JCTC Journal of Chemical Theory and Computation

Structural Determinants of Transmembrane β-Barrels

Themis Lazaridis*

Department of Chemistry, City College of New York/CUNY, 138th Street & Convent Avenue, New York, New York 10031

Received March 6, 2005

Abstract: The recognition of β-barrel membrane proteins based on their sequence is more challenging than the recognition of α-helical membrane proteins. This goal could benefit from a better understanding of the physical determinants of transmembrane β-barrel structure. To that end, we first extend the IMM1 implicit membrane model in a way that allows the modeling of membrane proteins with an internal aqueous pore. The new model (IMM1-pore) gives stable molecular dynamics trajectories for three β-barrel membrane proteins of different sizes and negative water-to-membrane transfer energies of reasonable magnitude. It also discriminates the correct fold for a pair of 10-stranded and 12-stranded transmembrane β-barrels. We then consider a pair of β-barrel proteins: OmpA, which is a membrane β-barrel with hydrophobic residues on the exterior and polar residues in the interior, and retinol binding protein, which is a water soluble protein with polar residues on the exterior and hydrophobic residues in the interior. By threading the sequence of one onto the structure of the other we make two pairs of structures for each sequence, one native and the other a decoy, and evaluate their energy. The energy function discriminates the correct structure. By decomposing the energy into residue contributions we examine which features of each sequence make it fold into one or the other structure. It is found that for the OmpA sequence the largest contribution to stability comes from interactions between polar residues in the interior of the barrel. The major factor that prevents the retinol binding protein sequence from adopting a transmembrane fold is the presence of polar/charged residues at the edges of the putative transmembrane β-strands as well as the less favorable interior polar residue interactions. These results could help design simplified scoring functions for fold recognition and structure prediction of transmembrane β-barrels.

Introduction

The known membrane protein structures belong to two categories: all α-helical and transmembrane β-barrels (TMBB). TMBBs occur in the outer membrane of Gram negative bacteria and presumably also of mitochondria and chloroplasts. They have from 8 to 22 β-strands (always an even number). Their N and C termini are on the periplasmic side and the loops on that side are short while those on the extracellular side can be very long.[1] These proteins perform a wide range of functions, such as allowing passive diffusion of ions and hydrophilic molecules, specific import of nutrients, energy dependent export of toxins, and cell adhesion.

Discriminating TMBBs from soluble β-proteins based on sequence is more challenging than discriminating helical membrane proteins because TMBBs lack the 20-residue hydrophobic stretches that characterize the latter. Early work utilized the alternate hydrophobicity of TM β strands;[2,3] however, this approach gives a lot of false positives. Over the past few years a number of bioinformatic approaches have been developed for the prediction and discrimination of TMBBs based on either neural networks[4−6,7] or Hidden Markov models.[8−11] The accuracy of these methods is typically around 80%. Wimley[12] developed an empirical score based on the relative abundance of amino acid types

* Corresponding author phone: (212)650-8364; fax: (212)650-6107; e-mail: tlazaridis@ccny.cuny.edu.

10.1021/ct050055x CCC: $30.25
Published on Web 05/27/2005

in TMBBs. Liu et al. attempted to discriminate TMBBs from soluble β sheet proteins based on their amino acid composition.[13] Other methods utilize a wider variety of properties,[14] including the presence of a signal peptide.[15]

While the above methods can be successful in recognizing a sequence as a TMBB or predicting the number of strands and the topology, they do not produce an atomic-level three-dimensional model. The methods available for doing so are homology modeling and fold recognition. Application of homology modeling requires significant sequence identity between the sequence under investigation and a protein of known structure, which is not often the case in TMBBs. Fold recognition appears more promising for TMBBs because these proteins have a limited number of folds.

The first goal of the present work was to develop an atomic-level energy function that can discriminate the native state of TMBBs. We do this by modification of IMM1, an implicit membrane effective energy function. IMM1 represents the membrane as a slab of nonpolar solvent. By embedding a cylinder of aqueous solvent into this slab we created a function that can be applied to TMBBs. The second goal was to use this energy function to obtain insights into the determinants of TMBB structure, i.e., investigate the features that make a certain sequence fold into a TMBB as opposed to a soluble β-protein. We do this by considering a pair of β-barrels of the same number of strands, one that is a TMBB (OmpA) and another that is a soluble β-barrel (retinol binding protein). We thread each sequence on the structure of the other sequence, evaluate the energy, and examine which residues contribute most to the energy difference between the correct and the incorrect structure.

Methods

IMM1-Pore. The origin of the energy function developed here is EEF1,[16] an effective energy function for soluble proteins

$$W_{\mathrm{EEF1}} = E + \Delta G^{\mathrm{slv}} \tag{1}$$

where E is the intramolecular energy of the protein, given by a modified form of the CHARMM polar hydrogen force field, and ΔG^{slv} is the solvation free energy, given by

$$\Delta G^{\mathrm{slv}} = \sum_i \Delta G_i^{\mathrm{slv}} = \sum_i \Delta G_i^{\mathrm{ref}} - \sum_i \sum_{j \neq i} g_i(r_{ij}) V_j \tag{2}$$

where $\Delta G_i^{\mathrm{slv}}$ is the solvation free energy of atom i, r_{ij} is the distance between i and j, g_i is the solvation free energy density of i (a Gaussian function of r_{ij}), V_j is the volume of atom j, and $\Delta G_i^{\mathrm{ref}}$ is the solvation free energy of an isolated atom. In addition, EEF1 employs a distance-dependent dielectric constant ($\epsilon = r$ in Å) and a neutralized form of the ionic side chains. A more recent version with updated parameters for the charged and polar side chains is referred to as EEF1.1.[17]

IMM1 is a generalization of EEF1 appropriate for modeling proteins in lipid bilayers.[17] In IMM1 $\Delta G_i^{\mathrm{ref}}$ is a linear combination of values for water and for cyclohexane (which models the nonpolar core of the membrane)

$$\Delta G_i^{\mathrm{ref}}(z') = f(z')\Delta G_i^{\mathrm{ref,wat}} + (1 - f(z'))\Delta G_i^{\mathrm{ref,chex}} \tag{3}$$

where $z' = |z|/(T/2)$ and T is the thickness of the nonpolar core of the membrane. The function $f(z')$ describes the transition from one phase to the other:

$$f(z') = \frac{z'^n}{1 + z'^n} \tag{4}$$

The exponent n controls the steepness of the transition. A value of 10 gives a region of 6 Å over which the environment goes from 90% nonpolar to 90% polar. The midpoint of the transition (f=0.5) corresponds to the hydrocarbon-polar headgroup interface (roughly the level of the ester carbonyls in phospholipids). The strengthening of electrostatic interactions in the membrane is effected by a modified dielectric screening function

$$\epsilon = r^{f_{ij}} \tag{5}$$

where r is the distance in Å (but ϵ is dimensionless), and f_{ij} depends on the position of the interacting atoms with respect to the membrane. Far from the membrane, f_{ij} is equal to 1, and we recover the linear distance-dependent dielectric model. For f_{ij} we employed the empirical model

$$f_{ij} = a + (1 - a)\sqrt{f(z_i)\ f(z_j)} \tag{6}$$

with a being an adjustable parameter. The value $a = 0.85$ was found to give membrane insertion or adsorption energies of the expected order of magnitude.[17]

To make IMM1 applicable to membrane proteins with aqueous channels we replace f in eq 3 and eq 6 by F

$$\Delta G_i^{\mathrm{ref}}(z', r') = F(z', r')\Delta G_i^{\mathrm{ref,wat}} + (1 - F(z', r'))\Delta G_i^{\mathrm{ref,chex}} \tag{7}$$

where

$$F(z', r') = f(z') + h(r') - f(z')h(r') \tag{8}$$

$$h(r') = 1 - \frac{r'^n}{1 + r'^n}, \quad r' = r/R, \quad r = \sqrt{x^2 + y^2} \tag{9}$$

where R is the radius of the aqueous pore. The definition of F ensures that $F = 0$ when f and h are zero and $F = 1$ when f, or h, or both of them are equal to 1. Figure 1 shows the contour of the line $F = 0.5$ on the xz plane. The same value of n is used for both f and h, but that could be changed. The code was implemented in the program CHARMM, version c32a1.

To test this energy function (IMM1-pore) we performed molecular dynamics (MD) simulations on the following proteins: OmpA (pdb code 1BXW[18]), the translocator domain of NalP (pdb code 1UYN[19]), and FhuA (pdb code 1BY3[20]). The extra Met residue at the N terminus of OmpA was omitted. The missing loop of NalP was built in an arbitrary conformation. The parameter R in eq 9 was set equal to the radius of the barrel at the level of Cα atoms. In one case it was varied to identify the value that gives the lowest

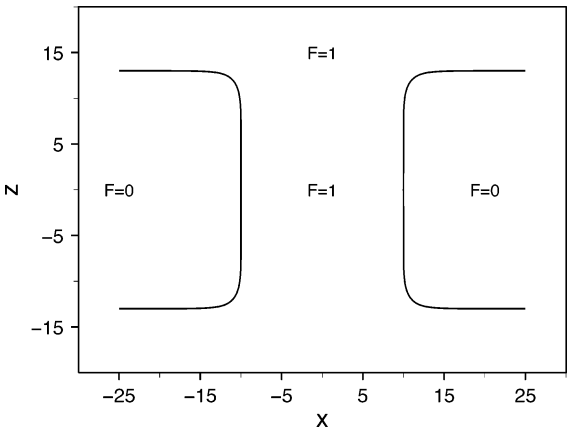

Figure 1. The boundary between nonpolar and aqueous regions in IMM1-pore (where *F* of eq 7 is equal to 0.5).

energy. All simulations proceeded with (a) generation of hydrogen coordinates using HBUILD, (b) a 300-step ABNR energy minimization, and (c) 1 ns MD simulation at 300 K using SHAKE, the Verlet algorithm, and 2 fs time step. In all cases the width of the membrane (*T*) was set to 26 Å.

Generation of the Decoys. To investigate the features that distinguish TMBBs from soluble β-barrels we first identified the closest possible water soluble analogue of OmpA, which seems to be retinol binding protein (RBP, pdb code 1AQB[21]). This protein has 8 β-strands and an additional helix at the C terminus (Figure 3). The structure is less regular than OmpA (the strands are shorter and more bent). The α helix was omitted in this study, i.e., only residues 1−141 were simulated. No disulfide bridges were built, to avoid having different numbers of bonds in the native structure and the decoy.

The decoys were generated using the program MODELLER 7v7.[22] The first hydrophobic residue of each β strand of RBP was aligned with the same from OmpA. The placement of the TM strands on RBP was adjusted slightly to maximize the hydrophobicity of the lipid-exposed residues. The final alignment is OMPA/RBP: 8/20, 38/39, 50/53, 78/67, 92/82, 120/103, 136/116, and 161/131. The maximal possible number of residues was aligned, and all insertions and deletions were placed in the middle of the loops. The MODELLER alignment files are given as Supporting Information.

Results

MD Simulations and Membrane Insertion Energies of TMBBs. Two criteria are used for testing the IMM1-pore energy function. First, MD simulations of TMBBs should give stable structures with small root-mean-square deviations (RMSDs) from the experimental structure. Second, the energy of insertion of a TMBB structure into the membrane (W in the membrane − W of the same conformation in water) should be negative.

The smallest TMBBs are 8-stranded. The best studied of these is OmpA.[18] The insertion energy of OmpA calculated by the standard IMM1 is positive, due to the unfavorable change in solvation free energy of the polar and charged residues in the interior of the barrel, which IMM1 treats as nonpolar. As a result, upon MD with the standard IMM1 the protein moves out of the membrane. Using IMM1-pore, the protein remains stably in the membrane. The energy in the membrane depends on the value of *R* and is lowest for $R = 11$ Å, which is close to radius of the cylinder formed by the Cβ atoms of the lipid-exposed residues. Using this value, the membrane insertion energy of the minimized

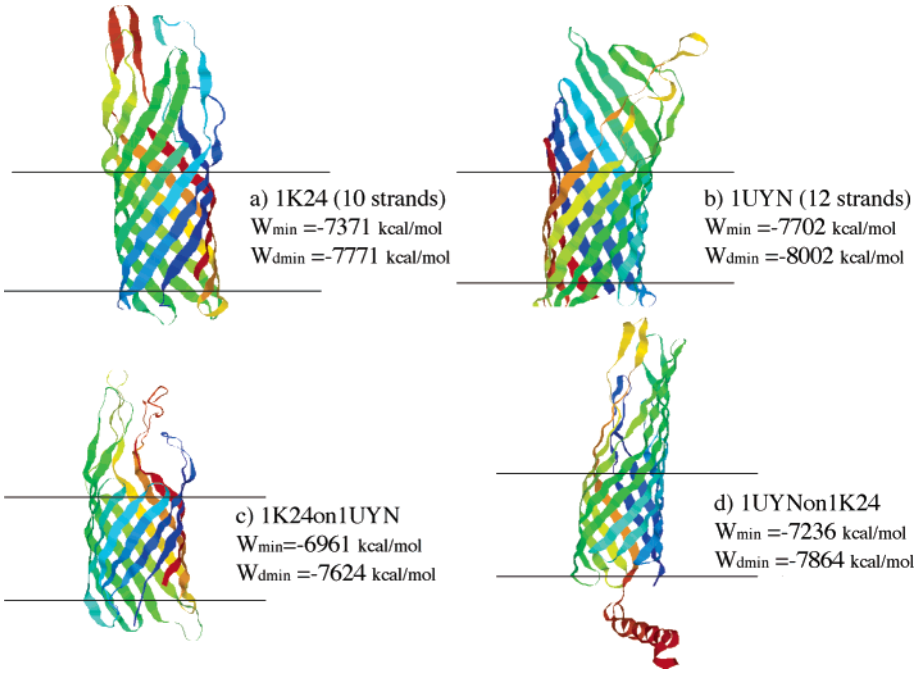

Figure 2. Native states and decoys for the fold recognition test: (a) native 1K24 (10 strands), (b) native 1UYN (12 strands), (c) the sequence of 1K24 threaded onto the structure of 1UYN, and (d) the sequence of 1UYN threaded onto the structure of 1K24. The structures shown are after 50 ps dynamics and minimization. Wmin are energies after minimization and Wdmin are energies after dynamics and minimization.

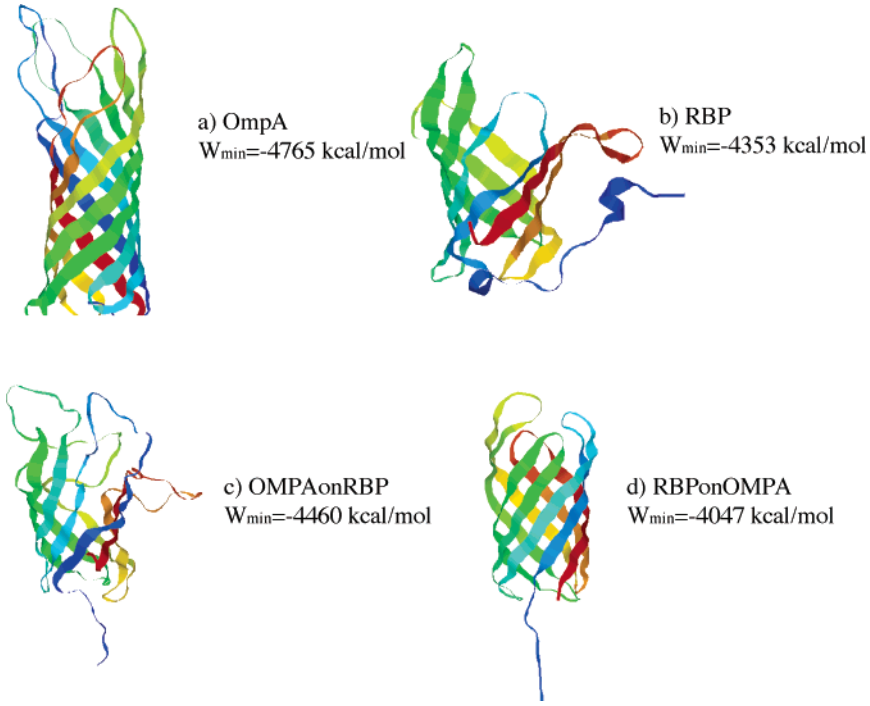

Figure 3. Native states and decoys: (a) native OmpA, (b) native RBP, (c) the sequence of OmpA threaded onto the structure of RBP, and (d) the sequence of RBP threaded onto the structure of OmpA. Wmin are energies after minimization.

OmpA structure is −14 kcal/mol. A 1 ns MD simulation with the same value of R gave a backbone RMSD of 2.9 Å (0.63 Å for the TM region only). These are comparable to previous explicit solvent results. Simulations of OmpA in an explicit lipid bilayer gave a Cα RMSD of about 2 Å for all residues and 0.75 Å for the TM region.[23] Somewhat larger deviations (3.8 Å loops, 1.5 Å TM region) were found in a more recent simulation of spontaneous DPC micelle formation around OmpA.[24] A simulation of OmpX, another 8-stranded TMBB, in water gave about 2 Å RMSD for all Cα atoms and about 1 Å for the TM region.[25] The 1 ns simulation of OmpA with IMM1-pore took 12.2 CPU h on a 3GHz Xeon processor. This is about 20 times faster than the explicit solvent simulations of OmpX.[25]

The translocator domain of the bacterial autotransporter NalP is a TMBB with 12 β-strands. It has been crystallized with an α-helix inside the lumen of the barrel.[19] The helix was omitted from this study. A 1 ns MD simulation of this TMBB gave a backbone RMSD of 1.9 Å for the TM region. The RMSD including the mobile loops was 4.2 Å (3.4 Å excluding the missing loop that was built in an arbitrary conformation). The membrane insertion energy for the minimized structure after dynamics was −15 kcal/mol.

A much larger TMBB is the FhuA receptor (pdb code 1BY3), which has 695 residues, 22 strands, and an N-terminal domain inside the β-barrel. A 1 ns MD simulation of this system with $R = 25$ Å gave overall backbone RMSD 3.9 Å. The backbone RMSD of the N-terminal domain alone was 3.4 Å and that of the TM β-barrel region 2.1 Å. For the final minimized structure, the optimal R is 28 Å and for that value of R the membrane insertion energy is −4 kcal/mol.

In summary, IMM1-pore shows the desired behavior: stable MD simulations, modest RMSDs, and favorable membrane insertion energies for TMBBs of different sizes.

Discrimination of TMBB Folds. Another important test of an effective energy function is whether it can discriminate the correct fold of a protein. For this test to be meaningful, one has to be able to create good-quality decoys. In the case of TMBBs, the sequence and the template must be of similar length, the lipid-exposed sides of all β-strands must be hydrophobic, and the extramembranous portions must be plausible. One pair where these conditions are met is the 10-stranded OpcA adhesin (pdb code 1K24) and the 12-stranded NalP (pdb code 1UYN, without the internal helix).

To convert OpcA into a 12-stranded TMBB we looked for two additional plausible TM β-strands, i.e., 10-residue stretches with good one-sided hydrophobicity. The best were found in the long loop between strands 3 and 4. The lipid-exposed residues on these two additional TM strands were N, I, L, T, E and K, V, L, T, and V. The Lys and Glu fall on the edge of the membrane, which is commonly observed. The two threonines are unfavorable in the hydrophobic core of a membrane but not excessively so. To convert NalP into a 10-stranded TMBB we removed the last two TM strands and built them into an extramembranous α-helix (when left unfolded, the effective energy was somewhat higher). Three-dimensional models for the decoys were built using MODELLER. The decoys and the native structures were energy-minimized and then subjected to a short (50-ps) MD simulation. R was 11 Å for the 10-stranded structures and 12 Å for the 12-stranded structures.

Figure 2 shows the native and decoy structures after MD and minimization together with their energies after minimization or after MD+minimization. They all remained folded in the membrane. Both the minimized and MD+minimized energies clearly discriminate the correct fold.

OmpA vs Retinol Binding Protein. Understanding of the sequence features that dictate the formation of a TMBB could

be enhanced by comparison of sequences that form TMBBs and sequences of similar length that form soluble β-barrels. TMBBs have nonpolar residues on the exterior and a mixture of polar and nonpolar residues in the interior. The opposite is true for water-soluble β-barrels. For OmpA the closest water soluble analogue seems to be retinol binding protein (RBP). It forms an eight-stranded β-barrel and has an additional helix attached on the exterior of the barrel.[21] In this work the helix was omitted. The first hydrophobic residue on each β-strand of RBP was approximately aligned with the first hydrophobic residue on each strand of OmpA. The alignment was provided to the program MODELLER to create two decoys: one for the sequence of OmpA adopting the structure of RBP (OMPAonRBP) and the other for the sequence of RBP adopting the structure of OmpA (RBPonOMPA). The energetics of the decoys was then compared to the energetics of the native structure for each sequence. The membrane-embedded forms (OmpA and RBPonOMPA) were simulated with IMM1-pore and the water soluble forms (RBP and OMPAonRBP) with EEF1.1 (which is equivalent to IMM1-pore far from the membrane). Native RBP gave a backbone RMSD of 2.65 Å upon 1 ns MD simulation with EEF1.1.

Figure 3 shows the decoys and the native structures. The loops in OMPAonRBP are clearly longer than those of a typical globular protein. In contrast, the loops of RBPonOMPA are a bit too tight. The minimized energies are shown in Figure 3. The native structures clearly have much lower energy than the decoys. The energy difference between natives and decoys may be exaggerated because the decoys are generated by an imperfect modeling procedure and may be unrefined compared to actual, crystallographic protein structures, especially in the loop regions. To examine the impact of this we made a model for OmpA based on the structure of another 8-stranded TMBB, NspA (pdb code 1P4T). The energy of the model (−4613 kcal/mol) was higher than that of native OmpA but still much lower than the energy of the soluble decoy (−4460 kcal/mol).

Unconstrained MD simulations of the decoys gave large RMSDs from the initial structure. RBPonOMPA moved completely out of the membrane in 0.5 ns and gave a backbone RMSD of 10.5 Å. OMPAonRBP gave a backbone RMSD of 6.1 Å. The final energies after dynamics were still higher than the energy of the native structure subjected to the same simulation protocol.

Having confirmed that the IMM1-pore energy function discriminates the correct structure for the OmpA and the RBP sequences, we proceed to analyze the contributions to the observed energy difference. The energy function, like its predecessors EEF1 and IMM1, is pairwise additive. Therefore, one can decompose the total energy into contributions from atoms, residues, or groups of residues. Here we consider residue contributions, including both side chain and backbone atoms. The contribution of a residue to the effective energy is equal to the sum of three terms: the intraresidue energy (SELF), one-half of the interaction energy of the residue with all other residues (INTE), and the solvation free energy of all atoms in the residue.[26] The sum of the residue contributions defined this way is equal to the total effective energy of the system. In practice, it is more convenient to include desolvation effects in the pairwise interactions. In that case, the third term is the *reference* solvation free energies of all the atoms in the residue (GREF). The latter make a contribution when membrane proteins are considered because the reference solvation free energies are not constant (see eq 7). This term accounts for the change in background environment. The residues are split into three categories: exterior (E, facing the lipid), interior (I, facing the aqueous pore), and loops (L). The contribution of each category is also reported. For this analysis the minimized structures were used because the decoys are unstable upon MD simulations.

Table 1. Contributions of Selected Residues to the Difference in Effective Energy Going from Native to Decoy for the OmpA Sequence (kcal/mol)[a]

		SELF	INTE	GREF	total	type
THR	152	−7.5	−6.4	0.0	−14.0	
ASN	145	2.6	0.2	−9.8	−7.0	E
LYS	12	−0.3	−0.1	−0.4	−0.8	I
ASP	56	−0.8	3.5	−0.1	2.7	I
LEU	83	0.3	4.0	2.7	7.0	E
GLN	78	2.0	6.4	−0.7	7.7	I
GLN	142	0.5	9.2	0.0	9.6	I
TYR	55	5.1	7.1	−1.5	10.7	E
total		81.2	251.5	−27.8	305	
E		45.4	48.4	11.4	105	
I		38.2	122.3	−25.7	135	
L		−2.4	80.8	−13.5	65	

[a] E and I denote barrel residues facing the lipid and the interior, respectively. All others are loop (L) residues. These values are based on the energy-minimized structures. A complete version of this table is given as Supporting Information.

Table 2. Contributions of Selected Residues to the Difference in Effective Energy Going from Native to Decoy for the RBP Sequence (kcal/mol)[a]

		SELF	INTE	GREF	total	type
ASN	66	−1.8	−10.0	0.0	−11.8	I
ASP	131	0.5	2.2	4.5	7.2	E
PHE	20	−4.1	9.0	2.8	7.7	E
TRP	24	0.4	6.5	1.0	7.9	E
LYS	30	−2.5	4.4	6.2	8.1	E
ASP	39	1.0	−3.1	12.7	10.7	E
ASP	126	−1.8	−3.0	17.3	12.5	E
ARG	139	−2.6	7.5	22.2	27.1	E
total		−14.3	186.3	133.5	306	
E		−11	87.3	75.3	152	
I		3.5	−11.2	38.6	31	
L		−6.8	110.3	19.6	123	

[a] E and I denote barrel residues facing the lipid and the interior, respectively. All others are loop (L) residues. These values are based on the energy-minimized structures. A complete version of this table is given as Supporting Information.

Tables 1 and 2 show the contributions of selected residues to the ΔW for the two sequences (the results for all residues are given as Supporting Information). The contributions are sorted from negative (favoring the decoy) to positive (favoring the native structure). For the OmpA sequence the largest contributions come from the I residues, then the E residues, and finally the loop residues. The I contribution comes primarily from interactions and secondly from self-

energy. The L contribution also comes primarily from interactions. Interactions and self-energy make equal contributions for E residues, with a smaller contribution coming from the change in environment (GREF). Overall, interactions make the largest contribution to ΔW, followed by self-energy, while the change in environment favors the water soluble decoy (polar residues are much "happier" in water).

For the 1AQB sequence the largest contributions to discrimination come from the E residues, followed by the L residues, with only a small contribution from the I residues. The E contribution comes from interactions and change in environment. The L contribution comes primarily from interactions. The I contribution comes primarily from change in environment.

For a more in-depth analysis we can look at individual residue contributions. For the OmpA sequence, Tyr55 makes the largest positive contribution, primarily because it hydrogen bonds to Gln75 in the native (N) structure and makes no such interactions in the decoy. The value of this contribution is probably exaggerated by the use of a single, energy-minimized conformation. A number of I residues make positive contributions because they lose favorable interactions from N to decoy. For example, Gln142 interacts with Gln78, Lys12, and Asp56 inside the pore in the N state but lacks such interactions in the decoy. Several E residues make positive contributions too, which come from a mixture of SELF, INTE, and GREF terms. For example, Leu83 loses some backbone interactions because the RBP β-sheet is less regular at the edges. It also loses some GREF energy because it goes from a position exposed to lipid to a position that is partly exposed to water (there is a cavity in the interior of the barrel in RBP which is presumably filled with water). The largest negative contributions are from loop residues that happen to form better interactions in the decoy. For example, Thr152 has a chance to make a hydrogen bond with Asn145 in the decoy but not in the N structure. Asn145's contribution is also negative because of change in environment; it goes from the edge of the membrane to an aqueous environment.

For the RBP sequence, the largest positive contribution comes from Arg139, primarily due to change in environment. In the decoy Arg139 is close to the edge of the membrane, whereas in the N structure it is exposed to water. Large positive contributions are made by other charged E residues that are on the edge of the membrane (e.g. Asp126, Asp39, Asp131, and Lys30). Some nonpolar E residues, such as Trp24 and Phe20, also make positive contributions because of favorable interactions in the soluble N state. Many of the large negative contributions are from I residues that happen to make good interactions in the TMBB decoy structure.

Discussion

A simple extension of the IMM1 implicit membrane model led to an energy function that can be used to study membrane proteins with an aqueous pore. MD simulations of a range of TMBBs with this function gave stable trajectories and favorable membrane insertion energies. In principle the function could be used with non-β-barrel MPs, such as ion channels. However, the simple cylindrical shape of the pore makes it best for cylindrical molecules. Other pore shapes could be accommodated by changing the definition of the function $h(r')$ in eq 9. For example, the pore cross section could be made elliptical, or it could be made to vary along the z axis.

One limitation is that the aqueous pore is static. Therefore, the energy function cannot be used to study, for example, concerted protein insertion and aqueous pore formation. This problem could conceivably be addressed by making the pore shape and size a dynamic variable in the MD simulation, in much the same way as the piston in constant pressure simulations[27] or the titration variables in constant-pH simulations.[28]

One important application of IMM1-pore stems from its ability to discriminate the native state of TMBBs. Because it is based on physics, it can be used to obtain insights into the features that drive or do not drive protein sequences into TMBB conformations. A first attempt to do so was made here by comparing OmpA with RBP. The major conclusions from this comparison are the following:

(1) The interior residues make a significant contribution to TMBB stability by engaging in favorable interactions with other interior residues. This effect may be largest for small barrels, where the interior residues are in contact with residues across the pore.

(2) Polar and charged residues at the edges of the exterior face of the β-strands destabilize the putative TMBB fold. For the positively charged residues this effect will depend on the nature of the lipids (zwitterionic or anionic).

The results have implications for bioinformatic approaches to TMBB structure prediction. Neural networks and Hidden Markov Models recognize amino acid composition and perhaps certain sequence patterns. Wimley's analysis of amino acid abundance at exterior and interior positions is essentially a one-body term. None of these approaches includes the effect of specific interactions between residues. It appears that inclusion of a pairwise score (as, for example, in soluble protein structure prediction[29,30]) should improve the discrimination.

Acknowledgment. This work was supported by the National Science Foundation (MCB-0316667). Some computational resources were provided by an RCMI grant from NIH (5G12RR003060).

Supporting Information Available: MODELLER alignment files and the complete versions of Tables 1 and 2. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Schulz, G. E. *Curr. Opin. Struct. Biol.* **2000**, *10*, 443.

(2) Vogel, H.; Jahnig, F. *J. Mol. Biol.* **1986**, *190*, 191.

(3) Schirmer, T.; Cowan, S. W. *Protein Sci.* **1993**, *2*, 1361.

(4) Jacoboni, I.; Martelli, P. L.; Fariselli, P.; De Pinto, V.; Casadio, R. *Protein Sci.* **2001**, *10*, 779.

(5) Gromiha, M. M.; Ahmad, S.; Suwa, M. *J. Comput. Chem.* **2004**, *25*, 762.

(6) Diederichs, K.; Freigang, J.; Umhau, S.; Zeth, K.; Breed, J. *Protein Sci.* **1998**, *7*, 2413.

(7) Natt, N. K.; Kaur, H.; Raghava, G. P. S. *Proteins: Struct., Funct., Bioinformatics* **2004**, *56*, 11.

(8) Martelli, P. L.; Fariselli, P.; Krogh, A.; Casadio, R. *Bioinformatics* **2003**, *18*, S46.

(9) Bagos, P. G.; Liakopoulos, T. D.; Spyropoulos, I. C.; Hamodrakas, S. J. *Bmc Bioinformatics* **2004**, *5*, 29.

(10) Bigelow, H. R.; Petrey, D. S.; Liu, J.; Przybylski, D.; Rost, B. *Nucleic Acids Res.* **2004**, *32*, 2566.

(11) Liu, Q.; Zhu, Y. S.; Wang, B. H.; Li, Y. X. *Comput. Biol. Chem.* **2003**, *27*, 69.

(12) Wimley, W. C. *Protein Sci.* **2002**, *11*, 301.

(13) Liu, Q.; Zhu, Y.; Wang, B.; Li, Y. *Comput. Biol. Chem.* **2003**, *27*, 355.

(14) Berven, F. S.; Flikka, K.; Jensen, H. B.; Eidhammer, I. *Nucleic Acids Res.* **2004**, *32*, W394.

(15) Zhai, Y.; Saier, M. H., Jr. *Protein Sci.* **2002**, *11*, 2196.

(16) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133.

(17) Lazaridis, T. *Proteins* **2003**, *52*, 176.

(18) Pautsch, A.; Schulz, G. E. *Nature Struct. Biol.* **1998**, *5*, 1013.

(19) Oomen, C. J.; Van Ulsen, P.; Van Gelder, P.; Feijen, M.; Tommassen, J.; Gros, P. *Embo J.* **2004**, *23*, 1257.

(20) Locher, K. P.; Rees, B.; Koebnik, R.; Mitschler, A.; Moulinier, L.; Rosenbusch, J. P.; Moras, D. *Cell* **1998**, *95*, 771.

(21) Zanotti, G.; Panzalorto, M.; Marcato, A.; Malpeli, G.; Folli, C.; Berni, R. *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54 (Pt 5)*, 1049.

(22) Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291.

(23) Bond, P. J.; Sansom, M. S. P. *J. Mol. Biol.* **2003**, *329*, 1035.

(24) Bond, P. J.; Cuthbertson, J. M.; Deol, S. S.; Sansom, M. S. P. *J. Am. Chem. Soc.* **2004**, *126*, 15948.

(25) Bockmann, R. A.; Caflisch, A. *Biophys J.* **2005**, *88*, 3191.

(26) Lazaridis, T.; Karplus, M. Microscopic basis of macromolecular thermodynamics. In *Thermodynamics in biology*; DiCera, E., Ed.; Oxford University Press: Oxford, 2001; p 3.

(27) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613.

(28) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. *Proteins* **2004**, *56*, 738.

(29) Sippl, M. J. *Curr. Opin. Struct. Biol.* **1995**, *5*, 229.

(30) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. *J. Mol. Biol.* **1997**, *268*, 209.

CT050055X