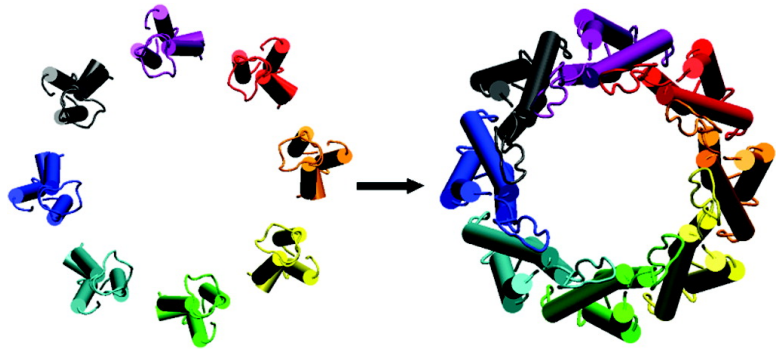
Communication

De Novo Prediction of the Structures of M. tuberculosis Membrane Proteins

Lintao Bu, and Charles L. Brooks III

J. Am. Chem. Soc., **2008**, 130 (16), 5384-5385 • DOI: 10.1021/ja710213p • Publication Date (Web): 02 April 2008



J|A|C|S
COMMUNICATIONS

Published on Web 04/02/2008

De Novo Prediction of the Structures of *M. tuberculosis* Membrane Proteins

Lintao Bu and Charles L. Brooks III*

Department of Molecular Biology and Center for Theoretical Biological Physics, The Scripps Research Institute, La Jolla, California 92037

Received November 10, 2007; E-mail: brookscl@umich.edu

Integral membrane proteins account for 30% of all proteins in the cell and play key roles in communication between a cell and its environment.[1] However, the experimental determination of three-dimensional structures of membrane proteins is extremely difficult. Among the approximately 30000 protein structures found in the Protein Data Bank (PDB),[2] only 0.2% are of membrane proteins. Due to their biological importance, a challenge to theory and computational biology is to assist in the experimental understanding of the structure and function of membrane proteins.

Mycobacterium tuberculosis is at the origin of most cases of tuberculosis, the leading cause of infectious mortality in the world. Structural analysis of the integral membrane proteins from *M. tuberculosis* could be very useful in understanding the molecular details of infection. Recently, one *M. tuberculosis* membrane protein, Rv1861, has been studied using solid-state NMR.[3] However, due to difficulties of sample preparation and labeling, to date there are only 10 membrane protein structures characterized by solid-state NMR in the PDB.

In this work, the structures of four *M. tuberculosis* membrane proteins, Rv2433, Rv1861, Rv1616, and Rv3069, are predicted using de novo methods. The number of transmembrane (TM) helices in these proteins varies from 2 to 4. Our computational methodology first identifies putative TM regions in each protein. Using the entire protein sequence, we generate a canonical α-helix, $\phi = -65°$ and $\psi = -40°$, which is forced to cross the membrane region in a manner that keeps the helix axis perpendicular to the membrane. The potential energy of the peptide is evaluated as a function of the residue number at the center.[4] The implicit solvent/implicit membrane GBSW[5] module in the CHARMM program[6] was used to represent the membrane environment. The all-hydrogen parameter set PARAM22[7] of the CHARMM force field was used. The parameters describing the membrane in our GB model were 0.04 kcal/(mol·Å^2) for the surface tension coefficient, representing the nonpolar solvation energy, 30 Å for the thickness of the membrane hydrophobic core, and 5 Å for a membrane smoothing length. The planar membrane is perpendicular to the *Z*-axis and centered at $Z = 0$. A local minimum in the energy profile coincides with the center of one TM helix at the center of the membrane. On this basis, the energy profile is used as a criterion to determine the center of each helix.

Second, we identified the beginning and ending residues for each helix. We used the sequence extending from the center of each individual helix, as identified with the energetic criterion described above, to the N-terminal and C-terminal domain by 15–20 residues to build a canonical α-helix. The distributions of $\phi-\psi$ angles for each residue from structures sampled at 300 K during a 10 ns replica exchange (REX)[8,9] MD simulation

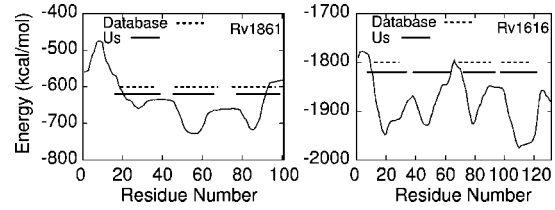


Figure 1. The potential energy of the proteins Rv1861 (A) and Rv1616 (B) as a single helix and as a function of the position of the helix. The short horizontal lines indicate the beginning and ending positions of each helix.

were used to provide information regarding the beginning and ending position of each helix, because residues in the aqueous phase or at the membrane−water interface are expected to fray significantly. (This method has been validated on bovine rhodopsin and applied to a calcium-sensing receptor in our previous studies.)[4,10]

Using these identified TM helical boundaries, initial configurations of the "disassembled" TM helical structures were constructed for sampling via RFDMD using the following protocol. Idealized α-helices were oriented along the membrane normal and in the membrane and then rotated by 11.25° around the *Z*-axis to produce 32 replicas, that is, we started from initial random helix−helix interface directions. Each helix was then translated by a certain distance from the *Z*-axis in the *XY* plane to generate an initial dissociated helical bundle structure. For example, the distance between any pair of neighboring helices was 20 Å, and all these helices are aligned in a linear array. REX MD simulations were performed for 5 ns to sample the structures. The CLUSTER facility in the MMTSB tool set[11] was used to cluster structures of each protein after convergence. The structure located at the center of the largest cluster was chosen as the predicted helical structure. The N-terminus, C-terminus, and loops of each protein were then added using the PREDICT facility in the MMTSB tool set. The whole constructed protein structure was then subjected to a new REX MD simulation for another 5 ns to refine and anneal the conformation.

The MMTSB package was used to control the REX MD simulations. A total of 16 replicas were distributed over an exponentially spaced temperature range, 300 to 600 K, to sample the configurational space of each individual helix and 32 replicas from 300 to 1000 K to sample the configurational space of the packed helices. A cylindrical harmonic restraint with a 50 Å radius and a force constant 1.0 kcal/(mol·Å^2) was applied to prevent the helices from diffusing apart. (Note that this is much larger than the radius of any of the TB proteins we studied.) A replica exchange was attempted every 1 ps, and the pairwise exchange ratio was around 40% for each run.

10.1021/ja710213p CCC: $40.75

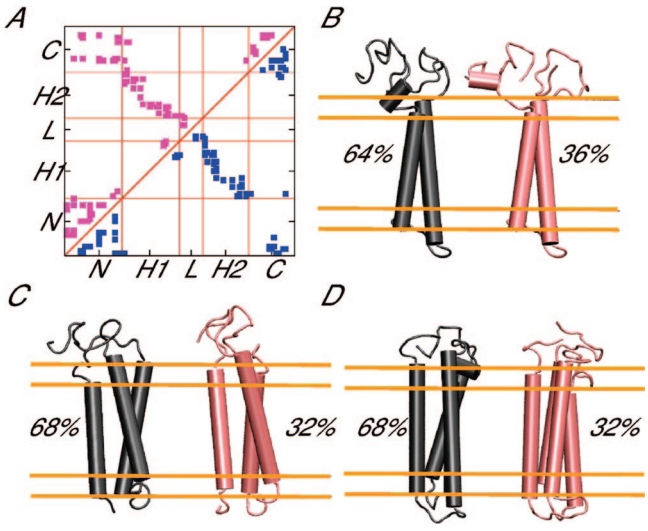


Figure 2. (A) Contact map of the C_α atoms in the two representatives of Rv2433c. The predicted structures for Rv2433c (B), Rv1861 (C), and Rv1616 (D) (the structure of Rv3069 is shown in Figure S2).

As illustrated in Figure 1, four TM helices are identified for protein Rv1616 using our approach, while the TB Structural Genomics Consortium database (http://www.doe-mbi.ucla.edu/TB) suggests only three TM helices for this gene product using sequence-only based measures. Although the predicted helix II sequence contains several polar residues, such as K44, T47, N50, R58, H61, and D62, which may be unfavorable in the membrane based on a sequence alone, our simulation results suggest that these polar residues are stabilized by forming intrahelical hydrogen bonds with backbone atoms (Figure S1, Supporting Information). Similarly, White et al. demonstrated that the isolated TM S4 helix from a voltage-gated K^+ channel can be stabilized via a hydrogen-bonded network of water and lipid phosphates around the Arg residues.[12] The predicted structure of Rv1616, shown in Figure 2D, clearly demonstrates the stability of helix II. For the other three *tuberculosis* proteins, our predicted helical regions are similar to those in the database.

We found two conformational clusters for Rv2433c (64 and 36%), Rv1616 (68 and 32%), and Rv1861 (68 and 32%). Three clusters (60, 22, and 18%) were found for Rv3069. The TM helices adopt similar structures across different clusters for each protein, but differ in the N-terminus, C-terminus, and loops. As illustrated in Figure 2A, the largest difference between the two predicted structures of Rv2433c corresponds to the interactions between the N-terminus and the C-terminus.

The solid-state NMR data for Rv1861 suggests that two of the three helices should have the same tilt angle, 37°, with respect to the membrane normal, in a mixture of lipids (DMPC/DMPG = 4:1).[3] Our predicted structure shows tilt angles of 9.1, 17.8, and 9.9°, respectively, for the three helices with a membrane thickness of 30 Å (corresponding to DOPC). If we reduce the membrane thickness to 25 Å (corresponding to DMPC), we observe an increase in the tilt angles 15.5, 21.1, and 15.0°. We note that the experiments also suggest that Rv1861 exists as an octamer.[3] To generate an octameric model of Rv1861, we imposed 8-fold rotational symmetry on the single protein using the IMAGE facility in CHARMM. Simulations to refine the octamer structure indicate that the tilt angle varies only ±1–3°, as shown in Figure S3, compared to the monomer. The RMSD of Rv1861 in an octamer is 1.7 Å relative to that in a monomer.

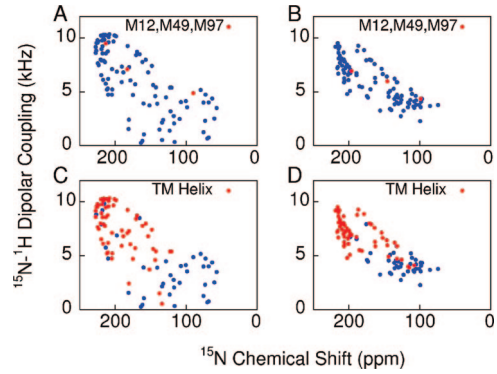


Figure 3. Computed PISEMA spectrum based on a single conformation of Rv1861 (A, C) and an ensemble average (B, D).

To further compare our predicted structure with the solid-state NMR measurements, we computed the PISEMA spectrum based on single conformation and an ensemble average of Rv1861, as shown in Figure 3. The computational protocol follows closely that of Im and Brooks.[13] Our results are in reasonable agreement with the experimental data, with regard to the overall shape of the spectrum.[3] The spectrum calculated based on a single conformation and an ensemble average are similar for TM helical residues but exhibit larger differences for other (loop) residues.

In this paper we have explored the de novo prediction of four *M. tuberculosis* membrane proteins using methods that have proven successful in studies of other integral membrane protein structures and oligomeric TM helices. We compare computed properties for one of these proteins, Rv1861, with recent NMR data and show that there is reasonable accord between the overall shape of the measured and computed PISEMA spectrum. Based on this comparison and our findings for other TM helical systems, we believe our predicted structures should serve as models for hypothesis generation for new experiments to probe the structure and functional roles of integral membrane structures of the *M. tuberculosis* system.

Acknowledgment. We are thankful for the generous support of this research by the NIH (Grant No. RR12255) and by the Center for Theoretical Biological Physics through funding from the NSF (Grant No. PHY0216576).

Supporting Information Available: Supplementary figures, complete ref 7, and predicted structures. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Wallin, E.; von Heijne, G. *Protein Sci.* **1998**, *7*, 1029–1038.
(2) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(3) Li, C.; Gao, P.; Qin, H.; Chase, R.; Gor'kovP. L.; Brey, W. W.; Cross, T. A *J. Am. Chem. Soc.* **2007**, *129*, 5304–5305.
(4) Bu, L.; Michino, M.; Wolf, R. M.; Brooks, C. L., III *Proteins* **2007**, *71*, 215–226.
(5) Im, W.; Feig, M.; Brooks, C. L., III *Biophys. J.* **2003**, *85*, 2900–2918.
(6) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(7) MacKerell, A. D. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(8) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
(9) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(10) Bu, L.; Im, W.; Brooks, C. L., III *Biophys. J.* **2007**, *92*, 854–863.
(11) Feig, M.; Karanicolas, J.; Brooks, C. L., III *J. Mol. Graph. Model.* **2004**, *22*, 377–395.
(12) Freites, J. A.; Tobias, D. J.; von Heijne, G.; White, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15059–15064.
(13) Im, W.; Brooks, C. L., III *J. Mol. Biol.* **2004**, *337*, 513–519.

JA710213P