

METHODS

Implicit model to capture electrostatic features of membrane environment

Rituparna Samanta¹, Jeffrey J. Gray^{1,2,3*}

1 Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America, **2** Program in Molecular Biophysics, Johns Hopkins University, Baltimore, Maryland, United States of America, **3** Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America

* jgray@jhu.edu

OPEN ACCESS

Citation: Samanta R, Gray JJ (2024) Implicit model to capture electrostatic features of membrane environment. *PLoS Comput Biol* 20(1): e1011296. <https://doi.org/10.1371/journal.pcbi.1011296>

Editor: Turkan Haliloglu, Bogazici University, TURKEY

Received: June 24, 2023

Accepted: December 13, 2023

Published: January 22, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1011296>

Copyright: © 2024 Samanta, Gray. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The F23 energy function is available in the Rosetta software suite (<https://www.rosettacommons.org>). Rosetta is available to noncommercial users for free and to commercial users for a fee. The scripts used to run

Abstract

Membrane protein structure prediction and design are challenging due to the complexity of capturing the interactions in the lipid layer, such as those arising from electrostatics. Accurately capturing electrostatic energies in the low-dielectric membrane often requires expensive Poisson-Boltzmann calculations that are not scalable for membrane protein structure prediction and design. In this work, we have developed a fast-to-compute implicit energy function that considers the realistic characteristics of different lipid bilayers, making design calculations tractable. This method captures the impact of the lipid head group using a mean-field-based approach and uses a depth-dependent dielectric constant to characterize the membrane environment. This energy function Franklin2023 (F23) is built upon Franklin2019 (F19), which is based on experimentally derived hydrophobicity scales in the membrane bilayer. We evaluated the performance of F23 on five different tests probing (1) protein orientation in the bilayer, (2) stability, and (3) sequence recovery. Relative to F19, F23 has improved the calculation of the tilt angle of membrane proteins for 90% of WALP peptides, 15% of TM-peptides, and 25% of the adsorbed peptides. The performances for stability and design tests were equivalent for F19 and F23. The speed and calibration of the implicit model will help F23 access biophysical phenomena at long time and length scales and accelerate the membrane protein design pipeline.

Author summary

Membrane proteins participate in many life processes. They constitute 30% of the human proteome and are targets for over 60% pharmaceuticals. Accurate and accessible computational tools to design membrane proteins will transform the platform to engineer membrane proteins for therapeutic, sensor, and separation processes. While soluble protein design has advanced, membrane protein design remains challenging due to the difficulties in modeling the lipid bilayer. Electrostatics plays an intimate role in the physics of membrane protein structure and function. However, accurately capturing electrostatic energies in the low-dielectric membrane often requires expensive calculations that are not scalable. In this work, we contribute a fast-to-compute electrostatic model that considers different

the benchmark tests are available on Git Hub (<https://github.com/Graylab/Implicit-Membrane-Energy-Function-Benchmark>).

Funding: RS and JJG are supported by the National Institute of Health through grant R35GM141881-03. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: JJG is an unpaid board member (co-director) of the Rosetta Commons. Under institutional participation agreements between the University of Washington, acting on behalf of the Rosetta Commons, Johns Hopkins University may be entitled to a portion of the revenue received on licensing Rosetta software including some methods described in this manuscript. JJG has a financial interest in Cyrus Biotechnology. Cyrus Biotechnology distributes the Rosetta software, which may include methods described in this manuscript. The results of the study discussed in this paper could affect the value of Cyrus Biotechnology. These arrangements have been reviewed and approved by the Johns Hopkins University in accordance with its conflict-of-interest policies.

lipid bilayers and their features, making design calculations tractable. We demonstrate that the updated energy function improves the calculation of the tilt angle of membrane proteins, stability, and confidence in designing charged residues.

This is a *PLOS Computational Biology* Methods paper.

Introduction

Electrostatic interactions are essential in membrane protein (MP) stability, structure, and function. Such interactions determine the way macromolecules interact with membranes: for example, adsorption and internalization of small peptides such as melittin and indolocidin are dependent on the lipid surface charge; [1] signaling and ion-transport through ion channels and active transporters are dependent on the electrostatic potential across the cellular membrane; [2] and association of many peripheral MPs involves electrostatic interactions with the membrane. [3, 4] Particularly, the lipid head groups can be charged, and the hydrophobic core of the bilayer can enhance interactions between charged particles giving rise to biophysical features that are less frequently observed in the aqueous phase. [5] Hydrophobicity scales the propensity of a given residue to be present in the hydrocarbon region. [6, 7] However, some show unusual behavior: for example, residues Arg and Lys snorkel out of the membrane interior to minimize the cost of inserting charged groups into the membrane; [8–10] aromatic residues Trp and Tyr mostly interact with the head-group region, but Phe is present in both membrane core as well as the head-group region. [11] Additionally, the non-polar bilayer increases the strength of traditional hydrogen bonds. [5, 12] Despite the importance of membrane electrostatic interactions, they are challenging to capture computationally, primarily due to the high computational cost incurred in sampling solvent and solute degrees of freedom.

One way for electrostatic calculations to overcome solute and solvent degrees of freedom is to treat the solvent as a continuous medium and ignore solvent-solvent interactions. A common approach in biological systems is solving the Poisson-Boltzmann (PB) equation with finite-difference methods representing lipids both explicitly and as continuum slabs; however, computational costs have been a bottleneck. [13–16] To overcome the cost, the generalized Born (GB) approximation represents the membrane environment through a switching function or heterogeneous dielectric slab. [17–19] The calculated GB energy includes a screening radius which in turn depends on its local environment and surrounding neighbors, however, it is still expensive and complicated. A third possibility is the heuristic Coulomb equation. In Rosetta soluble protein calculations, this model has yielded a significant speedup relative to GB calculations and complements explicit modeling of hydrogen bonds. [20, 21]

One approach to overcome the computational cost of lipid membrane complexity is to represent the membrane as a continuum. [22–26] To compare the computing time performances for an implicit and explicit membrane model, Ulmschneider and Ulmschneider [27] undertook a comprehensive analysis of sampling efficiency of WALP16 peptide folding using a GB based implicit membrane model and, an explicit DPPC lipid bilayer and an octane slab membrane mimic. A quantitative assessment of sampling efficiency, utilizing a wide range of performance metrics, revealed that the implicit membrane models were at least two orders of magnitude more efficient than the octane membrane mimic, the simplest of the two explicit

models. Lazaridis implicit membrane model (IMM1), [28, 29] is a Gaussian solvent-exclusion model that uses experimentally measured transfer energies for side-chain analogs in organic solvents to emulate the transfer energy into the cell membrane. IMM1 uses a water-to-membrane transition function-dependent dielectric constant to strengthen the electrostatic interactions in the membrane. IMM1 has been applied to various biomolecular modeling problems, including studies of peptides, [30] de-novo folding, [31] and de-novo design of transmembrane helical bundles. [32] However, organic solvent slabs differ from phospholipid bilayers because lipids are thermodynamically constrained to a bilayer configuration, resulting in a unique polarity gradient that influences side-chain preferences. [33, 34] Another alternative is to directly calculate amino-acid preferences by deriving statistical potentials from a database of known MP structures. Yet, statistical potentials do not capture varying physicochemical features of different membranes. [35–38]

The F19 energy function, which was inspired by IMM1, included the transfer energy of residues in the phospholipid bilayer based on the Moon and Fleming (MF) scale. [39, 40] However, F19 omitted the electrostatic interactions of the membrane environment, leading to errors in benchmarking tests. For example, in a test to calculate the tilt angle of adsorbed peptide cecropin A(1–8)-magainin 2(1–12) (PDB:1f0d), F19 oriented it at 56° with the membrane normal with its N-terminus Lys residues facing the aqueous phase. However, solution NMR found the tilt angle of the peptide at 90° , which allowed favorable attractive electrostatic interactions between the cationic Lys residues and the anionic phosphate lipid head groups of DLPC. A second issue in F19 energy function was an underestimated water-to-bilayer transfer energy $\Delta G_{w,l}$ for Asp and Glu at neutral pH. The transfer energies for Asp and Glu derived from the MF hydrophobicity scale were measured at a pH of 3.8. [40] Since this pH is close to the nominal pKa values of Asp and Glu side chains, the experimental transfer energy was influenced by the presence of protonated Asp and Glu (D^0 , E^0) residues, which are less polar than the deprotonated residues (D^{-1} , E^{-1}) at neutral pH. In this paper, our objective is to add the missing electrostatic features of the membrane to F19 energy function and test whether the new energy function, named Franklin2023 (F23), better captures features of MPs.

We explore three modifications to F19. First, we develop a fast-to-compute Coulomb-based electrostatics method that includes a low dielectric constant of the lipid bilayer. Second, we use a mean field-based calculation for the electrostatic potential due to the lipid membrane and water by solving Poisson's equation using a fast Fourier transform (FFT) based solver while avoiding any assumption or discrepancies due to ionic strength incurred in PB or its linearized version. [41, 42] Third, we used a modified value of $\Delta G_{w,l}$ for Asp and Glu residues at neutral pH.

Until very recently, the energy functions have been developed and tested for specialized tasks, questioning their generalizability. To confront the challenge of over-fitting and a specialized model, our group recently developed a 12-test benchmark suite that probes (1) protein orientation in the bilayer, (2) stability, (3) sequence, and (4) docking structures of membrane protein. [43] These tests form a platform to evaluate the strengths of an energy function and suggest areas of improvement. We evaluate the performance of F23 on predicting the (1) orientation of peptides in the membrane environment, (2) thermal stability due to point mutation, (3) transfer energy of peptides and (4) design evaluations. Our results have shown significant improvement. Relative to F19, F23 improves the calculation of the tilt angle of membrane proteins for 90% of WALP peptides of different lengths, 15% of transmembrane (TM)-peptides, and 25% of the adsorbed peptides. The performances for stability and design tests are equivalent to F19 and F23. Finally, we conclude the paper with perspectives on further improvements for implicit membrane energy functions.

Methods

Implicit model development

Electrostatic energy term based on lipid type. We calculate the electrostatic effect of lipid bilayers based on all-atom molecular dynamics (MD) simulation of lipid molecules and solvents alone. We derive the electrostatic potential $\Psi(x, y, z)$ by solving Poisson’s equation using a 3D-grid based FFT solver:

$$\nabla \cdot \nabla \epsilon \Psi = \sum_{l=1}^{l=\text{lipid}} \rho_l + \sum_{w=1}^{w=\text{water}} \rho_w + \sum_{s=1}^{s=\text{salt}} \rho_s, \tag{1}$$

where ρ_w, ρ_l and ρ_s is the position-dependent charge densities of water, lipid, and salt averaged over MD trajectories, [44] and $\epsilon = 8.854 \times 10^{-12} \text{ CV}^{-1}\text{m}^{-1}$ is the dielectric constant of vacuum. To extract the depth-dependent Ψ , we use non-linear regression to fit the piece-wise function $\tilde{\Psi}_w$ as shown in Fig 1a. We derived the fitting parameters for all simulated lipid bilayers (Table A in S1 Text) as:

$$\tilde{\Psi}_w = \begin{cases} A_1 + \frac{A_1 - A_2}{\{1 + \exp\left(\frac{|z| - A_3}{A_4}\right)\}} & \text{for } |z| \geq z_c \\ C_1|z|^4 + C_2|z|^3 + C_3|z|^2 + C_4|z| + C_5 & \text{for } |z| < z_c \end{cases}$$

where z_c is the maximum of the critical points for the function $\tilde{\Psi}_w$ for $|z| < z_c$. The fitting parameters of different lipid types are listed in Table A in S1 Text. The reference state of $\Psi = 0$ in water is set at depth $z \approx 40 \text{ \AA}$ and the potential difference $\Delta\Psi$ is obtained (as shown by the dashed lines in Fig 1a). We compare the electrostatic potential due to salt and no salt in Fig 1b. As expected, the potential difference is found to be lower due to the presence of salt. The electrostatic energy due to the protein is then:

$$\Delta G_{\text{lipid}} = \sum_{r=1}^{N_{\text{res}}} \sum_{a=1}^{N_{\text{atom}}} \Delta\Psi(z) q_{r,a}, \tag{2}$$

where the sums are over all residues and all atoms in each residue, $q_{r,a}$ is the charge of a protein atom, and $\Psi(z)$ is the electric potential of the empty water-bilayer system as a function of membrane depth. $\Psi(z)$ is an average of $\Psi(x, y, z)$ in the other directions.

Following our previous work, [44] the MD trajectories used in this analysis were generated from all-atom molecular dynamics simulations of phospholipids, water, and salt. The simulations were performed using the NAMD molecular dynamics engine at a constant pressure of 1 atm and a temperature of 37°C. The CHARMM36 force field was used for lipids, and TIP3 model for water. See the methods section of Alford et al., [44] for further details.

Bilayer depth-dependent dielectric constant. Due to the hydrophobic environment within the membrane, the dielectric constant is lower than that in the isotropic soluble region. To account for the dielectric variation, in IMM1 a water-to-membrane transition-function-dependent dielectric constant is substituted in EEF. [28]

Inspired by IMM1 in F23, for structure prediction and design, we use a membrane hydration f_{hyd} dependent dielectric constant. Membrane hydration (f_{hyd}) is a proxy of the membrane depth and is capable of capturing water cavity effects. [44] In the center of the membrane, when an atom is exposed to lipids, $f_{\text{hyd}} = 0$; whereas when an atom is fully exposed to water, $f_{\text{hyd}} = 1.0$. When an atom is in the membrane yet faces a water cavity, $0.0 < f_{\text{hyd}} < 1.0$. For an

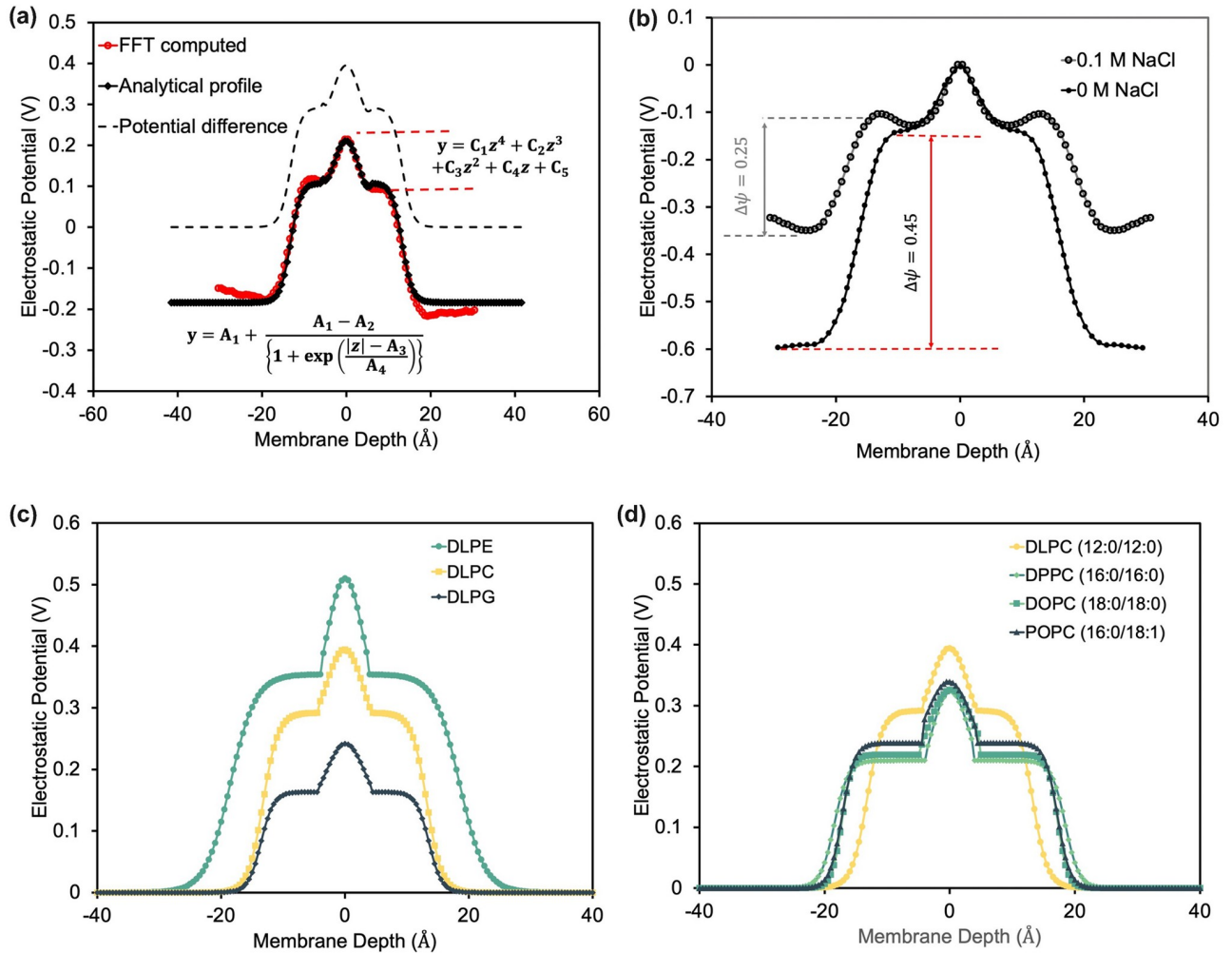


Fig 1. Features of electrostatic potential due to the lipid bilayers in the membrane. (a) Electrostatic potential as a function of depth and analytical fit. (b) Comparison of electrostatic potential in a 0.1 M salt solution and that calculated by Yu *et al.* in pure water. [45] Comparison of electrostatic potential as a function of membrane depth for (c) different lipid head groups and (d) lipid tail lengths.

<https://doi.org/10.1371/journal.pcbi.1011296.g001>

atom pair, we define the hydration fraction $f_{\text{hyd},ij}$ as the geometric mean of $f_{\text{hyd},i}$ and $f_{\text{hyd},j}$. Rosetta uses a distance-dependent dielectric constant for soluble proteins with sigmoidal function $\epsilon_{\text{sol}}(r_{ij})$ to transition between the protein core ($\epsilon_{\text{core}} = 6$) and surface ($\epsilon_{\text{surface}} = 80$). [46] To define the distance-dependent dielectric constant within the membrane, we use the same sigmoidal function $\epsilon_{\text{memb}}(r_{ij})$ to transition between protein core ($\epsilon_{\text{core}} = 3$) and surface ($\epsilon_{\text{surface}} = 10$) as summarized in Fig 2a. [47, 48] To combine the effect of distance dependence and membrane hydration, we use a linear mixture equation:

$$\epsilon(r_{ij}, f_{\text{hyd},ij}) = f_{\text{hyd},ij} \epsilon_{\text{sol}}(r_{ij}) + (1 - f_{\text{hyd},ij}) \epsilon_{\text{memb}}(r_{ij}) \quad (3)$$

Fig 2b shows $\epsilon(r_{ij}, f_{\text{hyd},ij})$ as a function of $f_{\text{hyd},ij}$ and atom pair distance r_{ij} . Based on the modified dielectric constant, the Coulomb energy is:

$$E_{\text{elec},ij}^{\text{dielectric}} = \frac{C_0 q_i q_j}{\epsilon(r_{ij}, f_{\text{hyd},ij})} \left[\frac{1}{r_{ij}} - \frac{1}{r_{\text{max}}} \right], \quad (4)$$

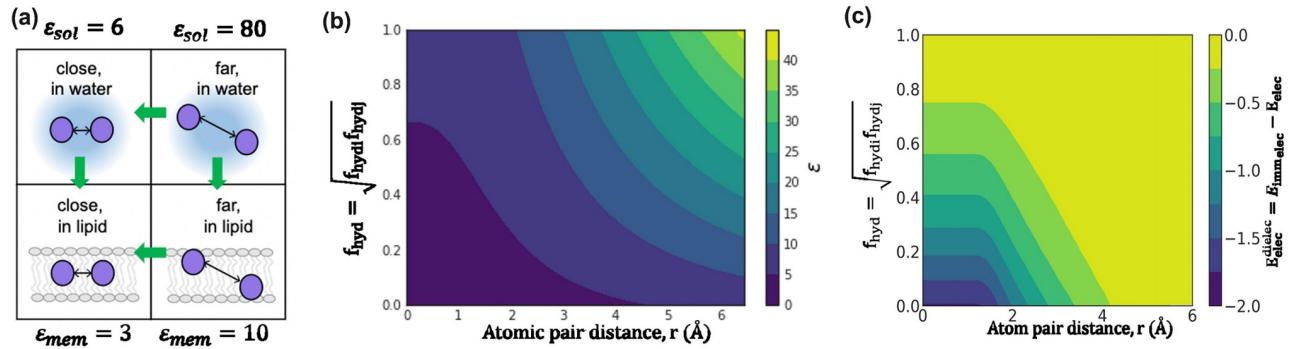


Fig 2. Coulomb electrostatic energy due to the lipid bilayers in the membrane. (a) Membrane-dependent electrostatics potential: The energy increases when entering the bilayer and at close atom pair distances. Green arrows indicate the direction of increasing energy. (b) Dependence of the dielectric constant ϵ on fraction of hydration f_{hyd} and atom-pair distance $r_{i,j}$. The different dielectric constants are colored from low (dark blue) to high (yellow). (c) Variation of the membrane dielectric-dependent electrostatics energy as a function of fractional hydration f_{hyd} and atom-pair distance $r_{i,j}$. Each grid point corresponds to an energy calculated for point charges with opposing signs, and the energy varies from strong (dark blue) to weak (yellow).

<https://doi.org/10.1371/journal.pcbi.1011296.g002>

where q_i, q_j are partial charges of protein atoms, C_0 is Coulomb’s constant ($322 \text{ \AA kcal/mol } e^{-2}$), where e is the elementary charge and the electrostatic energy is truncated at $r_{max} = 5.5 \text{ \AA}$. [39] To isolate the excess electrostatic energy due to the membrane, in this energy term, we subtract the electrostatic energy from the solution. In addition, nullifying the electrostatic energy ensures the reference energy of the unfolded protein in the soluble region remains unaffected. [49] The modified energy is given as:

$$\Delta E_{elec,ij}^{dielectric} = C_0 q_i q_j \left[\frac{1}{r_{ij}} - \frac{1}{r_{max}} \right] \left[\frac{1}{\epsilon(r_{ij}, f_{hyd,ij})} - \frac{1}{\epsilon(r_{ij})} \right] \quad (5)$$

Fig 2c presents the electrostatic energy due to the membrane as a function of pair fraction of hydration $f_{hyd,ij}$ and atom pair distance.

Modifying $\Delta G_{w,l}$ for Asp and Glu at neutral pH. F19 included the hydrophobicity of residues as per the Moon and Fleming (MF) hydrophobicity scale, [40] which reflects the water-to-bilayer partitioning of side chains in the context of a native transmembrane protein spanning a phospholipid bilayer. The MF hydrophobicity scale is linearly correlated to Wimley White (WW) scale with a higher slope, which is based on side-chain water-to-octanol transfer energies (as seen in Fig 3a). [40, 50] The two exceptions are Asp and Glu since the MF partition energy is measured at a pH close to their pK_a . The question arises, how do we calculate the values of Asp and Glu at neutral pH (D^{-1}, E^{-1})? To determine the MF-scale transfer energy of D^{-1}, E^{-1} , we use the linear best fit of MF and WW hydrophobicity scales and extrapolate the WW transfer energy of D^{-1} and E^{-1} .

Based on the MF scale transfer energy of D^{-1}, E^{-1} , we used linear regression to calculate the transfer energy of atoms $\Delta G_{w,l}^{atom}$ by solving $[A_{aa,atom}] \cdot \Delta G_{w,l}^{atom} = \Delta G_{w,l}^{aa}$, where $[A_{aa,atom}]$ is a matrix of atom type stoichiometry coefficients and the $\Delta G_{w,l}^{atom}$ and $\Delta G_{w,l}^{aa}$ represent the vectors of $\Delta G_{w,l}$ for different atom and amino acid types. We append $[A_{aa,atom}]$ and $\Delta G_{w,l}^{aa}$ to include the stoichiometry of D^{-1} and E^{-1} to recalculate $\Delta G_{w,l}^{atom}$. Table B in S1 Text shows all the equations and values of $\Delta G_{w,l}^{aa}$ we have used. Fig 3b compares $\Delta G_{w,l}^{atom}$ calculated with and without $\Delta G_{w,l}^{aa}$ for D^{-1} and E^{-1} used in F23 and F19 respectively. The additional rows for D^{-1} and E^{-1} led to renewed values of $\Delta G_{w,l}^{COO}$ and $\Delta G_{w,l}^{OOC}$ as seen in Fig 3a and 3b.

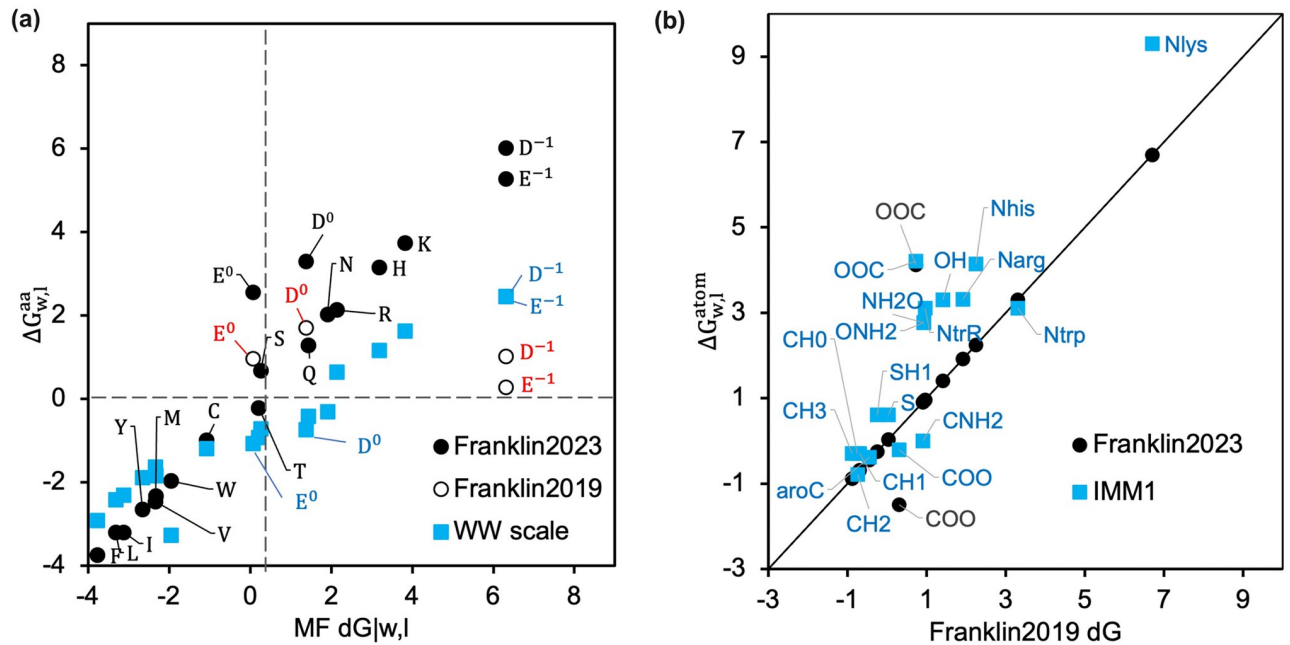


Fig 3. Values of $\Delta G_{w,l}^{atom}$ and $\Delta G_{w,l}^{aa}$ with different scales. (a) Comparison of $\Delta G_{w,l}^{aa}$ calculated using F19 and F23 relative to the experimental values from MF and WW scales. The MF $\Delta G_{w,l}^{aa}$ for D^{-1} and E^{-1} are estimated by extrapolation of a linear fit of the other WW data points. (b) Comparison of $\Delta G_{w,l}^{atom}$ calculated using F23 and IMM1 relative to F19.

<https://doi.org/10.1371/journal.pcbi.1011296.g003>

Integration of new terms into F23

We integrate two new energy terms (ΔG_{lipid} and $\Delta E_{elec,ij}^{dielectric}$) to F19 through linear combination of terms.

$$F23 = Ref2015 + w_{w,l} \Delta G_{w,l} + w_{lipid} \Delta G_{lipid} + w_{dielec} \Delta E_{elec,ij}^{dielectric}, \quad (6)$$

where Ref2015 is Rosetta's standard energy function for soluble proteins, [46] $\Delta G_{w,l}$ add the transfer energy from water to bilayer, [44] and ΔG_{lipid} and $\Delta E_{elec,ij}^{dielectric}$ are the energies due to the lipid head group and the membrane dielectric constant. We can recover the energy function F19 by setting $w_{w,l} = 0.5$, $w_{lipid} = 0$ and $w_{dielec} = 0$. We solved the simultaneous equations for the three unknown positive weights to maximize the correlation coefficient for the experimentally measured $\Delta \Delta G_{mut}$ for amino acids at the center of OmpLA protein [40] and mutations to aromatic side chains (Phe, Trp, Tyr) [51]. The resulting weights were $w_{w,l} = 1$, $w_{lipid} = 0.128$ and $w_{dielec} = 0.01$, achieved using L-BFGS-B algorithm within bounds of [0, 1]. Our choice of these bounds was to avoid negative weights and extremely large positive weights while recognizing potential biases due to the bounds and more data points at the center of the membrane. We omitted the matrix rows for Ala due to its role as a host residue for all experimental mutations, Pro to avoid steric clashes resulting in large values, Asp and Glu residues because the experimental values were measured at low pH.

Results and discussion

Biologically realistic membrane features

F23 captures the effects of the diverse lipid bilayer. We first investigate the effect of the lipid head and tail group (Fig 1c and 1d) by examining the electrostatic potential of the empty

lipid layer in the presence of counter-ions and salt solution. While this implicit method cannot capture the interface deformation or interaction of lipids with the residues, this electrostatic potential will be sensitive to different lipid types. Fig 1c shows the $\Delta\Psi$ for three different lipid head groups (DLPE, DLPC, and DLPG) with the same tail group, *i.e.* Lauric acid(12: 0) acyl chains (DL). Phosphatidylethanolamine (PE) and phosphatidylcholine (PC) have neutral head groups with the same overall composition, except that PE has a N^+H_3 and PC has a $N^+(CH_3)_3$ atom. Relative to PC, the smaller head group of PE increases the local charge density leading to higher electrostatic potential. [52, 53] Phosphatidylglycerol (PG), on the other hand, has anionic lipid head groups known to adsorb counterions to the bilayer interface strongly, screening the net electrostatic potential relative to PC. [54, 55] Fig 1d shows the $\Delta\Psi$ for four different lipid tail groups (DLPC, DPPC, DOPC, and POPC) with the constant head group PC and different length and degree of unsaturation. These calculated results match chain length trends observed in experimentally measured surface potential. [56] Using these different profiles, F23 can approximate different homogeneous lipid environments through the potential $\Delta\Psi$, and parameters for other mixed lipid compositions can be added with a short MD run.

Benchmark performance of F23

We evaluate the performance of F23 using five benchmark tests. The tests are designed to evaluate an energy function's ability to replicate measured membrane protein stability and perform structure prediction and design. We compare the performance of F23 to two other existing implicit models: (1) F19 and (2) M12, which is based on IMM1. [39, 57] We chose these models for their low computational cost which allows evaluation with structure prediction and design tests.

Test 1: Orientation of polyalanine and WALP peptides. MPs are thermodynamically stable in the bilayer because of favorable orientation and insertion energy. Thus, a key challenge for energy function and the focus for our first and second tests is to recapitulate this lowest energy orientation of membrane peptides. Following our prior work [43, 44, 58], we used a protocol that samples all possible orientations of the peptide relative to the implicit membrane within $\pm 60^\circ$ of the bilayer center (d), tilt angles relative to the membrane normal (θ) between $\pm 180^\circ$, rotation angles relative to the principal helical axis (ϕ) between $0 - 360^\circ$. The global energy minimum of all sampled positions is defined as the most stable predicted orientation. In this first test, we test the effect of the length of repeating helical peptides on the tilt angle.

Polyalanine. To test the effect of peptide length, we predict the tilt angle of polyalanine peptides with a varying number of alanine residues (ranging from 20–40). Polyalanine peptides have been shown to adopt predominantly helical conformations in lipid bilayers. We model the residues as α -helices with $\psi = -47^\circ$ and $\phi = -57^\circ$ with charged ends. We compare our results with orientations calculated using Sengupta *et al.*'s 5-slab continuum dielectric model with CHARMM36 force field [59] as shown in Fig 4a. Due to the lack of any lipid type mentioned in the 5-slab model, in this test, we calculated the tilt angles using DLPC for both F19 and F23. Similar to that in Sengupta *et al.*, [59] the tilt angles calculated by F23 and F19 are proportional to the length of the peptide. However, relative to the 5-layer slab model, F19 and F23 over-predict the tilt angle by about 10–15°. Irrespective of the size of the polyalanine, M12 places the peptides outside the membrane at a tilt near 90°.

WALP peptides. WALP peptides typically comprise two Trp residues at each end with a helical core composed of alternating Leu and Ala residues: GWW(LA)_nLWWA. They have proven to be opportune models for experimentally investigating lipid and peptide interactions. [60] The WALP peptides are modeled as ideal α -helices with $\psi = -47^\circ$ and $\phi = -57^\circ$. In this test, we calculate the tilt angle of WALP peptides of different lengths in DMPC lipid layers

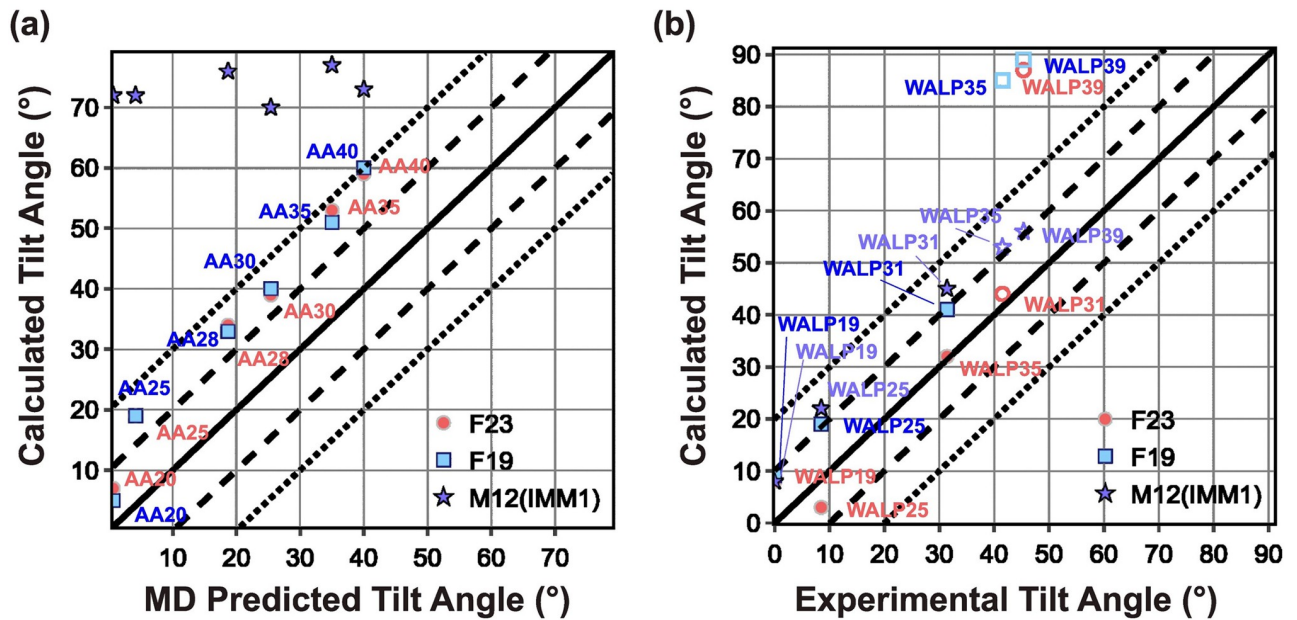


Fig 4. Test1: Orientation of polyaniline and WALP peptides. (a) Predicted tilt angle of polyaniline as a function of the number of residues. The polyaniline is named AAx, where x represents the number of residues. (b) Predicted tilt angle of WALP peptides as a function of the number of residues. The WALP peptides are presented as WALPx, where x represents the number of residues. Peptides for which experimental results are available are presented by filled markers, and those for which simulated data is available are shown by unfilled markers.

<https://doi.org/10.1371/journal.pcbi.1011296.g004>

(using M12, F19, and F23). We compare the results with experimental solid-state NMR (ssNMR) and spectroscopic measurements available for WALP19–31 (shown in Fig 4b by filled markers) and the rest of the results predicted by Sengupta *et al.*'s 5-slab model (shown in Fig 4b by unfilled markers). [59, 61, 62] Sengupta *et al.* calculated the tilt angle of fixed backbone peptide by determining the energy landscape as a function of depth and tilt angle that included the peptide energy using CHARMM potential and membrane solvation energy using electrostatics and non-polar terms.

Similar to ssNMR, F19 calculated the tilt angle of WALP to be directly proportional to the number of residues (Fig 4b). For shorter peptides, the tilt angle is in good agreement with the experiment; however, for the peptides longer than WALP31, F19 buries the peptides at the center of the membrane (tilt angle 90°). For longer peptides, F19 is unable to compensate the net hydrophobicity due to the non-polar residues by the energy due to polar residues, resulting in burying the peptide at the center of the membrane (see examples of WALP35 and WALP39 in Fig A(a-b) and A(d-e) in S1 Text). To demonstrate the impact of lipid composition, Fig E in S1 Text displays calculated WALP tilt angles using F23 in three different lipid compositions.

Relative to F19, F23 computed the tilt angle for WALP peptides closer to the experimental measures and those predicted by the 5-slab model. Particularly, F23 increased the accuracy of tilt angle prediction due to two factors: (a) lower hydrophobicity of the polar residues at the terminal (Fig A(c) in S1 Text), and (b) the additional membrane electrostatic energy. However, F23 still buried WALP39 with a tilt angle much higher than the angle predicted by the 5-slab model (Fig A(d) in S1 Text). Contrary to F23 and F19, M12 calculated the tilt angle of all peptides close to those by the 5-slab model (Fig 4b) because both the models use the WW hydrophobicity scale, which has a lower slope than the MF hydrophobicity scale used by F19 and F23 (Fig 3a). [40, 43] However, hydrophobicity alone cannot be the reason, as otherwise, the tilt angles for polyanilines would have been similar. M12 uses a knowledge term based on the

probability of finding a residue at a given depth of the membrane, which penalizes the Trp residues in the membrane but not the Ala residues. This heuristic term correlates well with the location of energy minima of the total energy Fig B(a-b) in [S1 Text](#).

In summary, with the combination of hydrophobic transfer energy and electrostatic interaction energy with the membrane relative to F19, F23 improves the accuracy of tilt angle calculation. However, F23 buries long WALP peptides due to higher hydrophobic energy relative to M12 (IMM1).

Test 2: Orientation of TM- and adsorbed peptides. Our second test also concerns the tilt angle for peptides. However, unlike Test 1, which explored related families in a membrane, these peptides are of diverse sequences and in different lipid environments, with some that adsorb on the surface of the membrane and some that pass through it. The different lipid types used for experiments and the parameters used for this test are listed in Table C in [S1 Text](#). The objective here is to see if the lowest energy orientation of the membrane peptides identified by the energy function corresponds to the native orientation. This test is the cornerstone of our benchmark because it was used for the validation of early implicit membrane models. [43, 63] The tilt angles of TM-peptides were measured previously using solid-state NMR experiments using different lipid compositions, including DPC micelles, DMPC vesicles, mixed DOPC: DOPG bilayers, and pure DOPC bilayers. [43, 63] While the experimental uncertainty is not available for all measurements, tilt angle measurements have a typical error range of $\pm 3\text{--}5^\circ$. The tilt angle for adsorbed peptides is measured by solution NMR in dodecyl phosphocholine micelles or trifluoroethanol with an uncertainty of $\pm 6\text{--}12^\circ$. [64] Since our previous publications have already reported a comparison of F19 with M12 and other energy functions, we focus here on the comparison of F23 with F19 ([Fig 5a and 5b](#)). [43]

F19 calculated the tilt angle within $\pm 20^\circ$ of the experimentally measured values for three out of four peptides. F23 improved the tilt angle calculation for WALP23 by 20° and NR1 subunit of the NMDA receptor (2nr1) peptides by 5° to result in tilt angles within $\pm 20^\circ$ of the

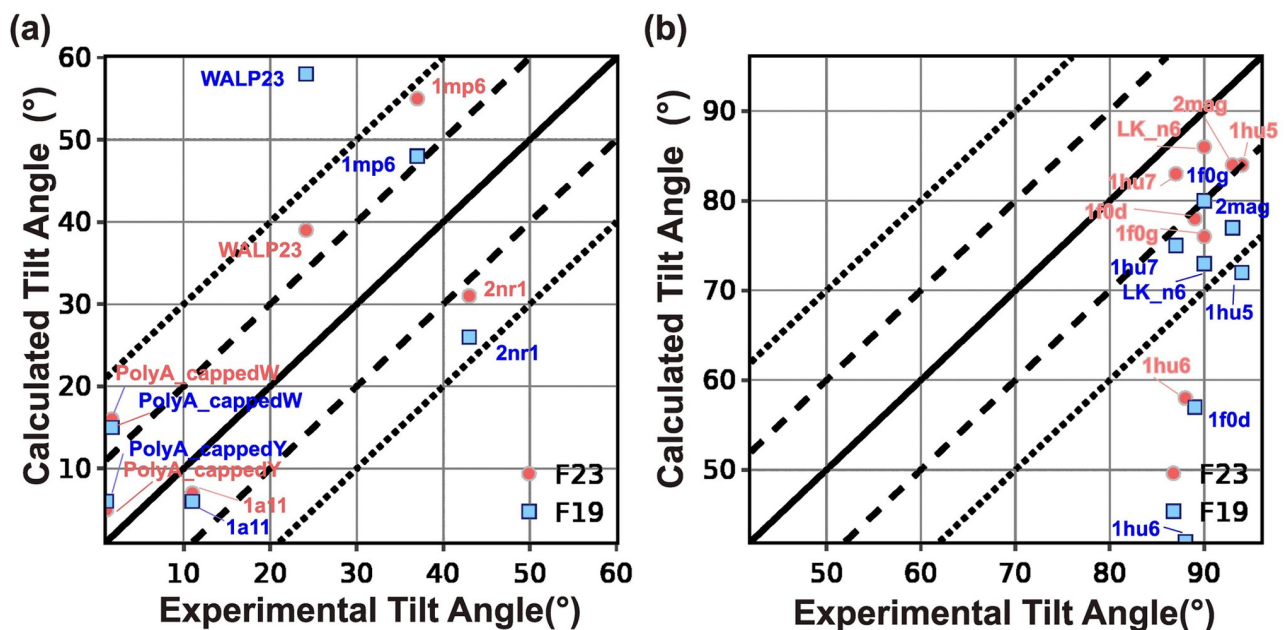


Fig 5. Test2: Orientation of TM and adsorbed peptides. (a) Predicted tilt angle of TM peptides compared with experimental measurements. (b) Predicted tilt angle of adsorbed membrane peptides and compared with experimental measurements.

<https://doi.org/10.1371/journal.pcbi.1011296.g005>

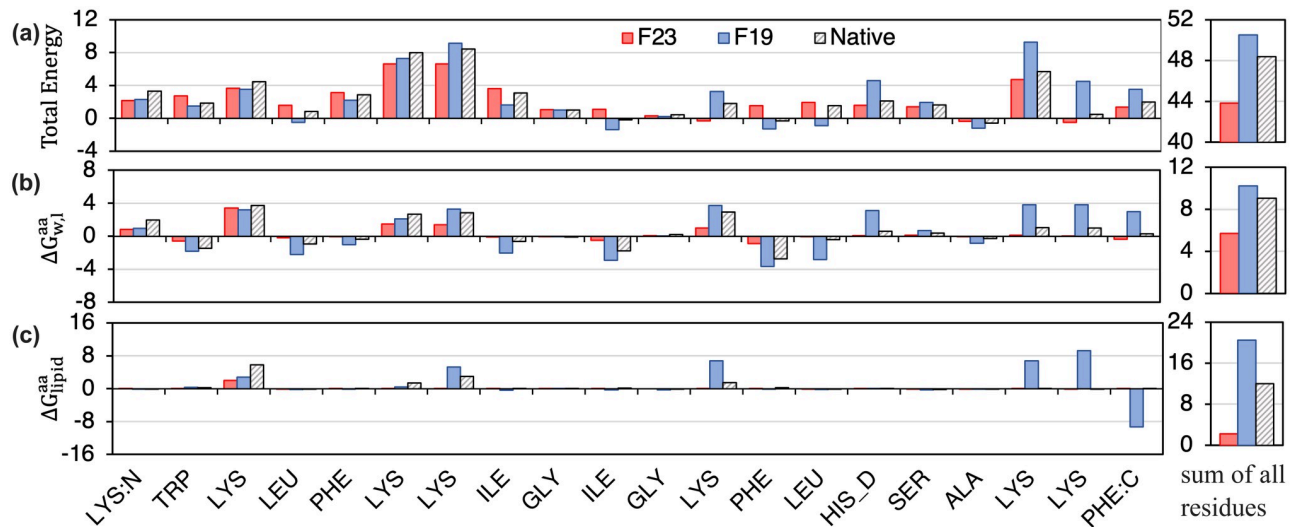


Fig 6. Per-residue F23 energy calculated at native, F23 and F19 calculated tilt angles. (a) F23 per-residue total energy (REU), (b) per-residue transfer energy to the bilayer, and (c) per-residue electrostatic energy due to the lipid layer at native (black hashed), F23 (red) and F19 (blue) calculated tilt angles. **Right:** Sum of all residues.

<https://doi.org/10.1371/journal.pcbi.1011296.g006>

experimentally measured values for all the four peptides (as seen in Fig 5a). F23 worsened the tilt angle for influenza A M2 (1mp6) peptides.

To understand the deviation between the experiment and calculated tilt angles of WALP23 and 1mp6, we examine the energy landscape under F23 and F19. F23 improved the tilt angle of WALP23 due to the lower hydrophobicity of the C-terminus. To illustrate the reasons for the difference in tilt angle for 1mp6, we present the energy landscape under F23 and F19 in Fig C(a-g) in S1 Text. F19 calculated the tilt angle at $z = -5 \text{ \AA}$ as 48° . Due to the lower hydrophobicity of C-terminus Asp and Leu residues, F23 further improved the tilt angle of 1mp6 at $z = -5 \text{ \AA}$ (Fig C(f) in S1 Text). However, electrostatic interactions between arginine and the lipid layer reduced the energy of F23 further by 0.1 REU and led to a worsening of the calculated tilt angle (Fig C(f-g) in S1 Text).

For four out of seven adsorbed membrane proteins, F19 calculated the tilt angle within $\pm 20^\circ$ of the experimentally measured values. F23 significantly improved calculation resulting in tilt angles within $\pm 20^\circ$ of the experimentally measured values for six out of seven peptides (Fig 5b). F23 reduced the error in tilt angle of magainin-cecropin hybrid (pdb:1f0d) and novispirin G10 (pdb:1hu6) by 21° and 16° respectively.

We compare the F23 per-residue energy of the magainin-cecropin hybrid (pdb:1f0d) at configurations predicted by F19, F23, and the native state (Fig 6b). F23 penalizes Lys residues at the tilt angle calculated by F19 and allows favorable attractive electrostatic interactions between the cationic side chains and the anionic phosphate lipid headgroups of DLPC. In addition, F23 assigned higher energy to the polar Lys residues present near the center of the membrane in the F19 configuration relative to the surface absorbed native state. To further establish the importance of the electrostatic interaction, we recalculate the residue with a special version of F23 (shown in Fig D(d-f) in S1 Text as dashed lines) in which the $\Delta G_{w,l}^{atom}$ parameters are reverted back to those of F19. Indeed, higher electrostatic repulsion at the center results in the magainin-cecropin hybrid peptide (pdb:1f0d) being placed near the surface of the membrane and obtaining a near-native tilt angle, confirming that the ΔG_{lipid} and $\Delta E_{elec,ij}^{dielectric}$ terms are responsible for the improvement.

To summarize, the tilt angle calculation by F23 for several TM-peptides and adsorbed peptides has improved due to the addition of the new electrostatic terms and the renewed values of $\Delta G_{w,l}^{\text{COO}}$ and $\Delta G_{w,l}^{\text{OOC}}$.

Test 3: $\Delta\Delta G_{\text{mutation}}$. Our third test evaluates how a change in the sequence of a membrane protein affects its overall thermostability. $\Delta\Delta G_{\text{mutation}}$ is the change in Gibbs free energy of folding from water to the lipid bilayer of each mutant relative to the native sequence.

$$\Delta\Delta G_{\text{mutation}} = [G_{\text{bilayer}}^{\text{mutant}} - G_{\text{water}}^{\text{mutant}}] - [G_{\text{bilayer}}^{\text{native}} - G_{\text{water}}^{\text{native}}], \quad (7)$$

To design MPs, $\Delta\Delta G_{\text{mutation}}$ is crucial in evaluating the effect of genetic mutations on protein functions. To evaluate $\Delta\Delta G_{\text{mutation}}$, we use a Rosetta fixed-backbone and fixed-orientation protocol. [44] We use two sets of $\Delta\Delta G_{\text{mutation}}$ measurements from the Fleming lab taken at equilibrium in DLPC vesicles and in the context of a β -barrel protein scaffold. The two datasets are mutations from Ala to all 19 remaining canonical amino acids at position 210 on outer membrane phospholipase A (OmpLA), [40] and at position 111 on an outer membrane palmitoyl transferase (PagP). [51] The experimental uncertainty of the measured $\Delta\Delta G_{\text{mutation}}$ values is $\pm 0.6 \text{ kcal mol}^{-1}$.

Fig 7a and 7b compares the experimental and calculated $\Delta\Delta G_{\text{mutation}}$ from Ala to all other residues for OmpLA and PagP proteins at a fixed height in the membrane. We also included D^{-1} and E^{-1} to emulate these residues at neutral pH. While F19 linearly correlates well for both OmpLA and PagP, F23 predicts the $\Delta\Delta G_{\text{mutation}}$ closer to the experimental values (e.g. in OmpLA H improves from 7.2 to 5.2 REU, K from 6.4 to 5.2 REU, N from 6.2 to 4.7 REU; in PagP R improves from 5.9 to 3.8 REU and K from 8.5 to 7 REU), evident from the slope of the

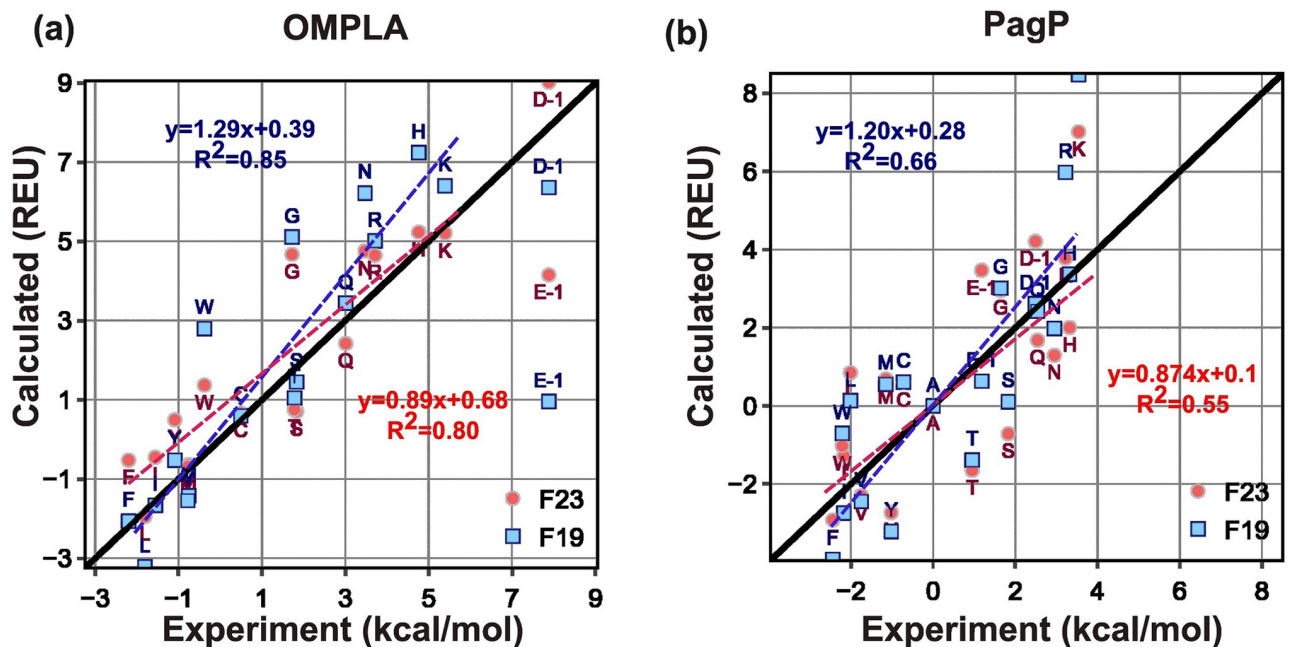


Fig 7. Test3: $\Delta\Delta G_{\text{mutation}}$ of alanine to all other residues. Predicted $\Delta\Delta G_{\text{mutation}}$ and experimental measurements for (a) OmpLA and (b) PagP protein. For the experimental values of $\Delta\Delta G_{\text{mutation}}$ for D^{-1} and E^{-1} for OmpLA, we have used their extrapolated MF scale values as described in Methods section. Since the experimental $\Delta G_{w,l}^{\text{aa}}$ based on the MF hydrophobicity, was measured at a pH of 3.8, where some of the Asp and Glu are in their protonated states, we removed them while calculating the slope and correlation coefficient of the best-fit lines (dashed lines) for the predictions relative to experimental values. The black line is the $y = x$ line. The red and blue dashed lines are linear best-fit lines for values predicted by F23 and F19, respectively. The equations for the best-fit lines for F23 and F19 are shown in red and blue, respectively.

<https://doi.org/10.1371/journal.pcbi.1011296.g007>

best-fit line (the fit excludes Asp and Glu, see the caption for further details). To assess the uncertainty of the predicted $\Delta\Delta G_{\text{mutation}}$, we conducted 10 repeated calculations, resulting in standard deviations of 0.008 (F23, OmpLA), 0.016 (F23, PagP), 0.086 (F19, OmpLA) and 0.009 (F19, PagP).

Test 4: $\Delta\Delta G_{\text{ins}}$. Next, we asked whether the energy function can recapitulate the thermodynamics of protein insertion into the membrane. This test set includes five poly-Leu peptides designed by Ulmschneider *et al.*, [65] four of which follow a GL_XRL_XG motif, where $X = 5, 6, 7, 8$. The fifth follows a different pattern by adding a flanking Trp: GWL_6RL_7G . We compare our results against transfer energies from MD simulations in POPC bilayers, which were validated against intrinsic fluorescence measurements. The experimental uncertainty of the measured $\Delta\Delta G_{\text{ins}}$ values was $\pm 1.4 \text{ kcal mol}^{-1}$. [65]

To evaluate the performance of energy functions, we compute the $\Delta\Delta G_{\text{ins}}$ (Eq 8) as the energy difference between the lowest energy orientation of the peptide in the lipid bilayer phase and in the aqueous phase:

$$\Delta\Delta G_{\text{ins}} = [G_{\text{lipid}} - G_{\text{ref}}] - [G_{\text{water}} - G_{\text{ref}}], \quad (8)$$

where G_{ref} is the Gibbs free energy of an unfolded protein in solution.

Fig 8 compares $\Delta\Delta G_{\text{ins}}$ calculated by F19 and F23 to that by MD simulation. The correlation coefficient for F23 is 0.01 higher than F19, which is insignificant. However, the slopes for the best-fit lines are about double for F23 relative to F19. Relative to F19, the additional electrostatic terms in F23 increased favorable energy due to insertion. F19 and F23 recognize the relative $\Delta\Delta G_{\text{ins}}$ among the design peptides; however, the energy of insertion is overestimated

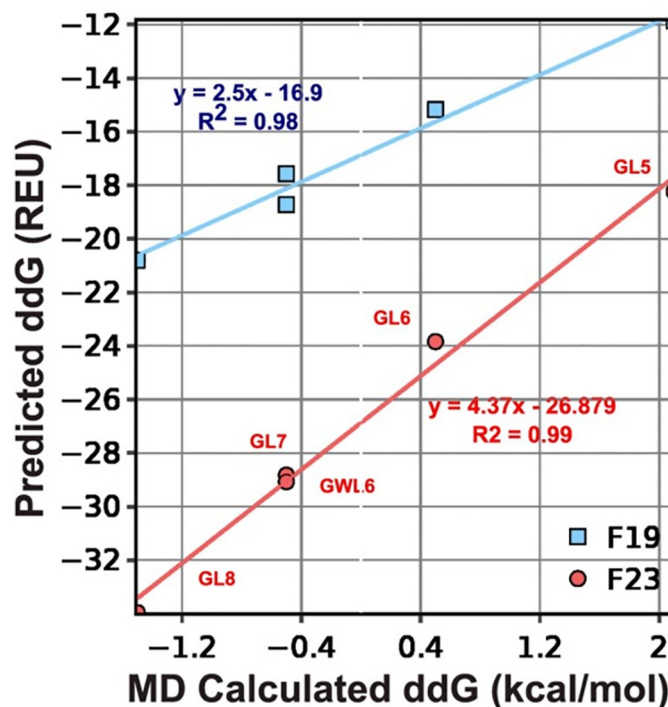


Fig 8. Test4: $\Delta\Delta G_{\text{ins}}$ of polyleucines. $\Delta\Delta G_{\text{ins}}$ is the energy difference between the lowest energy orientation of the peptide in the lipid bilayer phase and that in the aqueous phase. $\Delta\Delta G_{\text{ins}}$ is a measure of the affinity of a folded protein to be in the membrane phase relative to the aqueous phase. Predicted $\Delta\Delta G_{\text{ins}}$ are compared with experimental and values calculated by MD simulations.

<https://doi.org/10.1371/journal.pcbi.1011296.g008>

compared to the calculations from MD. As in many other Rosetta-based calculations, these energy functions reproduce trends but not the exact values.

Test 5: Sequence recovery. Finally, for the design applications, we evaluate the ability of an energy function to recover native MP sequences. We perform redesign using Rosetta's Monte Carlo fixed-backbone protocol that samples possible sequences using a full protein rotamer-and-sequence optimization and a multicool-simulated annealing. [66] Each protein is initialized and fixed in the orientation from the OPM database. [67] For each backbone conformation, we generate 50 designed sequences and chose the lowest scoring sequence for further analysis. We compute three metrics: (1) sequence recovery, which measures the fraction of native amino acids recovered in the redesigned sequence, (2) Kullback–Leibler (KL) divergence (D_{KL}), which measures the difference in the probability distribution of design and native residues, and (3) perplexity, which measures the confidence of a model in the designed sequences. To extract features relevant to the membrane, we compute these metrics for subsets of amino acids exposed to the aqueous phase (outside of membrane or pore facing), lipid phase, and interface region. The dataset for the test includes 204 MPs, including both α -helical bundles and β -barrels, with all entries having a resolution of 3.0 Å or better and less than 25% sequence identity. The native and host lipid compositions for these proteins vary widely and include compositions not yet covered by the F23 lipid parameters; thus, for simplicity, we perform all design calculations in the DLPC bilayer.

High sequence recovery has long been correlated with strong energy function performance for soluble proteins. [49] F19 designed sequences are 31.8% identical to their native sequences (Fig 9a). With the new electrostatics terms, F23 recovered 30.8% of native residues. To estimate the variation of designed sequences for a given backbone, Fig F(a-b) in S1 Text shows the mean and standard deviation of sequence recovery and total score of all 50 designs for top 20 performing PDB backbones. The soluble protein energy function (R15 [46]) and other previous implicit membrane energy function, which was parameterized from the behavior of side-chain analogs in organic solvents (M07 [31]), lagged in sequence recovery. To compare the influence of different solvent environments, we calculated sequence recovery over subsets of residues. First, we compare buried vs. solvent-exposed side chains (Fig 9b). Among buried side chains, F19 recovered 3.4% higher native residues than F23. On the surface, sequence recovery for M12 and F23 was 3% higher than F19. This improvement in the sequence recovery of surface-exposed residues in contrast to those buried by F23 may indicate a potential path for improvement that might include higher weights of the dielectric constant term. F23 recovered a higher fraction and more variety of native residues (above random selection) relative to F19 for both water-facing and lipid-facing residues (Fig 9c and 9d).

Amino acid distribution in design proteins is similar to that in native proteins. To evaluate the distribution of amino acid types in the designed proteins relative to that in native proteins, we calculated D_{KL} as:

$$D_{\text{KL}} = \sum_{s \in \text{AA}_{20}} p_s^{\text{des}} \log \left(\frac{p_s^{\text{des}}}{q_s^{\text{nat}}} \right), \quad (9)$$

where p_s^{des} , q_s^{nat} are the probabilities of amino acid type s in the design and native sequences respectively and AA_{20} is the set of all 20 canonical amino acids. p_s is calculated as:

$$p_s = \frac{\sum_{i=1}^{N_{\text{bb}}} \sum_{j=1}^{N_i^{\text{res}}} \mathbb{1}_{x_{i,j}}(s)}{\sum_{i=1}^{N_{\text{bb}}} N_i^{\text{res}}} \quad (10)$$

where N_{bb} is the total number of protein backbones in our design set, N_i^{res} is the length of the

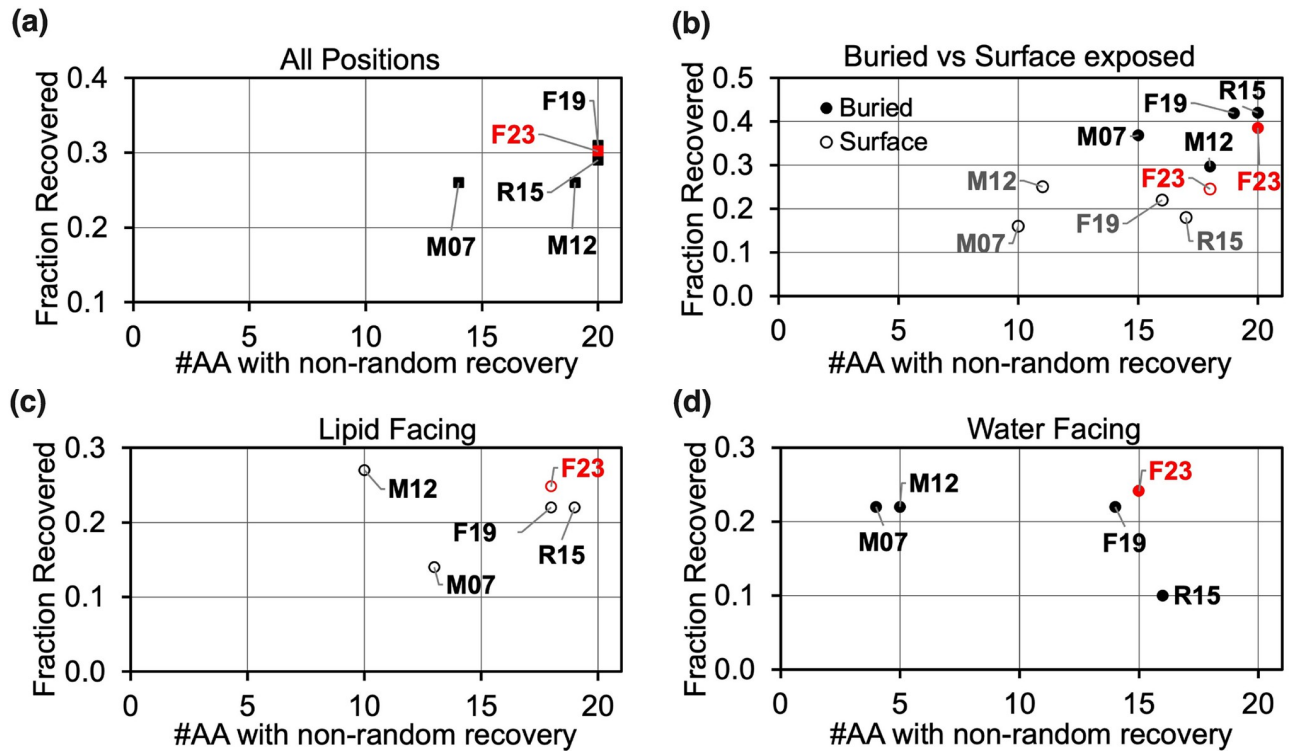


Fig 9. Test 5: Sequence recovery of membrane proteins. Plots show the fraction of native amino acids recovered on the y-axis and the fraction of amino acid types with individual recovery rates greater than 0.05 on the x-axis. An accurate energy function would have a high sequence recovery rate both overall and for the individual amino acid types. The results are shown for all positions in panel (a), buried (in filled circles) vs. surface-exposed (in open circles) positions in (b), lipid-exposed positions in (c), and water exposed in (d).

<https://doi.org/10.1371/journal.pcbi.1011296.g009>

i^{th} protein, $x_{i,j}$ is the amino acid type of the j^{th} residue in the i^{th} protein and, $\mathbb{1}_{x_{i,j}}(s) := 1$ if $x_{i,j} = s$ or 0 otherwise.

For residue level understanding, we calculate $D_s = \log\left(\frac{p_s^{\text{des}}}{q_s^{\text{nat}}}\right)$, which is the difference in log-likelihood of amino acid type s in designed and native sequences, and their sum $D = \sum_{s \in \text{AA}_{20}} D_s$. A negative D_s indicates that the amino acid s is under-enriched in designed sequences, and a positive D_s indicates that s is over-enriched. An ideal energy function will have $D_{\text{KL}} = D_s = D = 0$.

F19 outperforms Ref15 and other implicit energy functions M12 and M07 with a low and negative divergence $D = -2.7$ and $D_{\text{KL}} = 0.06$. F23, on the other hand, under-performed F19 with a $D = -4.8$ and $D_{\text{KL}} = 0.27$, which is still lower than previous energy functions (Fig 10a and 10b). Both F19 and F23 can design proteins with amino acid distribution comparable to native MPs. However, with the additional electrostatic terms, F23 over-enriched charged residues and under-enriched polar and aromatic residues.

Residue level probability distribution of designed sequences and perplexity. Natural protein sequences are constrained by evolution and explore only a part of the possible sequence space. *De novo* design methods, on the other hand, may explore all possible combinations guided by the energy functions to generate new proteins. With sequence recovery and divergence measurements, we have explored similarities between designed and native sequences. However, the question arises, for a given native residue, what is the range of residues the model considers plausible? Does the designed residue imitate the physicochemical

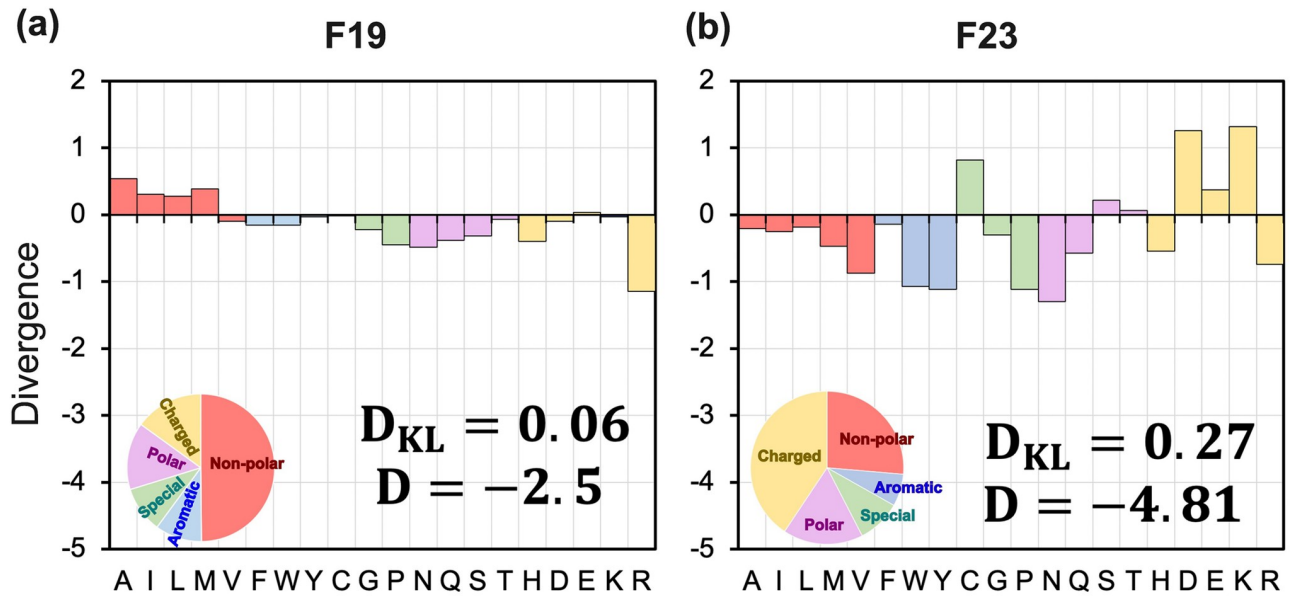


Fig 10. Distribution of residue divergence in the redesigned membrane proteins. Plots show the difference in log-likelihood (D) and KL-divergence (D_{KL}) of designed residue distribution relative to the native sequences. 204 MPs are redesigned with fix-bb and membrane orientation using the (a) F19 and (b) F23 energy functions. A positive divergence indicates that an amino acid is over-enriched, whereas a negative indicates that an amino acid is under-enriched. Values are given on a logarithmic scale. An amino acid composition pie chart for the sequence designed by each energy function is also shown in the bottom left-hand corner of the divergence plots. The color red is for non-polar, blue is for aromatic, violet is for polar, yellow is for charge, and green is for special residues.

<https://doi.org/10.1371/journal.pcbi.1011296.g010>

properties of the native residue? And how confident is the model in its design prediction? To do so, we explored the probability distribution of designed amino acids at all positions with a given native residue (Fig 11). To measure the model’s confidence in the prediction, we calculated perplexity per residue type (Fig 11d–11f):

$$pp_r = 2^{\sum_{s \in AA_{20}} -p(s|r) \log_2 p(s|r)} \tag{11}$$

where, pp_r is the perplexity for a native amino acid type r , $p(s|r)$ is the probability of amino acid s being designed to replace a native residue r . The probability $p(s|r)$ is given as:

$$p(s|r) = \frac{\sum_{i=1}^{N_{bb}} \sum_{j=1}^{N_i^{res}} \mathbb{1}_{x_{i,j}^{des}}(s, r)}{\sum_{i=1}^{N_{bb}} \sum_{j=1}^{N_i^{res}} \mathbb{1}_{x_{i,j}^{nat}}(r)} \tag{12}$$

where the superscripts *des* and *nat* denote designed and native sequences. The indicator function in the numerator counts native residues r that convert to s : $\mathbb{1}_{x_{i,j}^{des}}(s, r) := 1$ if $(x_{i,j}^{des} = s \wedge x_{i,j}^{nat} = r)$ and 0 otherwise, and the denominator indicator function counts native residues of type r : $\mathbb{1}_{x_{i,j}^{nat}}(r) := 1$ if $x_{i,j}^{nat} = r$ and 0 otherwise. When a model assigns a high probability to a given residue, imply its confidence in the decision (meaning less perplexed). To elaborate, a perplexity of 20 means the model is confused with 20 residue choices for design, and a perplexity of 2 means the model is confused with 2 residue choices. However, low perplexity alone can be misleading, and it is important for the designed residue choices to be of the right kind. Thus, an ideal energy function will have a confusion matrix of residue types on design outcomes with a high probability for diagonal elements (near diagonal elements represent the native residue or of similar chemical properties) and low perplexity.

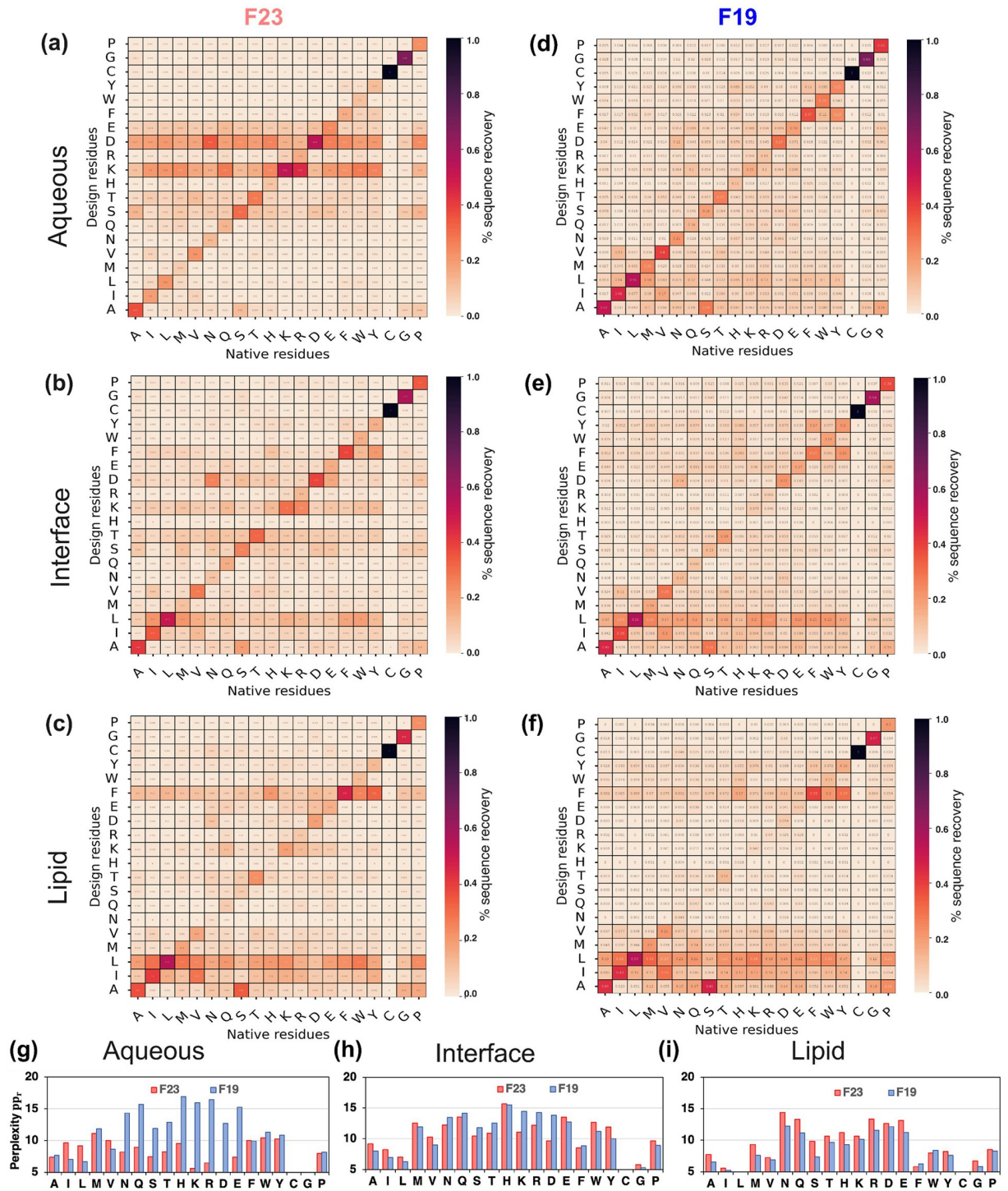


Fig 11. Redesigned membrane proteins exhibit native-like sequences. (a-c) Confusion matrices presenting probability distribution ($p_{i|r}$) of design residues i replacing a native residue r by F23 energy function. (d-f) Confusion matrices presenting probability distribution ($p_{i|r}$) of design residues i replacing a native residue r by F19 energy function. (g-i) Per-residue perplexity (pp_r) by F19 and F23 energy functions. In the confusion matrices, black denotes the highest (1.0), and light peach denotes the lowest (0.0) probabilities respectively. The criteria to distinguish between the membrane region is based on relative hydration in the membrane (f_{hyd}) and geometry of the water-exposed pore of the protein (f_{pore}). [44] Lipid: $f_{hyd} < 0.25, f_{pore} < 0.50$; Interface: $f_{hyd} \in [0.25, 0.75], f_{pore} < 0.50$; Aqueous: $f_{hyd} > 0.75, f_{pore} > 0.50$.

<https://doi.org/10.1371/journal.pcbi.1011296.g011>

F19 has low perplexity for designing non-polar and special residues and higher perplexity for charged and polar residues (Fig 11g–11i). F23 follows the trends of F19; however, it improved (reduced) the perplexity of the charged residues significantly for the aqueous region. For other polar residues, F23 has low perplexity; however, it often selects charged residues in place of other polar residues.

To sum, F19 can capture native protein-like features slightly better than F23. The perplexity and confusion matrix of F23 shows higher confidence and accuracy for designing charged residues.

Conclusion

We developed a new implicit energy function Franklin2023 (F23) by adding electrostatic energy terms to the existing energy function Franklin2019 (F19) and evaluated its performance based on five different tests calculating structure, stability, and design calculations. F23 has three modifications relative to F19. First is the fast-to-compute Coulomb-based electrostatic energy, including the low dielectric constant of the lipid bilayer. Second is the mean-field-based electrostatic energy term capturing the effect of the lipid head group in the membrane environment. Third is the modified value of $\Delta G_{w,l}^{aa}$ for Asp and Glu residues at neutral pH. As a result, F23 captures electrostatic and non-polar properties of diverse homogeneous lipid layers of variable thickness, distinguishes between zwitterionic and anionic lipid head groups, and identifies pores in MP structures.

Fig 12 summarizes the performance of F19 and F23 on the diverse benchmark test set. Relative to F19, F23 has improved the calculation of the tilt angle of membrane proteins for 90% of WALP peptides of different lengths, 15% of TM-peptides, and 25% of the adsorbed peptides. To calculate the tilt angles, we used a fixed backbone for the peptides. Experimental results have shown that the peptide backbone is quite flexible; we hypothesize allowing backbone flexibility will further improve the tilt angle calculation. F19 and F23 predicted stability equally well in $\Delta\Delta G_{\text{mutation}}$ and $\Delta\Delta G_{\text{ins}}$. F19 identifies native protein-like residues better than F23.

Test	Description	Experiment error	Computational metric	F19	F23
Protein orientation in the bilayer					
1a	Polyalanines tilt angle	—		100	100
1b	WALP peptides tilt angle	$\pm 5^\circ$	%peptides with error	60	80
2a	TM-peptide tilt angle	$\pm 5^\circ$	$\leq \pm 20^\circ \uparrow$	85	100
2b	Adsorbed peptide tilt angle	$\pm 12^\circ$		60	85
Stability					
3a	$\Delta\Delta G_{\text{mutation}}$ (OmpLA)		RMSE \downarrow	6.40	4.88
3b	$\Delta\Delta G_{\text{mutation}}$ (PagP)		RMSE \downarrow	7.09	7.10
4	$\Delta\Delta G_{w,l}$ for insertion		$R^2 \uparrow$	0.98	0.99
Design					
5	Sequence Recovery		$D_{\text{KL}} \downarrow$	-2.5	-4.8
5	Sequence Recovery		Recovery \uparrow	0.31	0.30

Fig 12. Summary of F19 and F23 performance on MP benchmark tests. Darker boxes show better performances.

<https://doi.org/10.1371/journal.pcbi.1011296.g012>

However, F23 significantly improved the confidence and accuracy of designing charged residues. For simplicity and lack of exact lipid types, we performed all design calculations in the DLPC bilayer; this is a broadly simplifying assumption that can be addressed by using the protein-matched membrane types.

The results of the benchmark tests show that, relative to F19, F23 has improved the ability of implicit membrane models to capture more relevant features of MPs. However, F23 did not improve predictions of stability and design performance as expected, possibly due to an incomplete representation of the membrane environment in the model. There are now several future steps that can be taken to improve the implicit models. Some potential steps are adding the effect of pH to account for membrane-induced pKa shifts, [68, 69] using robust optimization techniques to ameliorate double counting of physical effects due to use of empirical functions, [70] adding mixed and asymmetric lipid layers, [71, 72] and adding deformation of the lipid interface as it is perturbed by charged amino acids [73, 74]. F23 can serve as a good foundation model for including pH or lipid mixtures, as it includes additional electrostatic features and more characteristics of the lipid features. To capture the interface deformation or interaction of lipids with the residues, we have to move beyond the slab-like rigid representation of the membrane. Approaches such as those that include a defined curvature [75] or elastic bending constants of lipid membranes [76] along with other robust models can serve as foundation models for this purpose. We hypothesize that adding the above-mentioned features will capture the characteristics of realistic membranes. Configurations generated by implicit models can be used as an input or a filter for further investigations, including all-atom models to study phenomena mediated by lipid and water interaction. With continued improvement in MP feature representation, we anticipate that the implicit membrane model will enable reliable, high-throughput, and high-resolution MP structure prediction and design.

Supporting information

S1 Text. Implicit model to capture electrostatic features of membrane environment. Fig A. WALP35 peptide at the center of the lipid membrane. (a) The total energy by F23 and $\Delta G_{w,l}^{aa}$ using F19 and F23 calculated as a function of tilt angles. (b) The net energy components due to the presence of the lipid membrane as a function of tilt angle calculated using F19 (including $\Delta G_{w,l}$), F23 (including $\Delta G_{w,l}^{aa}$, ΔG_{lipid}^{aa} and M12 (including hydrophobicity, solvation, and knowledge terms). **WALP39 peptide in the lipid membrane.** (c) The per-residue $\Delta G_{w,l}^{aa}$ calculated using F23 and F19 for WALP39 at a depth = 10Å, tilt-angle = 90° and azimuthal-angle = 0°. (d) The total energy by F23 and $\Delta G_{w,l}^{aa}$ using F19 and F23 for WALP39 calculated as a function of tilt angles. (e) The net energy components due to the presence of the lipid membrane at depth = 0Å and as a function of tilt angle calculated using F19, F23 and M12. The arrows show the corresponding axes for the plot. The ones for which arrows are not shown have values represented in the other direction. **Fig B. Comparing the energy landscape for WALP25 and WALP35 by M12 at the center of the lipid membrane.** The M12 total energy and other membrane-based energy terms were calculated for WALP25 and WALP35 peptides as a function of tilt angle at a depth = 0Å and minimized over rotation angle. The different score terms are as follows: (a) The total energy due to the membrane environment includes hydrophobicity, solvation, and knowledge terms. (b) the knowledge term is a statistical measure of the propensity of a residue to be at a particular membrane depth. (c) hydrophobicity energy and (d) solvation energy. **Fig C. Energy landscape of influenza A M2 peptide (pdb: 1mp6).** (a) Comparing the native tilt angle (gray) with that calculated by F19 (blue) and F23 (magenta and light pink). The Asp and c-term Leu for which the $\Delta G_{w,l}^{aa}$ is modified by F23 are shown in sticks. The total energy landscape of 1mp6 is shown as calculated by (b) F19 and (c) F23 as a function of depth

in the membrane and tilt angle. The energy at each tilt angle is minimized over all azimuthal angles. (d) The per residue $\Delta G_{w,l}^{aa}$ of Imp6 calculated using F19 and F23 at membrane depth $d = 10\text{\AA}$, tilt-angle = 0° and azimuthal-angle = 0° . The trend of (e) total energy, (f) $\Delta G_{w,l}^{aa}$, and (g) ΔG_{lipid}^{aa} as a function of tilt angles as calculated by F19 at membrane depth $d = 5\text{\AA}$ and F23 at membrane depths $d = 5$ and 9\AA . **Fig D. Energy landscape of magainin-cecropin hybrid peptide (pdb: 1f0d).** (a) Comparing the native tilt angle (gray) with that calculated by F19 (blue) and F23(magenta). The total energy landscape of 1f0d is shown as calculated by (b) F19 and (c) F23 as a function of depth in the membrane and tilt angle. The energy at each tilt angle is minimized over all azimuthal angles. The trend of (d) total energy, (e) $\Delta G_{w,l}^{aa}$, and (f) ΔG_{lipid}^{aa} as a function of membrane depth as calculated by F19 and F23 minimized over all tilt-angle and azimuthal-angle. **Fig E. Test1:Orientation of WALP in different lipid layers.** Predicted tilt angle of WALP peptides as a function of the number of residues. The WALP peptides are represented as WALPx, where x is the number of residues. The different markers indicate tilt angles calculated using F23 and different lipid types. The experimental tilt angles are measured in the DMPC lipid types. Filled markers present peptides for which experimental results are available, and those for which simulated data is available are shown by unfilled markers. **Fig F. Test6: Variation in the Rosetta designs for a given PDB backbone.** (a) Comparing the sequence recovery for a PDB backbone. The bar shows the mean and the error bar shows the variation in the sequence recovery of 50 Rosetta designs for a given PDB. (b) Comparing the total Rosetta score (REU) for a PDB backbone. The bar shows the mean and the error bar shows the variation in the total score of 50 Rosetta designs for a given PDB. **Table A.** Electrostatic potential parameters fit from all-atom molecular dynamics data. **Table B.** Equation to calculate the $\Delta G_{w,l}^{atom}$ transfer energy. **Table C.** Lipid composition parameters for α -helical peptide tilt-angle calculations. (PDF)

Acknowledgments

We appreciate Dr. Rebecca Alford for sharing some of the code for the dielectric constant part of the work. We acknowledge Dr. Hope Woods and Dr. Rocco Moretti for reviewing our codes. We thank Prof. Damien Thevenin, Prof. Alexey Ladokhin for insightful discussions regarding WALP peptides. We also thank Dr. Patrick Fleming, Prof. Jeff Klauda, and Yulan Yu for their insightful discussions regarding electrostatic field verification. This work was carried out using facilities at the Advanced Research Computing at Hopkins (ARCH).

Author Contributions

Conceptualization: Rituparna Samanta.

Data curation: Rituparna Samanta.

Formal analysis: Rituparna Samanta.

Funding acquisition: Jeffrey J. Gray.

Investigation: Rituparna Samanta.

Methodology: Rituparna Samanta.

Project administration: Jeffrey J. Gray.

Software: Rituparna Samanta.

Supervision: Jeffrey J. Gray.

Validation: Rituparna Samanta.

Visualization: Rituparna Samanta.

Writing – original draft: Rituparna Samanta.

Writing – review & editing: Jeffrey J. Gray.

References

1. Langner M, Kubica K. The electrostatics of lipid surfaces. *Chemistry and Physics of Lipids*. 1999; 101(1):3–35. [https://doi.org/10.1016/S0009-3084\(99\)00052-3](https://doi.org/10.1016/S0009-3084(99)00052-3) PMID: 10810922
2. Howarth C, Gleeson P, Attwell D. Updated Energy Budgets for Neural Computation in the Neocortex and Cerebellum. *Journal of Cerebral Blood Flow & Metabolism*. 2012; 32(7):1222–32. <https://doi.org/10.1038/jcbfm.2012.35> PMID: 22434069
3. Cho W, Stahelin RV. Membrane-Protein Interactions in Cell Signaling and Membrane Trafficking. *Annual Review of Biophysics and Biomolecular Structure*. 2005; 34(1):119–51. <https://doi.org/10.1146/annurev.biophys.33.110502.133337> PMID: 15869386
4. Mulgrew-Nesbitt A, Diraviyam K, Wang J, Singh S, Murray P, Li Z, et al. The role of electrostatics in protein–membrane interactions. *Biochimica et Biophysica Acta (BBA)—Molecular and Cell Biology of Lipids*. 2006; 1761(8):812–26. <https://doi.org/10.1016/j.bbalip.2006.07.002> PMID: 16928468
5. Senes A, Ubarretxena-Belandia I, Engelman DM. The C α -H...O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions. *Proceedings of the National Academy of Sciences*. 2001; 98(16):9056–61. <https://doi.org/10.1073/pnas.161280798>
6. White SH, Wimley WC. Hydrophobic interactions of peptides with membrane interfaces. *Biochimica et Biophysica Acta (BBA)—Reviews on Biomembranes*. 1998; 1376(3):339–52. [https://doi.org/10.1016/S0304-4157\(98\)00021-5](https://doi.org/10.1016/S0304-4157(98)00021-5) PMID: 9804985
7. Granseth E, von Heijne G, Elofsson A. A Study of the Membrane–Water Interface Region of Membrane Proteins. *Journal of Molecular Biology*. 2005; 346(1):377–85. <https://doi.org/10.1016/j.jmb.2004.11.036> PMID: 15663952
8. Öjemalm K, Higuchi T, Lara P, Lindahl E, Suga H, von Heijne G. Energetics of side-chain snorkeling in transmembrane helices probed by nonproteinogenic amino acids. *Proceedings of the National Academy of Sciences*. 2016; 113(38):10559–64. <https://doi.org/10.1073/pnas.1606776113> PMID: 27601675
9. Ulmschneider MB, Ulmschneider JP, Freites JA, von Heijne G, Tobias DJ, White SH. Transmembrane helices containing a charged arginine are thermodynamically stable. *European Biophysics Journal*. 2017; 46(7):627–37. <https://doi.org/10.1007/s00249-017-1206-x> PMID: 28409218
10. Verbeek SF, Awasthi N, Teiwes NK, Mey I, Hub JS, Janshoff A. How arginine derivatives alter the stability of lipid membranes: dissecting the roles of side chains, backbone and termini. *European Biophysics Journal*. 2021; 50(2):127–42. <https://doi.org/10.1007/s00249-021-01503-x> PMID: 33661339
11. Killian JA, von Heijne G. How proteins adapt to a membrane–water interface. *Trends in Biochemical Sciences*. 2000; 25(9):429–34. [https://doi.org/10.1016/S0968-0004\(00\)01626-1](https://doi.org/10.1016/S0968-0004(00)01626-1) PMID: 10973056
12. Cao Z, Hutchison JM, Sanders CR, Bowie JU. Backbone Hydrogen Bond Strengths Can Vary Widely in Transmembrane Helices. *Journal of the American Chemical Society*. 2017; 139(31):10742–9. <https://doi.org/10.1021/jacs.7b04819> PMID: 28692798
13. Luo R, David L, Gilson MK. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J Comput Chem*. 2002 Oct; 23(13):1244–53. <https://doi.org/10.1002/jcc.10120> PMID: 12210150
14. David L, Luo R, Gilson MK. Comparison of generalized born and poisson models: Energetics and dynamics of HIV protease. *Journal of Computational Chemistry*. 2000; 21(4):295–309. [https://doi.org/10.1002/\(SICI\)1096-987X\(200003\)21:4%3C295::AID-JCC5%3E3.0.CO;2-8](https://doi.org/10.1002/(SICI)1096-987X(200003)21:4%3C295::AID-JCC5%3E3.0.CO;2-8)
15. Roux B, Bernèche S, Im W. Ion Channels, Permeation, and Electrostatics: Insight into the Function of KcsA. *Biochemistry*. 2000 Nov; 39(44):13295–306. <https://doi.org/10.1021/bi001567v> PMID: 11063565
16. Murray D, Honig B. Electrostatic Control of the Membrane Targeting of C2 Domains. *Molecular Cell*. 2002; 9(1):145–54. [https://doi.org/10.1016/S1097-2765\(01\)00426-9](https://doi.org/10.1016/S1097-2765(01)00426-9) PMID: 11804593
17. Im W, Feig M, Brooks CL. An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins. *Biophysical Journal*. 2003; 85(5):2900–18. [https://doi.org/10.1016/S0006-3495\(03\)74712-2](https://doi.org/10.1016/S0006-3495(03)74712-2) PMID: 14581194

18. Tanizaki S, Feig M. A generalized Born formalism for heterogeneous dielectric environments: Application to the implicit modeling of biological membranes. *J Chem Phys*. 2005 Mar; 122(12):124706. <https://doi.org/10.1063/1.1865992> PMID: 15836408
19. Feig M, Im W, Brooks CL. Implicit solvation based on generalized Born theory in different dielectric environments. *J Chem Phys*. 2003 Dec; 120(2):903–11. <https://doi.org/10.1063/1.1631258>
20. O'Meara MJ, Leaver-Fay A, Tyka MD, Stein A, Houlihan K, DiMaio F, et al. Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. *Journal of Chemical Theory and Computation*. 2015 02; 11(2):609–22. <https://doi.org/10.1021/ct500864r> PMID: 25866491
21. Morozov AV, Kortemme T, Baker D. Evaluation of Models of Electrostatic Interactions in Proteins. *The Journal of Physical Chemistry B*. 2003 03; 107(9):2075–90. <https://doi.org/10.1021/jp0267555>
22. Ben-Tal N, Honig B, Peitzsch RM, Denisov G, McLaughlin S. Binding of small basic peptides to membranes containing acidic lipids: theoretical models and experimental results. *Biophysical Journal*. 1996; 71(2):561–75. [https://doi.org/10.1016/S0006-3495\(96\)79280-9](https://doi.org/10.1016/S0006-3495(96)79280-9) PMID: 8842196
23. Ben-Tal N, Ben-Shaul A, Nicholls A, Honig B. Free-energy determinants of alpha-helix insertion into lipid bilayers. *Biophysical Journal*. 1996; 70(4):1803–12. [https://doi.org/10.1016/S0006-3495\(96\)79744-8](https://doi.org/10.1016/S0006-3495(96)79744-8) PMID: 8785340
24. Kessel A, Cafiso DS, Ben-Tal N. Continuum Solvent Model Calculations of Alamethicin-Membrane Interactions: Thermodynamic Aspects. *Biophysical Journal*. 2000; 78(2):571–83. [https://doi.org/10.1016/S0006-3495\(00\)76617-3](https://doi.org/10.1016/S0006-3495(00)76617-3) PMID: 10653772
25. Schlenkrich M, Brickmann J, MacKerell AD, Karplus M. In: Merz KM, Roux B, editors. *An Empirical Potential Energy Function for Phospholipids: Criteria for Parameter Optimization and Applications*. Boston, MA: Birkhäuser Boston; 1996. p. 31–81.
26. Kessel A, Ben-Tal N. Free energy determinants of peptide association with lipid bilayers. In: *Current Topics in Membranes*. vol. 52. Academic Press; 2002. p. 205–53.
27. Ulmschneider JP, Ulmschneider MB. Sampling efficiency in explicit and implicit membrane environments studied by peptide folding simulations. *Proteins: Structure, Function, and Bioinformatics*. 2009; 75(3):586–97. <https://doi.org/10.1002/prot.22270> PMID: 19003985
28. Lazaridis T. Effective energy function for proteins in lipid membranes. *Proteins: Structure, Function, and Bioinformatics*. 2003; 52(2):176–92. <https://doi.org/10.1002/prot.10410> PMID: 12833542
29. Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins: Structure, Function, and Bioinformatics*. 1999; 35(2):133–52. [https://doi.org/10.1002/\(SICI\)1097-0134\(19990501\)35:2%3C133::AID-PROT1%3E3.0.CO;2-N](https://doi.org/10.1002/(SICI)1097-0134(19990501)35:2%3C133::AID-PROT1%3E3.0.CO;2-N) PMID: 10223287
30. Lazaridis T, Leveritt JM, PeBenito L. Implicit membrane treatment of buried charged groups: Application to peptide translocation across lipid bilayers. *Biochimica et Biophysica Acta (BBA)—Biomembranes*. 2014; 1838(9):2149–59. Interfacially active peptides and proteins. <https://doi.org/10.1016/j.bbamem.2014.01.015> PMID: 24525075
31. Barth P, Schonbrun J, Baker D. Toward high-resolution prediction and design of transmembrane helical protein structures. *Proceedings of the National Academy of Sciences*. 2007; 104(40):15682–7. <https://doi.org/10.1073/pnas.0702515104> PMID: 17905872
32. Lu P, Min D, DiMaio F, Wei KY, Vahey MD, Boyken SE, et al. Accurate computational design of multi-pass transmembrane proteins. *Science*. 2018; 359(6379):1042–6. <https://doi.org/10.1126/science.aag1739> PMID: 29496880
33. MacCallum JL, Bennett WFD, Tieleman DP. Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophysical Journal*. 2008 May; 94(9):3393–404. <https://doi.org/10.1529/biophysj.107.112805> PMID: 18212019
34. Franks NP, Abraham MH, Lieb WR. Molecular Organization of Liquid n-Octanol: An X-ray Diffraction Analysis. *Journal of Pharmaceutical Sciences*. 1993 May; 82(5):466–70. <https://doi.org/10.1002/jps.2600820507> PMID: 8360823
35. Koehler Leman J, Lyskov S, Bonneau R. Computing structure-based lipid accessibility of membrane proteins with mp_lipid_acc in RosettaMP. *BMC Bioinformatics*. 2017; 18(1):115. <https://doi.org/10.1186/s12859-017-1541-z> PMID: 28219343
36. Hsieh D, Davis A, Nanda V. A knowledge-based potential highlights unique features of membrane α -helical and β -barrel protein insertion and folding. *Protein Science*. 2012; 21(1):50–62. <https://doi.org/10.1002/pro.758> PMID: 22031179
37. Schramm C, Hannigan B, Donald J, Keasar C, Saven J, DeGrado W, et al. Knowledge-Based Potential for Positioning Membrane-Associated Structures and Assessing Residue-Specific Energetic Contributions. *Structure*. 2012; 20(5):924–35. <https://doi.org/10.1016/j.str.2012.03.016> PMID: 22579257
38. Senes A, Chadi DC, Law PB, Walters RFS, Nanda V, DeGrado WF. E(z), a Depth-dependent Potential for Assessing the Energies of Insertion of Amino Acid Side-chains into Membranes: Derivation and

- Applications to Determining the Orientation of Transmembrane and Interfacial Helices. *Journal of Molecular Biology*. 2007; 366(2):436–48. <https://doi.org/10.1016/j.jmb.2006.09.020> PMID: 17174324
39. Alford RF, Koehler Leman J, Weitzner BD, Duran AM, Tilley DC, Elazar A, et al. An Integrated Framework Advancing Membrane Protein Modeling and Design. *PLOS Computational Biology*. 2015 09; 11(9):1–23. <https://doi.org/10.1371/journal.pcbi.1004398> PMID: 26325167
 40. Moon CP, Fleming KG. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proceedings of the National Academy of Sciences*. 2011; 108(25):10174–7. <https://doi.org/10.1073/pnas.1103979108> PMID: 21606332
 41. Fogolari F, Brigo A, Molinari H. The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition*. 2002 2023/02/22; 15(6):377–92. <https://doi.org/10.1002/jmr.577> PMID: 12501158
 42. Fogolari F, Zuccato P, Esposito G, Viglino P. Biomolecular Electrostatics with the Linearized Poisson–Boltzmann Equation. *Biophysical Journal*. 1999; 76(1):1–16. [https://doi.org/10.1016/S0006-3495\(99\)77173-0](https://doi.org/10.1016/S0006-3495(99)77173-0) PMID: 9876118
 43. Alford RF, Samanta R, Gray JJ. Diverse Scientific Benchmarks for Implicit Membrane Energy Functions. *Journal of Chemical Theory and Computation*. 2021; 17(8):5248–61. <https://doi.org/10.1021/acs.jctc.0c00646> PMID: 34310137
 44. Alford RF, Fleming PJ, Fleming KG, Gray JJ. Protein Structure Prediction and Design in a Biologically Realistic Implicit Membrane. *Biophysical Journal*. 2020; 118(8):2042–55. <https://doi.org/10.1016/j.bpj.2020.03.006> PMID: 32224301
 45. Yu Y, Krämer A, Venable RM, Brooks BR, Klauda JB, Pastor RW. CHARMM36 Lipid Force Field with Explicit Treatment of Long-Range Dispersion: Parametrization and Validation for Phosphatidylethanolamine, Phosphatidylglycerol, and Ether Lipids. *Journal of Chemical Theory and Computation*. 2021; 17(3):1581–95. <https://doi.org/10.1021/acs.jctc.0c01327> PMID: 33620194
 46. Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*. 2017; 13(6):3031–48. <https://doi.org/10.1021/acs.jctc.7b00125> PMID: 28430426
 47. Stern HA, Feller SE. Calculation of the dielectric permittivity profile for a nonuniform system: Application to a lipid bilayer simulation. *The Journal of Chemical Physics*. 2003; 118(7):3401–12. <https://doi.org/10.1063/1.1537244>
 48. Gramse G, Dols-Perez A, Edwards MA, Fumagalli L, Gomila G. Nanoscale measurement of the dielectric constant of supported lipid bilayers in aqueous solutions with electrostatic force microscopy. *Biophysical Journal*. 2013; 104(6):1257–62. <https://doi.org/10.1016/j.bpj.2013.02.011> PMID: 23528085
 49. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. *Proceedings of the National Academy of Sciences*. 2000; 97(19):10383–8. <https://doi.org/10.1073/pnas.97.19.10383> PMID: 10984534
 50. Wimley WC, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology*. 1996; 3(10):842–8. <https://doi.org/10.1038/nsb1096-842> PMID: 8836100
 51. McDonald SK, Fleming KG. Aromatic Side Chain Water-to-Lipid Transfer Free Energies Show a Depth Dependence across the Membrane Normal. *Journal of the American Chemical Society*. 2016; 138(25):7946–50. <https://doi.org/10.1021/jacs.6b03460> PMID: 27254476
 52. Casal HL, Mantsch HH. Polymorphic phase behaviour of phospholipid membranes studied by infrared spectroscopy. *Biochimica et Biophysica Acta (BBA)—Reviews on Biomembranes*. 1984; 779(4):381–401. [https://doi.org/10.1016/0304-4157\(84\)90017-0](https://doi.org/10.1016/0304-4157(84)90017-0) PMID: 6391546
 53. Hauser H, Pascher I, Pearson RH, Sundell S. Preferred conformation and molecular packing of phosphatidylethanolamine and phosphatidylcholine. *Biochimica et Biophysica Acta (BBA)—Reviews on Biomembranes*. 1981; 650(1):21–51. [https://doi.org/10.1016/0304-4157\(81\)90007-1](https://doi.org/10.1016/0304-4157(81)90007-1) PMID: 7020761
 54. Zhao W, Róg T, Gurtovenko AA, Vattulainen I, Karttunen M. Atomic-Scale Structure and Electrostatics of Anionic Palmitoyloleoylphosphatidylglycerol Lipid Bilayers with Na⁺ Counterions. *Biophysical Journal*. 2007; 92(4):1114–24. <https://doi.org/10.1529/biophysj.106.086272> PMID: 17114222
 55. Dickey A, Faller R. Examining the contributions of lipid shape and headgroup charge on bilayer behavior. *Biophysical Journal*. 2008; 95(6):2636–46. <https://doi.org/10.1529/biophysj.107.128074> PMID: 18515396
 56. Smaby JM, Brockman HL. Surface dipole moments of lipids at the argon-water interface. Similarities among glycerol-ester-based lipids. *Biophysical Journal*. 1990; 58(1):195–204. [https://doi.org/10.1016/S0006-3495\(90\)82365-1](https://doi.org/10.1016/S0006-3495(90)82365-1) PMID: 2383632

57. Woetzel N, Karakaş M, Staritzbichler R, Müller R, Weiner BE, Meiler J. BCL::Score—Knowledge Based Energy Potentials for Ranking Protein Models Represented by Idealized Secondary Structure Elements. *PLOS ONE*. 2012 11; 7(11):1–17. <https://doi.org/10.1371/journal.pone.0049242> PMID: 23173051
58. Koehler Leman J, Lyskov S, Lewis SM, Adolf-Bryfogle J, Alford RF, Barlow K, et al. Ensuring scientific reproducibility in bio-macromolecular modeling via extensive, automated benchmarks. *Nature Communications*. 2021; 12(1):6947. <https://doi.org/10.1038/s41467-021-27222-7> PMID: 34845212
59. Sengupta D, Meinhold L, Langosch D, Ullmann GM, Smith JC. Understanding the energetics of helical peptide orientation in membranes. *Proteins: Structure, Function, and Bioinformatics*. 2005; 58(4):913–22. <https://doi.org/10.1002/prot.20383> PMID: 15657932
60. Holt A, Rougier L, Réat V, Jolibois F, Saurel O, Czaplicki J, et al. Order Parameters of a Transmembrane Helix in a Fluid Bilayer: Case Study of a WALP Peptide. *Biophysical Journal*. 2010; 98(9):1864–72. <https://doi.org/10.1016/j.bpj.2010.01.016> PMID: 20441750
61. Van der Wel PCA, Strandberg E, Killian JA, Koeppe RE. Geometry and Intrinsic Tilt of a Tryptophan-Anchored Transmembrane α -Helix Determined by 2H NMR. *Biophysical Journal*. 2002; 83(3):1479–88. [https://doi.org/10.1016/S0006-3495\(02\)73918-0](https://doi.org/10.1016/S0006-3495(02)73918-0) PMID: 12202373
62. de Planque MRR, Goormaghtigh E, Greathouse DV, Koeppe RE, Kruijtz JAW, Liskamp RMJ, et al. Sensitivity of Single Membrane-Spanning α -Helical Peptides to Hydrophobic Mismatch with a Lipid Bilayer: Effects on Backbone Structure, Orientation, and Extent of Membrane Incorporation. *Biochemistry*. 2001; 40(16):5000–10. <https://doi.org/10.1021/bi000804r> PMID: 11305916
63. Ulmschneider MB, Sansom MSP, Di Nola A. Evaluating Tilt Angles of Membrane-Associated Helices: Comparison of Computational and NMR Techniques. *Biophysical Journal*. 2006; 90(5):1650–60. <https://doi.org/10.1529/biophysj.105.065367> PMID: 16339877
64. Bechinger B, Zasloff M, Opella SJ. Structure and orientation of the antibiotic peptide magainin in membranes by solid-state nuclear magnetic resonance spectroscopy. *Protein Science*. 1993; 2(12):2077–84. <https://doi.org/10.1002/pro.5560021208> PMID: 8298457
65. Ulmschneider MB, Ulmschneider JP, Schiller N, Wallace BA, Heijne Gv, White SH. Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion. *Nature Communications*. 2014; 5 (4863).
66. Leaver-Fay A, O'Meara MJ, Tyka M, Jacak R, Song Y, Kellogg EH, et al. Scientific benchmarks for guiding macromolecular energy function improvement. *Methods in Enzymology*. 2013; 523:109–43. <https://doi.org/10.1016/B978-0-12-394292-0.00006-0> PMID: 23422428
67. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic acids research*. 2012; 40(Database issue): D370–6. <https://doi.org/10.1093/nar/gkr703> PMID: 21890895
68. Panahi A, Brooks CLI. Membrane Environment Modulates the pKa Values of Transmembrane Helices. *J Phys Chem B*. 2015 Apr; 119(13):4601–7. <https://doi.org/10.1021/acs.jpcc.5b00289> PMID: 25734901
69. Vasquez-Montes V, Gerhart J, Thévenin D, Ladokhin AS. Divalent Cations and Lipid Composition Modulate Membrane Insertion and Cancer-Targeting Action of pHLIP. *Journal of Molecular Biology*. 2019; 431(24):5004–18. <https://doi.org/10.1016/j.jmb.2019.10.016> PMID: 31689432
70. Park H, Bradley P, Greisen P, Liu Y, Mulligan VK, Kim DE, et al. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *Journal of Chemical Theory and Computation*. 2016; 12(12):6201–12. <https://doi.org/10.1021/acs.jctc.6b00819> PMID: 27766851
71. Haleva E, Ben-Tal N, Diamant H. Increased Concentration of Polyvalent Phospholipids in the Adsorption Domain of a Charged Protein. *Biophysical Journal*. 2004 Apr; 86(4):2165–78. [https://doi.org/10.1016/S0006-3495\(04\)74276-9](https://doi.org/10.1016/S0006-3495(04)74276-9) PMID: 15041657
72. Heimburg T, Angerstein B, Marsh D. Binding of Peripheral Proteins to Mixed Lipid Membranes: Effect of Lipid Demixing upon Binding. *Biophysical Journal*. 1999 May; 76(5):2575–86. [https://doi.org/10.1016/S0006-3495\(99\)77410-2](https://doi.org/10.1016/S0006-3495(99)77410-2) PMID: 10233072
73. Vorobyov I, Bekker B, Allen TW. Electrostatics of Deformable Lipid Membranes. *Biophysical Journal*. 2010; 98(12):2904–13. <https://doi.org/10.1016/j.bpj.2010.03.046> PMID: 20550903
74. Vorobyov I, Allen TW. On the role of anionic lipids in charged protein interactions with membranes. *Biochimica et Biophysica Acta (BBA)—Biomembranes*. 2011; 1808(6):1673–83. <https://doi.org/10.1016/j.bbamem.2010.11.009> PMID: 21073855
75. Woods HM, Leman JK, Meiler J. Modeling membrane geometries implicitly in Rosetta. *bioRxiv*. 2023.
76. Marsh D. Elastic curvature constants of lipid monolayers and bilayers. *Chemistry and Physics of Lipids*. 2006; 144(2):146–59. <https://doi.org/10.1016/j.chemphyslip.2006.08.004> PMID: 17045578