomplished using simulated annealing torsion angle refinement of S_1 with moderate computational temperature and with stereochemical restraints only, leading to an average RMS difference of approximately 5 Å for main chain atoms, without loss of good stereochemistry. This perturbed structure we denote as S_2. For the bacteriorhodopsin structure there were 42 dipolar sign changes, and for ompA there were 45 dipolar sign changes produced by the perturbation.

The optimal weighting of the two types of data is not clear. Since our goal was to correct dipolar signs, we chose a large weight for w_{dp} (100). In some refinements

w_{cs} is set to 1, minimizing the impact of chemical shift restraints. In others the chemical shift restraints were also emphasized by increasing w_{cs} to 100. For each refinement, we list the weights as (w_{dp}, w_{cs}). Conjugate gradient refinement of the perturbed bacteriorhodopsin model with CNS-SS02 and weighting (100, 1) corrected 86% of the incorrect dipolar signs when the data were correlated (Fig. 8, left). In contrast, only 57% of the incorrect dipolar signs were corrected by refinement when data were uncorrelated (Fig. 8, uncor.). It is possible that increasing the weight on chemical shift restraints in the potential energy function might help correct dipolar signs, even though the data are not used directly to resolve signs as in Fig. 2a. This is because the chemical shift data contain information about the orientation of bonded atoms in the peptide plane, and this orientational information alone can potentially resolve the dipolar sign degeneracy, as long as the chemical shift energy term is given sufficient weight. We tested this by refining with the weighting (100, 100), but without explicitly correlating to resolve dipolar signs. There was no difference in the fraction of corrected dipolar signs.

Refinement of the perturbed ompA model with weighting (100, 1) again confirmed that refinement corrects most of the dipolar signs when data are correlated, and again outperforms refinement with uncorrelated data (Fig. 8, right). However, when the chemical shift weighting was increased to 100, the vast majority of the dipolar signs are corrected whether or not the data are correlated. Thus, there are examples in which use of correlated data is critical in significantly improving poor models, and other examples in which judicious weighting between dipolar couplings and chemical shifts may be all that is required.

A second potential use of the correlated data would be in the improvement of models that are already a reasonable approximation. To test this second hypothesis, smaller perturbations were applied to the S_1 models of bacteriorhodopsin and ompA, such that all dipolar
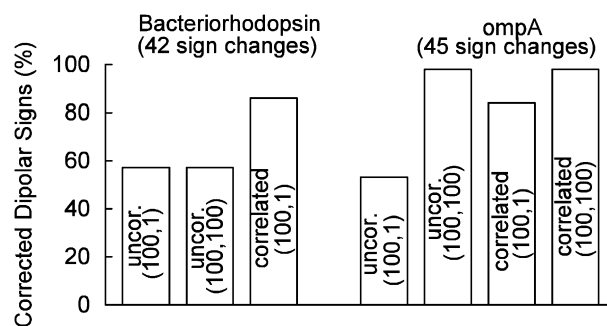

Fig. 8. Results of refinement using uncorrelated and correlated orientational data. Refinement weights are given in parentheses for dipolar coupling and chemical shift terms, respectively, labeled with weighting values. Results show fraction of the incorrect dipolar signs in S_2 that are corrected in the refined S_3 model.
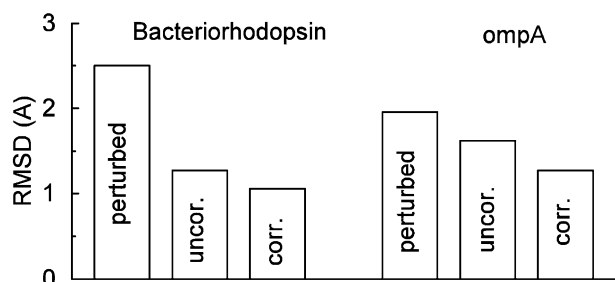
Fig. 9. The results of conjugate gradient refinement using either un-correlated or correlated data. RMSD is the RMS difference in main chain atom locations between the initial S_1 structure and the perturbed (S_2) or the final refined structure. Weights used for bacteriorhodopsin were $w_{dp} = 1$, $w_{cs} = 0.00001$, and for ompA, $w_{dp} = 1$, $w_{cs} = 0.001$.

terms remained of the correct sign. Perturbations were through simulated annealing with stereochemical restraints alone, at a lower computational temperature than before. For bacteriorhodopsin, the main chain RMS difference between the perturbed S_2 structure and the S_1 structure was 2.50 Å. When conjugate gradient refinement was performed using uncorrelated data the RMS difference was reduced by 50% (Fig. 9, left). When correlated data were used in the refinement the RMS difference was reduced by 58%. For ompA, the initial main chain RMS difference of 1.96 Å was reduced by 17 and 35% by refinement with uncorrelated and correlated data, respectively (Fig. 9, right). Thus, in both cases, correlated data yielded superior refinements even when the initial models were close enough approximations to be void of any errors in dipolar signs.

5. Discussion

There are several degeneracies in dipolar coupling data, due to symmetries in the dipolar tensor [10]. Resolving any of these degeneracies aids in the determination of peptide plane orientation relative to the fixed external magnetic field. We have demonstrated that the dipolar sign degeneracy can typically be resolved by correlating chemical shift and dipolar coupling data. Using various samplings of the unit sphere and simulated orientational data derived from α-helical and β-sheet proteins, we found that between 60 and 90% of the dipolar signs are resolved when data are correlated, versus 10% to 70% when data are uncorrelated.

Atomic refinement can correct dipolar signs that are initially in error. The fraction of corrected signs can depend on how dipolar coupling and chemical shift restraints are weighted, and whether the two data types are correlated. We have shown that refinement with correlated data is very effective at correcting dipolar signs, even if only the dipolar coupling restraints receive significant weighting. When both restraints are weighted

but data are not explicitly correlated, it is sometimes possible to correct as many dipolar signs as when data are correlated. However, strategies for correcting signs that depend on weighting are not ideal, because large weights that force sign correction will do so at the expense of other terms in the optimization, leading to poor stereochemistry. Explicit correlation of dipolar coupling and chemical shift data allows weights to be selected that allow a good fit to experimental data to be achieved with a stereochemically reasonable model.

The shape of the potential energy function is a key determinant for the success of an atomic refinement. We have shown that degeneracy in the dipolar sign leads to a dipolar coupling energy landscape with low energy barriers and shallow basins about minima. Resolution of the dipolar sign greatly improves the shape of the landscape, and should, in principle, lead to improved refinements. Using two test structures, bacteriorhodopsin and ompA, we have demonstrated that conjugate gradient refinement with correlated data yields structures that are closer to the target structures (i.e., those from which data were simulated) than refinement with uncorrelated data.

Our strategy for determining the dipolar sign is based solely on the location of the correlated data point within the doublet powder pattern (Fig. 2). This approach is amenable to software implementation, and is included in the package CNS-SS02. (Available free of charge to academic institutions at www.sb.fsu.edu/~rsref.) However, it is sometimes possible to determine the dipolar sign of a data point through other means. For example, residues contained within α-helices map to PISA wheels in the frequency plane, provided that the helix is sufficiently regular [7,11,20]. If a data point is clearly part of a PISA wheel, then it is likely to have the same dipolar sign as other points in the wheel. For example, a clearly defined PISA wheel may be present for an α-helix making a small angle with $\mathbf{B}_0$, in which case most of the elements of the wheel will lie in region A of the powder pattern. If a few elements of the wheel lie in the degenerate region C, they should be assigned a positive dipolar sign. Provision has been made in CNS-SS02 to allow the user to override the automatic determination of a dipolar sign, so if one knows the dipolar sign of a point lying in a degenerate region, this sign can be entered as input to the program.

Although correlating two types of data, here ^{15}N–^1H dipolar couplings and ^{15}N chemical shifts, resolves most of the dipolar signs, there are some that remain degenerate. These remaining degeneracies can be eliminated by correlating three data types. Suppose that for one residue three pieces of data are collected, say ^{15}N–^{13}C and ^{15}N–^1H dipolar couplings and ^{15}N chemical shift data. In the worst case, suppose that (ν_{NH}, σ_N) and (ν_{NC}, σ_N) lie in degenerate regions of the appropriate frequency planes (Fig. 2). Consider first (ν_{NH}, σ_N). Since

it is in a degenerate region it could have been mapped from a point A on the unit sphere such that $v_{NH} > 0$, or from a point B on the unit sphere such that $v_{NH} < 0$ (Fig. 10). (Due to symmetries in the dipolar coupling and chemical shift tensors there are actually as many as 8 points on the sphere that map to a single point (v_{NH}, σ_N) in the frequency plane, half with a positive dipolar coupling and the other half with a negative coupling.) Similarly, (v_{NC}, σ_N) can be mapped from two points on the unit sphere, A and C (Fig. 10). For example, point A might have $v_{NC} < 0$ while point C has $v_{NC} > 0$. There is only one point on the sphere, point A, that maps to both points (v_{NH}, σ_N) and (v_{NC}, σ_N), so the dipolar signs are resolved: $v_{NH} > 0$ and $v_{NC} < 0$. Thus, while it may not be possible to determine the dipolar signs from the individual frequency planes, the signs are resolved when information from both frequency planes are correlated.

To verify that all dipolar signs can be resolved using three-dimensional correlations, we simulated ^{15}N–^{13}C dipolar, ^{15}N–^{1}H dipolar, and ^{15}N chemical shift data from the structure of bacteriorhodopsin. As shown in Fig. 5, 9% of the ^{15}N–^{1}H dipolar signs are degenerate with two-dimensional correlation. Also, 25% of the ^{15}N–^{13}C dipolar signs are degenerate (not shown). By correlating all three data types as outlined in Fig. 10, we were able to correctly determine the dipolar signs of all data points. Thus, three-dimensional correlations are effective at removing all dipolar sign degeneracies. While obtaining the necessary data for three-dimensional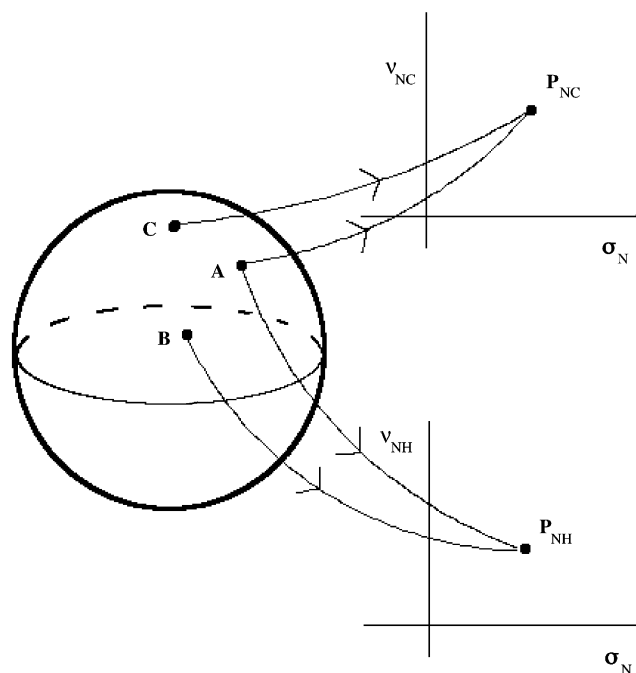 correlations is difficult, we have demonstrated that the removal of dipolar sign degeneracies leads to a better atomic refinement.

References

[1] P.J. Wallin, G. Von Heijne, Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms, Protein Sci. 7 (1998) 1029–1038.
[2] F.M. Marassi, S.J. Opella, NMR structural studies of membrane proteins, Curr. Opin. Struct. Biol. 8 (1998) 640–648.
[3] R. Fu, T.A. Cross, Solid-state NMR investigation of protein and polypeptide structure, Annu. Rev. Biophys. Biomol. Struct. 28 (1999) 235–268.
[4] R.R. Ketchem, W. Hu, T.A. Cross, High resolution conformation of gramicidin A in a lipid bilayer by solid-state NMR, Science 261 (1993) 1457–1460.
[5] R.R. Ketchem, B. Roux, T.A. Cross, High-resolution polypeptide structure in a lamellar phase lipid environment from solid-state NMR derived orientational constraints, Structure 5 (1997) 1655–1669.
[6] S.J. Opella, J. Gesell, A.P. Valente, F.M. Marassi, M. Oblatt-Montal, W. Sun, A. Ferrer-Montiel, M. Montal, Structural studies of the pore-lining segments of neurotransmitter-gated ion channels, Chemtracts-Biochem. Mol. Biol. 10 (1997) 153–174.
[7] S.J. Opella, F.M. Marassi, J.J. Gesell, A.P. Valente, Y. Kim, M. Oblatt-Montal, M. Montal, Structures of the M2 channel-lining segments from nicotinic acetylcholine and NMDA receptors by NMR spectroscopy, Nat. Struct. Biol. 6 (1999) 374–379.
[8] J. Wang, S. Kim, F. Kovacs, T.A. Cross, Structure of the transmembrane region of the M2 protein H$^+$ channel, Protein Sci. 10 (2001) 2241–2250.
[9] F.M. Marassi, A simple approach to membrane protein secondary structure and topology based on NMR spectroscopy, Biophys. J. 80 (2001) 994–1003.
[10] J.R. Quine, T.A. Cross, Protein structure in anisotropic environments: unique structural fold from orientational constraints, Concepts Magn. Reson. 12 (2000) 71–82.
[11] J.K. Denny, J. Wang, T.A. Cross, J.R. Quine, PISEMA powder patterns and PISA wheels, J. Magn. Reson. 152 (2001) 217–226.
[12] F.M. Marassi, S.J. Opella, A solid-state NMR index of membrane protein structure and topology, J. Magn. Reson. 144 (2000) 156–161.
[13] C.H. Wu, A. Ramamoorthy, S.J. Opella, High-resolution heteronuclear dipolar solid-state NMR spectroscopy, J. Magn. Reson. 109 (1994) 270–282.
[14] R. Tycko, P.L. Stewart, S.J. Opella, Peptide plane orientations determined by fundamental and overtone ^{14}N NMR, J. Am. Chem. Soc. 108 (1986) 5419–5425.
[15] S.J. Opella, P.L. Stewart, K.G. Valentine, Protein structure by solid-state NMR spectroscopy, Quart. Rev. Biophys. 19 (1987) 7–49.
[16] R.R. Ketchem, B. Roux, T.A. Cross, Computational refinement through solid state NMR and energy constraints of a membrane bound polypeptide, in: K.M. Merz, B. Roux (Eds.), Membrane Structure and Dynamics, Birkhauser, Boston, MA, 1996, pp. 299–322.

Fig. 10. Illustration of how the correlation of three pieces of information about a peptide plane can resolve dipolar signs.

[17] R. Bertram, J.R. Quine, M.S. Chapman, T.A. Cross, Atomic refinement using orientational restraints from solid-state NMR, J. Magn. Reson. 147 (2000) 9–16.

[18] 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, G.L. Warren, Crystallography & NMR system: a new software suite for macromolecular structure determination, Acta Crystallogr. D 54 (1998) 905–921.

[19] A. Royant, K. Edman, T. Ursby, E. Peybay-Peyroula, E.M. Landau, R. Neutze, Helix deformation is coupled to vectorial proton transport in bacteriorhodopsin's photocycle, Nature 406 (2000) 645–648.

[20] J. Wang, J. Denny, C. Tian, S. Kim, Y. Mo, F. Kovacs, Z. Song, K. Nishimura, Z. Gan, R. Fu, J.R. Quine, T.A. Cross, Imaging membrane protein helical wheels, J. Magn. Reson. 144 (2000) 162–167.

[21] K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Moto-shima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, M. Miyano, Crystal structure of rhodopsin: a G protein-coupled receptor, Science 289 (2000) 739–745.

[22] Y. Zhou, J.H. Morais-Cabral, R. Mackinnon, Chemistry of ion coordination and hydration revealed by a K^+ channel–Fab complex at 2.0 Å resolution, Nature 414 (2001) 43–48.

[23] A. Pautsch, G.E. Schulz, Structure of the outer membrane protein A transmembrane domain, Nat. Struct. Biol. 5 (1998) 1013–1017.

[24] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macro-molecular energy, minization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[25] G.M. Clore, A.M. Gronenborn, N. Tjandra, Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude, J. Magn. Reson. 131 (1998) 159–162.

[26] Q. Teng, T.A. Cross, The in situ determination of the ^{15}N chemical-shift tensor orientation in a polypeptide, J. Magn. Reson. 85 (1989) 439–447.