JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

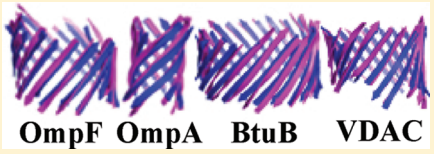
Predicting Three-Dimensional Structures of Transmembrane Domains of β-Barrel Membrane Proteins

Hammad Naveed, Yun Xu, Ronald Jackups Jr., and Jie Liang*

Department of Bioengineering, University of Illinois at Chicago, 835 South Wolcott Avenue, Chicago, Illinois 60607, United States

Supporting Information

ABSTRACT: β-Barrel membrane proteins are found in the outer membrane of gram-negative bacteria, mitochondria, and chloroplasts. They are important for pore formation, membrane anchoring, and enzyme activity. These proteins are also often responsible for bacterial virulence. Due to difficulties in experimental structure determination, they are sparsely represented in the protein structure databank. We have developed a computational method for predicting structures of the transmembrane (TM) domains of β-barrel membrane proteins. Based on physical principles, our method can predict structures of the TM domain of β-barrel membrane proteins of novel topology, including those from eukaryotic mitochondria. Our method is based on a model of physical interactions, a discrete conformational state space, an empirical potential function, as well as a model to account for interstrand loop entropy. We are able to construct three-dimensional atomic structure of the TM domains from sequences for a set of 23 nonhomologous proteins (resolution 1.8−3.0 Å). The median rmsd of TM domains containing 75−222 residues between predicted and measured structures is 3.9 Å for main chain atoms. In addition, stability determinants and protein−protein interaction sites can be predicted. Such predictions on eukaryotic mitochondria outer membrane protein Tom40 and VDAC are confirmed by independent mutagenesis and chemical cross-linking studies. These results suggest that our model captures key components of the organization principles of β-barrel membrane protein assembly.



INTRODUCTION

Membrane proteins comprise approximately one-third of all proteins encoded in a genome.[1,2] Among the two classes of membrane proteins, β-barrel membrane proteins are found in the outer membrane of gram-negative bacteria, mitochondria, and chloroplasts. They participate in many biological functions, including pore formation, membrane anchoring, and enzyme activity.[3−5] β-barrel membrane proteins are responsible for relaying extracellular signals to intracellular processes of many gram-negative bacterial pathogens.[4] Extensive studies have been carried out to understand the biophysical properties of β-barrel membrane proteins.[6−12]

β-barrel membrane proteins are often responsible for bacterial virulence and are promising targets for developing anti-infectious drugs.[13] Synthesized derivatives of β-cyclodextrin were designed on the basis of the symmetric nature of α-hemolysin and β-cyclodextrin. These derivatives have been shown to inhibit the activity of α-hemolysin, a major virulence factor in staphylococcal infection that ruptures host cell membranes upon folding into a β-barrel to gain access to iron released from these cells.[13]

Understanding the organizational principles of the transmembrane (TM) domains of β-barrel membrane proteins also has important engineering implications in developing protein nanopores for stochastic sensing and ultrarapid DNA sequencing, as well as nanoreactors for single molecule reaction and chemistry.[14−16] For example, mutant OmpG with altered voltage gating has been engineered with the goal to identify and quantify ADP molecules.[17] Mutant porin protein MspA has been engineered so that it can detect single strand DNA[14] In addition, an artificial β-barrel membrane protein has been constructed by duplicating the sequence of 8-strand OmpX. The resulting protein has a pore size similar to that of a 16-strand porin based on single-channel conductance measurements.[18]

However, such efforts are hindered by the limited availability of structural data for β-barrel membrane proteins, and the lack of an overall quantitative theoretical understanding of the stability of β-barrel membrane proteins. β-Barrel membrane proteins are sparsely represented in the protein structure database, as high resolution structural determination remains a challenging task.[19] Although computational studies have led to important insight, including prediction of β-barrel membrane proteins from genomic sequences,[20−22] prediction of transmembrane segments from the sequence,[23] identification of sequence and spatial (anti)motifs,[24,25] and characterization of their ensemble structural properties,[26] predicting the structures of β-barrel membrane proteins has not been successful.

Anfinsen's dogma, also known as the thermodynamic hypothesis, was an important breakthrough for understanding folding of a protein to its native structure. It states that the native structure of the protein is determined fully by its amino acid sequence.[27,28] However, complete understanding of this relationship has been elusive, and success has been reported only on small globular proteins.[28] Understanding the relationship between the protein sequence and structure has been the focus of many investigations.[29−32] Study of large proteins such

Received: October 20, 2011

   dx.doi.org/10.1021/ja209895m | J. Am. Chem.Soc. XXXX, XXX, XXX−XXX

as membrane proteins will require experimental determination or computational prediction of their structures, at least in the near future.

Template-based (or comparative modeling) methods that build three-dimensional models have had many successes in protein structure prediction.[33,34] These methods require knowledge of the structure of a template protein, which must share detectable sequence similarity to the query protein. However, recent structures of the voltage-dependent anion channel (VDAC) in mitochondria[35] and the usher protein PapC from *P. pilus*[36] contain 19 and 24 transmembrane β-strands, respectively. They violate the long-held beliefs that β-barrel membrane proteins consisted of even and between 8 and 22 strands.[37] As protein structures of these novel folds have never been observed in nature, template-based structure prediction will not work as no template structures homologous to VDAC and PapC exist. As techniques for membrane protein crystallization improve, it is expected that many additional structures of β-barrel membrane proteins with novel topology will emerge. As an alternative, free modeling methods for structure prediction do not require any template structures.[38−40] However, they cannot predict structures of β-barrel membrane proteins accurately, as these proteins are often of large size, with the number of residues ranging from 170 to 700.

Predicting β-barrel membrane protein structures is challenging. Although they are constrained by a well knit hydrogen bond network between adjacent transmembrane (TM) β-strands,[37] the presence of a large number of polar residues in the TM region, the uneven lengths of the TM-strands, the presence of autonomously self-folded in-plugs and clamps necessary to stabilize the protein,[11,41] the possible presence of protein−protein interactions in the TM region, and the nondiscriminating nature of generic H-bond interactions all make it challenging to predict accurately the three-dimensional structures of β-barrel membrane proteins.

In this study, we introduce a template-free method called 3D-SPoT (3D-Structure Predictor of Transmembrane β-barrels) for predicting the three-dimensional structures of the transmembrane domain of β-barrel membrane proteins. Our approach is based on the statistical mechanical model first introduced in ref 11. This model was successful in identifying weakly stable TM regions, in revealing general mechanisms of their stabilization, in predicting the oligomerization state, and in locating protein−protein interfaces.[11,42] Here, we first predict the interstrand hydrogen bond register between adjacent TM strands from the sequences. This is based on the empirical potential function TMsip derived from combinatorial analysis of known structures and a model for interstrand loop entropy. The full three-dimensional atomic structures of the TM region are then predicted on the basis of principles from differential geometry. In a blind test of 23 proteins, our prediction can generate accurate three-dimensional structures of the TM regions, with a median main chain rmsd of 3.9 Å.

As our method is based on physical interactions and does not require the availability of a template structure, it can be applied to predict structures of proteins with novel fold, including those from mitochondria of eukaryotes. As examples, we describe successes in predicting the structure of the VDAC protein, as well as insight gained from predicted structures of eukaryotic translocase in the outer mitochondrial membrane Tom40 and translocon of the outer chloroplast protein Toc75. We further discuss independent experimental verification of our predictions.

RESULTS

An antiparallel β-barrel has a well knit hydrogen bond network, in which each residue is hydrogen bonded to a residue on the adjacent strand (Figure 1). We use a physical model that

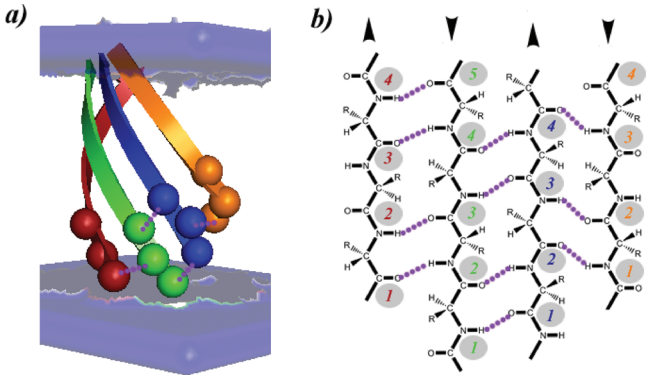


Figure 1. Pattern of hydrogen bonding between adjacent strands is shown as a cartoon diagram in (a) and as a detailed stick model in (b). Residues are represented as spheres in (a). Each residue in a strand is hydrogen bonded to a residue on the adjacent strand. Strand register is an index used to describe this hydrogen bonding pattern. For example, strand register is −1 when the first residue from the periplasmic side of the left (red) strand is hydrogen bonded to the second residue of the right (green) strand, 0 when the first residue of the left (green) strand is hydrogen bonded to the first residue of the right (blue) strand (b), and +1 when the second residue of the left strand (blue) is hydrogen bonded to the first residue of the right strand (orange).

accounts for strong H-bonds, weak H-bonds, and side-chain interactions[43,44] between neighboring strands in the transmembrane domain.[11,44,45] The energy of each type of interaction is quantified by an updated version of the original TMsip empirical energy function, which is based on extensive combinatorial analysis of known structures of TM strands of β-barrel membrane proteins.[25,45] Residues in a β-strand are modeled as candidate residues to be located in the transmembrane domain. In addition, we have incorporated a term accounting for interstrand loop entropy. Further details can be found in the Methods and in the Supporting Information.

For the task of structure prediction of β-barrel membrane proteins, we proceed in two steps: the prediction of the correct strand registration, and the prediction of three-dimensional coordinates of TM residues.

Predicting Strand Registration. We use a discrete state model to represent the configurational space of the strands, in which the relative position between a pair of neighboring strands can adopt $2L - 1$ different registrations, where L is the length of the strand (Supporting Information, Figure 1). For each TM β-strand, we generate all possible configurations in the discrete state model, each with a different registration of hydrogen bonds with its next sequential neighboring strand (Supporting Information, Figure 1). The energy for each configuration is evaluated by summing up the contribution from terms representing all strand-interaction types (strong hydrogen bonds, weak hydrogen bonds, and side chain interactions), a term for the loop entropy, and a term for bias toward right-handedness. For a strand pair, the registration with the lowest energy is selected as the predicted configuration.

To illustrate, we discuss the determination of the interstrand constraints of the integral outer membrane protein X (OmpX)

dx.doi.org/10.1021/ja209895m | J. Am. Chem.Soc. XXXX, XXX, XXX−XXX

from *E. coli*. OmpX is an eight-stranded monomeric β-barrel membrane protein implicated in the process of adhesion and entry of the bacterium into the host cells. It also confers resistance to attacks from the host immune system.[46] The strand registrations are predicted by taking the configurations of lowest energy. For 6 out of the 8 strands, the predicted strand registrations are the same as those observed in the X-ray structure of OmpX (more details in the Supporting Information, Figure 4).

The results of prediction of strand registrations for 23 β-barrel membrane proteins showed that overall, the registration of 248 out of 338 strand pairs are predicted correctly, representing an accuracy of 73% (Supporting Information, Table 2). This represents a significant improvement over previous β-barrel membrane protein specific contact prediction methods (~48%)[34,44] and the general contact prediction methods (~34% as reported in CASP9).[47]

Predicting Three-Dimensional Structures of TM Domains of β-Barrel Membrane Proteins. *Overall Shape of β-Barrels.* The extracellular loops of most membrane proteins are flexible and conformational entropy is likely to play an important role. The transmembrane domain is under physical constraints imposed by the lipid bilayer, as a result it frequently takes up the shape of a regular cylinder. Although some proteins are of an oval shape, the average ratio between the major and minor axes of the oval cylinder is small (1.17). As a first step toward determining the three-dimensional structures of the TM domains of β-barrel membrane proteins, we model the shape of the TM domains approximately as a proper cylindrical barrel instead of an elliptic cylindrical barrel.

Geometric Model for Three-Dimensional Atomic Structures. We use a geometric model of an intertwined coil to represent the shape of a β-barrel membrane protein (Supporting Information, Figure 2). In this model, each coil represents a β-strand, which is wrapped around a cylinder. The radius of the cylinder and the tilt of each coil with respect to the vertical axis are computed from the number of strands n and the shear number s.[48] Each coil in our model is represented by a parametric time-curve, and the position of the C_α atoms on these curves are computed through interstrand H-bond geometry. More details can be found in the Methods section and in the Supporting Information.

Predicting Three-Dimensional Atomic Structures. We use the C_α atoms to construct the main chain atoms using Gront et al.'s algorithm.[49] The side chains are subsequently added by the " autopsf" extension of VMD.[50] The orientation of side chains is corrected by calculating the energy of each strand using single body propensities only (details can be found in ref 44). There are two possible conformations for each strand. The side chain of the first residue of each strand can either face the interior or the exterior of the barrel. The minimum energy conformation among the two conformations is selected. This is followed by 200 steps of energy minimization using gradient descent method found in the software NAMD.[51]

Figure 2a–c depicts the predicted structures of the TM domains of protein OmpF, OmpA, and BtuB, which are shown superimposed on experimentally determined structures. The rmsd of the main chain atoms between the computed and measured structures is 2.7, 2.3, and 3.4 Å for OmpF, OmpA, and BtuB, respectively. The experimentally determined structures of these proteins are at the resolution 2.4, 2.5, and 2.0 Å, respectively. Overall, the experimental structures in our data set have a resolution of 2.3 ± 0.3 Å; at this resolution it is not possible to resolve every atom from the electron density map. As a result, most side chain atoms are modeled.[52] 3D-SPoT models the structures of the TM domains of 23 β-barrel membrane proteins well, with a median rmsd of 3.9 Å for main chain atoms and 5.3 Å for all atoms, respectively. The accuracy of predicted structures is maintained for large proteins. This is in contrast to other prediction method, where there is considerable deterioration in the quality of predicted structures (Table 1, Supporting Information Table 3).

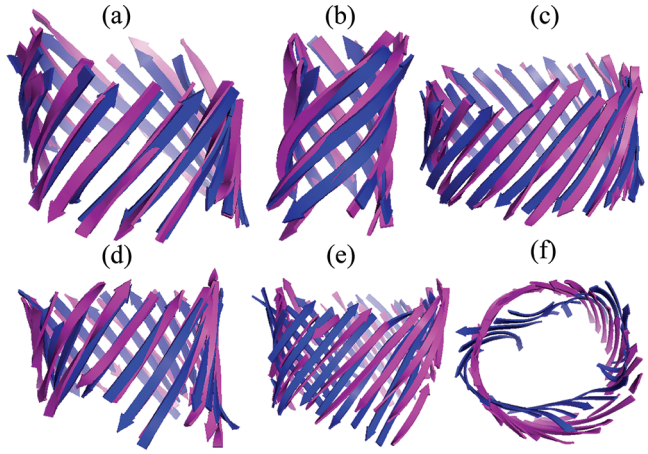


Figure 2. Predicted structures of the transmembrane domains (in magenta) superimposed on experimentally determined structures (in blue): (a) OmpF (pdb id: 2omf), (b) OmpA (pdb id: 1bxw), (c) BtuB (pdb id: 1nqe), (d) VDAC1 (pdb id: 3emn), and (e) PapC (pdb id: 2vqi). The main chain rmsd between the modeled and real structure is 2.7 Å for (a) OmpF, 2.3 Å for (b) OmpA, 3.4 Å for (c) BtuB, 3.9 Å for (d) VDAC1, and 8.2 Å for (e) PapC. The top view of the aligned real and modeled structures of PapC is shown in (f).

Table 1. Average RMSD for the Prediction of Three-Dimensional Structure of TM Domains of β-Barrel Membrane Proteins[a]

	$rmsd_{TM}$			
method	small	medium	large	all
TMBpro-server	6.0	6.3	11.8	7.3
3D-SPoT	5.4	6.0	5.7	5.6
$3D\text{-}SPoT_{MainChain}$	3.9	4.5	4.0	4.1

[a]3D-SPoT can predict the structure of TM domains with an average RMSD of 4.1 Å and a median of 3.9 Å for main chain atoms. Its accuracy does not deteriorate for large proteins. For comparison, results from template-based server TMBpro are also listed.

Predicting Structure of β-Barrel Membrane Protein with Novel Topology and Experimental Support. It is challenging to predict the structures of β-barrel membrane proteins with novel topologies. The voltage-dependent anion channel (VDAC) in mitochondria[35] has a unique topology unseen in any other known structures of β-barrel membrane proteins. Template-based prediction methods either fail to build any model or generate very poor structures. As our approach is based on a physical model capturing the basic organizing principles of transmembrane β-strands, it can be used to predict β-barrel membrane proteins of novel topology such as VDAC.

The predicted structure of the TM domains of the VDAC protein is shown in Figure 2d, with the experimentally determined structure superimposed. Although 3D-SPoT and

dx.doi.org/10.1021/ja209895m | J. Am. Chem.Soc. XXXX, XXX, XXX–XXX

the TMsip potential function were developed without the benefit of the knowledge of the topology and structure of VDAC, the predicted structure is accurate, with a main chain rmsd and an all atom rmsd of 3.9 and 5.8 Å, respectively.

Further insight can be gained on the basis of the predicted structure, the calculated energy profile of individual β-strands showed that there are four weakly stable regions in VDAC protein. Strands 1, 2 and 7, 8, 9 comprise the weakly stable regions I and II, respectively. Strands 13 and 17 form weakly stable regions III and IV, respectively. The calculated oligomerization index (see ref 11 for details) ρ is 2.47, which is greater than the threshold of 2.25 for monomers,[11] indicating that this protein likely forms oligomers. We predicted that these weakly stable regions will form the protein−protein interaction interface of VDAC protein.

Detailed analysis showed that residues S43 in region I, T116, C127, and G140 in region II, and Q249 in region IV were the most unstable residues and are likely located in the protein−protein interaction interface.[11] Region III has two modestly unstable residues G192 and Q196, and the crystal structure of VDAC (PDB id: 3EMN) revealed a lipid bound to G192, suggesting that this region is stabilized by protein−lipid interactions, a mechanism also observed in FhuA protein.[42,53]

Our predictions were confirmed experimentally by introducing a cysteine at defined positions in cysteine-less VDAC1 mutants, which was cross-linked with BMOE (cysteine-specific cross-linker) subsequently. The results indicate that weakly stable regions I, II, and IV are part of the protein−protein interaction interface at different times under normal and apoptopic conditions.[54] Moreover, the results suggest that upon apoptosis induction, VDAC1 undergoes conformational changes and that its oligomerization precedes through a series of interactions involving two distinct interfaces.[54]

Predicted Structures and Properties of Tom40 and Toc75. An important goal of protein structure prediction is to gain insight into the mechanism of protein function. Here we discuss prediction of protein structures and inference of key properties of two important eukaryotic β-barrel membrane proteins, Tom40 and Toc75.

Majority of the chloroplast and mitochondrial proteins are encoded in the nucleus and need to be transported from the cytosol across the biological membranes of the organelles. Tom40 is the central protein of the TOM complex (Translocase in the Outer Mitochondrial membrane), which is responsible for the insertion or translocation across the outer mitochondrial membrane of most mitochondrial proteins.[55−59] The TOM complex is a multicomponent system. However, neither the structure nor the protein−protein interaction interfaces of any of them is known. Similarly, the TOC complex (Translocon of the Outer Choloroplast)[60,61] is responsible for translocating proteins into the chloroplast. Toc75 protein in the TOC complex is the protein that forms the channel across the chloroplast membrane. Toc75 has a number of POTRA (POlypeptide TRansport Associated) domains at its *N*-terminus. As in the case of Tom40, Toc75 interacts with other components of the TOC complex, but their structures and the protein−protein interaction interfaces are not known.

Predicted Structure and Protein−Protein Interaction of Tom40 with Experimental Support. We have constructed the three-dimensional structure of the TM domains of human Tom40 and *Pisum sativum* (pea) Toc75 (Figure 3, see the Supporting Information for more details). Tom40 is known to

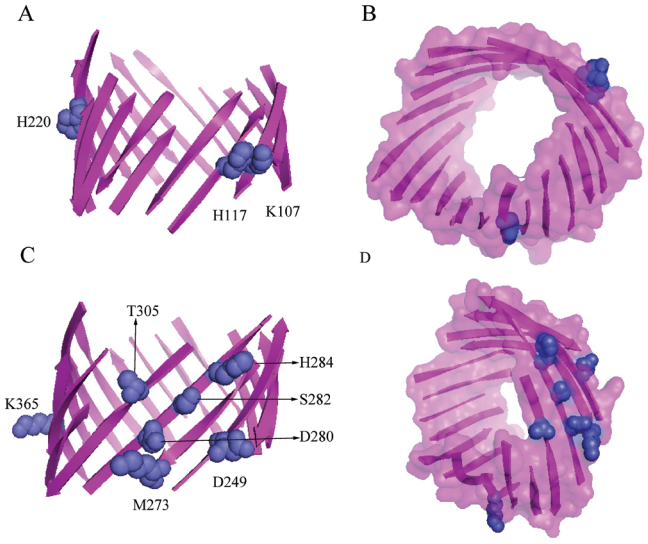


Figure 3. Predicted structures of the TM domains of Tom40 and Toc75. (a) Side view and (b) top view of the predicted structure of Tom40, along with the spatial location of unstable residues K107, H117, and H220. The triple mutant (K107L, H117L, and H220L) is predicted to be much more stable than the wild type monomeric protein; (c) Side view and (d) top view of the predicted structure of Toc75, along with the spatial location of the unstable residues D249, M273, D280, S282, H284, T305, and K365. Mutation D280L is predicted to have the most stabilizing effect on the monomeric form of Toc75 protein. The unstable residues in these proteins are likely to participate in protein−protein interactions.

interact with other proteins in the TOM complex and possibly with itself. The predicted Tom40 structure sheds some light on the location of the protein−protein interaction interface. The calculated energy profile of individual strands showed that strands 1, 2, and 9 are weakly stable (Supporting Information, Figure 5). They likely form two distinctive sites that may be involved in protein−protein interactions in the TM region: one by strands 1−2 and the other by strand 9. Indeed, the calculated oligomerization index ρ is 2.48, indicating that this protein likely forms oligomers.

We predicted that K107, H117, and H220 residues as most likely to participate in protein−protein interactions. We further predicted that the triple mutant K107L/H117L/H220L would disrupt these interactions and would significantly stabilize the monomeric form of human Tom40. The oligomerization index of this triple mutant $\rho_{K107L/H117L/H220L}$ was 1.37, suggesting that this mutant protein will be stable with a much lower propensity to participate in protein−protein interactions. Moreover, the oligomerization indices for single and double mutants (ρ_{K107L} = 2.08, ρ_{H117L} = 2.01, ρ_{H220L} = 2.03, $\rho_{K107L/H117L}$ = 1.91, $\rho_{K107L/H220L}$ = 1.71, and $\rho_{H117L/H220L}$ = 1.68) suggest that they are all more stable than the wild-type protein but less stable than the triple mutant.

Site-directed mutagenesis with thermal and chemical denaturation experiments have confirmed that the resistance of the mutated proteins to thermal and chemical perturbation was increased, particularly resistance of the triple mutant (K107L/H117L/H220L) to thermal denaturation was increased by 11 °C, and to chemical denaturation by an increased amount of 1.7 M GnHCl. The triple mutant was also verified to be primarily monomeric compared to the wild type that existed in dimeric and trimeric forms as well.[62] These experimental studies were motivated by our predictions, and their results are

dx.doi.org/10.1021/ja209895m | J. Am. Chem.Soc. XXXX, XXX, XXX−XXX

in agreement with our calculations. Moreover, the wild-type protein unfolded through a multistate mechanism, whereas the stabilized mutants follow a two-state conformational transition, suggesting that an intermediate stable state was eliminated due to these mutations.[62]

Predicted Structure of Toc75. From the predicted structure of Toc75, we found a large weakly stable region in the TM domain consisting of strands 5, 6, 7, and 8 (Figure 3). The calculated oligomerization index ρ is 2.21, which is indicative of a monomeric protein. However, residues D249, M273, D280, S282, H284, and T305 are unstable and might be a part of a weak/transient protein−protein interaction interface. We predict that mutations D249L, M273W, D280L, S282L, H284L, and T305L, along with perhaps the most dramatic mutation of D280L, will stabilize the monomeric form of Toc75 protein and reduce the propensity for Toc75 to be involved in protein−protein interaction. The calculated oligomerization index for all the single mutants is lower than that of the wild type, with ρ_{D280L} = 1.9 being the lowest, suggesting that these mutant proteins are more stable and less likely to participate in protein−protein interactions.

Furthermore, we found that strands 11 and 12 also have relatively high energy and might form a second protein−protein interaction site. Specifically, residue K365 in strand 11 is very unstable and is likely to be important for protein−protein interactions. We expect that future experimental studies similar to the ones carried out on Tom40 and VDAC will help elucidate the oligomerization state, and the protein−protein interaction interfaces of Toc75.

DISCUSSION

Computational Modeling of Structures of β-Barrel Membrane Proteins. Due to difficulties in experimental determination of membrane protein structures, there are limited number (<35 as of July 2011) of structures of nonhomologous β-barrel membrane proteins. However, it is estimated that there are many more β-barrel membrane proteins across different genomes.[10,22] In addition, a large number of new β-barrel membrane proteins will be discovered from large proteome, genome, and metagenome sequencing efforts.[63−65]

Computational prediction has the promise to provide a viable solution for understanding the structural basis of function and mechanism of these β-barrel membrane proteins. By taking known structures as templates and with the aid of a newly developed customized scoring matrix,[66] we expect the structures of about 5−10 times more proteins that are homologue of β-barrel membrane proteins with known structures can be modeled. The 3D-SPoT method described here is the first free modeling method for structure prediction of membrane proteins that goes beyond what has been seen structurally. It is well-suited to predict the TM structures of β-barrel membrane proteins that are of novel topology or have no detectable evolutionary relationship with any existing structures. It can build high quality structures and is not restricted by the number of strands, the oligomerization state, and the topology of the β-barrels.

Anfinsen's dogma provides the basis for understanding how the structure of a protein is determined from its sequence. A full understanding of the physical principles governing the folding and assembly of β-barrel membrane proteins will be valuable for their structure predictions. It will also help to design and engineer improved and novel peptides or proteins with desirable biophysical properties.

Eukaryotic β-Barrel Membrane Proteins. 3D-SPoT can also be used to study eukaryotic β-barrel membrane proteins, whose importance is increasingly being recognized. As proteins important for translocation across cellular membranes, they are key players of many cellular processes, including mitochondrial membrane permeability to small molecules and ions, Ca^{2+} homeostasis, translocation of unfolded proteins across the membrane, as well as apoptosis.[67,68] Knowledge of their structures would help to understand the mechanisms of these important cellular processes. As an example, the VDAC protein plays prominent roles in regulating permeability of ions and metabolites, in mediating cross-talks of subcellular processes, and in controlling apoptosis. Although 3D-SPoT was developed on the basis of structures of bacterial outer membrane proteins, our results show that predicted VDAC structure is accurate. This suggests that our model captures key elements of the organization principles of β-barrel membrane protein assembly, and β-barrel membrane proteins with no significant homology to any known proteins can now be modeled. Furthermore, structural characterization of predicted structures of Tom40 and Toc75 reveal important insight about the locations of protein−protein interaction sites.

Importance of Loop Entropy. The incorporation of loop entropy contributes significantly to the accuracy of the predicted structures, suggesting it is an important determinant of the structure of β-barrel membrane protein. As an example, the strand registrations for OpcA from *N. meningitidis*, an outer membrane protein facilitating adhesion and invasion of epithelial and endothelial cells,[69] was predicted correctly only for 4 out of 10 strands when loop entropy was disregarded. With the consideration of the loop entropy, the number of correctly predicted strand registration increases to 8. Overall, the gain in accuracy for all proteins by incorporating loop entropy alone is significant, the accuracy increases by 23% in prediction of strand-registration.

Repositioning during Biological Insertion and Strand Registration. At an overall accuracy of 73%, our prediction of strand registration is not perfect. It is possible that our model does not fully capture all the important interactions, the parameters in the empirical potential function may have room for improvement. It is also possible that other molecular interactions may be at play during the biological insertion of β-barrel membrane proteins, and repositioning due to these interactions may be necessary for arrival at the final registration. An analogous phenomenon has been observed in helical membrane protein glutamate transporter Glt_{ph}, in which the experimentally determined optimal embedding of TM helices matches those obtained from energy calculation but differs significantly from those observed in structures.[70] It was suggested that many TM helices are repositioned during folding and oligomerization.[70] It is conceivable that chaperons and β-barrel assembly machinery may also be factors that are significant in determining the final structures of β-barrel membrane proteins.

Multiple and Transient Oligomerization State of Multichain β-Barrels. 3D-SPoT does not yet predict accurately structures of β-barrel membrane proteins whose TM β-barrel is formed by multiple chains, such as that of α-hemolysin from *Staphylococcus aureus*. There are a number of technical challenges. For example, the exact number of chains that will aggregate to form a barrel structure is often unclear. α-

dx.doi.org/10.1021/ja209895m | J. Am. Chem.Soc. XXXX, XXX, XXX−XXX

Hemolysin assembles into a homoheptamer in the outer membrane as shown in the crystal structure, after being secreted as a water-soluble monomeric protein.[5] However, AFM studies by Czajkowsky et al. suggested that this protein may also exist in a hexameric form.[71] A preliminary examination of the transmembrane segments assembled as pentamer, hexamer, heptamer, and octamer have shown no major changes in stability of the barrel structure as measured by the relative melting temperature (see ref 11 for definition and details). These results suggest that multichain β-barrels may exist in several oligomeric states.

There is some evidence of the transient nature of oligomerization of some β-barrel membrane proteins. For example, FhuA may form a transient oligomer,[72] and PhoE, normally a stable trimer, may also exist in monomeric form.[73] The effects of these transient oligomeric states on the structure of β-barrel membrane protein are not yet well-known.

FUTURE DEVELOPMENT

There are many places for improvement. For example, the modeled structure of PapC has large rmsd (8.2 Å, Figure 2e,f), as the current geometric model does not take into full account the heterogeneity in interactions between TM strands and does not model the elliptic cylindrical barrel effectively. Furthermore, β-bulges pose additional challenges. In a regular β-strand, the orientation of the side chain alternates between external lipid space and internal barrel space. However, this pattern is disrupted by β-bulges where side chains of consecutive residues either face the internal space or the lipid space. These structural motifs have been shown to have functional significance. For example, outer membrane long-chain fatty acid transporter FadL from *E. coli* uses a β-bulge to mediate substrate passage from the extracellular environment into the lipid bilayer, from which the substrate can diffuse into the periplasm.[74] Identification of β-bulges from sequence should result in significant improvement in modeled structures.

As the composition of the lipid membrane can influence the structure of membrane proteins and specific lipid−protein interaction can provide significant stabilization as seen in FhuA,[42,53] incorporation of detailed lipid−protein interactions may further improve the accuracy of our predicted structures. Similarly, explicit consideration of interactions between the β-barrel domain and the in-plug/out clamp domains should also help to refine predicted structures.

SUMMARY

In this study, we showed that the three-dimensional structures of the TM domains of β-barrel membrane proteins can be predicted accurately. As our method is based on basic organizational principles and requires no template structures, TM domains of proteins with novel topology can also be modeled effectively, as evidenced by the study of VDAC and Tom40 proteins. This approach opens the possibility of structural studies of many β-barrel membrane proteins, including those in eukaryotic mitochondria.

Predicted TM-domain structures can provide important biological information. For example, sites for protein−protein interactions and mechanism of oligomerization can be inferred from predicted structures. Predicted structures can also be useful for design and engineering of membrane proteins with desired stability. This would facilitate crystallization of challenging membrane proteins by altering the stability of the TM domain. Predicted structure will also help in engineering proteins with controlled pore size and geometry, which would provide the structural foundation for fine-tuned biological functions such as ion permeability and channel selectivity. Further studies combining computational prediction and experimental measurement in these areas are likely to be fruitful.

METHODS

We use 25 β-barrel membrane proteins with known structures as our data set. Each pair of proteins has less than 32% pairwise-sequence identities (see Supporting Information). Predictions are only made for the 23 proteins, after excluding multichain β-barrels (TolC and α-hemolysin) to avoid over estimation of repeated interaction types. We take the canonical model of membrane β-strands based on the physical interactions between β-strands described in refs 43 and 44. The energetic contribution of each residue $E(i)$ has five components: the interaction energy with a neighboring strand through the backbone strong H-bond E_{SH}, the side-chain interaction E_{SC}, the backbone weak H-bond E_{WH}, loop entropy, and a penalty for left handedness.

Assuming a reduced conformational space for a pair of neighboring strands, we fix one strand and allow the second strand to slide up or down. A strand of length L can therefore have $2L - 1$ different registrations with its neighbor. The conformation with the minimal energy is then selected. The geometric model in this study is that of an intertwined coil (Supporting Information, Figure 3). Each coil represents a β-strand, which is wrapped around a hypothetical cylinder. Each coil in our model is modeled by a parametric time-curve, and the position of the C_α atoms on these curves are computed through interstrand H-bond geometry. The main chain atoms are constructed using Gront et al.'s algorithm.[49] The side chains are then added using the "autopsf" extension of the software VMD.[50] The orientation of side chains is corrected by calculating the energy of each strand using single body propensities only (details about single body propensity can be found in ref 44). Additional details can be found in SI.

ASSOCIATED CONTENT

Supporting Information

Data set and cross-validation discussion. Model for interstrand interactions. Strand register prediction and its results. Three-dimensional structure prediction and its results. Energy profiles. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*To whom correspondence should be addressed E-mail: jliang@uic.edu

ACKNOWLEDGMENTS

We thank Drs. Larisa Adamian, Linda Kenney, Stephan Nussberger, Varda Shoshan-Barmatz, and Dennis Gessmann for helpful discussions. This work was supported by NIH grants GM079804, NSF grants DMS-0800257 and DBI-1062328, and ONR N00014-09-1-0028. H.N. is supported by the Fulbright Fellowship and the Higher Education Commission of Pakistan.

REFERENCES

(1) Wallin, E.; von Heijne, G. *Protein Sci.* **1998**, *7*, 1029−1038.
(2) Arkin, I.; Brunger, A. *Biochim. Biophys. Acta* **1998**, *1429*, 113−128.
(3) Delcour, A. *J. Mol. Microbiol. Biotechnol.* **2002**, *4*, 1−10.
(4) Bishop, R. *Biochim. Biophys. Acta* **2008**, *1778* (9), 1881−96.
(5) Song, L.; Hobaugh, M.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. *Science* **1996**, *274*, 1859−1866.

(6) Hong, H.; Joh, N.; J.U., B.; L.K., T. *Methods Enzymol.* **2009**, *455*, 213−36.
(7) Tamm, L.; Arora, A.; Kleinschmidt, J. *J. Biol. Chem.* **2001**, *276*, 32399−32402.
(8) Tamm, L. K.; Hong, H.; Liang, B. *Biochim. Biophys. Acta* **2004**, *1666*, 250−263.
(9) Burgess, N.; Dao, T.; Stanley, A.; Fleming, K. *J. Biol. Chem.* **2008**, *283* (39), 26748−58.
(10) Wimley, W. C. *Curr. Opin. Struct. Biol.* **2003**, *13*, 404−411.
(11) Naveed, H.; Jackups, R. Jr.; Liang, J. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12735−12740.
(12) Valavanis, I.; Bagos, P.; Emiris, I. *Comput. Biol. Chem.* **2006**, *30*, 416−424.
(13) Karginov, V.; Nestorovich, E.; Schmidtmann, F.; Robinson, T.; Yohannes, A.; Fahmi, N.; Bezrukov, S.; Hecht, S. *Bioorg. Med. Chem.* **2007**, *15*, 5424−5431.
(14) Butler, T.; Pavlenok, M.; Derrington, I.; Niederweis, M.; Gundlach, J. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20647−20652.
(15) Adiga, S.; Jin, C.; Curtiss, L.; Monteiro-Riviere, N.; Narayan, R. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **2009**, *1*, 568−581.
(16) Banerjee, A.; Mikhailova, E.; Cheley, S.; Gu, L.; Montoya, M.; Nagaoka, Y.; Gouaux, E.; Bayley, H. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8165−8170.
(17) Chen, M.; Khalid, S.; Sansom, M.; Bayley, H. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 6272−6277.
(18) Arnold, T.; Poynor, M.; Nussberger, S.; Lupas, A.; Linke, D. *J. Mol. Biol.* **2007**, *366*, 1174−1184.
(19) Baker, M. *Nat. Methods* **2010**, *7*, 429−434.
(20) Ou, Y.; Gromiha, M.; Chen, S.; Suwa, M. *Comput. Biol. Chem.* **2008**, *32*, 227−231.
(21) Ou, Y.; Chen, S.; Gromiha, M. *J. Comput. Chem.* **2010**, *31*, 217−223.
(22) Wimley, W. C. *Protein Sci.* **2002**, *11*, 301−312.
(23) Ou, Y.; Chen, S.; Gromiha, M. *J. Comput. Chem.* **2010**, *31*, 217−223.
(24) Jackups, R. Jr.; Cheng, S.; Liang, J. *J. Mol. Biol.* **2006**, *363*, 611−623.
(25) Jackups, R. Jr.; Liang, J. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2010**, *7*, 524−536.
(26) Waldispuhl, J.; O'Donnell, C.; Devadas, S.; Clote, P.; Berger, B. *Proteins* **2008**, *71*, 1097−1112.
(27) Sela, M.; White, F. Jr.; Anfinsen, C. *Science* **1957**, *125*, 691−692.
(28) Anfinsen, C. *Science* **1973**, *181*, 223−230.
(29) Dill, K.; Bromberg, S.; Yue, K.; Fiebig, K.; Yee, D.; Thomas, P.; Chan, H. *Protein Sci.* **1995**, *4*, 561−602.
(30) Dill, K. *Protein Sci.* **1999**, *8*, 1166−1180.
(31) Brooks, C. 3rd; Gruebele, M.; Onuchic, J.; Wolynes, P. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 11037−11038.
(32) Onuchic, J.; Wolynes, P. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70−75.
(33) Venclovas, C.; Ginalski, K.; Fidelis, K. *Proteins* **1999**, No. Suppl. 3, 73−80.
(34) Randall, A.; Cheng, J.; Sweredoski, M.; Baldi, P. *Bioinformatics* **2008**, *24*, 513−520.
(35) Ujwal, R.; Cascio, D.; Colletier, J.; Faham, S.; Zhang, J.; Toro, L.; Ping, P.; Abramson, J. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17742−17747.
(36) Remaut, H.; Tang, C.; Henderson, N.; Pinkner, J.; Wang, T.; Hultgren, S.; Thanassi, D.; Waksman, G.; Li, H. *Cell* **2008**, *133*, 640−652.
(37) Schulz, G. E. *Biochim. Biophys. Acta* **2002**, *1565*, 308−317.
(38) Ben-David, M.; Noivirt-Brik, O.; Paz, A.; Prilusky, J.; Sussman, J.; Levy, Y. *Proteins* **2009**, *77* (Suppl 9), 50−65.
(39) Wu, S.; Skolnick, J.; Zhang, Y. *BMC Biol.* **2007**, *5*, 17.
(40) Kaufmann, K.; Lemmon, G.; Deluca, S.; Sheehan, J.; Meiler, J. *Biochemistry* **2010**, *49*, 2987−2998.
(41) Bonhivers, M.; Desmadril, M.; Moeck, G.; Boulanger, P.; Colomer-Pallas, A.; Letellier, L. *Biochemistry* **2001**, *40*, 2606−2613.
(42) Adamian, L.; Naveed, H.; Liang, J. *Biochim. Biophys. Acta* **2011**, *1808*, 1092−1102.
(43) Ho, B.; Curmi, P. *J. Mol. Biol.* **2002**, *317*, 291−308.
(44) Jackups, R. Jr; Liang, J. *J. Mol. Biol.* **2005**, *354* (4), 979−93.
(45) Jackups, R. Jr.; Liang, J. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2006**, *1*, 3470−3473.
(46) Vogt, J.; Schulz, G. *Structure* **1999**, *7*, 1301−1309.
(47) Monastyrskyy, B.; Fidelis, K.; Tramontano, A.; Kryshtafovych, A. *Proteins* **2011**, *79*, 119−125.
(48) McLachlan, A. *J. Mol. Biol.* **1979**, *128*, 49−79.
(49) Gront, D.; Kmiecik, S.; Kolinski, A. *J. Comput. Chem.* **2007**, *28*, 1593−1597.
(50) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14* (33−8), 27−8.
(51) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(52) PDB, PDB-101. http://www.pdb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/resolution.html.
(53) Ferguson, A.; Hofmann, E.; Coulton, J.; Diederichs, K.; Welte, W. *Science* **1998**, *282*, 2215−2220.
(54) Geula, S.; Naveed, H.; Liang, J.; Shoshan-Barmatz, V. *J. Biol. Chem.* **2011**, DOI: 10.1074/jbc.M111.268920.
(55) Neupert, W. *Annu. Rev. Biochem.* **1997**, *66*, 863−917.
(56) Voos, W.; Martin, H.; Krimmer, T.; Pfanner, N. *Biochim. Biophys. Acta* **1999**, *1422*, 235−254.
(57) Gabriel, K.; Buchanan, S.; Lithgow, T. *Trends Biochem. Sci.* **2001**, *26*, 36−40.
(58) Gabriel, K.; Egan, B.; Lithgow, T. *EMBO J.* **2003**, *22*, 2380−2386.
(59) Endo, T.; Yamano, K. *Biochim. Biophys. Acta* **2010**, *1803*, 706−714.
(60) Andres, C.; Agne, B.; Kessler, F. *Biochim. Biophys. Acta* **2010**, *1803*, 715−723.
(61) Li, H.; Chiu, C. *Annu. Rev. Plant Biol.* **2010**, *61*, 157−180.
(62) Gessmann, D.; Mager, F.; Naveed, H.; Arnold, T.; Weirich, S.; Linke, D.; Liang, J.; Nussberger, S. *J. Mol. Biol.* **2011**, *413*, 150−161.
(63) Ayalew, S.; Confer, A.; Hartson, S.; Shrestha, B. *Proteomics* **2010**, *10*, 2151−2164.
(64) Pinne, M.; Haake, D. *PLoS One* **2009**, *4*, e6071.
(65) Boyce, J.; Cullen, P.; Nguyen, V.; Wilkie, I.; Adler, B. *Proteomics* **2006**, *6*, 870−880.
(66) Jimenez-Morales, D.; Liang, J. *PLoS One* **2011**, *6*, e26400.
(67) Shoshan-Barmatz, V.; De Pinto, V.; Zweckstetter, M.; Raviv, Z.; Keinan, N.; Arbel, N. *Mol. Aspects Med.* **2010**, *31*, 227−285.
(68) Werhahna, W.; Jänschb, L.; Braun, H. *Plant Physiol. Biochem.* **2003**, *41*, 407−416.
(69) Zhu, P.; Klutch, M.; Derrick, J.; Prince, S.; Tsang, R.; Tsai, C. *Gene* **2003**, *307*, 31−40.
(70) Kauko, A.; Hedin, L.; Thebaud, E.; Cristobal, S.; Elofsson, A.; von Heijne, G. *J. Mol. Biol.* **2010**, *397*, 190−201.
(71) Czajkowsky, D.; Sheng, S.; Shao, Z. *J. Mol. Biol.* **1998**, *276*, 325−330.
(72) Locher, K.; Rosenbusch, J. *Eur. J. Biochem.* **1997**, *247*, 770−775.
(73) Van Gelder, P.; Tommassen, J. *J. Bacteriol.* **1996**, *178*, 5320−5322.
(74) Hearn, E.; Patel, D.; Lepore, B.; Indic, M.; van den Berg, B. *Nature* **2009**, *458*, 367−370.