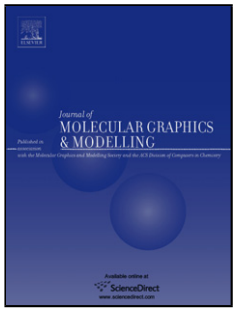


Title: Continuum electrostatic approach for evaluating positions and interactions of proteins in a bilayer membrane

Author: Chirayut Supunyabut Sunit Fuklang id="aut0015" orcid="0000-0002-8029-332X"> Pornthep Sompornpisut

PII: S1093-3263(15)00069-8
DOI: http://dx.doi.org/doi:10.1016/j.jmgm.2015.04.003
Reference: JMG 6531

To appear in: *Journal of Molecular Graphics and Modelling*

Received date: 29-8-2014
Revised date: 2-4-2015
Accepted date: 3-4-2015

Please cite this article as: C. Supunyabut, S. Fuklang, P. Sompornpisut, Continuum electrostatic approach for evaluating positions and interactions of proteins in a bilayer membrane, *Journal of Molecular Graphics and Modelling* (2015), http://dx.doi.org/10.1016/j.jmgm.2015.04.003

Title: Continuum electrostatic approach for evaluating positions and interactions of proteins in a bilayer membrane

Authors: Chirayut Supunyabut, Sunit Fuklang and Pornthep Sompornpisut*

Affiliation: Computational Chemistry Unit Cell, Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

***Corresponding author**: Pornthep Sompornpisut,
Department of Chemistry, Faculty of Science,
Chulalongkorn University, Bangkok 10330, Thailand.
Phone: 662-2187604; Fax: 662-2187598
E-mail: Pornthep.S@chula.ac.th

ABSTRACT

Orientations of proteins in the membranes is crucial to their function and stability. Unfortunately the exact position of these proteins in the lipid bilayer are mostly undetermined. Here, the spatial orientation of membrane proteins within the lipid membrane was evaluated using a Poisson-Boltzmann solvent continuum approach to calculate the electrostatic free energy of the protein solvation at various orientations in an implicit bilayer. The solvation energy was obtained by computing the difference in electrostatic energies of the protein in water and in lipid/water environments, treating each as an implicit solvent model. The optimal position of transmembrane proteins (TMP) in a lipid bilayer is identified by the minimum in the "downhill" pathway of the solvation energy landscape. The energy landscape pattern was considerably conserved in various TMP classes. Evaluation of the position of 1060 membrane proteins from the Orientations of Proteins in Membranes (OPM) database revealed that most of the polytopic and β-barrel proteins were in good agreement with those of the OPM database. The study provides a useful scheme for estimating the membrane solvation energy made by lipid-exposed amino acids in membrane proteins. In addition, our results tested with the bacterial potassium channel model demonstrated the potential usefulness of the approach in assessing the quality of membrane protein models. The present approach should be applicable for constructing transmembrane proteins-lipid configuration suitable for membrane protein simulations and will have utility for the structural modelling of membrane proteins.

Keywords:

Transmembrane protein, Poisson–Boltzmann, Implicit solvent model, Orientations of Proteins in Membranes database, Electrostatic energy of membrane solvation

Abbreviations:

TMP : Transmembrane protein; NMP : Non-transmembrane protein; OPM : Orientations of Proteins in Membranes; PB : Poisson–Boltzmann; RMSD : Root-mean-square deviation

1. Introduction

Membrane proteins play important roles in many biological processes, such as cell communication, signal transduction and regulation, cell adhesion and catalysis. They constitute about one-third of the proteins encoded in prokaryote to eukaryote genomes [1]. Based on the nature of their association with the lipid bilayer, membrane proteins can be grouped into the three major categories of (i) integral, (ii) peripheral and (iii) lipid-anchored membrane proteins. Integral or transmembrane proteins (TMP) are the major class of membrane proteins and exhibit key structural features, such as secondary and tertiary structures as well as orientation and packing of transmembrane segments, to satisfy the interactions with the phospholipid bilayer. Therefore, their spatial position with respect to biological membranes are vital to their function, folding and structure stability [2-6].

A variety of experimental techniques, including, x-ray diffraction XRD, solid-state and solution NMR, spin-labelling electron paramagnetic resonance (EPR), electron microscopy and fluorescence spectroscopy amongst others, have been used to solve the three-dimensional (3D) structures of proteins at various levels of resolution [7-11]. As can be seen by the growth of released entries in Protein Data Bank (PDB), the number of membrane protein structures increases every year. However, because the conventional methods of structure determination of membrane proteins (solution NMR and x-ray crystallography) are typically carried out in detergent micelles, the PDB structure coordinates rarely contain explicit information regarding the orientation of the protein structure with respect to the lipid bilayer. In molecular dynamics (MD) simulations, the generation of reliable trajectory is critically affected by the quality of starting configuration [12]. It would be useful to have a systematic tool for setting up a proper orientation of protein in membrane, whereby the starting configuration is examined

energetically without expensive computation. Although the spectroscopic techniques, such as NMR [13], site-directed spin labelling with EPR [14, 15], cysteine accessibility mutagenesis [16, 17] and tryptophan scanning mutagenesis with fluorescence spectroscopy [18, 19], are powerful methods to address the orientation as well as the position of TMP in the bilayer, they require considerable time, effort and expense to study each protein.

Computational methods play an increasing role in studying membrane proteins. A number of computational approaches, such as the use of all-atoms or coarse-grained molecular dynamics (MD) simulation with explicit water-lipid solvents [20-22], solvation free energy calculation in implicit continuum multidielectric treatment [23-26], or the PPM/OPM empirical functions [27, 28], have been used to optimize the spatial arrangement of proteins in the membrane. While MD is time consuming and computationally expensive, the latter two approaches allow a rapid assessment for positioning TMP. In particularly, the results of the PPM/OPM method provide a useful guide about positioning proteins in membrane. Interestingly, Choe et al employed continuum electrostatics in combination with the treatment of membrane deformation and non-polar energy terms to calculate the total insertion energy of transmembrane helical peptide in a membrane [29]. These promising results have stimulated the feasibility of using the continuum electrostatic approach for studying interactions and orientations of proteins in implicit descriptions of the membrane environment which is, however, still not well-established.

The solvent continuum method is a powerful tool for estimating the free energy of binding/solvation in biological systems. Among these methods, the Poisson–Boltzmann (PB) equation is widely used for electrostatic calculations of protein molecules in a continuous dielectric solvent [30-32]. The PB equation allows one to obtain accurate descriptions of the

solvent effect on the protein. There have been a number of software packages that implement this technique. Among these popular software is the Adaptive Poisson–Boltzmann Solver (APBS) program. It can be used together with the graphical interface APBSmem, which provides a simple and efficient protocol for calculating the electrostatic solvation energy of membrane protein systems [33, 34]. In this protocol, the aqueous and lipid bilayer models are implicitly treated as high- and low-dielectric continuum regions, respectively. The protein orientation in the implicit solvent environments can be graphically monitored and managed by the users. The program offers the "Protein Solvation" module to compute the electrostatic free energy of membrane solvation for transferring the protein from the surrounding aqueous environment to the mixed water-membrane slabs. The available features of the software has made it possible to perform a routine test for searching the optimal positions of membrane proteins within the membrane.

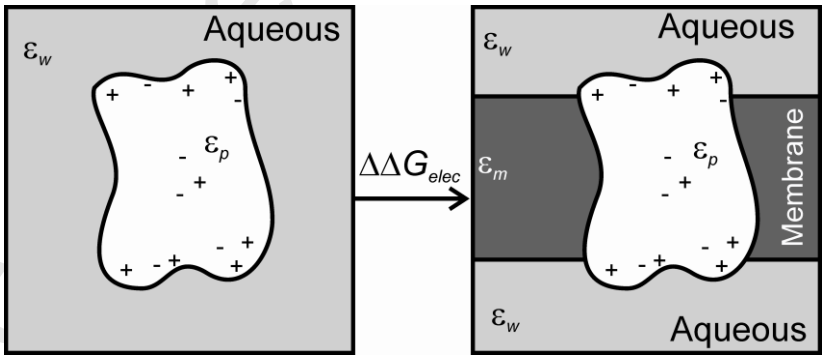


Figure 1. Schematic representation of the electrostatic free energy ($\Delta\Delta G_{elec}$) of solvation for a protein transferring from an aqueous environment to an aqueous-membrane environment when treated as continuum solvents. The influence of the solvent is

represented by the dielectric continuum, $\varepsilon_m = 2$ for the membrane region and $\varepsilon_w = 80$ for water.

In this study, the PB solvent continuum method was employed to evaluate membrane protein positions in the lipid bilayer. Unless otherwise specified, the term "energy" or any related term used here involve with the electrostatic contribution. For a test case, we initially examined the free energy landscape of membrane solvation for five membrane protein ion channels, whose spatial orientation is fairly well known. The method was then validated with water soluble and peripheral membrane proteins. A total of 1060 membrane proteins from Orientations of Proteins in Membranes (OPM) database (http://opm.phar.umich.edu) was examined for their optimal position by means of the PB approach. As described above, the strategy was to calculate the solvation free energy differences for a given protein immerged in the two solvent environment (water and water-lipid) presented by high (for water) and low (for membrane) continuum dielectric slabs (Figure 1). The electrostatic free energies of membrane solvation were used to evaluate a variety of positions of proteins within the membrane. Furthermore, the per-residue energy decomposition analysis gave a general picture of the range of the electrostatic solvation cost of 20 common amino acids in membrane. We also showed that the present approach can be used to discriminate 3D protein structure of native fold from incorrect folds.

2. Background

The strategy was to compute the electrostatic free energy of proteins inserted into a lipid bilayer at various orientations and depths to identify the spatial position of the most favorable energy. In an implicit-solvent description of the solvation, we employed the PB

method to estimate electrostatic contributions, which is the energy arising from the dielectric response to the electric field produced by the charged molecules. The PB equation representing the relation between the electrostatic potential and the charge distribution in the presence of an electrolyte solution is presented in Eq. (1);

$$\nabla \cdot [\varepsilon(\vec{r})\nabla\phi(\vec{r})] - \bar{\kappa}^2(\vec{r})\sinh[\phi(\vec{r})] = -e\beta 4\pi\rho(\vec{r}), \quad (1)$$

where ε is a dielectric permittivity coefficient, ϕ is the reduced electrostatic potential, κ^2 is the Debye-Hückel screening parameter, e is the unit charge, β = $1/k_BT$ is the inverse temperature and ρ is the charge distribution of the protein. For a given protein configuration, the ionic strength of the salt solution and the surrounding solvent of a known dielectric coefficient, the electrostatic potential can be obtained from solution of the PB equation. In the present study, calculations of the electrostatic free energy of solvation for the protein are carried out with the linearized form of the PB equation according to,

$$\Delta G_{elec} = \frac{1}{2}\sum \Phi(\vec{r})\rho(\vec{r}) \, . \quad (2)$$

3. Methods

3.1. Testing ion channel proteins

Initially, the calculations of the electrostatic solvation energy of proteins inserted into the membrane were performed on five membrane protein channels of KcsA (PDB ID: 1K4C) [35], Kv1.2-2.1 chimera (PDB ID: 2R2R) [36], large-conductance mechanosensitive channel

(PDB ID: 2OAR) [37], the voltage-gated sodium channel NaVAb (PDB ID: 3RVY) [38] and the CorA Mg channel (PDB ID: 2IUB) [39]. Starting structure configurations (Figure S1) were generated by orienting the channels pore with the z-axis and by placing the center of transmembrane segments at the origin of the Cartesian coordinate system (Figure 2).

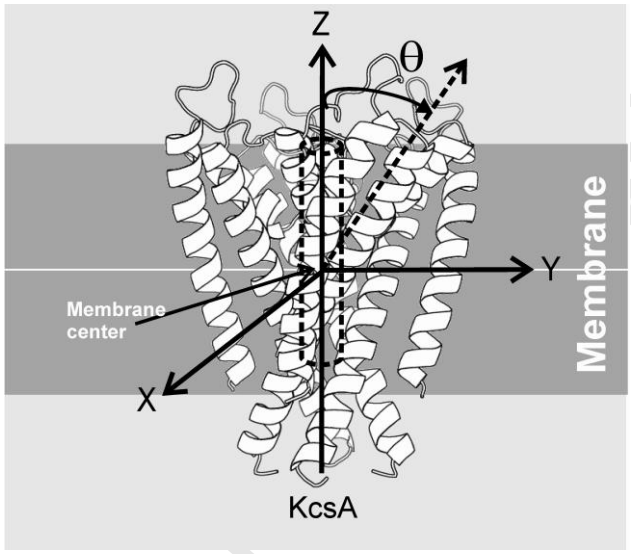


Figure 2. Starting configuration of the model (KcsA potassium channel) in the membrane slab. The pore of the channel is oriented to align with the z-axis and the centre of the transmembrane segments is located at the membrane centre. This reference position corresponds to the z-position = 0 and the rotation (θ) = 0º.

The electrostatic free energy of solvation was calculated for various configurations at various positions and angles with respect to the membrane region. The solvent boundaries were generated using a ± 20 Å translation about the z-axis at a 1 Å step size and up to 60º of rotation (θ) along the x- or y-axis with a 10º step size. Transformation of the atomic coordinates was performed using the program VMD and the tcl-command scripts [40]. For each configuration, the electrostatic free energy of solvation was computed according to Eq.

(3) in section 3.2. It should be noted that the examined range of protein positions captures different depth levels for proteins penetrating from the aqueous phase into the membrane-like environment.

3.2. Calculating the electrostatic free energy of membrane solvation

All ions, ligands, and non-amino acids that are present in the crystal structures were removed prior to running the PB calculation. The program PDB2PQR was used to add hydrogen atoms and assign atomic charges and radii to each generated configuration [41, 42]. Partial atomic charges and radii of the protein were taken from the PARSE parameter sets [41, 42].

The electrostatic contribution to the solvation free energy was performed using the program APBS 1.2.1 [34]. Unless otherwise specified, the thickness of the bilayer membrane (hydrophobic slab) used in the PB calculation was 30 Å. The solvent dielectric values were set to 2 ($\boldsymbol{\varepsilon}_m$) and 80 ($\boldsymbol{\varepsilon}_w$) for the membrane region and the water region, respectively [43]. For the protein, a dielectric value of 2 was assigned to points within the solvent accessible surface. The solvent ionic strength was 0.1 M with coulomb charge +1 and -1, and radius 2.0 Å. The water probe radius was 1.4 Å. Dielectric (ε), charge (ρ), and ion-accessibility maps (κ) of the proteins were generated with the sequential focusing multigrid algorithm, which consisted of three resolution maps (300 × 300 × 300 Å for coarse, 200 × 200 × 200 Å for medium, and 100 × 100 × 100 Å for fine resolution). Grid points of 161 × 161 × 161 were employed for all maps, and APBSmem was used to set up the input parameters (Figure S2) [33]. The electrostatic energy of solvation for each protein was computed with the linearized PB equation for both in the implicit aqueous model and for the water-membrane model (Figure 1),

giving rise to $\Delta G_{elec,water}$ and $\Delta G_{elec,membrane}$, respectively. For a given protein configuration, we computed the electrostatic free energy of membrane solvation ($\Delta\Delta G_{elec}$) by subtrating the solvation energy for a rigid protein between the two environments. $\Delta\Delta G_{elec}$ can be calculated according to Eq. (3),

$$\Delta\Delta G_{elec} = \Delta G_{elec,membrane} - \Delta G_{elec,water}. \quad (3)$$

4. Results

4.1 Positioning protein channels in membrane

According to a previous study from Grabe Lab [29], the free energy for inserting α-helices into membranes has been derived from the three energy terms: electrostatic, membrane deformation and nonpolar. Here, our test aimed to use the PB method for evaluating the orientation of proteins within a lipid bilayer with a constant hydrophobic thickness. As an initial test, the evaluation was conducted on five pore-forming ion channels since it should be relatively easy to assess their optimal orientation in the membrane. The 3D energy landscapes of membrane solvation showed a "downhill" pathway to the minimum near the bilayer centre (indicated by the white arrow in Figure 3A) similar to all five proteins. This suggested that the optimal orientation of the proteins in membranes falls within the region enclosed by the z-position and the rotation ≈ 0.

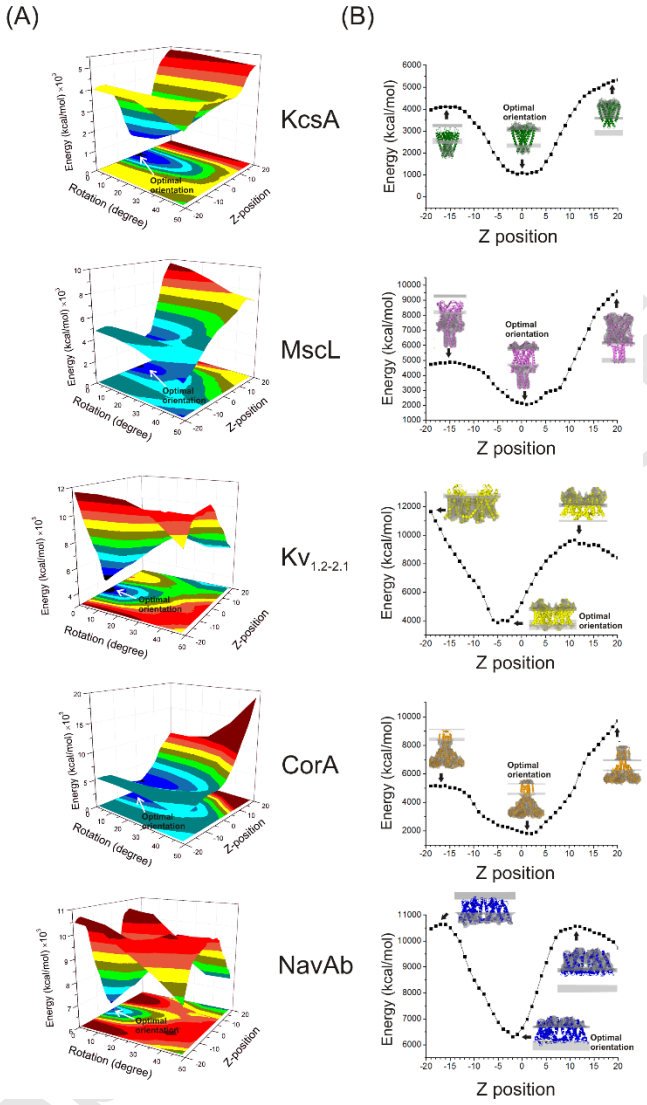


Figure 3. The electrostatic free energy of membrane solvation associated with the orientation and the position of five ion-pore forming membrane proteins within the lipid bilayer. Shown are (A) $\Delta\Delta G_{elec}$ landscapes and (B) 1-dimenstional $\Delta\Delta G_{elec}$ profiles as a function of the z-position at a fixed rotation angle ($\theta = 0^o$). Inset arrows and membrane-protein models indicated configurations with the lowest energy and with the z-position ~ ±15Å.

Detailed energy landscapes also suggested that the $\Delta\Delta G_{elec}$ of membrane solvation decreases upon the increased depth of membrane insertion of transmembrane domains (Figure 3). The reduction of the solvation energy barrier can be described by the interactions of hydrophobic transmembrane to the low dielectric of the lipid. At the optimal rotation value ($\theta = 0^o$), the 1D energy profiles of the five pore-forming proteins showed a shallow curve, which allows for easily spotting the minimum energy position (Figure 3B). One can see that the energy cost for membrane solvation decreased with increasing the surface area of hydrophobic transmembrane segments exposure to the membrane environment. A visual inspection of the protein positioning in the hydrophobic slab for the highest- and lowest-energy configurations revealed a good fit of the transmembrane domains within a low dielectric surrounding. From these results, the optimal orientation of all the examined proteins obtained corresponds to the protein configuration where the z-position = 0 and the rotation (θ) = 0^o. At this configuration, most apolar or aromatic residues of membrane-exposed transmembrane segments are essentially inside the hydrophobic bilayer and the pore axis runs approximately perpendicular to the membrane. These results are in good agreement with expected orientation of pore-forming transmembrane proteins, suggesting that our approach is a meaningful way to spatially orient and position membrane proteins in a bilayer.

4.2. The effect of the membrane thickness

Different organisms have various lipid compositions which could influence the bilayer thickness, one of the most important factors that affects the protein function and assembly. In addition, the length of bilayer thickness can be altered by the effects of bilayer deformation induced by the protein [44, 45]. The $\Delta\Delta G_{elec}$ calculation of section 4.1 was based on a

hydrophobic thickness of 30 Å. This value was approximated from the average thickness (≈29 Å) for transmembrane proteins estimated from high-resolution structures[44]. Here, the examined thickness of the hydrophobic core varied in the range of 10 to 50 Å, which is slightly extended from the membrane deformation energy calculation of Kessel and BenTal [46]. Figure 4 demonstrates the effect of membrane thickness on the $\Delta\Delta G_{elec}$ calculation of the optimal configuration (z_{opt}) and other four configurations corresponding to the z-position = ±5 and = ±15 Å denoted as z_{opt+5}, z_{opt-5}, z_{opt+15} and z_{opt-15}.

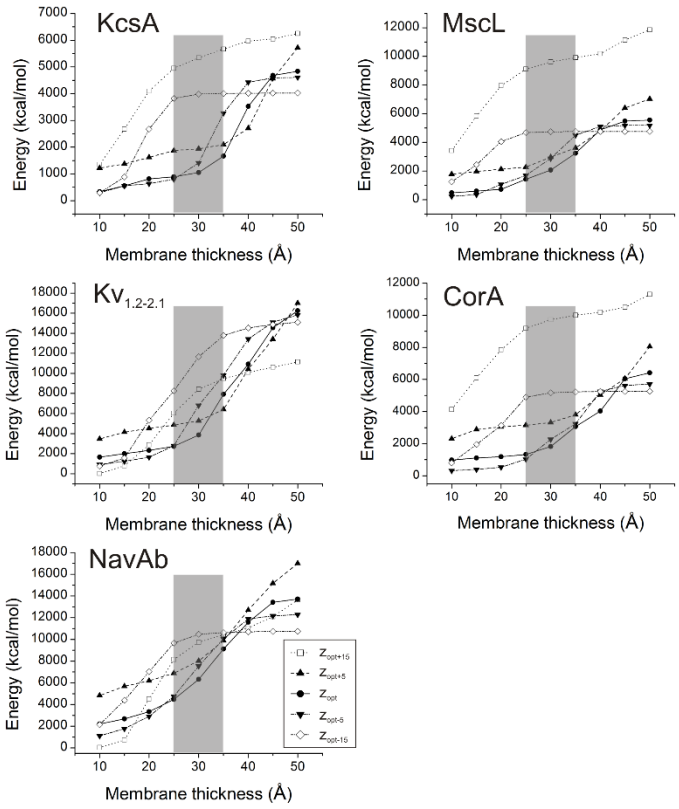


Figure 4. The $\Delta\Delta G_{elec}$ profiles as a function of the membrane thickness for the optimal and non-optimal configurations of the five ion-pore transmembrane proteins. The energy profile of the optimal configuration (z_{opt}) is represented as a line with filled circles. The discontinuous lines belong to the configurations of z_{opt+5} (up triangles), z_{opt-5} (down triangles), z_{opt+15} (unfilled squares) and z_{opt-15} (unfilled diamonds), which correspond to ±

5 and ± 15 Å from the optimal configuration (taken from Figure 3B). The gray bar highlighted the range of the thickness that the method can provide reliable prediction.

The positions of z_{opt+5} and z_{opt-5} were arbitary but designed to capture the variation between and the length of membrane thickness and the extent of transmembrane segments. The z_{opt+15} and z_{opt-15} are, however, the outlier position of proteins in membrane and energetically unfavorable because their transmembrane domains are more exposed to water region than z_{opt}. The calculation with a given thickness of membrane fails, if the $\Delta\Delta G_{elec}$ values of either z_{opt+15} or z_{opt-15} are lower than that of z_{opt}.

Detailed analysis was focused on the comparison of $\Delta\Delta G_{elec}$ values between the most favorable z_{opt} and the two outlier positions. It appears that the calculations failed to discriminate the correct and incorrect positions of proteins in lipid bilayer for the thickness range from 10 to 20 Å and from 40 to 50 Å (Figure 4). The results of the calculations with the thickness of 25 and 35 Å (the gray bar) may be considered for the qualification because the protein position is well predicted with a small deviation. The reason for this uncertainty may well be attributed to the effect of bilayer deformation or hydrophobic mismatch or both. Nevertheless, from the results, the use of a 30 Å thick of lipid bilayer appears, at best, to reproduce the correct position of proteins without the need to treat nonpolar and membrane deformation energies. The resulting value is considered to be closed to the ~27-28Å thickness of the hydrophobic core of the dioleoylphosphatidylcholine bilayer, one of the most common lipid model for experimental and simulation studies [44].

4.3. Barrier pattern of membrane solvation for soluble proteins

Next, the same methodology was appiled to investigate the electrostatic energy landscape of membrane solvation for water soluble proteins. In this test, we selected three soluble proteins (PDB ID: 2HU4, IHXB and 3LOX). The parameters used in the $\Delta\Delta G_{elec}$ calculation were identical to those used for the five protein channels (section 4.1), and the results are shown in Figure 5.

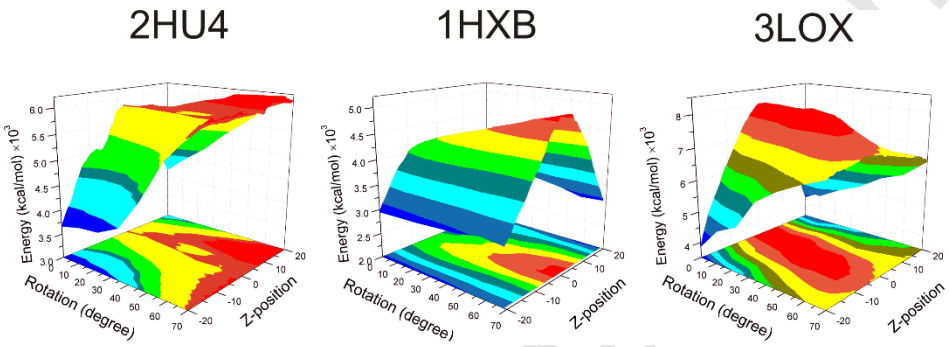


Figure 5. The $\Delta\Delta G_{elec}$ landscapes of three water-soluble proteins.

As expected, the electrostatic energy of membrane solvation dramatically increased at the expense of structural exposure to the low dielectric environment of the membrane. From the plots, the solvation energy of the soluble proteins in the membrane region tends to increase as the depth for protein penetrating into the bilayer increases. Apparently, the energy barrier landscape for the water-soluble proteins has no apparent minimum in the lipid bilayer, but instead became energetically less favorable with increasing the level of membrane depth. Of particular interest is the observation that the soluble proteins gave the opposite pattern of the $\Delta\Delta G_{elec}$ landscapes compared to what had been found for the pore-forming transmembrane proteins (Figure 3A). For a particular orientation, the z-position of a soluble protein (PDB ID: 3LOX) seems to produce an "uphill" pathway pattern that the $\Delta\Delta G_{elec}$ became most unfavorable at the middle of the bilayer. These results reflect the natural preference of the

solvent environment for which these proteins are located. In the section 4.4, we have carried out an extensive investigation of the $\Delta\Delta G_{elec}$ features for transmembrane proteins (TMP) and non-transmembrane proteins (NMP) to explore more any distinctive patterns of the electrostatic energy of membrane solvation.

4.4. Examining positions of proteins in OPM database

4.4.1 Representative proteins of five classes

OPM database is a collection of transmembrane and membrane-associated proteins, whose structural orientations with respect to the lipid bilayer were pre-determined by minimizing the transfer energy of the protein atoms from water to the membrane slab [27]. Here, we generated the $\Delta\Delta G_{elec}$ landscape for fifteen representative membrane proteins of five classes categorized according to the OPM classification (Figure 6). These include α-helical polytopic (PDB ID: 2Z73, 3QAK and 3L1L), α-helical bitopic (PDB ID: 2KV5, 2KWX and 1ITA) and β-barrel (PDB ID: 1QJ8, 2QOM and 1KMO), α-helical peptide (PDB ID: 1RON, 2XDC and 2JXF) and monotopic/peripheral (PDB ID: 1N69, 3NA0 and 1NR6). The first three classes are TMP as they contain transmembrane domains while the last two classes are NMP. Detailed calculations for $\Delta\Delta G_{elec}$ of the OPM protein models were performed with the same scenario used for the pore-forming proteins (section 4.1), except that the starting configuration was taken from the atomic coordinates of the OPM proteins instead of searching for it. This is because we wanted to reduce the amount of configurations to be examined for a large number of proteins in the database. Secondly, the orientation and position of the OPM proteins had been pre-optimized based on the Lomize method[27]. It was found that the positions given by the OPM database are the same or almost identical to our results tested with the pore-forming

membrane proteins and other examined TMPs. Thus, they also serve as an ideal reference for comparison.

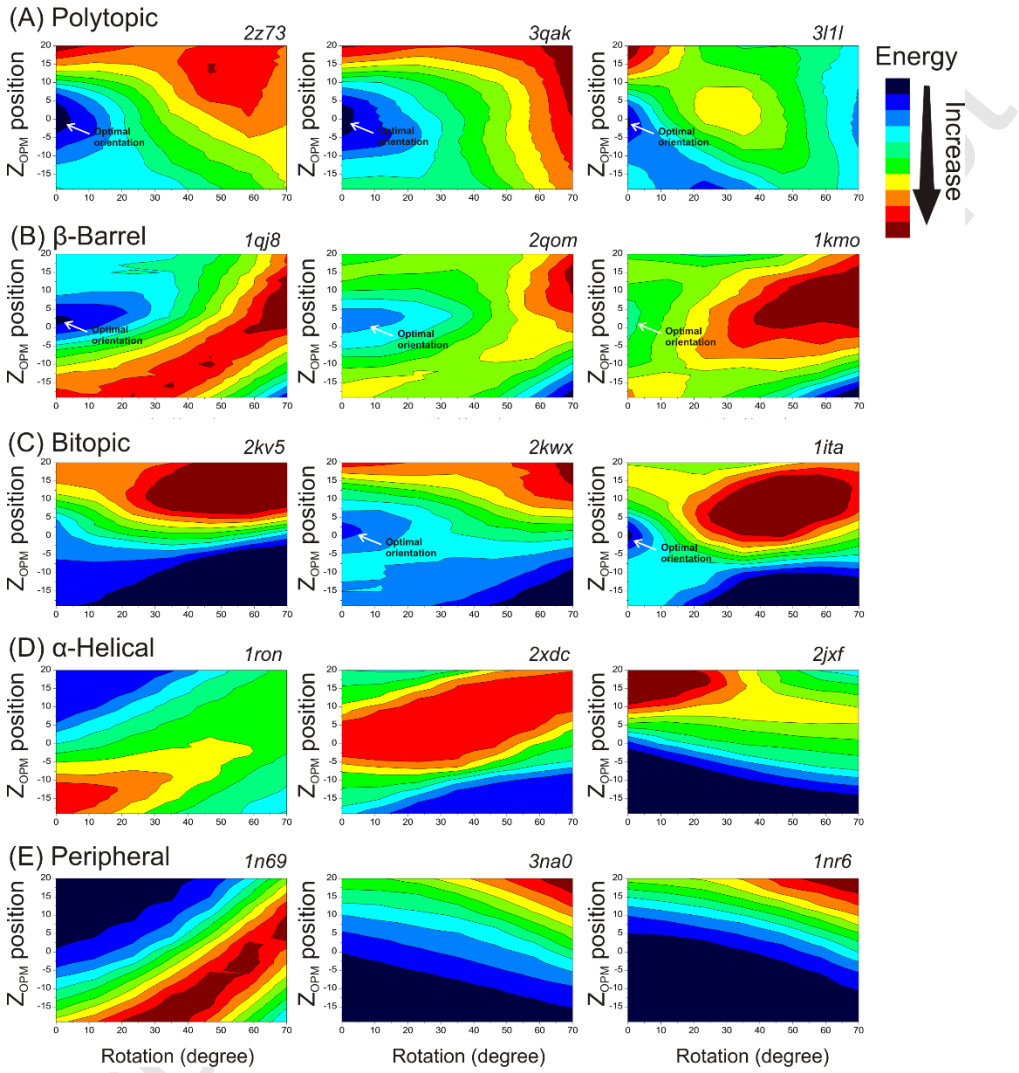


Figure 6. The $\Delta\Delta G_{elec}$ contours of the five classes of membrane proteins: (A) polytopic, (B) β-barrel, (C) bitopic, (D) α-helical peptide and (E) monotopic/peripheral proteins. For each type, three representative proteins were examined. Protein PDB codes are given.

From an analysis of the $\Delta\Delta G_{elec}$ landscapes of membrane solvation, we were able to identify the minimum energy in a lipid bilayer for transmembrane proteins of α-helical polytopic, α-helical bitopic and two of the β-barrel (2KWX and 1ITA) classes (Figure 6A-6C). As the membrane insertion increases, the energy decreases and approaches a minimum value near the centre of hydrophobic slab. The pattern of the $\Delta\Delta G_{elec}$ landscape is similar to that observed in the pore-forming membrane proteins, indicating that there is a conserved pattern of the solvation energy barrier among various TMP classes. The optimal configuration of most proteins in this set had $\theta \approx 0^{o}$ and $-5 \leq z \leq 5$, implying no significant shift in the orientation and position with respect to the starting configuration. This suggested the optimal configuration obtained from the PB approach is consistent with that obtained from the OPM database. In other words, the two different approaches yield generally similar results for the case of TMP. For 2KV5 (Figure 6C), a peculiar pattern of the $\Delta\Delta G_{elec}$ landscape cannot yet be described. Because this protein is a membrane-bound toxin peptide, the detailed structure of this protein may not belonged to the TMP class .

In case of NMP, the energy pattern of the membrane solvation (Figure 6D and 6E) is essentially different from that of the TMP class. All the energy landscapes of membrane solvation did not manifest the minimum energy configuration within a membrane. The absence of the optimal configuration of NMP can be understood from the perspective that the proteins of this class do not contain any transmembrane segment. These results are similar to water-soluble proteins in a way that the insertion of proteins into a membrane environment results in a more unfavorable electrostatic energy of solvation.

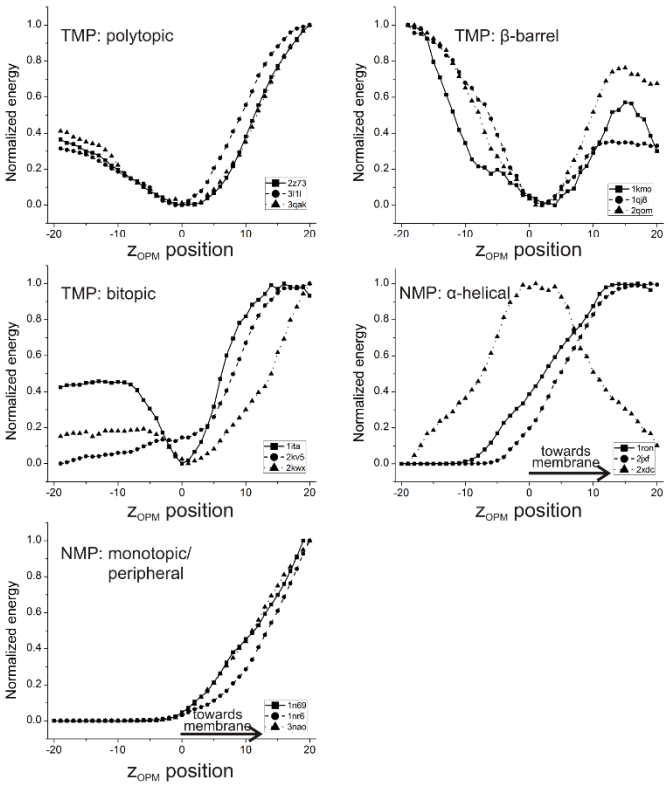


Figure 7. The 1D $\Delta\Delta G_{elec}$ profiles of transmembrane proteins (TMP) and non-transmembrane proteins (NMP) at a fixed rotation angle (θ = 0). Inset arrow indicates the direction when moving the protein towards the membrane region.

It is helpful to provide a simple illustration of how the minimum energy can be located through a one dimensional (1D) energy profile, as shown in Figure 7. According to the results, consistent with the OPM orientation, the energy profiles of all 15 proteins were at the best orientation of θ = 0º (no rotation). The energy values are normalized to their maximum value. For the α-helical polytopic proteins, most of the $\Delta\Delta G_{elec}$ profiles showed a V-like curve around the minimum, which is pronounced near the starting configuration of the OPM proteins (z_{OPM} = 0). A similar minimum-energy position was also observed in the $\Delta\Delta G_{elec}$ profiles of β-

barrel and bitopic (except for 2KV5) proteins. Unlike the transmembrane proteins, α-helical peptides and peripheral/monotopic proteins have no canonical amphipathic topology with sufficiently long enough to span a lipid bilayer. The results of the NMP appear to be as expected, in that the $\Delta\Delta G_{elec}$ profile does not have an apparent minimum. According to visual inspection, the structures of the representative NMP are either partially embedded in the lipid bilayer or just remain intact on the membrane surface. Thus, the $\Delta\Delta G_{elec}$ value of these proteins monotonously increases as the protein moves into the membrane.

4.4.2 Comparison with OPM proteins

The electrostatic free energy of membrane solvation for representative membrane proteins of five classes revealed a distinct patterns between TMP and NMP (Figure 4, 5 and 6). We further examined membrane protein positions for a total of 1060 OPM proteins by using the PB method (Table S1). This protein set was comprised of 285 polytopic proteins, 33 β-barrel proteins, 64 bitopic proteins, 214 α-helical peptide and 464 monotopic/peripheral proteins. The proteins in this set are TMPs and NMPs having a variety of structural architectures. The 1D-$\Delta\Delta G_{elec}$ profile along the z-axis was used to suggest protein positions and discriminate between TMP and NMP. As shown by the results of section 4.4.1, TMP has a distinct pattern of the solvation energy from NMP. The pattern is defined, whether in the energy profile, the solvation energy decreases or increases as a function of the z-position and there exists the minimum energy of membrane solvation in the bilayer region. The criteria for accounting a consistent pattern of the solvation energy of TMP is that the $\Delta\Delta G_{elec}$ profile belonging to polytopic, β-barrel and bitopic proteins must exhibit a V-like shape with a

minimum on the z-position axis. On the other hand, the energy profiles of α-helical peptide and monotopic/peripheral proteins has the pattern of NMP.

Table 1. Summary of the examined position of 1060 OPM proteins in each of five classes

Protein Class	Number of proteins[1]/Total
TMP	
α-Helical polytopic	248/285
β-Barrel	26/33
α-Helical bitopic	21/64
NMP	
α-Helical peptide	202/214
Monotopic/peripheral	461/464

[1] Number of proteins that are identified to have a consistent pattern of the $\Delta\Delta G_{elec}$ profile. For TMP, the profile proceeds in a downhill direction as a function of membrane insertion depth and has the minimum point in the bilayer. For NMP, the profile does not have the feature described in the TMP.

The energy pattern analysis revealed that 87, 79 and 33% of the polytopic, β-barrel and bitopic proteins, respectively, produced a $\Delta\Delta G_{elec}$ profile consistent with transmembrane proteins (Table 1). From the results, one can see that the lowest percent of the consistency pattern was to the proteins in the bitopic dataset. This may be mainly a result of the missing structural subunits for the proteins in this dataset. It has been reported that many bitopic proteins have been characterized as a functional homo- or heteromultimer, and share a structural domain consisting of only one transmembrane helix[47]. Therefore, protein models extracted from database are missing at least one structural domain from the other subunits which alters our calculation. In addition, a closer look at the structure and spatial arrangement of the bitopic α-helical transmembrane proteins used in the study revealed a variety of structural forms with some being monomeric or dimeric membrane anchors of a single

transmembrane helix but with many being extracellular and/or cytoplasmic parts without a transmembrane segment. This is in contrast to the polytopic and β-barrel datasets which contain more complete structure information and organization of transmembrane bundle compared to bitopic proteins. Although the biologically relevant oligomeric structures of many proteins may be extracted from the "biological unit" provided in the Protein Data Bank or Protein Quaternary Server, these information are mostly based on the prediction of single crystal structure and still controversial[48, 49]. The detailed physiologically relevant structure and packing of bitopic proteins would need more investigation to obtain more conclusive understanding about the energy pattern of this class.

The analysis of the $\Delta\Delta G_{elec}$ profiles of NMP found that 94 and 99% of α-helical peptide and peripheral proteins, respectively, have no minimum-energy position in the bilayer. The energy profiles are somewhat similar to what has been observed in Figure 6D and 6E. In case of α-helical peptide, the 3D structure contains only a single peptide and is partially embedded or anchored to the membrane from one side. Therefore, the pattern of $\Delta\Delta G_{elec}$ profiles are more similar to peripheral/monotopic class where they mostly attach at the membrane surface.

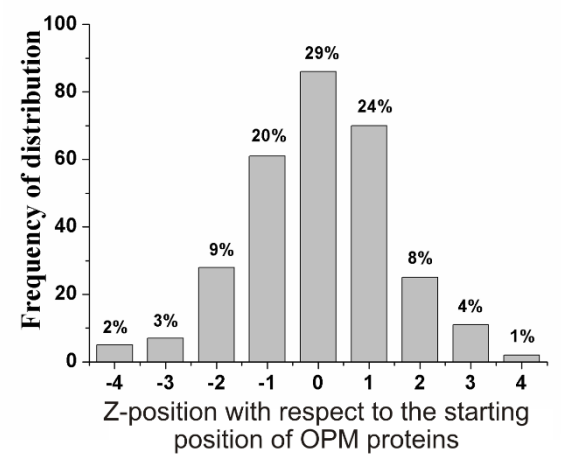


Figure 8. Frequency distribution histogram for the position difference between the minimum-energy Z-position obtained from the present approach and that from the OPM models.

A comparison between the minimum-energy positions obtained by the PB approach and the reference position of the OPM proteins was carried out. A total of 295 (248 + 21 + 26) transmembrane proteins that had been verified as the consistency pattern by this present method and the OPM database (Table 1) were analyzed for the deviation of z-position. The results are illustrated as a frequency distribution histogram (Figure 8). In the x-axis, the relative position equal to zero indicates that the present and the OPM models gave the same optimal position. About 29% of the tested transmembrane proteins have the position of proteins in lipid bilayer identical to the OPM proteins. The consistency of the z-position between the two methods increased to 73 and 90% by taking into account a deviation of 1 and 2 Å, respectively. As the results, the present method achieved nearly the same results as the OPM approach.

4.3 Membrane solvation penalty of 20 amino acids

The solvent exposed surface of water-soluble proteins have generally a hydrophilic property whereas membrane exposed residues of transmembrane segments are more hydrophobic in order to reduce the energy cost of solvation [50, 51]. For a quantitative understanding of energy contribution of individual amino acid, we further performed an energy decomposition on a per-atom basis to compute the membrane solvation energy of individual amino acids that are exposed to the lipid environment. Using the energy data of the energy-minimum position of about 250 non redundant TMP models, the per-atom contributions were summed over the atom group of residues in protein sequence. Furthermore, buried residues were eliminated to avoid large variations of the electrostatic solvation energy due to different degree of surface exposure of individual residues [52]. Buried and exposed residues were identified by the ratio of the solvent accessible surface area (SASA) of the residue in the protein to that in the extended models. SASA of extended Gly-Xxx-Gly peptide models were determined in order to estimate the exposure of amino acid residues without interfering with neighbouring side chains. The program ACCESS was used to compute SASA per residue [53]. Hydrogen atoms were omitted in the SASA calculation and a water-probe radius (1.4 Å) was used. For a given structure, a residue whose SASA is less than 50% with respect to that in a fully extended conformation, is not considered exposed. The cut-off for lipid-exposed residues is based on the previous analysis of helix-helix packing patterns in membrane proteins [54].

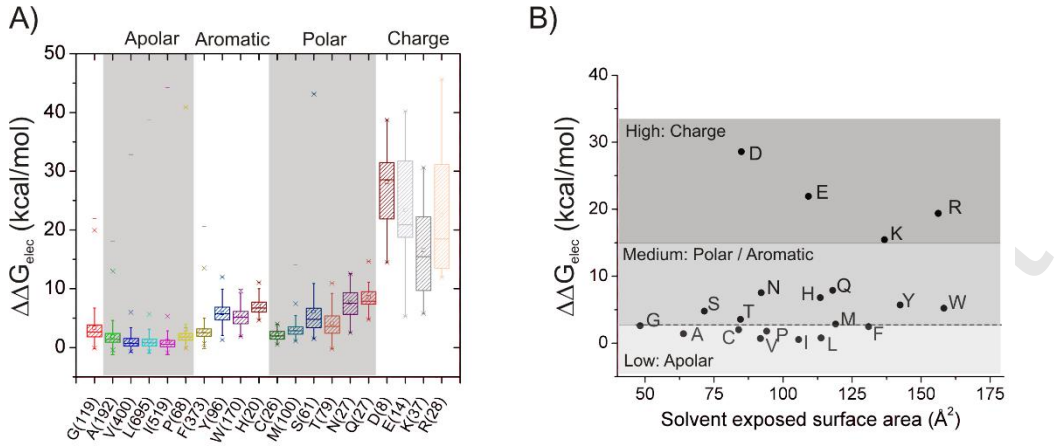


Figure 9. Statistical results of electrostatic contribution of the 20 amino acids at non-buried site. (A) Energy contribution showing their centre and range as box-and-whisker plots. The bottom and the top of the box are 25 and 75 percentile. The median is the straight line, and the mean is plotted as a square. Whiskers are two vertical lines extending from the top and bottom of the box. Numbers in parenthesis along the x-axis are the total number of data used for the statistical analysis. (B) $\Delta\Delta G_{elec}$ versus solvent exposed surface area of the 20 amino acids. Based on a general classification, all amino acids except for glycine are categorized into 4 groups. The ionizable residues (D, E, K and R) are present in the charged state.

The statistics of the distribution of the electrostatic solvation energy barrier for 20 amino acids were shown as boxplots in Figure 9A and three basic descriptive statistics were summarized in Table 2. The spread and the center of energy data in the boxplot were found to be compatible within each amino acid class (Figure 9B). For charged amino acid residues, the electrostatic solvation penalty exhibited the largest spread with the most disfavoured solvation energy (from 20 to 30 kcal/mol on the median values) as expected from unfavourable

interactions of charged side chains such as carboxylate, ammonium and guadinidinium groups in the low dielectric of the membrane. As has been known that most of these charged residues in proteins are not completely exposed themselves to the lipid environment [51]. They are usually found to be at a partially exposed site. In contrast, apolar residues produced the smallest electrostatic energy cost for membrane solvation with the median values about 0-3 kcal/mol. However, apolar residues in some side-chain orientation could make a favorable energy contribution with a few kcal/mol, as indicated by the negative values within the lower outlier of the boxplots. The energy cost for membrane solvation of aromatic residues was somewhat similar to that of polar residues, with the exception that the spread and the center of $\Delta\Delta G_{elec}$ for Phe, Cys and Met fell within the energy range of apolar residues. It is noteworthy that the membrane solvation penalty of Gly is slightly higher than that of apolar residues. This is owing to the fact that the $\Delta\Delta G_{elec}$ calculation for Gly was mainly determined from the backbone, where unfavorable electrostatic potentials on oxygen and nitrogen atoms are greater than those on methyl, methylene and methine carbons. As indicated by the atomic partial charges in the PARSE parameter set, the more polar backbone of Gly are exposed to the low dielectric environment, giving rise the energy penalty greater than that of the hydrophobic sidechain of apolar residues. Similarly to aromatic residues containing the polar oxygen and nitrogen heteroatoms, the membrane solvation of Tyr, Trp and His is associated with the energy penalty larger than Phe.

Table 2 First quartile (Q_1), median and third quartile (Q_3)[1] of $\Delta\Delta G_{elec}$ of 20 amino acids whose lipid-exposed surface is greater than 50% with respect to the extended models.

	$\Delta\Delta G_{elec}$(kcal/mol)			SASA(Å)^2			
	Q_1	Med	Q_3	Q_1	Med	Q_3	Total[2]
GLY	1.8	2.6	3.8	45	48	55	86
ALA	0.9	1.4	2.4	60	64	70	116
VAL	0.2	0.7	1.5	85	92	99	161
LEU	0.3	0.8	1.4	107	114	123	202
ILE	0.1	0.5	1.2	99	106	114	187
PRO	1.3	1.8	2.4	87	94	104	162
PHE	1.9	2.5	3.2	121	131	144	224
TYR	4.7	5.7	6.9	132	142	154	242
TRP	4.0	5.2	6.1	142	158	178	263
HIS	6.1	6.8	8.0	107	114	124	202
CYS	1.4	2.0	2.7	79	84	86	147
MET	2.3	2.8	3.6	107	119	134	198
SER	3.4	4.8	6.7	68	71	80	130
THR	2.4	3.6	5.4	81	85	93	155
ASN	5.7	7.5	9.3	88	92	116	169
GLN	7.3	7.9	9.5	102	118	126	195
ASP	24.6	28.6	32.7	82	85	101	159
GLU	18.8	21.9	31.8	99	109	124	188
LYS	9.7	15.4	22.2	122	137	144	232
ARG	13.6	19.4	31.4	144	156	171	264

[1]These descriptive statistics are not as heavily influenced by outliers.

[2] Total solvent accessible surface area of amino acid residues were determined from extended Gly-Xxx-Gly peptide models.

One can see that the population of amino acid residues in transmembrane domains (at the x-axis label in Figure 9A) is predictably affected by the common hydrophobic properties; apolar amino acids are the most commonly found residues, followed by aromatic, polar and finally charged residues. It appears that Asp and Glu residues are rarely found. Aromatic residues are more frequently found than polar residues. The results may be common but have suggested that the selected proteins in the dataset and the cut-off value are adequate for a reliable statistical assessment. Except for Asp and Glu, an amount of the amino acid data is

large enough for a statistical analysis to estimate the energy contribution to the solvation energy on a per-residue basis.

Figure 9B showed the plot of the median values of $\Delta\Delta G_{elec}$ and the lipid-exposed surface area. The scatter of data indicated that the membrane solvation penalty may be affected not only by the amount of the exposed surface area but also the magnitude of atomic partial charges. Although there was no direct linear relationship of the two variables, the plot combined with the classification of amino acid groups provides a useful scheme for estimating the solvation energy contribution made by amino acids. Depending on the surface exposure in membrane, the solvation penalty scale may be ranked from low for apolar hydrophobic residues, through medium for polar and aromatic residues, to high for charged residues. The results strongly agree with a general observation about the frequency of amino acid compositions found in transmembrane domains. Therefore, the energy penalty estimation should be applicable to membrane proteins of various structural architectures and will provide utility for the structural modelling of membrane proteins.

4.4 Solvation energy change by mutation and decoy discrimination

To demonstrate the advantageous of our calculation, two specific studies were carried out. From the results obtained in section 4.1, we selected the optimal $\Delta\Delta G_{elec}$ of KcsA, a representative protein, and performed the energy decomposition analysis for the per residue membrane solvation penalty. Figure 10A illustrates the molecular surface of KcsA colored in accordance with the $\Delta\Delta G_{elec}$ values of each residue as shown in Figure 10B. While the energy penalty on the outside membrane is mostly negligible, the most apparent energy difference took place for residues in membrane and membrane-water interface regions. In comparison of

the $\Delta\Delta G_{elec}$ distribution within the protein, most residues in a lipid bilayer gave a wide range of membrane solvation penalty ranging from 0.5 to 6 kcal/mol. Residues that exhibited a relatively high penalty energy in the bilayer region correspond to the polar or charged residues such as T33 and Y45 of TM1, E71 of P-helix, T74 and T75 near the selectivity filter (SF) and C90 and T107 of TM2. It should be noted that our calculations did not treat the neutral form of the E71 residue because this residues is expected to be protonated[55]. From this plot, we also found that the burial or partial exposure (SA < 0.5) of some hydrophobic residues in TM1 and TM2 helices made favourable energy contribution to $\Delta\Delta G_{elec}$ up to -1 kcal/mol.

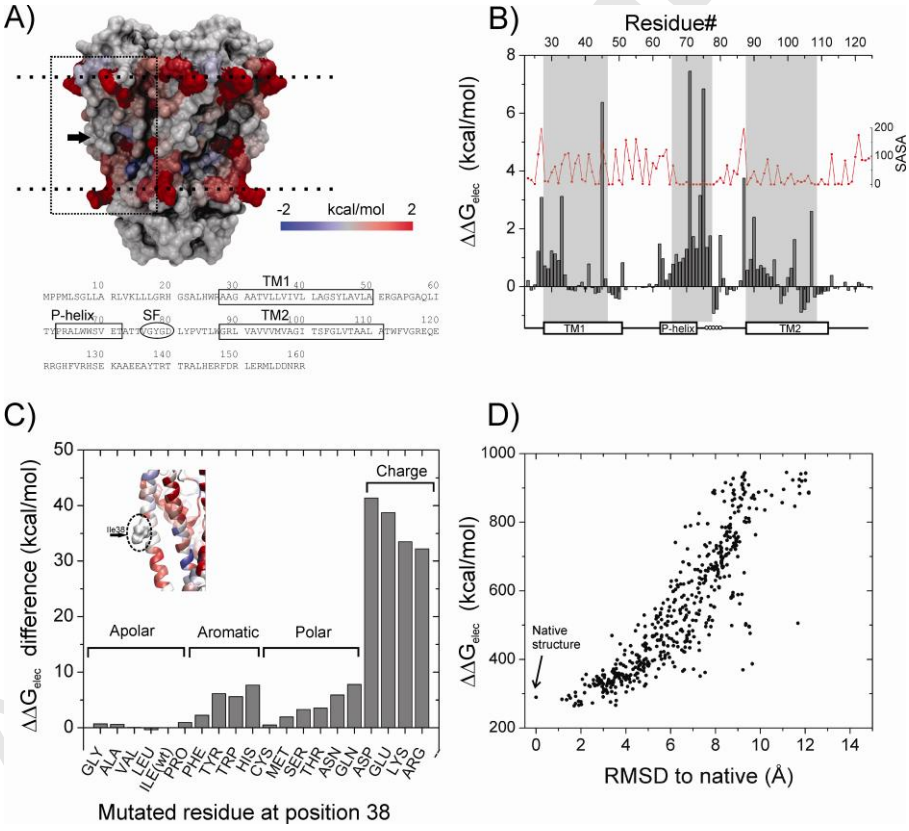


Figure 10. (A) $\Delta\Delta G_{elec}$ values were colored coded and mapped on the crystal structure of KcsA whose amino acid sequence is listed below. The colors represent the per residue $\Delta\Delta G_{elec}$ values where blue, grey and red are at -2, 0 and 2 kcal/mol, respectively, with linear interpolation between values. (B) $\Delta\Delta G_{elec}$ (bar graph) and solvent accessible

surface area (line) per residue along the KcsA sequence. The membrane region (the shaded areas) is mapped with transmembrane segments TM1 and TM2, pore helix and selectivity filter shown below the plot. (C) Membrane solvation penalty of substituted residues at the position 38 and inset showing the substituted position where the I38 sidechain faces to membrane (highlighted by the black arrow in (A)). (D) Scatter plot of $\Delta\Delta G_{elec}$ vs. RMSD to native for 550 decoys, along with a representative protein structure.

To examine a change in electrostatic solvation energy due to amino acid substitution, a set of structure models of "in silico" KcsA mutants modified at the membrane-exposed surface corresponding to the position Ile38 (Figure 10B) was built and subsequently subjected to the PB calculation to extract the $\Delta\Delta G_{elec}$ values. The results show that the largest change in the $\Delta\Delta G_{elec}$ value are for the substitution with charged residues. When Ile38 is replaced by non-aromatic apolar residues, the solvation energy seems to be not significantly perturbed. The solvation energy has increased with a range from 3 to 8 kcal/mol for the substituted residues with either polar or aromatic groups on the surface of the protein. This is except for the substitution of cysteine which gave $\Delta\Delta G_{elec}$ nearly equivalent to that of apolar group. Figure 10C shows that change in $\Delta\Delta G_{elec}$ due to the amino acid substitution is correlated well with the statistical results for electrostatic contribution of the membrane-exposed amino acids (Figure 9).

The second study is concerned with the identification of the native structure from decoys using $\Delta\Delta G_{elec}$. We generated a representative dataset corresponding to the KcsA x-ray structure. This dataset contains more than 500 decoys that were generated by treating the rigid-body movements of the two transmembrane helices of KcsA. By employing the steered

molecular dynamics (SMD) simulations on the obtained optimal configuration (section 4.1), this decoy test set consisted of distorted conformations or partially unfolded KcsA structures with the RMSD range from 1 to 12 Å relative to the native conformation. The SMD simulations were performed with NAMD [56]. Subsequently, each decoy was subjected to the PB calculations to obtain $\Delta\Delta G_{elec}$ of the representative protein structure.

Figure 10D shows $\Delta\Delta G_{elec}$ vs. RMSD for the decoys along with the native conformation. A good linear correlation between the two variables was observed with R^2 of 0.76. Correct or near-native folded protein conformations have lower value of $\Delta\Delta G_{elec}$ than partially unfolded structures. Based on the plot, $\Delta\Delta G_{elec}$ using the PB calculations did perform consistently well to distinguish between good and bad folded conformations, implying its potential application as a solvation parameter constraint for membrane protein structure prediction. However, the results need to be carefully interpreted to avoid overstate of efficiency due to a limited sampling size in terms of decoy number, only one representative protein and structure perturbation on specific transmembrane region. These issues would be addressed in future to provide a common perception.

5. Discussion

A considerable number of studies have been conducted previously that evaluated the membrane protein position relative to the lipid bilayer [22-24, 26-28, 57-59]. These studies were performed in various ways and presented different degrees of success. Here we have demonstrated that the exact electrostatic description of protein-membrane interactions by the PB method allows us to identify the optimal position of membrane proteins in a lipid bilayer, particularly for transmembrane proteins. The evaluation of the OPM proteins with the PB

solvation energy revealed definitive patterns to discriminate transmembrane proteins from non-transmembrane proteins.

Based on the solvent continuum approach, all the $\Delta\Delta G_{elec}$ values of proteins were positive, which suggested that the total electrostatic contribution for transferring proteins from the aqueous phase to a lower-dielectric membrane is always unfavourable. Because $\Delta\Delta G_{elec}$ is the electrostatic contribution from the polar part and purely enthalpic, other contributions such as non-polar and entropic energies can generally improve an accurate prediction of the solvation free energy[29]. MD simulations with explicit solvent molecules offers the way to explore its significance in the determinant of protein's orientation in a lipid bilayer but they are computational demanding. We have, however, demonstrated that the positioning of proteins in a lipid bilayer can be judged and verified by the relative comparison of PB energy differences. The continuum electrostatic approach gave the results with sufficient reproducibility and the omission of other contributions to solvation energy do not prohibit its usefulness for searching membrane protein's position in membranes.

It should be noted that in our calculations, the solvent model for both extracellular and intracellular sides was treated as a high dielectric constant with the same value equal to that of water. This is based on an assumption that water is a major component present in outer and inner cells. As a consequence, the solvent-exposed residues of the protein experience a similar environment at an external and internal of cells. The present approach is, therefore, not suitable for distinguishing the extracellular and intracellular orientations of the protein since it does not account for an influence of the difference between extracellular and intracellular fluids on the spatial arrangement of the protein. Considering a variety of physiological environments, this issue is difficult to address with precision due to a variety

of factors such as salt concentrations, pH or lipid compositions etc. It needs a further careful study, which is beyond the present scope of this work.

A great feature for using the present method is that the calculation is physically-based model. It is expected to have a good transferability. We showed that the comparison of free energy difference of membrane solvation by the continuum electrostatics is able to capture the correct position and conformation of various protein structures in membrane without the need to treat the shape of bilayer deformation and nonpolar energy contribution. Therefore, the PB continuum method can be considered to be a useful tool for optimizing protein positions in lipid bilayer. Another potential application of the present approach is to use it as a tool for assessing the quality of membrane protein structure models obtained by either theoretical or low-resolution experimental techniques. The proposal is that it be routinely used for setting up an initial configuration for performing molecular dynamics simulations of membrane protein systems. A number of studies have shown a dramatic effect on the stability and structure quality of the simulations due to hydrophobic mismatch [12]. Based on a preliminary test using all-atom MD simulations of a potassium KcsA channel embedded in POPC bilayer, we observed an attempt of the system to reorient the protein toward the optimal orientation when the initial configuration for the simulations is shifted from the position with the lowest $\Delta\Delta G_{elec}$ value (Supplementary Information). It is anticipated that the present approach will be of great assistance in the proper position placement of transmembrane proteins in a lipid bilayer. In addition, this may have a future application for assisting in structure refinement of crystallographic studies.

Since this study used the charged model, the energy values (in Figure 9 and Table 2) could be overestimated for Asp and Glu. It has been proposed that Asp and Glu would

preferably be in the neutral form in all membrane region whereas Lys and Arg may remain their charged state in the membrane due to the formation of water defects [60-62]. Because of the variation of membrane hydrophobicity, those lipid-facing Lys and Arg could, nevertheless, change in their ionization state depending upon how depth they are positioning within the membrane. In particular, the free energy to transfer Arg from water to the center of the membrane between the charged or neutral forms is not significantly different. In this regard, it is hard to determine the major species present in the system with the present approach. Interestingly, the predicted free energies from MD studies are of 14-17 kcal/mol for Arg [60, 63], which is comparable to that obtained in this study.

The membrane solvation penalty of amino acids has been derived computationally through statistical analysis of known structures of membrane proteins. We believe that although the solvation penalty values were based on an approximation of continuum models, it is useful for qualitative prediction. It has a potential application to predict the transmembrane topology of membrane proteins. However, it should be used great care when quantitative predictions are involved.

Our results support the notion that the Poisson-Boltzmann continuum solvent model provides a useful means for evaluating positions and interactions of proteins in the membrane. As compared to other existing methods such as PPM[64], TMDET[57] and IMPALA[65] where the empirical energy function is employed, our approach does not require any empirical parameters or training sets, other than atomic charges and radii. Thus this method is widely applicable to a variety of membrane protein systems especially those bound with metal ions or ligand (drug, toxin etc.) molecules, of which the empirical methods have neglected or excluded in the calculations. It provides the option to perform

energy decomposition on a per atom basis which allows to predict the influence of point mutations on protein stability. According to the results obtained from this study, the penalty scale of membrane solvation of each amino acid should be generally useful to predict the effect of mutations on the thermodynamic stability of transmembrane proteins. In this light, a detailed analysis of the electrostatic contribution, specific orientations and degrees of lipid-exposure of individual residues or transmembrane segments in the membrane region is currently being investigated. This should provide more insight into the nature of amino acid propensities in membrane proteins.

The major weakness of our method is that solving the finite-difference PB equation is still computationally demanding. It takes about 3-5 minutes of a large protein (408 residues) on a personal computer Intel Core I7 at 3.4 GHz with 8 GB of memory. Nevertheless, the method is versatile allowing for it to be used in wide range of applications. As have been shown in Figure 10D, it is capable of making significant contribution to structure prediction by allowing effective discrimination between decoys and native structures. Additionally, it holds great potential for structure-function studies since many membrane proteins undergo structural transitions in order to switch between closed (or inactive) and open (or active) forms. The pathway associated with such conformational changes in transmembrane regions could be mapped with the energy landscape of membrane solvation.

6. Conclusions

We promote the use of a Poisson–Boltzmann continuum electrostatic model to evaluate membrane protein positions within a lipid bilayer. The approach involves in the

calculation of the different electrostatic free energies of the protein in water and lipid environments treated as implicit solvent models. The protein orientation and membrane thickness were examined based on the $\Delta\Delta G_{elec}$ landscapes and profiles of five ion channel membrane proteins and the results are in good experimental agreement. The results highlight different characteristic patterns of the solvation energy landscape as a function of penetration depth between transmembrane proteins and soluble proteins. Furthermore, overall good agreement between our results and structural data of the OPM database is suggestive of the potential implication for use of the PB method in fine-tuning membrane protein position within a lipid membrane. The energy decomposition analysis for membrane proteins of well-defined positions gave a general scheme useful for estimating the membrane solvation penalty of individual amino acid residues. The present approach can be used as guidelines for setting up the MD system or discriminating correct from incorrect folds in protein structure prediction.

Acknowledgments

The authors would like to thank Olivier Dalmas for critical reading of this manuscript. This research has been supported by the Ratchadaphiseksomphot Endowment Fund 2013 of Chulalongkorn University (CU-56-408-HR) to PS and the 100th Anniversary Chulalongkorn University Fund for Doctoral Scholarship to CS.

Supplementary data

Supplemental data include the schematic representation of ion channel proteins and membrane proteins of five classes (Figure S1 and S3, respectively), the APBSmem module used for

setting up the calculation (Figure S2), validation by MD simulations (Figure S4), hydropathy plots (Figure S5) and a list of the PDB ID codes of the 1060 tested OPM proteins (Table S1). Structure coordinates where the optimum position in membrane can be identified can be accessed from our web server at http://atc.atccu.chula.ac.th/PB-membrane.

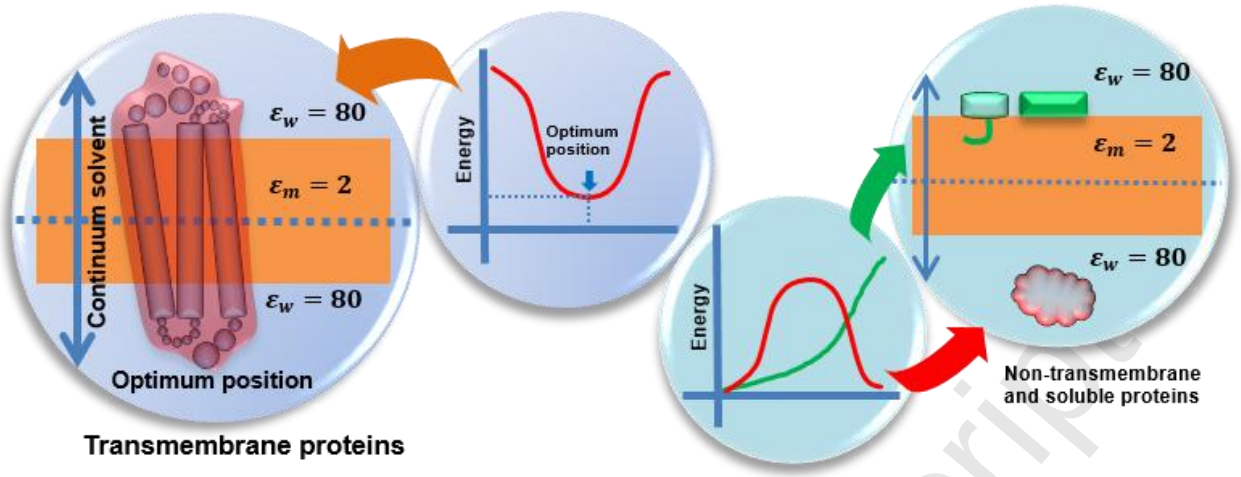
References

[1] E. Wallin, G. von Heijne, Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms, Protein Sci, 7 (1998) 1029-1038.
[2] S.H. White, W.C. Wimley, Membrane protein folding and stability: physical principles, Annu Rev Biophys Biomol Struct, 28 (1999) 319-365.
[3] C. Hunte, Specific protein-lipid interactions in membrane proteins, Biochem Soc Trans, 33 (2005) 938-942.
[4] A.G. Lee, How lipids affect the activities of integral membrane proteins, Biochim Biophys Acta, 1666 (2004) 62-87.
[5] J.U. Bowie, Stabilizing membrane proteins, Curr Opin Struct Biol, 11 (2001) 397-402.
[6] W.F. DeGrado, H. Gratkowski, J.D. Lear, How do helix-helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles, Protein Sci, 12 (2003) 647-665.
[7] J.J. Lacapere, E. Pebay-Peyroula, J.M. Neumann, C. Etchebest, Determining membrane protein structures: still a challenge!, Trends Biochem Sci, 32 (2007) 259-270.
[8] J.W. Taraska, Mapping membrane protein structure with fluorescence, Curr Opin Struct Biol, 22 (2012) 507-513.
[9] D.T. Murray, N. Das, T.A. Cross, Solid State NMR Strategy for Characterizing Native Membrane Protein Structures, Acc Chem Res, (2013).
[10] H.S. Mchaourab, P.R. Steed, K. Kazmier, Toward the Fourth Dimension of Membrane Protein Structure: Insight into Dynamics from Spin-Labeling EPR Spectroscopy, Structure, 19 (2011) 1549-1561.
[11] A. Arora, Solution NMR spectroscopy for the determination of structures of membrane proteins in a lipid environment, Methods Mol Biol, 974 (2013) 389-413.
[12] R.J. Law, C. Capener, M. Baaden, P.J. Bond, J. Campbell, G. Patargias, Y. Arinaminpathy, M.S.P. Sansom, Membrane protein structure quality in molecular dynamics simulation, J Mol Graph Model, 24 (2005) 157-165.
[13] S.J. Opella, F.M. Marassi, Structure determination of membrane proteins by NMR spectroscopy, Chem Rev, 104 (2004) 3587-3606.
[14] G.E. Fanucci, D.S. Cafiso, Recent advances and applications of site-directed spin labeling, Curr Opin Struct Biol, 16 (2006) 644-653.
[15] W.L. Hubbell, C. Altenbach, C.M. Hubbell, H.G. Khorana, Rhodopsin structure, dynamics, and activation: a perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking, Adv Protein Chem, 63 (2003) 243-290.
[16] L. Guan, H.R. Kaback, Lessons from lactose permease, Annu Rev Biophys Biomol Struct, 35 (2006) 67-91.
[17] S. Frillingos, M. Sahin-Toth, J. Wu, H.R. Kaback, Cys-scanning mutagenesis: a novel approach to structure function relationships in polytopic membrane proteins, FASEB J, 12 (1998) 1281-1299.
[18] E. London, A.S. Ladokhin, Measuring the depth of amino acid residues in membrane-inserted peptides by fluorescence quenching, Curr Top Membr, 52 (2002) 89-115.
[19] A.M. Powl, J.N. Wright, J.M. East, A.G. Lee, Identification of the hydrophobic thickness of a membrane protein using fluorescence spectroscopy: Studies with the mechanosensitive channel MscL, Biochemistry, 44 (2005) 5713-5721.

[20] B. Roux, T. Allen, S. Berneche, W. Im, Theoretical and computational models of biological ion channels, Q Rev Biophys, 37 (2004) 15-103.
[21] W.L. Ash, M.R. Zlomislic, E.O. Oloo, D.P. Tieleman, Computer simulations of membrane proteins, Biochim Biophys Acta, 1666 (2004) 158-189.
[22] K.A. Scott, P.J. Bond, A. Ivetac, A.P. Chetwynd, S. Khalid, M.S. Sansom, Coarse-grained MD simulations of membrane protein-bilayer self-assembly, Structure, 16 (2008) 621-630.
[23] M.B. Ulmschneider, M.S. Sansom, A. Di Nola, Properties of integral membrane protein structures: derivation of an implicit membrane potential, Proteins, 59 (2005) 252-265.
[24] M.B. Ulmschneider, M.S. Sansom, A. Di Nola, Evaluating tilt angles of membrane-associated helices: comparison of computational and NMR techniques, Biophys J, 90 (2006) 1650-1660.
[25] F. Basyn, B. Charloteaux, A. Thomas, R. Brasseur, Prediction of membrane protein orientation in lipid bilayers: a theoretical approach, J Mol Graph Model, 20 (2001) 235-244.
[26] F. Basyn, B. Spies, O. Bouffioux, A. Thomas, R. Brasseur, Insertion of X-ray structures of proteins in membranes, J Mol Graph Model, 22 (2003) 11-21.
[27] A.L. Lomize, I.D. Pogozheva, M.A. Lomize, H.I. Mosberg, Positioning of proteins in membranes: a computational approach, Protein Sci, 15 (2006) 1318-1333.
[28] M.A. Lomize, A.L. Lomize, I.D. Pogozheva, H.I. Mosberg, OPM: orientations of proteins in membranes database, Bioinformatics, 22 (2006) 623-625.
[29] S. Choe, K.A. Hecht, M. Grabe, A continuum method for determining membrane protein insertion energies and the problem of charged residues, J Gen Physiol, 131 (2008) 563-573.
[30] J. Warwicker, H.C. Watson, Calculation of the electric potential in the active site cleft due to alpha-helix dipoles, J Mol Biol, 157 (1982) 671-679.
[31] M.K. Gilson, M.E. Davis, B.A. Luty, J.A. Mccammon, Computation of Electrostatic Forces on Solvated Molecules Using the Poisson-Boltzmann Equation, J Phys Chem-Us, 97 (1993) 3591-3600.
[32] K.A. Sharp, B. Honig, Electrostatic interactions in macromolecules: theory and applications, Annual review of biophysics and biophysical chemistry, 19 (1990) 301-332.
[33] K.M. Callenberg, O.P. Choudhary, G.L. de Forest, D.W. Gohara, N.A. Baker, M. Grabe, APBSmem: a graphical interface for electrostatic calculations at the membrane, PLoS One, 5 (2010).
[34] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, Proc Natl Acad Sci U S A, 98 (2001) 10037-10041.
[35] Y. Zhou, J.H. Morais-Cabral, A. Kaufman, R. MacKinnon, Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 A resolution, Nature, 414 (2001) 43-48.
[36] S.B. Long, X. Tao, E.B. Campbell, R. MacKinnon, Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment, Nature, 450 (2007) 376-382.
[37] S. Steinbacher, R. Bass, P. Strop, D.C. Rees, Structures of the prokaryotic mechanosensitive channels MscL and MscS, Curr Top Membr, 58 (2007) 1-24.
[38] J. Payandeh, T. Scheuer, N. Zheng, W.A. Catterall, The crystal structure of a voltage-gated sodium channel, Nature, 475 (2011) 353-358.

[39] S. Eshaghi, D. Niegowski, A. Kohl, D. Martinez Molina, S.A. Lesley, P. Nordlund, Crystal structure of a divalent metal ion transporter CorA at 2.9 angstrom resolution, Science, 313 (2006) 354-357.
[40] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J Mol Graph, 14 (1996) 33-38, 27-38.
[41] T.J. Dolinsky, P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, N.A. Baker, PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, Nucleic Acids Res, 35 (2007) W522-525.
[42] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, N.A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations, Nucleic Acids Res, 32 (2004) W665-667.
[43] M. Grabe, H. Lecar, Y.N. Jan, L.Y. Jan, A quantitative assessment of models for voltage-dependent gating of ion channels, Proc Natl Acad Sci U S A, 101 (2004) 17640-17645.
[44] A.G. Lee, Lipid-protein interactions in biological membranes: a structural perspective, Biochim Biophys Acta, 1612 (2003) 1-40.
[45] O.S. Andersen, R.E. Koeppe, 2nd, Bilayer thickness and membrane protein function: an energetic perspective, Annu Rev Biophys Biomol Struct, 36 (2007) 107-130.
[46] A. Kessel, N. BenTal, Free energy determinants of peptide association with lipid bilayers, ACADEMIC PRESS INC, 525 B STREET, SUITE 1900, SAN DIEGO, CA 92101-4495 USA, 2002.
[47] I.T. Arkin, Structural aspects of oligomerization taking place between the transmembrane alpha-helices of bitopic membrane proteins, Biochim Biophys Acta, 1565 (2002) 347-363.
[48] K. Henrick, J.M. Thornton, PQS: a protein quaternary structure file server, Trends Biochem Sci, 23 (1998) 358-361.
[49] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J Mol Biol, 372 (2007) 774-797.
[50] D.M. Engelman, G. Zaccai, Bacteriorhodopsin is an inside-out protein, Proc Natl Acad Sci U S A, 77 (1980) 5894-5898.
[51] D.C. Rees, L. DeAntonio, D. Eisenberg, Hydrophobic organization of membrane proteins, Science, 245 (1989) 510-513.
[52] W.C. Wimley, T.P. Creamer, S.H. White, Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides, Biochemistry, 35 (1996) 5109-5124.
[53] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, J Mol Biol, 55 (1971) 379-400.
[54] M. Gimpelev, L.R. Forrest, D. Murray, B. Honig, Helical packing patterns in membrane and soluble proteins, Biophys J, 87 (2004) 4075-4086.
[55] S. Berneche, B. Roux, The ionization state and the conformation of Glu-71 in the KcsA K(+) channel, Biophys J, 82 (2002) 772-780.
[56] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, Scalable molecular dynamics with NAMD, J Comput Chem, 26 (2005) 1781-1802.
[57] G.E. Tusnady, Z. Dosztanyi, I. Simon, TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates, Bioinformatics, 21 (2005) 1276-1277.
[58] G.E. Tusnady, Z. Dosztanyi, I. Simon, PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank, Nucleic Acids Res., 33 (2005) D275-D278.

[59] C.A. Schramm, B.T. Hannigan, J.E. Donald, C. Keasar, J.G. Saven, W.F. DeGrado, I. Samish, Knowledge-Based Potential for Positioning Membrane-Associated Structures and Assessing Residue-Specific Energetic Contributions, Structure, 20 (2012) 924-935.
[60] J.L. MacCallum, W.F. Bennett, D.P. Tieleman, Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment, The Journal of general physiology, 129 (2007) 371-377.
[61] L. Li, I. Vorobyov, T.W. Allen, The different interactions of lysine and arginine side chains with lipid membranes, J Phys Chem B, 117 (2013) 11906-11920.
[62] J.L. MacCallum, W.F. Bennett, D.P. Tieleman, Distribution of amino acids in a lipid bilayer from computer simulations, Biophysical journal, 94 (2008) 3393-3404.
[63] S. Dorairaj, T.W. Allen, On the thermodynamic stability of a charged arginine side chain in a transmembrane helix, Proceedings of the National Academy of Sciences of the United States of America, 104 (2007) 4943-4948.
[64] M.A. Lomize, I.D. Pogozheva, H. Joo, H.I. Mosberg, A.L. Lomize, OPM database and PPM web server: resources for positioning of proteins in membranes, Nucleic acids research, 40 (2012) D370-376.
[65] F. Basyn, B. Charloteaux, A. Thomas, R. Brasseur, Prediction of membrane protein orientation in lipid bilayers: a theoretical approach, J Mol Graph Model, 20 (2001) 235-244.



Highlights

- We investigated Poisson-Boltzmann method for positioning proteins in membrane.
- We analyzed the solvation energy vs protein positions for 1060 OPM proteins.
- The optimal position can be identified for most of transmembrane proteins
- Energy profiles of transmembrane and non-transmembrane proteins are different.
- new hydrophobic scale has a potential application in membrane protein modeling