Complete Cross-Validation and *R*-Factor Calculation of a Solid-State NMR Derived Structure

Sanguk Kim,[†,§] J. R. Quine,[†,||] and T. A. Cross*[,†,§,‡]

Contribution from the Center for Interdisciplinary Magnetic Resonance, National High Magnetic Field Laboratory (NHMFL), Institute of Molecular Biophysics, Department of Chemistry, and Department of Mathematics, Florida State University, Tallahassee, Florida 32310

Received September 14, 2000. Revised Manuscript Received May 9, 2001

Abstract: Cross-validation of a solid-state NMR-derived membrane polypeptide structure is demonstrated. An initial structure has been achieved directly from solid-state NMR derived orientational restraints based on a variety of anisotropic nuclear spin interactions. Refining the molecular structure involves setting up a penalty function that incorporates all available solid-state NMR experimental data and an energy function. A validation method is required to choose the optimal weighting factor for the total penalty function to balance the contribution from the experimental restraints and the energy function. Complete cross-validation has been used to avoid over-fitting the orientational restraints. Such cross-validation involves partitioning of the experimental data into a test set and a working set followed by checking the free *R*-value during the refinement process. This approach is similar to the method used in crystallography and solution NMR. Optimizing the weighting factor on the penalty function by cross-validation will increase the quality of the refined structure from solid-state NMR data. The complete cross-validation and *R*-factor calculation is demonstrated using experimental solid-state NMR data from gramicidin A, a monovalent cation channel in lipid bilayers.

Introduction

With the advent of high-resolution macromolecular structure determination by solid-state NMR,[1−4] methods are needed to validate the structural quality. In this way overfitting of the experimental data is avoided and a method is established that permits the comparison of structures derived from different biophysical techniques. Here a complete cross-validation of the gramicidin A structure is described. Not only is this a unique three-dimensional structure determined by solid-state NMR, but it has been determined in a hydrated lipid bilayer environment above the gel to liquid-crystalline phase transition of the lipids. Moreover, this is a high-resolution structure that has led to numerous functional insights about conductance efficiency and specificity.[5] Additional functional questions with relevance to a whole class of cation channels could be addressed with further refinement and validation of this structure. Consequently, this is a unique membrane-bound structure in the Protein Data Bank (Accession code: 1MAG).

Gramicidin A is a 15-residue polypeptide with alternating D and L stereochemistry that as a dimer forms a monovalent cation selective channel across lipid bilayers. Its structure has been determined by utilizing solid-state NMR derived orientational restraints obtained from uniformly aligned bilayers of dimyristoyl-phosphatidylcholine (DMPC). The structure is a β-strand in which all of the side chains are on one side forcing the strand into a helix with 6.5 residues per turn resulting in a central pore, 4 Å in diameter. To span the bilayer two monomers are joined end to end with formyl blocked amino termini at the bilayer center.[6] Six intermolecular β-sheet hydrogen bonds stabilize this dimeric structure.

Orientational restraints are derived from anisotropic nuclear spin interactions in samples uniformly aligned with respect to the magnetic field axis. Each of these restraints orients the molecular frame of the spin interaction site with respect to both the magnetic field axis and a unique molecular axis through a $P^2\cos\theta$ dependence.[7] Furthermore, each of these precise restraints results in a set of potential molecular frame orientations.[8,9] In part, through a considerable set of restraints these degeneracies can be eliminated. For the gramicidin monomer, the experimental restraints used for defining an initial structure and for refinement are 19 ^{15}N and 2 $^{13}C_1$ anisotropic chemical shifts, 14 ^{15}N−$^{13}C_1$ and 19 ^{15}N−^{1}H dipolar couplings, 12 C_α−^{2}H, and 54 other quadrupolar couplings for a total of 120 orientational restraints.

A unique molecular fold has been defined from this set of orientational restraints that describes a right-handed β-helix and a unique set of hydrogen bonds.[1,10] Recently, both the inter- and intramolecular hydrogen bonding patterns have been

* Address correspondence to this author. E-mail: cross@magnet.fsu.edu. Phone: 850-644-0917. Fax: 850-644-1366.
† NHMFL.
§ Institute of Molecular Biophysics.
‡ Department of Chemistry.
|| Department of Mathematics.

(1) Ketchem, R. R.; Hu, W.; Cross, T. A. *Science* **1993**, *261*, 1457−1460.
(2) Ketchem, R. R.; Roux, B.; Cross, T. A. *Structure* **1997**, *5*, 1655−1669.
(3) Opella, S. J.; Marassi, F. M.; Gesell, J. J.; Valente, A. P.; Kim, Y.; Oblatt-Montal, M.; Montal, M. *Nat. Strut. Biol.* **1999**, *6*, 374−379.
(4) Davis, J. H.; Auger, M. *Prog. Nucl. Magn. Reson. Sp.* **1999**, *35*, 1−84.
(5) Tian, F.; Cross, T. A. *J. Mol. Biol.* **1999**, *285*, 1993−2003.
(6) Fu, R.; Cotten, M.; Cross, T. A. *J. Biomol. NMR* **2000**, *16*, 261−268.
(7) Cross, T. A.; Opella, S. J. *J. Am. Chem. Soc.* **1983**, *105*, 306−308.
(8) Quine, J. R.; Brenneman, M.; Cross, T. A. *Biophys. J.* **1997**, *72*, 2342−2348.
(9) Teng, Q.; Nicholson, L. K.; Cross, T. A. *J. Mol. Biol.* **1991**, *218*, 607−619.
(10) Ketchem, R. R.; Lee, K.-C.; Huo, S.; Cross, T. A. *J. Biomol. NMR* **1996**, *8*, 1−14.

10.1021/ja003380x CCC: $20.00
Published on Web 07/04/2001

confirmed by distance measurements.[6] The initial structure did not resolve all of the structural degeneracies: a set of degeneracies known as chirality ambiguities[8] have slightly different peptide plane orientations, but in all possible combinations they result in the same hydrogen bonding pattern and structural fold. The initial structure used to demonstrate the cross-validation method here was the same as the "−/+" structure in Ketchem et al.[10]

Refinement of the experimentally defined structure resolves the chirality ambiguities, relaxes the covalent geometry, and optimizes van der Waals interactions. Such refinement against the orientational restraints, hydrogen bond distances, and the CHARMM potential energy results in a unique three-dimensional structure for both the backbone and side-chains.[2] This refinement has led to a high resolution structure for a bilayer bound polypeptide or protein. However, the penalty function used in the refinement has resulted in a question about choosing the appropriate weighting factor to achieve the optimal balance between experimental restraints and the energy function. Here we have used the previously published structural restraints and refinement protocol[2] to calculate a free *R*-factor and a complete cross-validation to determine the optimal weighting factor.

R-factors represent a scaled measure of the difference between experimental and calculated observables in X-ray crystallography[11−13] and solution NMR spectroscopy.[14−16] Cross-validation is an extension of the *R*-factor method involving the partitioning of the experimental data set into a test set and a working set. Refinement of the structure is only based on the working set; and the *R*-factor calculation, known here as free *R*, is based only on the test set. Moreover, a "complete cross-validation" has been recently demonstrated for the NOE[17] and residual dipolar coupling restraints in solution NMR.[18] Unlike crystallography it is not appropriate to define one working and one test set for NMR data, because the structural restraints reflect local structure rather than global packing of proteins into a crystalline array. Consequently, the complete cross-validation involves random establishment of 10 test sets over the complete experimental data set. Cross-validation with each of these test sets results in a complete cross-validation, as we have performed here.

It is important to recognize that the restraints used in solution NMR are different from those in solid-state NMR and consequently the refinement protocol is different. Unlike residual dipolar restraints the alignment tensor in the solid-state NMR experiment is both fixed and known, and the accuracy of the restraints is high because of the large magnitude of the anisotropic interaction. Furthermore, the interpretation is also accurate, because the global motions are insignificant and local motions are quantified in three dimensions. While these are advantages for solid-state NMR, the high degree of accuracy also causes a high penalty barrier between structural possibilities making it difficult to search the conformational space completely. With relatively few restraints from solid-state NMR, an appropriate balance between experimental restraints and the potential energy function is even more important to achieve. As in solution NMR the goal here is to assess the quality of the refined structure, but as described above the critical factors that govern the quality of the structure are entirely different.

Methods

Structure Refinement. The initial structures have been refined against a generalized global penalty function including all of the orientational restraints and the CHARMM force field energy function.[19] The penalty function used to control the structural refinement is

$$\text{total penalty} = \sum_{i=1}^{M}(\lambda_i \times \text{structural penalty}_i) + \lambda_e \times \text{energy} \quad (1)$$

where M is the number of structural penalties and λ is a weighting factor. The individual structural penalties are calculated as

$$\text{structural penalty} = \sum_{j=1}^{N} \frac{1}{2}\left(\frac{\text{calculated}_j - \text{observed}_j}{\text{experimental error}_j}\right)^2 \quad (2)$$

where N is the number of measurements of a specific data type.

The use of the experimental error in the definition of the total penalty serves several purposes. In so doing the various data types are normalized by their frequency units and the magnitude of their error bar. In other words, the magnitude of each experimental error is relative to the quality of the observed interaction and, therefore, division by the error has the result of scaling the different data types so that they contribute appropriately to the total penalty. Here, ^{15}N chemical shifts have a maximum experimental error of 5 ppm; ^{15}N−^{1}H dipolar couplings have an experimental error of 2 kHz; ^{15}N−^{13}C dipolar couplings have an experimental error of 0.1 kHz; and ^{2}H quadrupolar couplings have an experimental error of 5 kHz.

The gramicidin A structure was refined using the experimentally characterized intramolecular β-strand type hydrogen bonds by including hydrogen bond distances in the penalty function, as routinely done in solution NMR structural refinement.[20] During the refinement, the intramolecular hydrogen bond distances were refined against ideal H−O and N−O distances for β-sheet structure (1.96 and 2.91 Å, respectively).[21] These distances were used with a substantial error bar of ±0.3 Å to model the range of distances observed in experimental structures.

The refined structures are obtained through a geometrical search in which the NMR observables and conformation parameters are calculated for each structural modification and compared with the observed data and the energy of the previous structure. The conformational search and evaluation is particularly difficult with the precise orientational restraints. The possible conformations are separated by very high penalty barriers, and therefore, an adequate search of the conformational space is difficult to achieve. Simulated annealing is used to perform the minimization of the penalty function and to generate a structure with minimized energy and optimized fit to the experimental data.[22,23] Modifications to the structure are made by allowing the complete geometry of the polypeptide to vary through modifications of the atomic coordinates and changes in peptide plane orientation.[24]

To search the necessary conformational and local structural space, two different types of modifications are used, since atom moves alone are not sufficient to introduce the structural modifications necessary to obtain structural agreement with the experimental data. Refinement by torsion moves allows for structural changes without consideration

(11) Hendrickson, W. A. *Methods Enzymol*. **1985**, *115*, 252−270.
(12) Brunger, A. T. *Acta Crystallogr*. **1993**, *D49*, 24−36.
(13) Brunger, A. T. *Methods Enzymol.* **1997**, *277*, 366−404.
(14) Yip, P.; Case, D. A. *J. Magn. Reson.* **1989**, *83*, 643−648.
(15) Nilges, M.; Habazettl, J.; Brunger, A. T.; Holak, T. A. *J. Mol. Biol.* **1991**, *219*, 499−510.
(16) James, T. L. *Methods Exnymol.* **1994**, *239*, 416−439.
(17) Brunger, A. T.; Clore, G. M.; Gronenborn, A. M.; Saffrich, R.; Nilges, M. *Science* **1993**, *261*, 328−331.
(18) Clore, G. M.; Garrett, D. S. *J. Am. Chem. Soc*. **1999**, *121*, 9008−9012.
(19) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem*. **1983**, *4*, 187−217.
(20) Case, D. A.; Wright, P. E. Determination of high-resolution NMR structures of proteins. In *NMR of Proteins*; CRC Press: Boca Raton, FL, 1993.
(21) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Systems*; Springer-Verlag: Berlin, Germany, 1994.
(22) Metropolis, W. L.; Kurtz, B. *J. Phys. Chem*. **1953**, *21*, 1087−1092.
(23) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. *Science* **1983**, *220*, 671−680.
(24) Ketchem, R. R.; Roux, B.; Cross, T. A. *Biological Membranes, A Molecular Perspective from Computation and Experiment*; Birkhauser, Boston, MA, 1996; pp 299−322.

of the covalent interactions and thus allows for larger structural modifications. Torsion moves are generated as compensating changes to ψ_i and ϕ_{i+1} of equal magnitude and opposite sign[25] by a random amount up to $\pm 3°$ per step. Changes in the torsion angles, which have the effect of modifying the peptide plane as a unit and providing a means by which the peptide plane conformational space is readily searched without greatly distorting the overall structure and energetic parameters, are implemented symmetrically about the chosen bond. Also random atom moves are used to relax the covalent geometry and to help minimize the global penalty. The Cartesian coordinates are altered with small diffusion parameters of 5×10^{-4} Å in each dimension. By using relatively small displacements, atom moves do not search a large conformation space but are used to generate changes in the structural penalty and to introduce minor atomic modifications to find a better match to the experimental data and energy parameters.

Acceptance of an attempted move is controlled by both the temperature and the difference in the penalty before and after the attempted move. A move that causes a decrease in the penalty is always accepted. A move that increases the penalty is only accepted if a random number between 0 and 1 is less than $\exp(-\Delta\text{penalty}/T)$. The simulated annealing refinement procedure is controlled by a temperature parameter and an annealing schedule. The Monte Carlo algorithm generates configurations corresponding to the Boltzmann distribution of canonical ensemble at a given temperature. The global minimization is controlled by an annealing schedule, i.e., the rate at which the temperature is lowered during the course of the refinement. The refinement strategy used in this study is to introduce minor structural modifications to the initial structure. Large changes would lead to conformational space that has already been shown to be excluded through the development of the initial structure. Therefore, the initial value of the temperature is set at 300 K for refinement by both torsion and atom moves. The system configuration undergoes 2000 modificatoins or 200 successful modifications, whichever is first, before the temperature is lowered by 1%. The refinement is terminated when no successful structural modification is found at a particular temperature after 2000 attempts, or after 500 temperature steps.

Increasing the weighting factor for the structural penalty leads to the over-interpretation of experimental data and decreases the structural fidelity based on the chemical knowledge of the system. Also, increasing the weighting factor for the structural penalty decreases the number of accepted modifications. During the refinement the structural penalty and energy penalty were both reduced, but they compete for control of the structural modifications. Therefore, the weighting factor for the structural penalty can increase the structural precision, here estimated by evaluating the atomic rms differences from their mean positions.[26]

***R*-Factor Calculation for Solid-State NMR.** Traditionally, the *R*-factor employed in crystallography is defined as $\sum||\text{obs}| - |\text{calc}||/\sum|\text{obs}|$ where "obs" and "calc" are the observed and calculated experimental values, respectively. For normalizing the frequency units of the various experimental data types, a more appropriate definition of the *R*-factor for solid-state NMR data is given by

$$R = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|\text{calculated}_i - \text{observed}_i|}{\text{experimental error}_i}\right)^2 \qquad (3)$$

where N is the number of structural penalties. This definition results in a larger *R* value because the magnitude of the error is much less than the magnitude of the observable. But, it certainly helps to normalize the *R* value of the various data types with their frequency units.

To obtain a high-quality structure using the solid-state NMR data, we need to choose an optimal weighting factor for the structural penalty. The *R* value for an atomic model refined against the total penalty is a function of the weighting factor. The *R* value decreases in proportion to the weighting factor. Thus, the *R* value is not particularly useful for evaluating the optimal magnitude of the weighting factor. But, the free *R* is the agreement between the subset of observed and calculated experimental data, which do not participate in the refinement. Using free *R* enables us to assess how well the chemical shifts and the dipolar couplings in the test set are predicted by those in the working set. Therefore, we can estimate the quality of the fit to the experimental data.

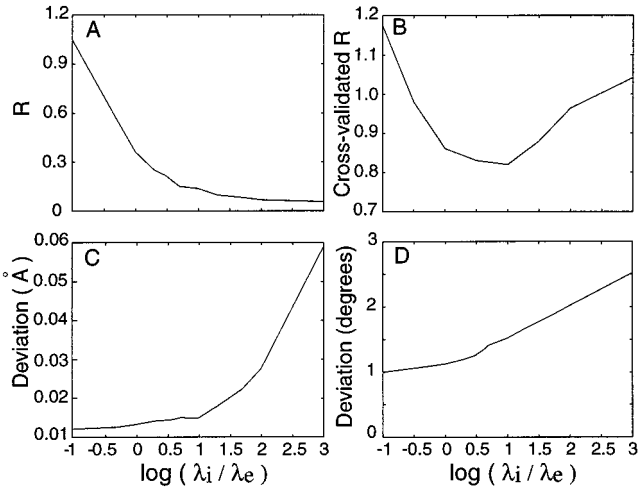

Figure 1. Influence of the structural penalty weighting factor λ_i on the refinement of the gramicidin A structure using solid-state NMR derived orientational restraints. $\lambda_e = 1$ kcal mol^{-1} and the structural penalties have dimensionless units. The orientational restraints are partitioned into 10 pairs of test and working sets comprising 10 and 90% of the data, respectively. (A) Weighted *R* value. (B) Cross-validated weighted *R* value. (C) Deviation of backbone bond length from ideality. (D) Deviation of backbone bond angles from ideality.

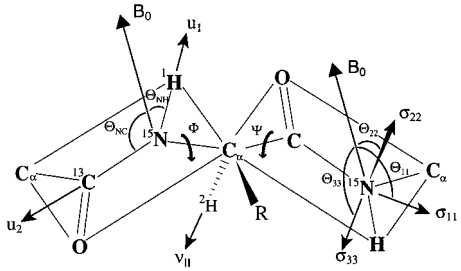

Figure 2. Orientational restraints in a peptide plane. The primary structural restraints for the polypeptide backbone are derived from the $^{15}N-^{1}H$ and $^{15}N-^{13}C$ dipolar interactions, the $C_\alpha-^{2}H$ quadrupolar interaction, and the anisotropic ^{15}N chemical shift. The peptide plane orientation can be initially determined by two bond vectors, N−H and N−C. Once the orientations of the individual peptide planes have been determined with respect to the external magnetic field, the orientations of the planes with respect to each other are determined by joining the plane through the shared C_α carbon which has a well-defined covalent geometry. This allows for the determination of the (ϕ,ψ) torsion angles.

Results and Discussion

Complete Cross-Validation of Solid-State NMR Data. Figure 1 shows the results from a series of refinements with 10 pairs of working and test data sets chosen from the backbone and partitioned in a ratio of 90% to 10% in which the weighting factor, λ_i for the chemical shifts and dipolar couplings, is varied from 0.1 to 1000 kcal mol^{-1}. Complete cross-validation consists of a partitioning of the chemical shifts and dipolar couplings into 10 test sets of roughly equal size. In this effort, ^{15}N chemical shifts and $^{15}N-^{1}H$, $^{15}N-^{13}C$ dipolar couplings from the backbone structure (Figure 2) have been divided for the test sets. When the test sets are partitioned, no more than two experimental restraints are selected from any one peptide plane for a given test set. Each working set is defined by excluding just one of the test sets from the whole data set. All ^{2}H quadrupolar interactions from the side chains are used in each working set during cross-validation of the backbone structure.

(25) Peticolas, W. L.; Kurtz, B. *Biopolymers* **1980**, *19*, 1153−1116.
(26) Havel, T. F.; Wüthrich, K. *J. Mol. Biol.* **1985**, *182*, 281−294.

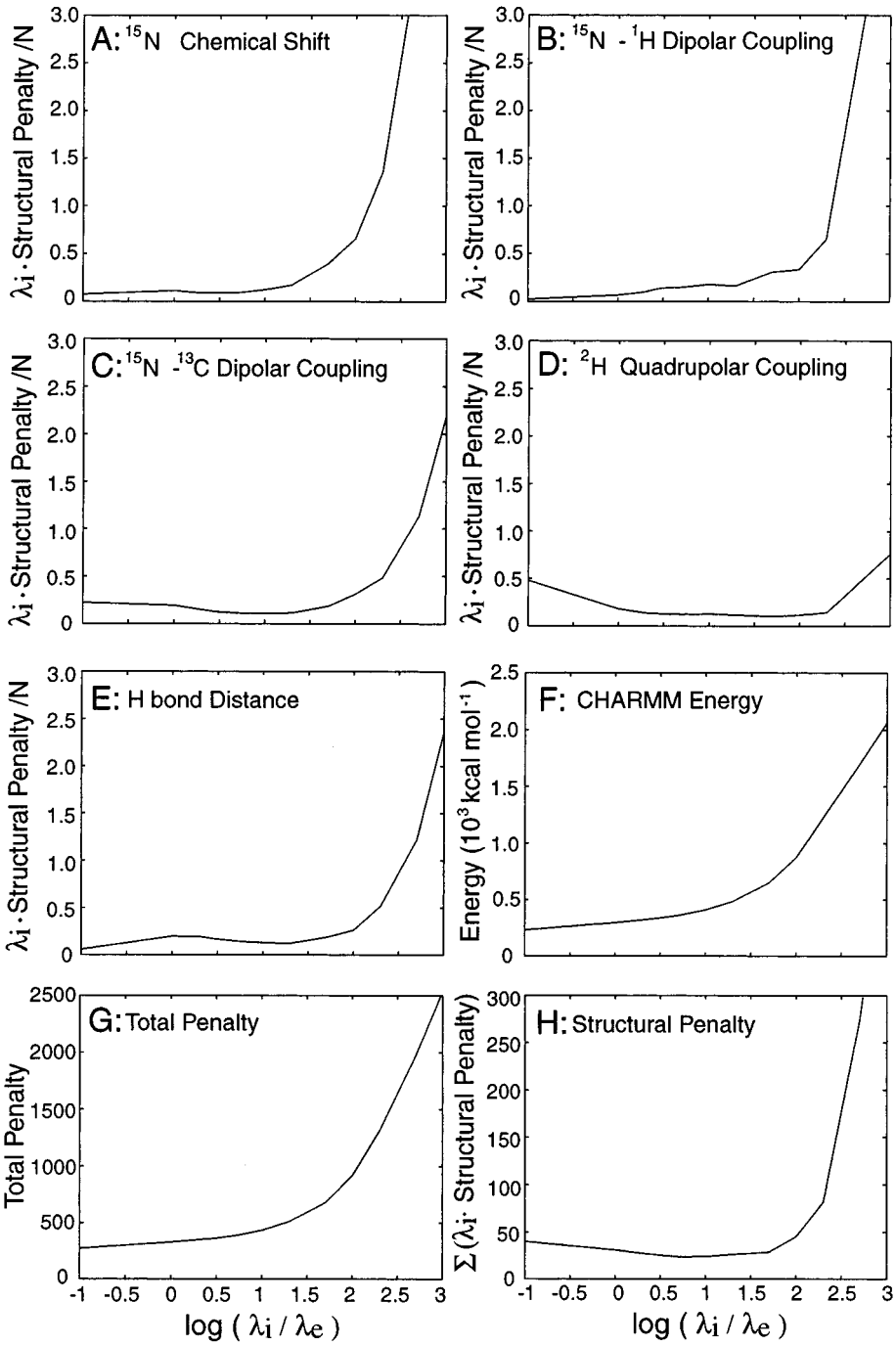

Figure 3. Influence of the weighting factor λ_i for the individual structural penalties, calculated energy values, and the total penalties on the refined structures. The λ_i weighted structural penalties are calculated for values of -1 to 3 of $\log(\lambda_i/\lambda_e)$ and divided by the number of restraint types used for the refinement. The penalty values multiplied by the λ values reflect the final penalty used during refinement and not the calculated penalty of the individual restraints. (A) ^{15}N chemical shifts. (B) ^{15}N$-^{1}$H dipolar couplings. (C)^{15}N$-^{13}$C dipolar couplings. (D) ^{2}H quadrupolar couplings. (E) Hydrogen bond distance. (F) The potential energy calculated from the refined structures. (G) The total penalty values [$\sum(\lambda_i \times$ structural penalty) $+ \sum(\lambda_e \times$ energy)]. (H) The structural penalty.

In this way, the effect of restraints on the *R*-factor for the backbone and side chains are separated.

The *R* value (Figure 1A) decreases as a function of λ_i, whereas the deviation from ideality for the bond lengths (Figure 1C) and bond angles (Figure 1D) in the refined structures increases as a function of λ_i. A λ_i value of 10 kcal mol^{-1} optimizes the cross-validated *R* value (Figure 1B). These data show that the structural quality is significantly decreased above the cross-validated *R* value minimum. In the observation of bond length and angle standard deviations in atomic resolution crystal structures of small molecules,[27,28] 0.02 Å for bond length deviations and 2° for bond angle deviations are reasonable. At a λ_i value of 10 kcal mol^{-1}, the average backbone structural violations are under this range. Overfitting at large λ_i results in decreased backbone structural fidelity. Thus, the cross-validated *R* value shows the optimal weighting factor for the structural penalty to avoid overfitting.

Figure 3 analyzes the penalty function contributions as a function of the weighting factor by the weighted individual

(27) Engh, R. A.; Huber, R. *Acta Crystallogr.* **1991**, *A47*, 392−400.
(28) Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.

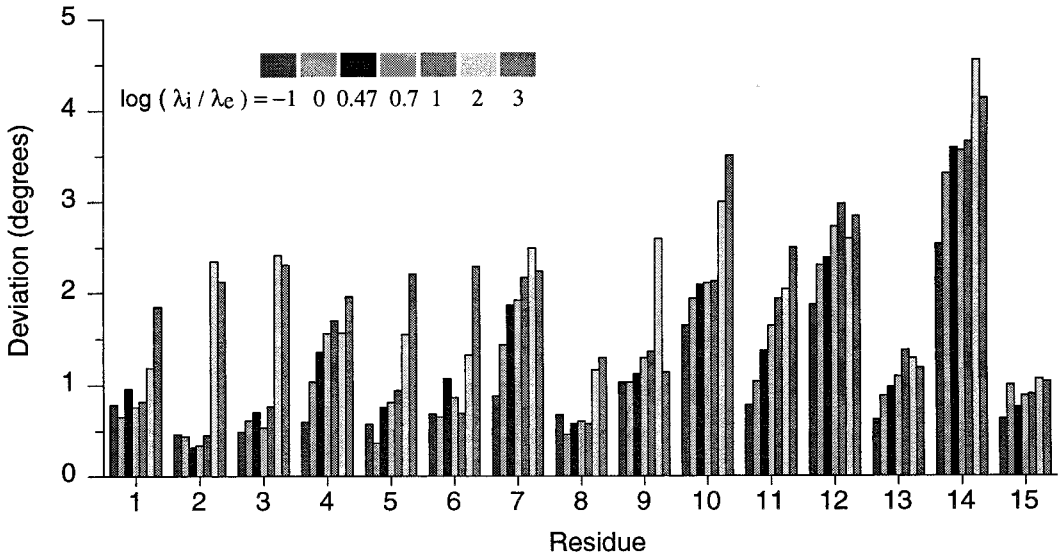

Figure 4. (A) Average residual bond angle deviations from ideality monitored for values of −1 to 3 of log(λ_i/λ_e). Gramicidin A residues: CHO-LVal$_1$-Gly$_2$-LAla$_3$-DLeu$_4$-LAla$_5$-DVal$_6$-LVal$_7$-DVal$_8$-LTrp$_9$-DLeu$_{10}$-LTrp$_{11}$-DLeu$_{12}$-LTrp$_{13}$-DLeu$_{14}$-LTrp$_{15}$-NHCH$_2$CH$_2$OH.

orientational restraint types from the total penalty. These structural penalties multiplied by λ_i for ^{15}N chemical shifts, ^{15}N−^{1}H dipolar couplings, ^{15}N−^{13}C dipolar couplings, and ^{2}H quadrupolar couplings are minimized well during the refinement process for a λ_i value of approximately 10 kcal mol^{-1}. But, above this optimum, the refinement process does not effectively minimize the experimental restraints. Below the optimal value the penalty for experimental restraints typically decreases with increasing weighting factor. At the optimal weighting factor for ^{15}N chemical shifts (Figure 3A) and ^{15}N−^{1}H dipolar couplings (Figure 3B) the penalty values are much smaller than the summed error. Also, ^{15}N−^{13}C dipolar couplings (Figure 3C) and ^{2}H quadrupolar couplings (Figure 3D) show a minimal weighted penalty value near the optimal λ_i value. The λ_i weighted hydrogen bond distance (Figure 3E) penalty also shows a minimum at the optimal λ_i. At a small value of the weighting factor the hydrogen bond penalty increases with λ_i, because ideal bond distances were forced to change by incorporating other structural restraints. But again, the hydrogen bond penalty is minimized at an optimal weighting of the structural penalties. In the final model using an optimal weighting factor, the rms difference for all intramolecular hydrogen bonds from ideality is 0.09 Å.

Although inclusion of the CHARMM structural energy has been important for correcting the structure in terms of covalent geometry and nonbond interactions, it influences the structure in a way that may not correspond exactly to the influence from the experimental data. The energy is calculated for the molecule in a vacuum while the experimental data have been obtained from a hydrated lipid bilayer environment. The penalty due to the experimental data is reduced to a point that is very close to zero during refinement. While the calculated energy is significantly reduced during refinement, it can be reduced further in the absence of the experimental restraints.[2] This indicates that although the penalties from the two different types of restraints are both necessary and are reduced during refinement, they compete for control of the structural modifications. Panels F and G in Figure 3 trace the calculated energy and the total penalty value from the refined model. These values increase with increasing λ_i for the structural penalties, because the potential energy calculation of the structure uses empirical energy functions, in this study, CHARMM. Overfitting of the experimental restraints leads to structural violations, including bond and nonbond interactions. The difference between the total penalty value and the energy is the total structural penalty shown in Figure 3H that has a minimum at $\lambda_i = 10$ kcal mol^{-1}. As shown in Figure 1B in which the cross-validated R has a minimum, there is a good correlation between the weighted structural penalty and the cross-validated *R*. Thus, the analysis of the λ_i weighted structural penalty could be used as another indicator to check the optimal weighting factor for experimental data.

Structural Quality Assessment. Figure 4 analyzes the averaged residual angle deviations from ideality as a function of residue and weighting factor. These angle deviations include all the backbone angles: C−N−C$_\alpha$, C$_\alpha$−C−N, C$_\alpha$−C−O, C$_\beta$−C$_\alpha$−C, N−C$_\alpha$−C, N−C$_\alpha$−C$_\beta$, and O−C−N. All residues display small violations at the optimal weighting value of the structural penalty, except for several D-Leu residues that have slightly higher deviations. Each of these leucines has tryptophans on either side resulting in tight packing of the side chains in this region of the polypeptide. These leucine spacers between the tryptophans have recently been recognized as important residues for ion permeability and channel lifetime, presumably through aliphatic/aromatic interactions.[29,30] These results suggest that for functional reasons there may be induced strain in the backbone structure.

Moreover, the leucine side chain data in the form of ^{2}H quadrupolar interactions represent very challenging data to interpret. This challenge is embedded in the task of separating dynamic and structural influences on the spectroscopic data. The small-amplitude librational motions in the backbone have been well characterized by both variable-temperature powder pattern analysis and field-dependent ^{15}N relaxation rates.[31,32] As this analysis moves out onto the side chains, it becomes more complex with additional and larger amplitude motions about the χ_1 and χ_2 torsion angles.[33,34] The results here suggest

(29) Jude, A. R.; Greathouse, D. V.; Koeppe, R. E.; Providence, L. L.; Andersen, O. S. *Biochemistry* **1999**, *38*, 1030−1039.
(30) Koeppe, R. E.; Hatchett, J.; Jude, A. R.; Providence, L. L.; Anderson, O. S. *Biochemistry* **2000**, *39*, 2235−2242.
(31) North, C. L.; Cross, T. A. *Biochemistry* **1995**, *34*, 5883−5895.
(32) Lazo, N. D.; Hu, W.; Cross, T. A. *J. Magn. Reson.* **1995**, *107B*, 43−50.
(33) Lee, K.-C.; Cross, T. A. *Biophys. J.* **1994**, *66*, 1380−1387.
(34) Lee, K.-C.; Huo, S.; Cross, T. A. *Biochemistry* **1995**, *34*, 857−867.

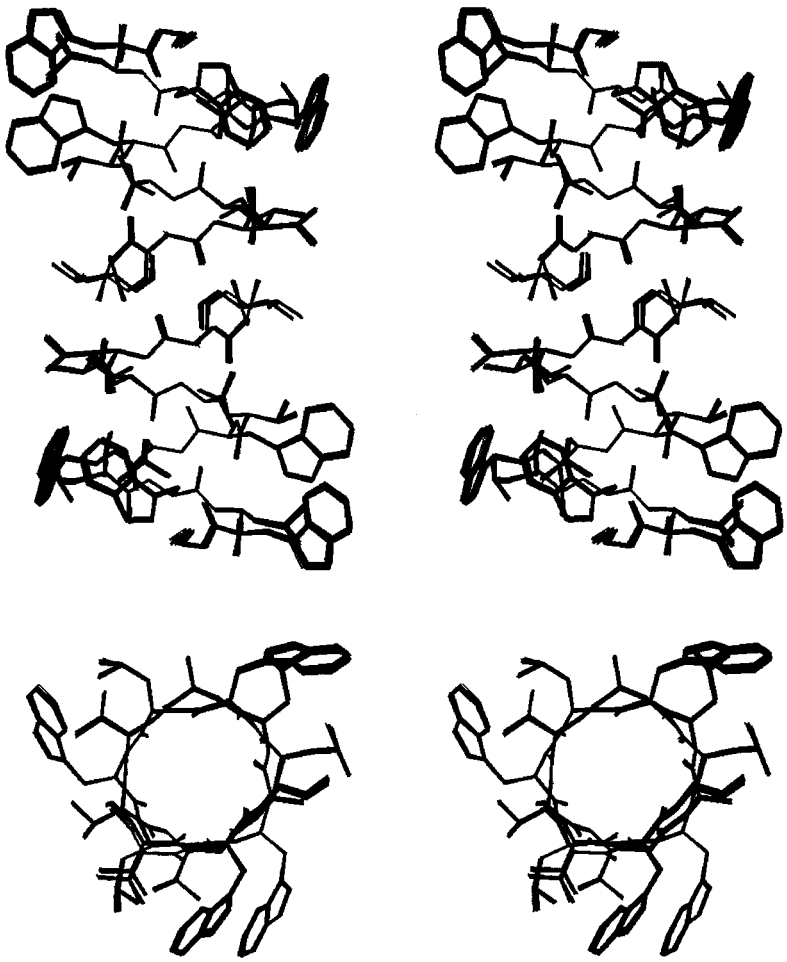

Figure 5. A stereoview of the refined structure of gramicidin A. The 40 simulated annealing structures are superimposed (heavy atoms are shown). The weighting factor for structural penalty, $\lambda_i = 10$ kcal mol^{-1}, and the weighting factor for potential funciton, $\lambda_e = 1$ kcal mol^{-1}, are used for the calculation. The root-mean-square deviation of the backbone coordinates from the refined ensembles of structures is 0.11 Å. This backbone root-mean-square deviation is calculated between individual refined structures and the mean coordinate structure. The side view (top) shows the symmetric dimer that spans the lipid bilayer. The end view (bottom) of a monomer clearly shows the channel pore that supports a single file column of water molecules.

that the dynamic and structural analysis is not yet optimally resolved and consequently backbone distortions are induced. Clearly the weighted structural penalty for ^{2}H orientational restraints (Figure 3D) and the energetics of the backbone sites associated with many of these side chain restraints are in potential conflict and may be in need of slight adjustment.

Figure 5 shows a set of 40 simulated annealing structures of gramicidin A obtained with the optimal weighting factor. The side view of the symmetric dimer and the end view of the monomer are shown to illustrate the lipid bilayer spanning structure and channel pore formation. The rms deviation of the backbone atoms between individual refined structures from their mean coordinate positions is just 0.11 Å. The precision of the orientational restraints is very effective in restraining this structure.

The orientational restraints correspond to first-order average Hamiltonian quantities and, consequently, the orientational restraints generate accurate and precise structural restraints. Another advantage of the orientational restraints is that the errors associated with each of the restraints do not sum as the structure is assembled.[10,35] This is because each restraint orients the molecule with respect to the laboratory frame of reference, an absolute restraint, whereas distances between atomic sites within a macromolecule, such as NOE restraints, represent relative restraints. Therefore, each orientational restraint is an independent structural restraint, while distance restraints are dependent structural restraints.

In solution NMR, residual dipolar couplings measured on partially oriented protein samples have been introduced recently into structural determination.[36,37] Each dipolar coupling serves as an orientational restraint for defining the orientation of a structural element relative to the alignment axis fixed in the molecular frame. It has been shown that using dipolar couplings from the protein backbone provides restraints that narrowly restrict the range of accessible ψ and ϕ backbone torsion angles.[38,39] From refinement ensembles, it has been shown that the precision of the refined backbone structure is greatly improved with dipolar couplings compared to the calculation with NOE restraints alone.[40] In addition, it has recently been shown that through the use of residual dipolar coupling restraints

(35) Cross, T. A.; Ketchem, R. R.; Hu, W.; Lee, K.-C.; Lazo, N. D.; North, C. L. *Bull. Magn. Reson.* **1992**, *14*, 96−101.

(36) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci.* **1995**, *92*, 9279−9283.

(37) Tjandra. N.; Bax, A. *Science* **1997**, *278*, 1111−1114.

(38) Bewley, C. A.; Gustafson, K. R.; Boyd, M. R.; Covell, D. G.; Bax, A.; Clore, G. M.; Gronenborn, A. M. *Nat. Struct. Biol.* **1998**, *5*, 571−578.

(39) Cai, M.; Huang, Y.; Zheng, R.; Wei, S.-Q.; Ghuirlando, R.; Lee, M. S.; Graigie, R.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1998**, *5*, 903−909.

(40) Clore, G. M.; Starich, M. R.; Bewley, C. A.; Cai, M.; Kuszewski, J. *J. Am. Chem. Soc.* **1999**, *121*, 6513−6514.

from the alignment media a unique structure can be achieved without the use of NOE data.[41] It has been shown previously that such a unique structure can be achieved with a substantial set of orientational restraints from a single alignment preparation in solid-state NMR.[1,2] These solid-state NMR restraints are so effective because of the lack of global motion and the well-characterized local motions. Here the orientational restraints are even more effective because of the near-maximal order parameters in our samples.

Conclusion

Complete cross-validation of a structure entirely characterized by orientational restraints has been achieved for the first time. A minimum in the free *R*-factor as a function of the relative weighting of experimental and energy restraints has resulted in an optimal weighting factor such that neither under- nor over-fitting of these restraints occurs during refinement. The cross-validation result further justifies the assessment of the structural quality. The rms deviation of 0.11 Å among the backbone atoms represents high precision for a membrane protein or polypeptide structural characterization in a lipid bilayer environment.

Acknowledgment. This work was primarily supported by the National Science Foundation, DMB 9986036 (T.A.C. and J.R.Q.), and the work was largely performed at the National High Magnetic Field Laboratory supported by Cooperative Agreement (DMR-9527035) and the State of Florida.

JA003380X

(41) Hus, J. C.; Marion, D.; Blackledge, M. *J. Am. Chem. Soc*. **2001**, *123*, 1541−1542.