Refinement of Protein NMR Structure under Membrane-like Environments with an Implicit Solvent Model

JunGoo Jee* **and Hee-Chul Ahn***

Center for Priority Areas, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0373, Japan
**E-mail: jee-jungoo@tmu.ac.jp*
**Advanced Analysis Center, Korea Institute of Science and Technology, Sungbuk-gu, Seoul 136-791, Korea*
Received February 18, 2009. Accepted April, 6, 2009

Refinement of NMR structures by molecular dynamics (MD) simulations with a solvent model has improved the structural quality. In this study, we applied MD refinement with the generalized Born (GB) implicit solvent model to protein structure determined under membrane-like environments. Despite popularity of the GB model, its applications to the refinement of NMR structures of hydrophobic proteins, in which detergents or organic solvents enclose proteins, are limited, and there is little information on the use of another GB parameter for these cases. We carried out MD refinement of crambin NMR structure in dodecylphosphocholine (DPC) micelles (Ahn *et al.*, *J. Am. Chem. Soc.* **2006**, *128*, 4398-4404) with GB/Surface area model and two different surface tension coefficients, one for aquatic and the other for hydrophobic conditions. Our data show that, of two structures by MD refinement with GB model, the one refined with the parameter to consider hydrophobic condition had the better qualities in terms of precision and solvent accessibility.

Key Words: NMR, Implicit solvent, Molecular dynamics simulation, Generalized Born model, Force field

Introduction

Rapid advances in calculation power and algorithms have made it practicable to employ molecular dynamics (MD) simulations with a sophisticated force field and a solvent model to refine NMR structures. However, the use of an explicit solvent model, in which solvent molecules such as water are added to a solute, is still demanding because of too large computing time to calculate an NMR ensemble comprising several ten structures. Instead, an implicit solvent model, where continuous media mimic the solvent's effects, has been successfully exploited. The generalized Born (GB) model, an implicit solvent model in most cases, has improved the qualities of NMR structures to levels similar to those achieved with explicit water refinement.[1]

MD refinement with the GB model is effective especially to confine the regions that do not converge only with NOE and torsion angle restraints by applying forces between atoms in the solute, and the solute and the solvent. For instance, the regions where hydrogen bonds and electrostatic interactions are dominant often suffer from a lack of NOE restraints, and MD refinement is helpful in such cases. The fast exchange feature of the weak protein complex also hinders gaining sufficient NOE restraints to confine the geometries of the interfaces. Our previous data showed that MD refinement with an AMBER force field and the GB model revealed better geometries of the interfaces in weak protein complexes.[2-4] Furthermore, the GB model can be extended to find the correct 3D fold of proteins with only sparse NOE restraints by using the advanced sampling method, replica exchange.[5]

Similar difficulties confront us in determining the membrane protein structure using solution NMR spectroscopy. Even in NMR samples that show well-dispersed backbone 2D [^{15}N, ^{1}H] spectra, the signals of the side-chains are often invisible or severely overlapped due to the inhomogeneity of the samples under membrane-like environments. Many side-chains lack NOE restraints on which NMR is heavily dependent for the calculation of structure.[6] In several cases, extra long-range backbone restraints from paramagnetic relaxation enhancement (PRE) and residual dipolar coupling (RDC) experiments have aided the improvement of NOE-based structures.[7,8] Compared with those of other protein NMR structures in solution, however, the qualities are still worse, and unrefined conformations of side-chains impede clear understanding of their roles.

We have been inspired to apply MD refinement with the GB model to calculating membrane protein structures. However, only a limited number of membrane protein structures have been solved by NMR (http://www.drorlist.com/nmr/MPNMR.html). There is no report on refinement by the GB model, while there have been several MD simulation and folding studies of membrane proteins using the GB model.[9,10] Before applying MD refinement to membrane proteins, it is necessary to know whether the GB model with default parameters for hydrophilic conditions can improve NMR structures under membrane-like environments or whether a new hydrophobic parameter is needed.

We chose crambin in dodecylphosphocholine (DPC) micelles as a model protein. Crambin is a hydrophobic and water-insoluble protein consisting of 46 residues. Organic solvents or detergents such as DPC are required to solubilize crambin. The NMR structure of crambin under DPC micelles (crambin^{DPC}) was solved recently by multidimensional NMR experiments (PDB ID: 1YVA).[11] The paramagnetic NMR experiment using Mn^{2+} showed that most of crambin's regions are surrounded by DPC micelles and thus are hydrophobic, except for small parts of the loop region.[11] In this study, we carried out MD refinement of crambin NMR structure with GB/Surface area model and two different surface tension coefficients.

Experimentals

We extracted the experimental distance and torsion angle restraints (1YVA.mr) from PDB database (http://www.rcsb.org). The restraints consisted of 637 NOE, 76 torsion angles and 38 hydrogen bonds restraints.[11] We first used CYANA (version 2.1)[12] to generate 100 structures that satisfy experimental restraints. The torsion angle dynamics step in CYANA was 15,000. Then, we performed MD refinement with experimental NMR restraints, and an all-atom force field and the GB model. We used AMBER force field (ff03) and AMBER 9 software package.[13,14] This MD refinement consisted of three stages: energy minimization with 1500 steps, molecular dynamics of 20 ps, and energy minimization with 1500 steps.[2-4] Among the 100 structures, 20 structures showing both the lowest AMBER energies and no significant violation in distance (< 0.3Å) and torsion angle (< 3°) restraint were selected as the final ensemble. For the experimental distance and torsion angle restraints, we used the force constants of 50 kcal·mol^{-1}·Å^{-2} and 200 kcal·mol^{-1}·rad^{-2}, respectively.

Results and Discussion

The GB model for current study. AMBER 9 can implement 6 GB models (*igb* = 1,2,5,6,7,10). For the GB model in the current study, we selected the GB/Surface Area (GB/SA, *igb* = 2, GB^{OBC}) model.[15] Detailed explanation of all the GB models is beyond the scope of this study. We briefly describe the basic idea employed for the current study. In the GB/SA model, the free energy of solvation, ΔG_{solv}, can be decomposed further into polar, ΔG_{polar}, and non-polar, $\Delta G_{nonpolar}$, contributions as $\Delta G_{solv} = \Delta G_{polar} + \Delta G_{nonpolar}$. The ΔG_{polar} term is based on the generalized Born formula derived from the Poisson-Boltzmann equation. The $\Delta G_{nonpolar}$ term is the product of the surface tension coefficient (*surften*, kcal·mol^{-1}·Å^{-2}) and the surface area (*SA*, Å^2), and is expressed as $\Delta G_{nonpolar}$ = *surften***SA*. Therefore, by increasing the value of *surften*, we can increase the contribution of the non-polar term. By setting the values of *surften* to 0.005 and 0.04, we repeated MD refinement with the same experimental restraints. The default value of *surften* in AMBER 9 was 0.005 for the aquatic condition, and the value of 0.04 was adapted from the folding study of the membrane protein using the GB model.[16]

MD refinement improved the qualities of crambin structures. Table 1 shows the statistics of the final structures in crambin^{DPC}, *surften* = 0.005 (crambin^{GBaqu}), and *surften* = 0.04 (crambin^{GBmem}). The refined structures did not show a large deviation from the reference ones. The pairwise differences in the backbone RMSD values between crambin^{DPC} and crambin^{GBaqu}, and crambin^{DPC} and crambin^{GBmem} were 1.26 ± 0.26 Å and 1.26 ± 0.28 Å, respectively. Crambin^{GBaqu} and crambin^{GBmem} were very similar to each other, showing pairwise RMSD values of 0.90 ± 0.28 Å. Both structures by MD refinement exhibited improvements in precision and accuracy. The mean RMSD values of the backbone (heavy) atoms in residues 1-46 improved from 0.70 ± 0.11 Å (1.02 ± 0.11 Å) for crambin^{DPC} to 0.61 ± 0.19 Å (0.88 ± 0.17 Å) for crambin^{GBaqu} and 0.50 ± 0.16 Å (0.75 ± 0.14 Å) for crambin^{GBmem} (Table 1). Further, the most favored region in the Ramachandran analysis improved from 79.6% to 91.0% and 89.7% for crambin^{GBaqu} and crambin^{GBmem}, respectively (Table 1). This result demonstrates the robustness of including MD refinement for NMR structure calculations and is consistent with our previous data for water-soluble proteins.[2-4]

It is noticeable that the structures by MD with the GB model in the current study outdid those by the most popular NMR refinement protocols using DMSO (PDB ID: 2EYC) and water (PDB ID: 2EYD) explicit solvents of CNS/ARIA[17] in terms of qualities. The mean backbone RMSD values in the residues 1-46 were 0.89 and 0.94 by DMSO and water-refined crambin structures, respectively.[11] We reason that this refinement method with explicit solvents does not use sophisticated protocols in the cut-off value for non-bond interactions and the MD schedule due to the compromise between performance and CPU efficiency.

Residue-specific solvent accessible surface areas revealed the differences in each structure. To obtain detailed insight into the differences in these structures, we calculated the

Table 1. Statistics of crambin^{DPC}, crambin^{GBaqu}, and crambin^{GBmem}

	crambin^{DPC}	crambin^{GBaqu}	crambin^{GBmem}
Energy (kcal/mol)			
AMBER GB energy	Not	−1068 ± 3	−967 ± 4
Constraints energy	available	14 ± 2	14 ± 1
Maximum Violation			
Distance violation (Å)	0.430	0.265	0.252
Angle violation (degree)	0.94	0.0	0.0
Mean RMSD (Å)			
Backbone	0.70 ± 0.11	0.61 ± 0.19	0.50 ± 0.16
Heavy atoms	1.02 ± 0.11	0.88 ± 0.17	0.75 ± 0.14
Ramachandran (%)			
Most favored region	79.6	91.1	90.0
Additionally allowed region	17.7	8.9	10.0
Generously allowed region	2.0	0.0	0.0
Disallowed region	0.7	0.0	0.0

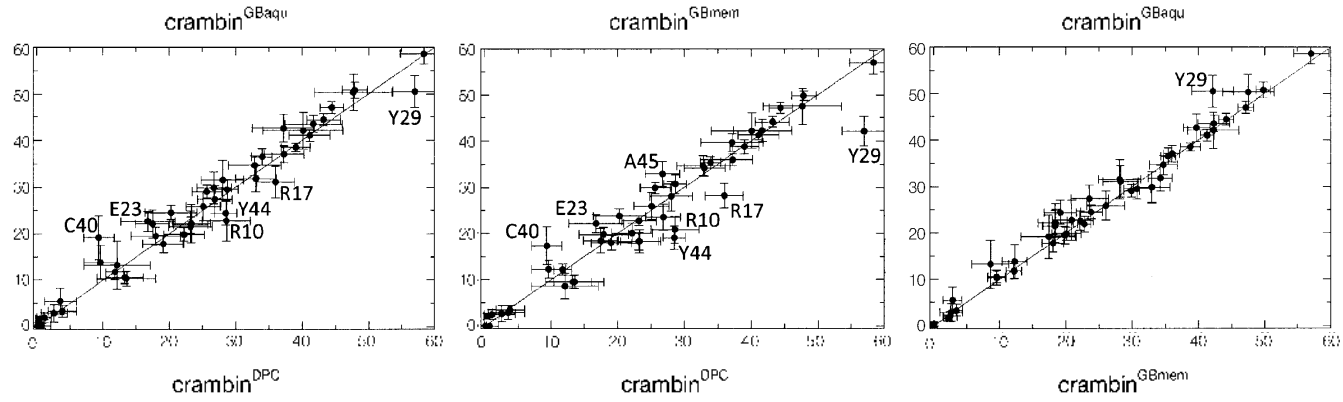

Figure 1. Pairwise comparisons of residue-specific solvent accessible surface area (RASA) values among crambin^{DPC} (PDB ID: 1YVA), crambin^{GBaqu}, and crambin^{GBmem}. MOLMOL[20] was used to calculate the RASA values. Both mean (dot) and standard deviations (bar) are shown. Labels for the x-axes are written at the bottoms of the figures, whereas the labels for the y-axes are drawn at the top. The "$y = x$" line is also drawn as a reference. Significant outliers are labeled. The unit of each axis is percentage (%).

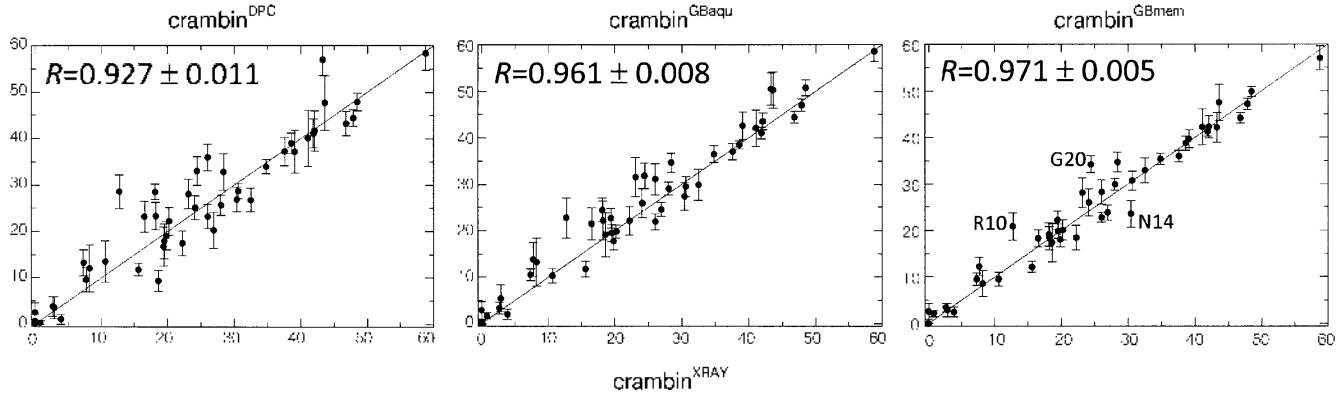

Figure 2. Pairwise comparisons of RASA values between crambin^{DPC}, crambin^{GBaqu} and crambin^{GBmem}, and crambin^{XRAY} (PDB ID: 1EJG). Figures were generated with the same notations as Figure 1. The Pearson correlation coefficient (R) values were calculated with the MATLAB package (Mathworks Inc., MA) with 1000 Monte Carlo cycles. For clarity, only outliers in crambin^{GBmem} are shown.

residue-specific solvent accessible surface area (RASA). If solvents can contact a residue freely, the RASA value is 100%. In contrast, if a residue is perfectly blocked from solvents, the RASA value is 0%. Therefore, RASA values are good indices to know the effects of an implicit solvent in the refinement. Pairwise comparisons of the RASA values among crambin^{DPC}, crambin^{GBaqu}, and crambin^{GBmem} were calculated (Figure 1). As predicted from the statistics (Table 1), there was a small difference in the pairwise RASA (pRASA) value of crambin^{GBaqu}-crambin^{GBmem}, while the pRASA value of crambin^{DPC}-crambin^{GBaqu} and crambin^{DPC}-crambin^{GBmem} revealed several outliers. However, it should be noted that there was a tendency for the RASA values of crambin^{GBaqu} to be slightly greater than those of crambin^{GBmem}, although the difference was not large (Figure 1). This implies that crambin^{GBaqu} favored solvents more, and shows the role of *surften* term in the GB/SA model. Among the outliers in the pRASA values between crambin^{DPC} and the structures refined by the GB model, the most distinct one was Tyr-29 because it is a part of the α-helix and structural changes due to MD refinement are usually not large in the α-helices and β-sheets. Tyr-29 was also the sole outlier in the pRASA of crambin^{GBaqu}-crambin^{GBmem}. Intriguingly, Tyr-29 was one of the residues that showed the largest chemical shift change in the 2D [^{15}N, ^{1}H] HSQC of crambin in DPC and dihexanoyl phosphatidylcholine (DHPC) micelles;[11] this implies that it was sensitive to environmental conditions even in actual situations. On the other hand, other big outliers, namely, Arg-17, Cys-40, and Tyr-44, are located in bent or loop regions, where conformations change occasionally due to MD refinement.

Residue-specific solvent accessible surface areas between crambin^{GBmem} and crambin^{XRAY} were in better agreement. We further compared the RASA values of these structures with that of the X-ray crambin structure (PDB ID: 1EJG, crambin^{XRAY}), which was crystallized under an ethanol and water mixture solution.[18] Crambin^{XRAY} is a fine structure with the best resolution (0.54 Å) among the available crambin structures. The RASA values of crambin^{GBmem} showed the best Pearson correlation coefficient (R) value of 0.971 ± 0.005 over crambin^{XRAY}, while crambin^{DPC} and crambin^{GBaqu} revealed the values of 0.927 ± 0.011 and 0.961 ± 0.008, respectively (Figure 2). In particular, most of the outliers that were found in the pRASA values of crambin^{DPC}-crambin^{GBmem} disappeared in that of crambin^{XRAY}-crambin^{GBmem}. This similarity between crambin^{XRAY} and crambin^{GBmem} can also be visualized by inspecting the side-chain orientations of aromatic residues. The orientations

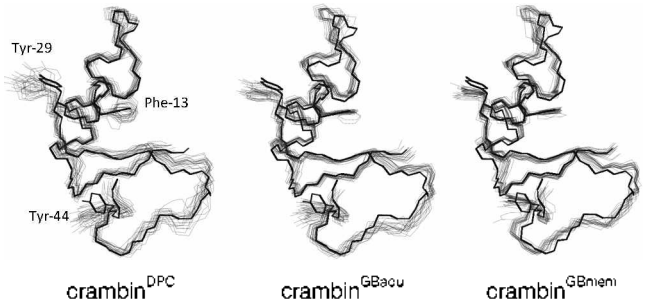

Figure 3. Overlay figures of crambin^{DPC}, crambin^{GBaqu}, and crambin^{GBmem} against crambin^{XRAY}. MOLMOL was used to generate these figures. Crambin^{XRAY} is presented with bold lines and the others with narrow lines. Side-chains of aromatic residues, Phe-13, Tyr-29, and Tyr-44, are also shown.

Table 2. χ1 angle statistics of aromatic residues

	crambin^{XRAY}	crambin^{DPC}	crambin^{GBaqu}	crambin^{GBmem}
Phe-13	179.6	154.6 ± 12.7	173.5 ± 6.7	171.4 ± 6.8
Tyr-29	161.3	−179.5 ± 24.2	168.0 ± 6.1	152.9 ± 5.5
Tyr-44	−69.5	147.6 ± 115.0	−48.5 ± 3.1	−53.2 ± 3.6

of the side-chains in all the aromatic residues (Phe-13, Tyr-29, and Tyr-44) converged well in both the GB-model-based structures compared with those in crambin^{DPC} (Figure 3). The values of the χ1 angles in crambin^{GBaqu} and crambin^{GBmem} were consistent with those in crambin^{XRAY}, whereas the rotation of Tyr-44 was totally different and those in Phe-13 and Tyr-29 were more deviated in crambin^{DPC} (Table 2). One may argue that detailed analysis is less meaningful because the solvents in crambin^{XRAY} are different from those in the current study. However, the purpose of comparison is to gain insight into the conformations of side-chains that are mainly influenced by solvents, and our analysis focused on the side-chains in a qualitative way.

For the general application of the GB model to proteins under membrane-like environments. One may criticize that crambin is a special case and most membrane proteins contain both hydrophobic and hydrophilic parts and that we cannot define the system with a global parameter, *surften*. However, recent developments in the GB model have enabled to describe a system into several layers of different surface tensions. Actually, it has been successfully applied to folding and several MD studies of membrane proteins.[9,10] We can apply the same algorithm to the NMR structure determination of real membrane proteins, as long as the boundaries of the membrane and aquatic parts are identified by experiments.

Conclusion

Our data, first, showed that MD refinement with an AMBER force field and the GB model improved the accuracy and precision of NMR structures. Second, the GB model with a parameter considering the membrane-like condition yielded more improvements than those with the normal GB parameter. We do not ignore the imperfectness of current GB model.[15] However, it should be again noted that both crambin^{GBmem} and crambin^{GBaqu} did not show any violation of the experimental restraints. Different from folding study, the conformational spaces to search in NMR structure refinement are narrow because of the experimental restraints, and the errors due to GB model will be much small. Our data will be an initial step toward structure determination of membrane proteins using NMR data.

Acknowledgments. JGJ appreciates Prof. Masahiro Shirakawa for his giving the initial motivation to study MD refinements.

References

1. Xia, B.; Tsui, V.; Case, D. A.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **2002**, *22*, 317-331.
2. Jee, J.; Byeon, I. J.; Louis, J. M.; Gronenborn, A. M. *Proteins* **2008**, *71*, 1420-1431.
3. Ohno, A.; Jee, J.; Fujiwara, K.; Tenno, T.; Goda, N.; Tochio, H.; Kobayashi, H.; Hiroaki, H.; Shirakawa, M. *Structure* **2005**, *13*, 521-532.
4. Fujiwara, K.; Tenno, T.; Sugasawa, K.; Jee, J. G.; Ohki, I.; Kojima, C.; Tochio, H.; Hiroaki, H.; Hanaoka, F.; Shirakawa, M. *J. Biol. Chem.* **2004**, *279*, 4760-4767.
5. Chen, J.; Im, W.; Brooks, C. L., 3rd *J. Am. Chem. Soc.* **2004**, *126*, 16038-16047.
6. Wuthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
7. Call, M. E.; Schnell, J. R.; Xu, C.; Lutz, R. A.; Chou, J. J.; Wucherpfennig, K. W. *Cell* **2006**, *127*, 355-368.
8. Roosild, T. P.; Greenwald, J.; Vega, M.; Castronovo, S.; Riek, R.; Choe, S. *Science* **2005**, *307*, 1317-1321.
9. Feig, M. *Methods Mol. Biol.* **2008**, *443*, 181-196.
10. Chen, J.; Brooks, C. L., 3rd; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140-148.
11. Ahn, H. C.; Juranic, N.; Macura, S.; Markley, J. L. *J. Am. Chem. Soc.* **2006**, *128*, 4398-4404.
12. Güntert, P.; Mumenthaler, C.; Wuthrich, K. *J. Mol. Biol.* **1997**, *273*, 283-298.
13. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668-1688.
14. Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27-85.
15. Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383-394.
16. Im, W.; Brooks, C. L., 3rd *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6771-6776.
17. Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M. *Proteins* **2003**, *50*, 496-506.
18. Jelsch, C.; Teeter, M. M.; Lamzin, V.; Pichon-Pesme, V.; Blessing, R. H.; Lecomte, C. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3171-3176.
19. Chen, J.; Brooks, C. L., 3rd *Phys. Chem. Chem. Phys.* **2008**, *10*, 471-481.
20. Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graph* **1996**, *14*, 51-55, 29-32.