Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides and proteins in membranes

Andrei L. Lomize[*], **Irina Pogozheva**, and **Henry I Mosberg**
Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, 428 Church St., Ann Arbor, MI, 48109-1065, USA

Abstract

A new computational approach to calculating binding energies and spatial positions of small molecules, peptides and proteins in the lipid bilayer has been developed. The method combines an anisotropic solvent representation of the lipid bilayer and universal solvation model, which predicts transfer energies of molecules from water to an arbitrary medium with defined polarity properties. The universal solvation model accounts for hydrophobic, van der Waals, hydrogen-bonding, and electrostatic solute-solvent interactions. The lipid bilayer is represented as a fluid anisotropic environment described by profiles of dielectric constant (ε), solvatochromic dipolarity parameter (π^*), and hydrogen bonding acidity and basicity parameters (α and β). The polarity profiles were calculated using published distributions of quasi-molecular segments of lipids determined by neutron and X-ray scattering for DOPC bilayer and spin-labeling data that define concentration of water in the lipid acyl chain region. The model also accounts for the preferential solvation of charges and polar groups by water and includes the effect of the hydrophobic mismatch for transmembrane proteins. The method was tested on calculations of binding energies and preferential positions in membranes for small-molecules, peptides and peripheral membrane proteins that have been experimentally studied. The new theoretical approach was implemented in a new version (2.0) of our PPM program and applied for the large-scale calculations of spatial positions in membranes of more than 1000 peripheral and integral proteins. The results of calculations are deposited in the updated OPM database (http://opm.phar.umich.edu).

Keywords

transfer energy; water permeation; polarity profile; membrane protein; database; Arg-rich peptide; hydrophobicity scale

---

[*]CORRESPONDING AUTHOR FOOTNOTE, Andrei L. Lomize, Ph.D. Associate Research Scientist, Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, 428 Church St., Ann Arbor, MI, 48109-1965, Tel: 734-615-7194; Fax: 734-763-5595almz@umich.edu.

INTRODUCTION

Biological membranes separate cells from the environment and create intracellular compartments. Many vital functions are associated with membranes, such as biosynthesis, signaling, transport, communication and energy conversion. Interactions and reciprocal influence of major membrane components, proteins and lipids, define properties of membranes, including their size, thickness, shape, charge, polarity and density[1–4]. In particular, the presence of proteins affects the lipid order, lateral diffusion, surface tension, and lipid curvature[3, 4]. On the other hand, lipid environment organizes and directs folding and co-translational or post-translational insertion in membranes of transmembrane α-helical[5–8] or β-barrel proteins[9], membrane-associated peptides, toxins, and peripheral proteins[10]. Modulation of lipids influences spatial orientation, assembly, targeting, conformational transitions and function of membrane-embedded proteins[2, 8, 11] and alter binding and distribution of small molecules, such as ligands, metabolites and drugs[12].

To understand and describe these processes, it is important to have an adequate representation of complex physicochemical properties of the lipid bilayer, which markedly change over short distances along the membrane normal, and to properly describe the energetics of protein-lipid interactions. The lipid matrix can be represented either explicitly, or as a fluid milieu with defined characteristics. The later approach is much more computationally efficient and practical for simulation of large multi-protein complexes. However, even more advanced continuum solvent models, which account for both non-electrostatic and electrostatic protein-solvent interactions[13, 14], usually apply a very simplistic approximation of the lipid bilayer by one or several hydrophobic slabs and implement arbitrary functions for describing the dielectric permittivity changes at the lipid-water interface[1]. Furthermore, these models focus mainly on the dielectric constant of solvent that defines the long-range electrostatics, but do not include the hydrogen bonding properties of the solvent that are known to be essential for describing interactions in the first solvation shell of polar molecules and ions[15–18].

In the preceding paper we presented empirical separation and parameterization of long-range electrostatics and the first solvation shell effects that account for hydrophobic, van der Waals, and hydrogen-bonding solute-solvent interactions. We also found the minimal set of solvent parameters required to reproduce transfer energies of molecules from water to any solvent: hydrogen bonding acidity and basicity parameters (α and β), dielectric constant ε, and solvatochromic dipolarity/polarizability parameter π*.

In this study we extended this model to quantify the energetics of protein-lipid interactions in artificial lipid bilayers and natural biological membranes. We introduced an advanced implicit solvent representation of the lipid bilayer that describes its anisotropic properties by transbilayer profiles of parameters α, β, ε and π*. The profiles were calculated from experimental distributions of lipid segments along the bilayer normal as determined by X-ray scattering and neutron diffraction of dioleoylphosphatidylcholine (DOPC) bilayers[19]. We also included the distribution of water in the lipid bilayer assessed by spin probes[20], preferential solvation of protein groups by water and lipid, and incorporated the lipid-protein hydrophobic matching condition[21].

This advanced model was applied for calculation of transbilayer energy profiles of amino-acid residues incorporated in an α-helix or a β-barrel inserted in the lipid bilayer. These physical energy profiles were compared with statistical energy profiles derived from distributions along the membrane normal of solvent-exposed residues in three-dimensional (3D) structures of membrane proteins. This allowed us to refine the parameters of water distribution in the lipid acyl chain region of biological membranes.

The proposed anisotropic solvent model of the lipid bilayer was integrated in a new version (2.0) of our PPM (Positioning of Proteins in Membranes) program[22]. This computational method is aimed at prediction of energetically optimal orientations and binding energy of diverse molecules ranging from small organic compounds to multi-protein complexes in membranes. PPM 2.0 was thoroughly tested and subsequently applied for calculations of spatial positions in membrane of more than 1000 integral and peripheral membrane proteins and peptides that were deposited in the updated OPM (Orientations of Proteins in Membranes) database[23] (http://opm.phar.umich.edu).

METHODS

Structure of the lipid bilayer

Volume probabilities (fractions) of different lipid segments along the DOPC bilayer, $\phi_m(z)$ were taken from the most recent X-ray and neutron scattering study[19]. The segments include CH=CH (CH), carbonyl+glycerol (CG), phosphate+choline (PCN), and methyl groups of choline (CholCH$_3$). (m=1, 3, 4 and 5, respectively):

$$\phi_m(z) = \frac{n_m V_m}{\sqrt{2\pi} A s_m} \left(\exp\left[-\frac{(z+Z_m)^2}{2s_m^2} \right] + \exp\left[-\frac{(z-Z_m)^2}{2s_m^2} \right] \right)$$

(1)

where A is the bilayer surface area per lipid molecule, V_m is the molecular volume of the fragment (in Å^3), Z_m and s_m define the locations and widths of corresponding Gaussians, respectively, and the number (n_m) is equal to 2 for CH(CH=CH) groups (there are two acyl chains in each lipid molecule) and to 1 for all other groups. The corresponding parameters are provided in Tables S1 and S2.

Volume fraction of total hydrocarbon (CH+CH$_2$+CH$_3$) was calculated using Gauss error function,

$$\phi_{HDC}(z) = 1/2 \left[erf(\frac{z+Z_{HDC}}{\sqrt{2} s_{HDC}}) - erf(\frac{z-Z_{HDC}}{\sqrt{2} s_{HDC}}) \right]$$

(2)

Volume fractions of CH$_2$+CH$_3$ groups and water were calculated as the differences:

$$\phi_{CH2+CH3}(z) = \phi_2(z) = \phi_{HDC}(z) - \phi_{CH}(z)$$

(3)

$$\phi_W^{X-Ray}(z) = \phi_6(z) = 1 - \sum_{m=1}^{5} \phi_m(z)$$

(4)

According to Electron Spin Resonance (ESR) studies, lipid bilayers contain an appreciable amount of water that is not detected by scattering methods. The distribution of water in the fluid DOPC bilayer determined by ESR probes was described by sigmoid function:

$$\phi_W^{ESR}(z) = \phi_{W,outer} + \frac{\phi_{W,mid} - \phi_{W,outer}}{1 + e^{(|z|-z_0)/\lambda}}$$

(5)

where $\phi_{W,outer}=0.037$ and $\phi_{W,mid}=0.07$ are volume fractions of water in the outer (more ordered) region of lipid acyl chains near glycerol groups and in the center of membrane, respectively, $\lambda=0.54$ Å, and $z_0=8.2$ Å[20, 24]. Molar water concentration can be defined as ϕ_W multiplied by bulk water concentration (55.5 M).

The concentration of water inside the membrane was taken as maximal of values determined from the X-ray and neutron scattering and ESR data because low concentrations of water are undetectable by the scattering ($\phi_W^{X-Ray}=0$ in the middle of membrane):

$$\phi_W(z)=\max\{\phi_W^{X-Ray}(z), \phi_W^{ESR}(z)\} \tag{6}$$

Mole fractions of components were calculated from the volume fractions:

$$x_i^{bil}(z)=\frac{\phi_i(z)/V_i}{\sum_{k=1}^{6}\phi_k(z)/V_k} \tag{7}$$

where V_i is partial volume of group i (Table S1; the partial volume of aliphatic groups for calculating molar fractions was taken as for $(CH_2)_4$ segment, 108 Å^3).

Calculation of polarity profiles

The bilayer was treated as a binary mixture of two weakly interacting solvents: (a) the hydrophilic (aqueous) solution of zwitterionic phosphatidylcholine groups in water and (b) the hydrophobic solution of lipid acyl chains including their glycerol groups. The preferential solvation occurs in the first solvation shell and includes the hydrophobic effect and direct hydrogen bonding with the solvent, as defined by parameters α and β. To describe this situation, we calculated six different profiles with α and β defined separately for the hydrophobic and aqueous components: $\alpha_{lip}(z)$, $\beta_{lip}(z)$, $\alpha_{aq}(z)$, $\beta_{aq}(z)$, $\pi_{bil}(z)$ or $\varepsilon_{bil}(z)$ and concentration of water $\phi_{aq}(z)$.

Polarity profiles were calculated from volume fractions of lipid segments and water determined by X-ray and neutron scattering (equations (1–6)) and using contributions of the corresponding groups to parameters α, β, π*, and ε. The values of α, β, π*, and ε for different groups were taken as for solvents with similar chemical structure (Table S1). However, we made two additional simplifications which helped to improve performance of the model: (a) ε of charged lipid groups dissolved in water (PCN and CholCh3 segments) were taken the same as ε of the water (78.4) based on the general observation that aqueous solutions of zwitterions have dielectric constants only slightly different from that of water[25]; and (b) parameters α and β of PCN and choline methyl fragments were taken as for the whole phosphatidylcholine group (that is 0.82 and 1.74, respectively).

The π* values were defined from molar fractions of the corresponding components[26, 27]:

$$\pi_{bil}^*(z)=\sum_{m=1}^{6}x_m^{bil}(z)\,\pi_m^* \tag{8}$$

Dielectric constant of the non-protic lipid part was calculated as a harmonic mean of the lipid components[27]

$$\varepsilon_{lip}(z)=\left[\sum_{k=1}^{5}\phi_m^{lip}(z)/\varepsilon_m\right]^{-1}$$

(9)

where volume fractions reflect composition of only lipid components, excluding water

$$\phi_m^{lip}(z)=\frac{\phi_m(z)}{\sum_{k=1}^{5}\phi_k(z)}$$

(10)

The subsequent averaging of dielectric constant for the mixture of lipid and water must be done in terms of mole fractions, since water is a protic solvent[27]:

$$\varepsilon_{mix}(x_w)=(1-x_w(z))\,\varepsilon_{lip}(z)+x_w(z)\,\varepsilon_w+F(x_w,\varepsilon_0,\varepsilon_1,\varepsilon_2)$$

(11)

where x_W and ε_W are the molar fraction and dielectric constant of water, respectively, and the polynomial function F describes the non-linear behavior of ε using empirical coefficients ε_0, ε_1 and ε_2 whose values depend on the specific co-solvent. These coefficients were determined for many different water-co-solvent pairs[27]. For co-solvents with dielectric constant lower than that of water, the $\varepsilon_{mix}(x_W)$ curves are usually convex because very low concentrations of water contribute little to the dielectric constant. Since the parameters for water-lipid mixtures are unknown, the convex curve (11) was approximated by a shifted linear dependence:

$$\begin{aligned}\varepsilon_{bil}(x_w) &=\varepsilon_{lip} \text{ if } x_w<\delta\\ \varepsilon_{bil}(x_w) &=\varepsilon_{lip}+(\varepsilon_w-\varepsilon_{lip})\frac{x_w-\delta}{1-\delta}, \text{ if } x_w\geq\delta\end{aligned}$$

(12)

Hence the presence of very small amounts of water does not affect the dielectric constant of lipid bilyer ($\varepsilon_{bil}=\varepsilon_{lip}$ at $x_W<\delta=0.1$), and ε_{bil} reaches the value of ε_W at $x_W=1$.

Parameters α, β were calculated using mole fractions[27]:

$$\alpha_{lip}(z)=\sum_{m=1}^{3}x_m^{lip}(z)\,\alpha_m$$

(13)

$$\beta_{lip}(z)=\sum_{m=1}^{3}x_m^{lip}(z)\,\beta_m$$

(14)

$$\alpha_{aq}(z)=\sum_{m=4}^{6}x_m^{aq}(z)\,\alpha_m$$

(15)

$$\beta_{aq}(z)=\sum_{m=4}^{6} x_m^{aq}(z)\,\beta_m$$

(16)

where α_m, β_m, π_m and ε_m are fragmental parameters of the corresponding lipid segments, CH(CH=CH), CH$_2$/CH$_3$, CG, PCN, and CholCH$_3$ (m=1,2,3,4 and 5, respectively) and water (m=6) shown in Table S1. The corresponding mole fractions, ϕ_m were defined separately for the lipid acyl chain and aqueous mixtures:

$$x_i^{lip}(z)=\frac{\phi_i(z)/V_i}{\displaystyle\sum_{k=1}^{3}\phi_k(z)/V_k}$$

(17)

$$x_i^{aq}(z)=\frac{\phi_i(z)/V_i}{\displaystyle\sum_{k=4}^{6}\phi_k(z)/V_k}$$

(18)

Universal solvation model for anisotropic environment

The transfer energy of a solute molecule from water to an arbitrary position in the lipid bilayers was calculated using the universal solvation model developed in our previous work, but including the ionizable groups and dependence of solvation parameters σ_i and η on the atom position along the bilayer normal (z_i):

$$\Delta G_{transf}(d,\phi,\tau)=\sum_{i=1}^{N}\sigma_i^{wat\to bil}(z_i)\,ASA_i+\sum_{j=1}^{M}\eta_j^{wat\to bil}(z_j)\,\mu_j+\sum_{k=1}^{L}\min\{\Delta E_k^{ion},\Delta E_k^{neutr}\}$$

(19)

where $\sigma_i(z_i)$ is solvation parameter that depends on type of atom and describes surface transfer energy (per Å^2) of atom i from bulk water to the point z_i, ASA_i is a solvent-accessible surface area of atom i, $\eta(z_j)$ is an energy cost of transferring the dipole moment of 1D from water to point z_j, μ_j is a dipole moment of group j, E_k^{ion} and E_k^{neutr} are energies of ionizable group k in ionized and neutral states, respectively, N is the number of atoms in the molecule (excluding atoms of ionizable groups), M is the number of group dipoles, L is the number of ionizable groups, and parameters d, ϕ and τ define spatial position of the molecule with respect to the lipid bilayer, as previously described[22]. Thus, each ionizable group was individually selected in a charged or uncharged state, whichever has a lower energy at the given depth in the membrane.

The transfer energy of ionizable group k in neutral state was calculated as a sum of ionization energy of the group and *ASA*-dependent transfer energies of the corresponding atoms in uncharged state:

$$\Delta E_k^{neutr}(z)=\Delta G_k^{ionization}+\sum_{l=1}^{L_k}\sigma_l^{wat\to bil}(z_l)\,ASA_l$$

(20)

where L_k is the number of atoms in ionizable group k, while the energy cost of deionization during transfer from water to nonpolar environment was defined by the Henderson-Hasselbalch equation:

$$\Delta G_k^{ionization} = 2.3RT(pH - pK_k) \tag{21}$$

The transfer energy in the ionized state was described by the following equation:

$$\Delta E_k^{ion}(z) = \frac{166\, e_{Born}}{r_k}[F_{Abe}^{wat} - F_{Abe}(z_k)] + \sum_{l=1}^{L_k} \sigma_{l,ion}^{wat \to bil}(z_l)ASA_l \tag{22}$$

where $\sigma_{l,ion}$ is solvation parameter of O or N atoms in charged state; e_{Born} is a weight factor of long-range electrostatic contribution to transfer energy; r_k is an ionic radius. The dielectric function for ions was described by the Born equation modified by Abe[28]:

$$F_{Abe}(z) = \frac{1}{\ln \varepsilon(z)} - \frac{1}{\varepsilon(z) \ln \varepsilon(z)} - 1 \tag{23}$$

The electrostatic energy was corrected for charged groups involved in ion pairs as described in equation (40). To simplify the calculations, the energy of an ionizable group was evenly divided between all atoms of the group (two oxygens in acids and three nitrogens in guanidine group). Dipole moments of ionizable groups were taken as in unionized state (i.e. as in the middle of the lipid bilayer).

Atomic and solvation radii were chosen as described in our companion paper. Accessible surface areas were calculated using the program SOLVA from NACCESS with probe radius of water (1.4 Å).

According to our universal solvation model, the values of solvation parameters σ_i for different types of atoms (equation 19) depend on polarity parameters of the lipid-water mixture (α, β and ε) that change along the bilayer normal:

$$\sigma_i^{wat \to bil}(z) = \sigma_i^0 - e_i(1/\varepsilon_{bil}(z) - 1/\varepsilon_{wat}) + a_i(\alpha_{bil}(z) - \alpha_{wat}) + b_i(\beta_{bil}(z) - \beta_{wat}) \tag{24}$$

where $\alpha_{bil}(z)$, $\beta_{bil}(z)$ and $\varepsilon_{bil}(z)$ are transmembrane profiles of hydrogen bonding donor and acceptor capacities and dielectric constant of the lipid bilayer, respectively, α_{wat}, β_{wat} and ε_{wat} are the corresponding values in water, and the linear regression coefficients e_i, a_i and b_i for different types of atoms (Table S3) were derived from water-solvent transfer energies of model compounds in the preceding publication. However, this equation was modified to describe the preferential solvation (see below).

Two alternative dielectric models were tested for describing the long-range dipolar contribution. This contribution was defined by parameter η that represents energy cost of transferring an electric dipole moment of $1D$ from water to the point of membrane with coordinate z. The first model was based on the use of solvatochromic dipolarity parameter, π^*:

$$\eta^{wat \to bil}(z) = e_{dip,\pi}\, [\pi_{bil}^*(z) - \pi_{wat}^*] \tag{25}$$

where $\pi^*_{bil}(z)$ and π^*_{wat} are values of this parameter in bilayer and water, respectively.

An alternative model for the dipolar contribution was based on the Block-Walker dielectric function of the media, $F_{BW}(\varepsilon)$[29]:

$$\eta^{wat \to bil}(z) = e_{dip,BW}\,(F_{BW}^{bil}(z) - F_{BW}^{wat}) \tag{26}$$

$$F_{BW} = \frac{3\varepsilon \ln \varepsilon}{(\varepsilon \ln \varepsilon - \varepsilon + 1)} - \frac{6}{\ln \varepsilon} - 2 \tag{27}$$

The required regression coefficients e_{Born}, σ^0, e_i, a_i, b_i, $e_{dip,\pi}$, and $e_{dip,BW}$ (equations 22, 24–26) were taken from our accompanying paper (Table S3). The signs were chosen to describe transfer from water considered as the reference phase.

Treatment of preferential solvation

The universal solvation model (equations 19–27) was modified to account for the preferential solvation of protein groups by water. We used a simple model of preferential solvation that operates with local mole fractions of solvents in the first solvation shell of a solute[27, 30]. The lipid bilayer was treated as a binary mixture of two weakly interacting phases: an aqueous solution of charged phosphatidylcholine groups (Aq), and a nonpolar solution of acyl chains including glycerol groups of lipids (L).

$$S_i \bullet Aq + L = S_i \bullet L + Aq \tag{28}$$

where Aq is a molecule of an aqueous component (primarily water), and L is a segment of lipid that replaces the molecule of water. This equation was used for each individual atom, S_i of the macromolecular solute.

The corresponding preferential solvation constant is then defined as follows[27, 30]:

$$K = \frac{x_{SAq}\,x_L}{x_{SL}\,x_{Aq}} = \exp(\Delta G_i(z)/RT) \tag{29}$$

where x are mole fractions of the corresponding species in the vicinity of solute group S_i

$$x_{SAq} + x_{SL} = x_L + x_{Aq} = 1 \tag{30}$$

and the free energy difference associated with Aq/L exchange is

$$\Delta G_i(z) = \sigma_i^{aq \to lip}(z)\,ASA_{Aq} = [\sigma_i^{wat \to lip}(z) - \sigma_i^{wat \to aq}(z)]\,ASA_{Aq} \tag{31}$$

where $\sigma_i^{aq \to lip}$ is an atomic solvation parameter describing transfer of protein atom i from aqueous to the lipid solution, and ASA_{Aq} is an area occupied by a molecule of an aqueous component at the surface of protein atom i. The average surface area occupied by a water molecule, ASA_{Aq}, was estimated 14 Å, based on the hydration numbers and ASA of ammonium, guanidine and carboxylete ions[31–33].

Each atomic solvation parameter was described by an average value for the lipid-bound and aqueous solution-bound states, rather than simply using equation (24):

$$\sigma_i^{wat \to bil}(z) = \sigma_i^{wat \to aq}(z) \, x_{SAq}(z) + \sigma_i^{wat \to lip}(z) \, x_{SL}(z)$$

(32)

where mole fractions x_{Aq} and x_{SL} were determined as follows:

$$x_{Aq}(z) = x_W(z) + x_{PCN}(z) + x_R(z)$$

(33)

$$x_{SL}(z_i) = \left[1 + \exp(\Delta G_i / RT) \frac{x_{Aq}(z_i)}{1 - x_{Aq}(z_i)} \right]^{-1}$$

(34)

and where x_W, x_{PCN} and x_R are molar fractions of water, lipid phosphatidyl segment and choline group, respectively.

Transfer energies from bulk water to lipid acyl chain solvent were calculated using equation (24):

$$\sigma_i^{wat \to lip}(z) = \sigma_i^0 - e_i(1/\varepsilon_{bil}(z) - 1/\varepsilon_{wat}) + a_i(\alpha_{lip}(z) - \alpha_{wat}) + b_i(\beta_{lip}(z) - \beta_{wat})$$

(35)

where ε_{bil}, α_{lip} and β_{lip} are calculated using equations (9–12), (13) and (14), respectively.

Transfer energies from bulk water to aqueous solution of phosphatidylcholine groups, were calculated similarly, but assuming no hydrophobic effect in this area $\left(\sigma_i^0 = 0 \right)$ and assuming dielectric constant to be the same as in water:

$$\sigma_i^{wat \to aq}(z) = a_i(\alpha_{aq}(z) - \alpha_{wat}) + b_i(\beta_{aq}(z) - \beta_{wat})$$

(36)

where α_{aq} and β_{aq} are calculated using equations (15) and (16), respectively.

The solvatochromic parameter π^* and dielectric constant describe the long-range electrostatic component of the transfer energy. They depend only on the bulk dielectric properties of the mixture (as described by ε_{bil} or π^*_{bil}), but not on the direct binding of the solvent in the first shell. Therefore, they are not a part of the preferential solvation.

This approach takes into account only interactions of the solute group S_i with both solvents, but it neglects the mutual interactions of the solvents Aq and L. It can be used because the aqueous and non-aqueous (acyl chain) solvents are poorly mixable. The phases could not be chosen as lipid and water because lipid phosphates are strongly hydrated.

Application of the model to proteins

The application of the method to proteins rather than to small molecules is facilitated by the existence of a limited number of standard amino acids, whose group dipole moments and ionization energies can be easily tabulated. The group dipole moments (1.8, 1.4 1.5, 3.7, 4.8 and 2.7 D for COOH, OH, SH, guanidine, imidazole and indole groups, respectively[34–36]) were assigned to the appropriate non-hydrogen atoms (O^η in Tyr, N^ε of Trp, O^γ of Ser and

Thr, N^ε of Lys, and N^δ of Arg, O^δ of Asp and Asn, O^ε of Glu and Gln, N^δ of Asn, N^ε of Gln, N^δ and N^ε of His). The dipole moments of the peptide groups were divided into contributions from NH and C=O groups (1.3 D and 2.3 D, respectively).

Many protein groups are buried from the solvent in the protein interior. They are not expected to significantly contribute to the transfer energy, because they remain in the same environment during transfer of the protein from water to the lipid bilayer. Hence, we took into account only the dipolar energy of groups exposed to the solvent. The individual dipolar contributions in (19) were smoothened to linearly decrease to zero when *ASA* of an atom associated with the dipole moment becomes smaller than the hydration area occupied by a single molecule of water ($ASA_{Aq} \sim 14$ Å, the same as in eq. (31)):

$$\Delta G_{dip,j} = \eta_j^{wat \to bil}(z_j)\mu_j, \text{ if } ASA_j > ASA_{Aq} \tag{37}$$

$$\Delta G_{dip,j} = \eta_j^{wat \to bil}(z_j)\mu_j(ASA_j/ASA_{Aq}), \text{ if } ASA_j \le ASA_{Aq} \tag{38}$$

The smoothening of Born energy for ionized groups (equation (22)) was done similarly:

$$\Delta E_k^{ion} = \frac{166\, e_{Born}}{r_k}[F_{Abe}^{wat} - F_{Abe}(z_k)]\, f_k + \sum_{l=1}^{L_k} \sigma_{l,ion}^{wat \to bil}(z_l)ASA_l, \text{ if } ASA_k > ASA_{Aq} \tag{39}$$

$$\Delta E_k^{ion} = \frac{166\, e_{Born}}{r_k}[F_{Abe}^{wat} - F_{Abe}(z_k)]\, (ASA_k/ASA_{Aq})f_k + \sum_{l=1}^{L_k} \sigma_{l,ion}^{wat \to bil}(z_l)ASA_l \quad \text{if} \quad ASA_k \le ASA_{Aq} \tag{40}$$

where coefficient f_k=0.5 for groups involved in ion pairs, and f_k=1 in all other cases. The value of 0.5 was chosen because an electrostatic self-energy of a dipole formed by an ionic pair is roughly of the same magnitude as the Born energy of a single individual ion[37]. The linear smoothing model should be improved in the future because it underestimates electrostatic energy of cationic groups with small *ASA*.

It has been evaluated that formation of the hydrogen bonds by $O^\gamma H$ groups of sterically constrained Ser or Thr residues in nonpolar environments leads to an energy gain of ~-1.5 kcal/mol[38]. Hence we included contributions of these H-bonds normalized by the hydrogen bonding donor capacity α of the environment:

$$E_{HB} = \sum_{i=1}^{N} E_0(\alpha_{wat} - \alpha_{bil}(z_i))/\alpha_{wat} \tag{41}$$

where E_0 was chosen to provide -1.5 kcal/mol in the middle of membrane. The H-bonds were defined as pairs of non-hydrogen donor and acceptor atoms with distance smaller than 3.5 Å and valence angles A–B…X larger than 90°.

Hydrophobic matching condition for transmembrane proteins

The thickness of a fluid lipid bilayer can adjust to match the hydrophobic thickness of a transmembrane peptide or protein. The corresponding energy cost was described by the simplified expression[21]:

$$E_{lip} = \Delta e_{def} N_{lip} \left| d_0 - d_{eq} \right| \tag{42}$$

where d_0 is intrinsic hydrophobic thickness of undisturbed bilayer, d_{eq} is the actual hydrophobic thickness of the protein and bilayer in the area of their direct contact, N_{lip} is the number of annular lipid molecules interacting with protein, and the constant $\Delta e_{def} = 0.029$ kcal/mol/Å was chosen based on experimental studies of hydrophobic mismatch (average value of $\Delta g_{mis} = 0.08 k_B T \ nm^{-2}$ [21]).

The values of d_0 were taken as average hydrophobic thicknesses of inner membranes and outer bacterial membranes: 30 and 23.6 Å, respectively, as estimated previously[22]. An additional ± 1Å deviation from the d_0 value was allowed without any energy penalty to account for a possible variation in the thickness of membranes in different organisms and cells. The numbers of annular lipids were estimated by calculating the transmembrane perimeter of a protein and dividing it by the width of an area occupied by a lipid molecule at the surface of the protein (it was found to be ~6 Å to reproduce the numbers of annular lipids for TM proteins[21]).

The equilibrium value of the thickness d_{eq} was determined for each individual transmembrane protein by minimizing total transfer energy of the protein (equation (19)) with added E_{lip}. Incorporation of the hydrophobic mismatch did not significantly change the calculated hydrophobic thickness for a vast majority of TM proteins (the differences were usually within the $\pm$ deviations observed in energy interval of 1 kcal/mol, as provided in the OPM database). This could be expected because the additional energy penalty (equation 42) is relatively small: 0.3 kcal mol^{-1}Å^{-1} for a single-helical protein (11 annular lipids) and 0.7 kcal mol^{-1}Å^{-1} for the seven-helical protein, rhodopsin (25 annular lipids). Larger changes in the hydrophobic thickness (of ~2–4 Å) were observed for single-helical or double-helical peptides and a few poor quality structures with missing helix ends and loops.

Optimization of spatial positions of molecules in membrane

The original version of our PPM program was modified to optimize positions of molecules in the lipid bilayer with the new solvation model. The computation requires only energy functions and parameters described in equations (1–42) and Tables S1, S3, S4. A solute molecule was considered as a rigid body whose spatial position was defined by three independent variables: two rotation angles and one translation along the bilayer normal (φ, τ and d, respectively). Transfer energy (equation 19) was optimized by combining grid scan and Davidon-Fletcher-Powell method for local energy minimization. Derivatives of transfer energy with respect to rigid body variables of the molecule were analytically calculated, as described previously[22]. Derivatives of energy with respect to z were calculated as finite differences with step of 0.01 Å. The hydrophobic thickness of TM proteins was optimized by grid scan for location of the hydrocarbon boundary (Z_{HDC} in eq. 2) with step of 0.05 Å. All other peaks of the lipid and water distributions were shifted accordingly during the optimization to remain at the same distance from the hydrocarbon boundary as in DOPC.

The program uses as input only a set of coordinate files in the PDB format. Unlike the previous version, it allows an automatic determination of transmembrane secondary structures without using any external software. The dipole moments and standard pKa

values of different groups are included in the library of amino acid residues or directly in the PDB files for small molecules.

Calculation of energy profiles and interfacial scale for amino acid residues

The transbilayer energy profiles of amino acid residues were calculated by placing each amino-acid into the α-helix or β-barrel that was arranged perpendicularly to the membrane and moved along the membrane normal with step of 1 Å. The backbone structures of the α-helix or a β-barrel were taken from crystal structures of membrane proteins (1kzu and 1qjp PDB files, respectively). All residues surrounding the "probe" residue in α-helix or β-barrel were replaced by alanine. Side-chain conformers of the probe residue that are sterically allowed in α-helix or β-sheet were taken from a standard rotamer library and minimized with CHARMM implemented in Quanta (Accelrys Inc.). Conformers with the lowest transfer energy were automatically selected for each location along the bilayer normal during calculations of energy profiles. The conformers are not the same in different points of the profile due to "snorkeling" of polar residues to the regions with a higher concentration of water and preferential burial of nonpolar residues in the hydrocarbon core. Only side-chain atoms of the probe residue were included in the calculation of transfer energy. The locations of hydrocarbon core boundary Z_{HDC} was adjusted to 30 and 23 Å, as corresponding to average hydrophobic thicknesses of α-helical and β-sheet proteins, respectively. The curves were symmetrized by averaging energies obtained for the positive and negative values of z because these energies are dependent on the overall (N- to C- terminus) orientation of the helix with respect to the membrane.

To make a comparison with the Wimley-White interfacial scale, the optimal water-membrane positions and transfer energy were calculated for N-methylacetamides of amino acids. The N-methylacetamides were taken in extended and folded conformations (φ= −120°, ψ=−120° and φ=−60°, ψ=−60°), with different side-chain conformers (χ^1=±60 and 180°), and the conformation with the lowest transfer energy was selected after optimization of the molecule as a rigid body using OPM 2.1. The α-helical conformation of Arg-rich S4 peptide was taken from the corresponding X-ray structure (PDB ID: 2kyh, residues 113–130) and was modified to include residues 113–114 in the helix.

Calculation of statistical energy profiles of amino-acid residues

Statistical energies were derived from distributions along the membrane normal of solvent exposed residues in 3D structures of 119 α-helical and 53 β-barrel transmembrane proteins from the OPM database that were oriented relative to the membrane plane by PPM 1.0[22] (lists of proteins are provided in Supplementary materials). We did not include residues that are buried from the lipid in the protein interior, because the statistical preferences for such residues are dependent on the solid-state intra-protein van der Waals interactions and hydrogen bonds, rather than on protein-lipid interactions of interest. The number of transmembrane protein structures was relatively small because we use structures of only non-homologous or remotely homologous proteins (with sequence identity lower than 40%) to avoid a systematic statistical bias due to evolutionarily relatedness of the proteins. A majority of lipid-facing residues are replaced in homologous proteins with identity of ~40% because of the much higher sequence variability of solvent-exposed residues. We also excluded proteins whose hydrocarbon core boundaries were poorly defined, as followed from calculations with PPM 1.0.

Statistical energy of an amino-acid residue in a slice [z−δ; z+δ] was calculated by analogy to the partition coefficient, but using surface rather than volume concentrations of residues in structures of α-helical or β-sheet transmembrane proteins:

$$\Delta G_i^{stat}(z) = -RT \ln C_i(z)/C_{i,water} \tag{43}$$

where $C_i(z)$ is an average surface concentration of the lipid-facing residue type i in the slice with coordinate $z \pm 0.5$Å, and $C_{i,water}$ is an average surface concentration of the residue type i facing the water, i.e. located beyond 25 Å distance from the membrane center. The sets of lipid-facing residues in transmembrane proteins (i.e. excluding water-facing residues in transmembrane channels and funnels) were calculated as previously described[22].

The surface concentration of a residue in the slice was determined by averaging ASA of the residue (side-chain only) in a set of proteins:

$$C_i(z) = ASA_i(z)/ASA_{total}(z) \tag{44}$$

where $ASA_i(z)$ is ASA of all side-chain atoms of residue in the slice, and $ASA_{total}(z)$ are total ASA of the slice, respectively, calculated for the set of proteins.

Reference surface concentration of the same residue in water was calculated similarly

$$C_{i,wat} = ASA_{i,wat}/ASA_{total,wat} \tag{45}$$

Parameters of permeating water in biological membranes ($\phi_{W,outer}$, $\phi_{W,mid}$ and λ in equation (5)) were optimized by minimizing the root mean square deviation of statistical and physical energy profiles calculated by PPM 2.0 for Tyr and Trp residues (all points from -30 to $+30$ Å with step of 1Å were taken simultaneously for both residues). Physical energy profiles were calculated for residues in individual α-helix moving across the membrane, as described above. These profiles were normalized to account for the difference in residue accessibility between polytopic proteins included in statistical analysis and an individual α-helix used for calculation of physical profiles, as explained in Results.

RESULTS

Polarity profiles of DOPC bilayer

The lipid bilayer was considered as an anisotropic fluid whose physico-chemical properties change along the membrane normal (Figure 1). These properties are described by profiles of the commonly used polarity parameters that are required for the universal solvation model: hydrogen bonding donor (α) and acceptor (β) capacity, solvatochromic dipolarity parameter (π*), and Block-Walker dielectric function $F_{BW}(\varepsilon)$ for the lipid-water mixture. The profiles of these parameters (Figure 1B) were calculated for DOPC and DOPS bilayers from published experimental distributions of lipid fragments and water[19, 39, 40] (Figure 1A) as described in Methods.

According to the published ESR studies[20, 39], the concentration of water quickly drops to 1–2.9 M in the carbonyl region ($z=14.8$ Å) and sharply decreases again to 0.1–0.4 M in the lipid double bond region ($z_0=8.2$ Å) of the DOPC bilayer (Figure 1A), as described by equation (5). The middle point of $z_0=8.2$ Å is close to the maximum of lipid double bond distribution (9.6 Å) that represents the natural boundary between the more ordered region with the higher concentration of defects for water and a more disordered, liquid-like region[41].

To account for the preferential solvation, we treated the lipid bilayer as a binary mixture of polar aqueous and nonpolar lipid phases (see Methods). Polar phase represents aqueous solution of charged lipid segments including phosphate and choline groups. Nonpolar phase represents lipid acyl chains with glycerol/carbonyl segment (Figure 1B). Parameters α and β, which define solute-solvent interactions in the first shell were calculated separately for the phases (α_{aq}, β_{aq} and α_{lip}, β_{lip}, respectively, see equations 13–18). The dielectric parameters (π^* and ε), which define long-range electrostatic interactions, were calculated for the whole mixture using equations (8–12).

The profiles of polarity parameters (α_{aq}, β_{aq}, π_{bil}^*, *and* ε_{bil}) and Block-Walker dielectric function $F_{BW}(\varepsilon_{bil})$ demonstrate two sharp transition regions: one at the hydrocarbon core boundary (14.8 Å from the interior), and another in the lipid double bond region (8.2 Å from the interior). This behavior of polarity parameters follows the distribution of water inside the DOPC bilayer. Parameter β_{aq}, which defines the hydrogen-bonding acceptor capacity of the aqueous phase, has an additional maximum at ~20 Å from the membrane center, due to the presence of phosphate groups. Parameter β_{lip}, which represents hydrogen-bonding acceptor capacity of the lipid phase, shows a maximum at the hydrocarbon core boundary due to the presence of lipid carbonyl groups. The shape of polarity profiles indicates the existence of three major regions: the polar head group region (i.e. ~15–30 Å from the interior), the "mid-polar"[10] region (~8–15 Å from the interior) and the innermost nonpolar membrane core (0–8 Å).

Comparison of results for DOPC or DOPS bilayers shows that polarity parameters depend on the lipid composition. For example, DOPS is characterized by higher values of β_{aq} (it is larger by ~20% at 20 Å distance, see Figure S1) due to the presence of serine carboxyl groups in DOPS.

The profiles of hydrogen bonding capacities and dielectric functions (Figure 1B) were used to calculate atomic solvation parameters $\sigma_i(z)$, which define transfer energy for each of ten atom types, and dipolar solvation parameter $\eta(z)$, which defines the energy of transferring an electric dipole moment of $1D$ from water (Figure 2).

Solvation parameter curves have two transition regions at ~15 Å and ~8 Å from the bilayer center. Further, parameters for OH and N^+ groups show minima at ~20 Å, due to the formation of H-bonds between these groups and lipid phosphate or carbonyl groups. Subsequent calculations of transfer energy (equation 19) are based on these profiles.

Energy profiles of amino-acid residues in DOPC bilayer

To verify the model, we calculated transfer energies of 18 residues (excluding Pro and Gly) incorporated into a poly-Ala α-helix moving along the membrane normal (Figures 3, 4). The profiles of calculated transfer (i.e. physical) energies were compared with statistical energies derived from distributions of corresponding residues in 3D structures of α-helical membrane proteins (Figure 5).

The analysis of energy profiles helped us to evaluate the role of permeating water in the ordered part of the lipid acyl chain region, the significance of the preferential solvation, and the performance of alternative dielectric models (Figure S2). Neglecting the presence of water in the lipid bilayer resulted in very narrow low-energy areas for Trp and Tyr in the "mid-polar" region, inconsistent with much wider statistical energy curves. Ignoring the preferential solvation led to a significant increase of transfer energies for polar and especially negatively charged residues, which also disagrees with statistical profiles. The Block-Walker dielectric model provided wider low-energy areas for Tyr and Trp residues than π^*-based model, which is more consistent with statistical distributions. Therefore, we

selected the Block-Walker dielectric model for describing electrostatic transfer energy of dipoles (equations 26, 27) and included the preferential solvation (equations 28–36) and water distribution parameters determined by spin probes (equation (5)) for all subsequent calculations.

Calculated energy profiles of nonpolar residues (Val, Leu, Ile, Phe, Ala, Met) (Figure 3A) have wide low-energy areas encompassing the whole lipid acyl chain region, while curves for polar (Cys, Ser, Thr, Asn, Gln, His), and charged (Lys, Arg, Asp, Glu) residues have wide plateaus in this region (Figure 3 B, C). All residues with significant dipole moments (His, Ser, Thr, Cys) have shoulders in the "mid-polar" region. Residues bearing OH and N^+ groups (Ser, Thr, Tyr, Arg, Lys) have additional shallow minima in the lipid head group region due to their H-bonding with lipid phosphate and carbonyl groups.

Patterns for Tyr and Trp residues are more complex, with broad minima in the "mid-polar" region (Figure 3D) and maxima in the central bilayer region. More specifically, the curve for Tyr residues has well-pronounced minima (−1.83 kcal/mol) at 14 Å, a shoulder at 8 Å, and a wide area with positive transfer energy (+0.94 kcal/mol) in the middle (0–6 Å from the interior). The curve for Trp residue has wider negative energy areas between 8 and 15 Å from the bilayer center. The minima for Trp and Tyr are in agreement with experimental studies that show preferential location of these residues at the lipid-water interfaces[42–44]. These minima arise in part due to the presence of 1–2 M water in the "mid-polar" region.

Similar calculations were performed for residues in α-helix inserted into DOPS bilayer. Results for DOPS slightly deviate from those for DOPC. The most significant differences were obtained for residues bearing N^+ (Lys, Arg) and OH-groups (Tyr, Ser) (Figure S3). Energy profiles for these residues show deeper minima in the head group region due to additional H-bonding interactions with carboxyls of phosphatidylserine in DOPS bilayer.

In order to better understand the influence of protein structural context on solvation energy, we compared the water-to-bilayer transfer energies of amino-acid residues in three different systems: (1) side-chain analogs; (2) amino-acid residues incorporated in a single transmembrane α-helix; (3) amino-acid residues incorporated in a hydrophobic β-barrel. Curves obtained for residues incorporated into the β-barrel are shown on Figures 4, S3, and S4.

Comparison of these three systems demonstrate that profiles of transfer energy for the same amino-acid residue follow a similar pattern, but show different absolute values of energies in each system (Figures 4, S4). The absolute values of transfer free energy depend on exposure of the corresponding residue to the solvent, which is different for isolated side chain analogs and residues in α-helix or in β-barrel. For example, the minimal absolute value of transfer energy of Trp side-chain analogue, 3-methylindole, was approximately 1.5 and 2.4 fold larger than for Trp residues in isolated α-helix and β-barrel, respectively, which is roughly proportional to their ASA of 293, 176 and 130 $Å^2$, respectively (Figure 4).

Comparison with statistical distributions of residues in transmembrane proteins

To further evaluate our model, we made a more detailed comparison of physical energy profiles, which were calculated for amino-acid residues in a transmembrane α-helix or a β-barrel, with statistical energies derived from distributions of solvent-exposed residues in α-helical and β-barrel transmembrane proteins, as described in Methods (Figures 5, 6 and S5).

The comparison of energy values in the middle of membrane shows that physical energies are approximately twice as large as statistical energies (Figure 6). This result may be attributed to the smaller accessibility of residues to solvent in polytopic proteins, compared

to that in an isolated α-helix. To verify this, we calculated average lipid-accessible ASA for different types of residues in a set of 119 transmembrane α-helical proteins. Indeed, the average ASA of residues in proteins were between 30 and 50% of that in an isolated α-helix. Hence, in order to compare statistical and physical energies, the later must be adjusted by multiplying them by a normalizing coefficient of ~0.5 (k=0.48 in Figure 6A) that accounts for the difference in the lipid-exposed areas (similar to that in Figure 4). It is noteworthy that the obtained statistical and normalized physical scales are comparable to the biological hydrophobicity scale (Figure 6B) that was obtained using translocon-assisted insertion of hydrophobic peptides into endoplasmic reticulum membranes (see Discussion)[45].

The normalized physical energy profiles reproduce the overall shape and the most important features of statistical energy profiles for the corresponding amino acids in α-helical and β-barrel proteins (Figures 5 and S5). These features include the negative transfer energies for nonpolar residues, the positive energies for polar and charged residues, the pronounced energy minima for Trp and Tyr in the "mid-polar" region, and the presence of two transition regions at ~15 and ~9 Å the membrane center for amphiphilic and polar residues. The internal, ~9 Å transition appears as a shoulder in statistical energy profiles of uncharged polar residues (Figure 5B), which is less pronounced for charged residues (Figure 5C) in part due to insufficient statistics.

At the same time, the low-energy area of Tyr and Trp calculated in DOPC were narrower and more outwardly shifted than minima of statistical energy profiles. This could be due to the different lipid and protein compositions of biological membranes and the model DOPC bilayer. Hence, we attempted to refine the parameters of implicit solvation model for biological membranes by fitting the profiles of statistical and physical energies for Tyr and Trp residues. These residues were chosen as molecular probes for the following reasons: (1) shape of their curves is highly complex and, therefore, most informative; (2) their ASA are fairly constant along the bilayer normal; and (3) these residues are involved in all types of solute-solvent interactions, including hydrophobic and H-bonding interactions in the first solvation shell and long-range electrostatics of their dipoles.

We found that energy profiles for Tyr and Trp residues were sensitive to distribution of water inside the lipid bilayer (equation (5)). Fitting statistical and normalized physical energy profiles for residues in α-helices led to the following parameters in equation (5): $\phi_{W,outer}$=0.066, $\phi_{W,mid}$=0.010, z_o=9.0 Å and λ=1.1 Å, assuming that inner and outer leaflets of membrane are symmetric. The optimal parameters for transmembrane β-barrel proteins were slightly different, suggesting a higher water concentration in outer bacterial membrane (Table S4), although the fit was relatively poor (rmsd of 0.33 kcal/mol), as compared to α-helical proteins (rmsd 0.21 kcal/mol) due to smaller statistics for β-barrel proteins. The parameters of water distribution obtained by fitting for Tyr and Trp residues were used for calculation of energy profiles for the remaining residues.

The adjustment of energy profiles for Tyr and Trp residues indicates a higher concentration of water in biological membranes bearing α-helical proteins (3.66 M in the "mid-polar" region and 0.55 M in the central region) as compared to that in DOPC (2.05 M and 0.4 M, respectively) and more gradual changes of water concentration between the regions (λ=1.1 Å in natural membranes *versus* λ=0.55 Å DOPC). Dimensions of different bilayer regions in natural membranes and DOPC are rather similar: they have hydrophobic semi-thicknesses of 15 and 14.8 Å, respectively, with semi-widths of the central nonpolar region (z_0) of 9 Å and 8.6, respectively, and widths of the "mid-polar" region of 6 and 6.2 Å, respectively (Table S4). The fitting shows that water concentration in outer bacterial membranes may be higher than in inner membranes (5.33 M in the "mid-polar" region and 0.55 M in the central region).

Further analysis shows that small differences of statistical and physical energies remain even after optimizing the concentration of water (Figure 5). These discrepancies are related to statistical biases not described by our model. Indeed, the statistical energies of Lys and Arg residues are lower and energies of Asp and Glu are higher than physical energies in the lipid head group region at the inner, but not at the outer side of membrane (Figure 5C). These differences reflect the increased amount of positively charged residues and the depletion of negatively charged residues in the inner leaflet of cellular membranes, known as the "positive inside" bias caused by topogenic signals and interactions of these residues with membrane lipids and the translocon machinery[8, 46]. The energy minima for aromatic residues (Phe, Tyr and Trp) are slightly wider on statistical energy curves than on physical energy curves in 15–20 Å region (Figure 5D). This might be due to stabilizing cation-πinteractions or other factors omitted in our model. For example, the calculations did not reproduce the higher occurrence of aromatic residues in β-barrel proteins in the inner leaflet of the outer membrane (Figure S5, A), which may be related to their participation in sequence motifs (e.g. YxF motif at C-terminus) essential for recognition of β-barrels by chaperones assisting translocation and membrane insertion[47]. However, all such energy preferences are dependent on the lipid composition and specific protein-protein interactions, which lies beyond the scope of the present work.

Binding and spatial locations of small molecules in phospholipid bilayers

The proposed anisotropic solvent model of the lipid bilayer was integrated in a new version of our PPM program for calculation of membrane binding energy and spatial arrangement of proteins or small molecules in membranes. Such application is a sensitive test to identify the existing limitations and deficiencies of the method. Thus, we calculated binding energies and preferred localizations in a phospholipid bilayer for a diverse set of small molecules, including polar, nonpolar, neutral and charged compounds, whose partitioning between water and bilayers or positioning in membranes have been experimentally determined (Table S5). However, to avoid analysis of conformational equilibrium, we selected a set of conformationally rigid organic molecules composed of N, C, O, S, H atoms.

Experimental binding affinities of small molecules were reproduced with correlation coefficient R^2=0.65 and rmse of 0.74 kcal/mol (Figure 7A). The performance of PPM 2.0 for small molecules should be further improved by additional parameterization and by including other factors that are currently missing in the model, such as immobilization energy, electrostatic components of ionic interactions, or steric factor[48].

As an additional test, we calculated spatial positions and orientations in membrane for another set of amphiphils, including indole analogs and three drugs (imipramine, nicotine and caffeine), whose interactions with phopsholipid bicelles have been studied by NMR[49–51]. Our calculations show that all such molecules are located near the lipid carbonyl area, with nonpolar aromatic ring penetrating to the lipid acyl chain region and their positively charged groups interacting with lipid phosphates (Figure 7B). The overall orientations and positions of the molecules, including the penetration of imipramine and nicotine aromatic rings into the hydrophobic environment, and the weak insertion and binding of caffeine are consistent with NMR studies of these drugs in phospholipid bicelles[50]. The calculated locations of indole analogs in the "mid-polar"/glycerol region are consistent with one of the binding modes observed by ^{1}H and ^{2}H NMR[49]. However, we did not detect the second mode of indole binding near choline groups, as observed in a bimodal distribution found in NMR studies[49]. Such results indicate that our model can properly balance different nonspecific interactions that define locations of small molecules in the interfacial region of lipid bilayer, but the method should be improved by including the cation-π interaction[52] to reproduce the second indole location[49].

Comparison with interfacial energies of amino acid residues

The method was tested by calculating spatial positions and transfer energies of N-methyl-acetamides of nonpolar amino acids, while N-methyl-acetamides of polar amino acids were predicted not to be bound. These energies were compared with membrane binding affinities that have been determined for the hydrophobic pentapeptide Ac-WL-X-LL and tripeptde AXA-O-*t*-Bu in the presence of POPC and DMPC bilayers, respectively[53, 54]. After subtraction of energy of the reference Ala residue, the relative binding energies of Tyr, Trp, Met, and Thr residues were reproduced with rmse of 0.5 kcal/mol and R^2=0.6 (Table S6, Figure 8). However, calculated energies of strongly hydrophobic residues (Ile, Leu, Val, Phe) were overestimated as compared with data for pentapeptides. Better agreement was attained with experimental data for tripeptides. The discrepancy for strongly hydrophobic residues may be explained by a partial hydrophobic collapse of residues in the unfolded peptides in the membrane-bound state, which may be more pronounced for larger peptides. Indeed, NMR studies of the Ala-Phe-Ala-O-*t*-Bu demonstrated the presence of intramolecular NOEs between the aromatic ring of Phe[2], H^β of adjacent Ala and C-terminal tert-butyl group in membrane-bound tripeptide[55].

Membrane binding affinities of peripheral proteins

Our previous evaluation of the performance of the PPM1.0 indicated that even the simplified version of the method was able to predict the membrane binding free energies of peripheral proteins that are anchored to the membranes via patches of solvent-exposed hydrophobic residues. However, the correlation between experimental and PPM1.0-calculated values of transfer free energy was rather poor (R^2=0.47 and rmse of 2.73 kcal/mol). The advanced PPM 2.0 method improved predictions of membrane binding energies (R^2=0.78 and rmse of 1.13 kcal/mol) for the same set of 16 peripheral proteins (Figure 9, Table S7). The agreement between experimental and calculated values was better when calculations were performed with water distribution parameters obtained for natural membranes.

PPM2.0 has recently been applied for the prediction of the effects of mutations of water-exposed hydrophobic residues of human α-tocopherol transfer protein[56]. Good correlations (R^2 from 0.83 to 0.89) were obtained between predicted changes in transfer free energy of mutant proteins and rates of protein binding to phospholipid bilayers, as well as with the relative rates of tocopherol delivery to PC vesicles.

Spatial positions in membrane of an Arg-rich peptide

We further evaluated our method on Arg-rich peptide [113]LGLFRLVRLLRFLRILLI[130] derived from the S4 helix of KvAP voltage sensor domain of the K-channel of archaebacteria *Aeropyrum pernix*,. Based on published studies, the energetic cost of burial of an Arg-residue in the nonpolar environment should be very high[57], which would prevent insertion of S4-peptide with four arginines into the lipid bilayer. However, studies of isolated S4 peptide in model membranes by solid-state NMR[58] demonstrated the transmembrane orientation of this peptide in α-helical conformation. This peptide also can be readily inserted into the membrane through the translocon machinery[59]. To reconcile the differences between theoretical expectations and the experiment, we calculated the spatial position of this peptide in the lipid bilayer.

The results of calculations performed by PPM 2.0 indicate that S4 peptide can form α-helix and associate with the lipid bilayer either in a surface orientation (ΔG_{transf} equal to −9.5 kcal/mol) or in transmembane orientation (ΔG_{transf} from −9.5 to −14 kcal/mol), depending on the bilayer thickness (Figure 10A). The transmembrane orientation becomes energetically preferred only in a bilayer with hydrophobic thickness smaller than 23.5 Å. The tilt angle with respect to the membrane normal of the peptide in transmembrane orientation varies

from 22 to 40° depending on the bilayer thickness, and it is ~73° in the alternative surface orientation.

In the transmembrane orientation of S4 α-helix, the guanidinium groups of R120 and R123 "snorkel" to the "mid-polar" region of the lipid bilayer which has an increased water concentration (3.66 M, as evaluated for natural membranes). The preferential solvation of guanidinium groups by water in this region significantly reduces the energy penalty for transferring Arg residues to this generally hydrophobic environment. R117 and R126 residues are inserted in the lipid head group regions from the both sides of the bilayer, where they may form ionic bridges with lipid phosphates (Figure 10B).

Our observations agree with results of solid-state NMR studies, which demonstrated that S4 peptide can be inserted into a DMPC bilayer at 40° tilt angle producing local bilayer thinning of ~9 Å[58]. Indeed, 9 Å-thinning of the DMPC bilayer, whose average hydrophobic thickness in fully hydrated fluid state was estimated from X-ray scattering at 25.4 Å[60], would result in a bilayer with a hydrophobic thickness of 16.4 Å. Our estimates indicate that S4 peptide would insert into the membrane at 40° tilt angle at this thickness. On the other hand, our calculations with the hydrophobic mismatch penalty (see Methods) yield the optimal hydrophobic thickness of the peptide at 21±6.8 Å. This thickness corresponds to the insertion of the S4 peptide at 22.5±11.4° tilt angle.

Thus, our results suggest that the relative proportion between two orientations depends on the membrane thickness: surface-bound S4-helices will be more abundant than transmembrane S4-helices for a hydrophobic thickness larger than 23.5 Å and *vice versa*. This is in line with the small energy penalty (~0.5 kcal/mol) obtained for translocon-mediated insertion of S4-helix in endoplasmic reticulum membranes[59], whose hydrophobic thickness was estimated at 27.5 Å[61].

Similar to recent MD simulations of the S4-helix[62], our method was able to predict membrane thinning and to identify the most important interactions of S4 peptide with lipid bilayer components, such as hydration of R120 and R123 by water in the "mid-polar" region and hydrogen bonds of R117 and R126 with lipid phosphates. However, the significant advantage of our method is its high computational efficiency and its ability to operate directly with transfer free energies that can be easily compared with experimental data.

Application of PPM2.0 to membrane proteins from the OPM database

After the successful testing of the developed method for small molecules, peptides and peripheral proteins, we recalculated the spatial positions in membranes of ~1000 membrane-associated peptides and proteins deposited in our OPM database[23]. Membrane polarity profiles were essentially the same as in the DOPC bilayer (Figure 1B) except the increased concentration of permeating water in natural membranes (Table S4).

We compared the newly calculated spatial positions with experimental data for 24 transmembrane and 42 peripheral proteins that were previously used for verification of the PPM 1.0 method[22, 63]. The previously predicted and newly obtained orientations of these proteins in membrane were rather similar and almost equally consistent with experimental data, in part because such data are usually approximate or qualitative. A more detailed comparison with all relevant experimental studies of peptides and proteins is beyond the scope of this work, although the corresponding information is included in comments for individual proteins in the updated OPM database[23].

The results of calculations were improved for proteins with incomplete 3D structures, transmembrane proteins with poorly defined hydrophobic boundaries and weakly bound

peripheral proteins, but remained mostly unchanged in other cases. For example, recalculation of prostaglandin H2 synthase-1 dimer (PDB code 1q4g) resulted in a fully symmetric membrane binding mode, unlike the previously obtained slightly tilted orientation. For several transmembrane proteins, such as diacylglycerol kinase, ferrous-iron efflux pump and acid-sensitive ion channel (PDB codes 2kdc, 2qfi and 3hgc, respectively), PPM 2.0 determined positions of both hydrophobic boundaries, while PPM 1.0 erroneously predicted insertion from one side with penetration depth of more than 20 Å. The hydrophobic thicknesses of several transmembrane proteins, such as ClC-chloride channel and calcium ATPase, increased and became closer to 30 Å. Hydrophobic thicknesses of these proteins were probably underestimated previously. The incorporation of a hydrophobic mismatch penalty allowed us to improve calculations of single transmembrane α-helices, having small and therefore poorly defined hydrophobic boundaries. Furthermore, PPM2.0 was able to predict the experimentally suggested membrane binding modes for some proteins with incomplete 3D structures (outer membrane complex of type IV secretion system, 3jqo) or in closed conformation (oxysterol-binding protein homolog, 1zhx), where PPM 1.0 failed.

A significant progress was observed for positioning of peripheral proteins in membranes, whose binding energies were calculated more accurately with PPM 2.0 (Figure 9). Some proteins reduced their membrane penetration depths or slightly changed orientation due to explicit incorporation of electrostatic energy in a new model. This effect is more pronounced for peptides and proteins bound to membranes via nonregular hydrophobic loops whose main chain peptide groups are exposed to solvent. The stronger anchoring of Tyr and Trp residues also plays a role.

DISCUSSION

Main features of the proposed model

In this work we present the development, testing and application of our improved method for calculation of membrane binding affinities and spatial positions in biological membranes of diverse molecules, ranging from small drug-like compounds and peptides to peripheral and integral membrane proteins. In our approximation, a solute is regarded as a rigid body, and the lipid environment is represented as an anisotropic fluid. The underlying physical model describes only changes in solvation energy, assuming that the internal energy of a solute does not change upon its transfer from water to the lipid bilayer. This solvation model can be further applied for analysis of conformational changes and protein-protein association, when combined with molecular mechanics force fields for describing internal energy of the system[38, 64], but this is beyond the scope of this work.

The underlying methodology includes two novel components: (a) an original universal solvation model for calculating transfer energy of molecules from water to an arbitrary isotropic or anisotropic environment with defined polarity parameters, (b) an anisotropic solvent representation of the lipid bilayer based on published experimental data.

According to the universal solvation model presented in the preceding paper, the transfer energy of a solute includes two terms: (1) an ASA-dependent term that accounts for van der Waals and H-bonding solvent-solute interactions and entropy of solvent molecules in the first solvation shell; and (2) an electrostatic term that describes solvation energy of dipoles and ions, ionic interactions, deionization penalty for charged groups in nonpolar environment. The universal solvation model was carefully parameterized to operate with free energy of solvation (as described in the accompanying paper). This allows us to make direct estimation of transfer free energies of membrane-associated peptides, proteins and small molecules that can be compared with their experimental membrane binding affinities.

The proposed anisotropic solvent model approximates complex properties of the lipid bilayer by profiles of polarity parameters (ϵ, π^*, α and β) that are used in empirical solvation models developed for small molecules[15], as well as for experimental characterization of membrane interfaces[65, 66]. The polarity profiles implemented in the current model (Figure 1B) are based on distributions of quasi-molecular segments of lipids in DOPC (Figure 1A) and DOPS, as determined from X-ray and neutron scattering data[19, 40]. The complex behavior of polarity parameters differs from that in hydrophobic slab models, where the water-lipid interface was characterized by the abrupt dielectric boundary with a smoothing length λ of 0.6 to 2 Å[22, 67–69]. Our new model also accounts for the permeation of water into the lipid bilayer, describes the preferential solvation in the lipid-water mixture, and includes the hydrophobic mismatch for transmembrane proteins[21].

Comparison of polarity profiles calculated for DOPC and DOPS bilayers demonstrated that hydrogen bonding capacity parameters are sensitive to the nature of lipid head group. Dependence of polarity parameters on the lipid composition has been previously demonstrated by spin or fluorescent probes[70, 71].

Properties of phospholipid bilayer

Observed changes of polarity parameters across the DOPC bilayer indicate the existence of three distinct regions: (1) the head group region formed by choline and phosphates (14.8 to 30 Å from the center of membrane), where water concentration changes from 55.5 M to 30 M; (2) the "mid-polar" region (8.2 to 14.8 Å) with octanol-like properties ($\epsilon \sim 11$) and intermediate water concentration (~ 1–3 M) that is formed by lipid carbonyls and ordered acyl chains including the double bonds; and (3) the internal hydrocarbon core region (within ~ 8 Å from the membrane center) with cyclohexane-like properties ($\epsilon \sim 2.5$) and low water concentration (< 0.5 M) that is formed by disordered acyl chains. The noticeable water concentration in the "mid-polar" region of the DOPC blayer was experimentally detected by spin probes[20, 39, 72]. The presence of water in the high density acyl tail/head group regions was supported by MD simulations of phospholipid bilayers[73]. Thus, the water concentration changes in a wide interfacial zone, which combines head groups of the lipids and the ordered part of their acyl chains. This aqueous gradient gives rise to transmembrane polarity profiles that can be detected by physical methods[71, 72] or using statistical analysis of protein structures[74]. A similar structure of the lipid bilayer, with nonpolar, mid-polar and highly polar phosphatidylcholine regions was identified in MD simulations[12].

The dielectric properties of phospholipid bilayers in our model generally agree with estimations based on EPR and fluorescence studies, especially for the lipid carbonyl and acyl chain regions [48, 71]. At the same time, the dielectric constant was assumed to be about the same in the bulk water and in the lipid head group region but abruptly change at the hydrocarbon core boundary (Figure 1B), while the fluorescence data suggest a more gradual decrease of the dielectric constant from $\epsilon \sim 60$ to $\epsilon \sim 30$ over the entire lipid head group region[48, 75]. The relatively high dielectric field in the lipid head group area can be supported by the observation that aqueous solutions of zwitterions have dielectric constants slightly higher than water[25]. In general, the dielectric constant for the electrolyte solution may decrease or increase with rising electrolyte concentration due to the hydration of ions or formation of ion pairs, respectively[76]. Higher dielectric constant in the lipid head group region would facilitate penetration of charged groups inside this region. This agrees with the high occurrence of charged amino-acids in the lipid head group region, where they may form hydrogen bonds and ionic pairs with lipids.

Our evaluation of properties of biological membranes from the distribution of residues in structures of membrane proteins indicates that the hydrophobic thicknesses vary from ~ 28– 30 Å in membranes with α-helical proteins to ~ 22–24 Å in outer bacterial membranes with

β-barrel proteins. Accordingly, widths of the "mid-polar" regions vary from 6.0–6.2 Å to 4–4.3 Å in membranes with α-helical and β-barrel proteins, respectively.

The "mid-polar" region is composed of lipid acyl chains with double bonds including glycerol and carbonyl segments. It is also known as soft polymer region with a relatively high lipid order[73]. The abundance of acyl chains in this region provides hydrophobic interactions, while the 1–3 M of water is sufficient for the preferential hydrogen bonding with polar groups of the protein, which reduces their dehydration penalty in this region. This effect is most pronounced for the charged carboxylate groups of Asp and Glu residues. Furthermore, the electrostatic energy cost for transferring dipolar and ionized groups to this region is relatively small because Block-Walker and Abe dielectric functions reach only half of their maximal values (Figure 1B).

The distinct physicochemical properties of the "mid-polar" region favor the accumulation of Trp and Tyr residues, which was experimentally demonstrated for membrane proteins and amino acid analogues [42, 43, 49, 77]. This can be understood by analyzing the relative contributions of hydrophobic, H-bonding and dipolar interactions of Trp and Tyr residues in different regions of the lipid bilayer (Figure S6). This analysis shows that energy minima of Tyr and Trp residues in the "mid-polar" region originate from the favorable hydrophobic interactions that are only slightly counterbalanced by electrostatic penalty for transferring their dipoles and by the cost of breaking hydrogen bonds between their polar groups and bulk water. The higher energy maximum in the membrane center for Tyr *versus* Trp residue can be attributed to the higher dehydration penalty of the Tyr OH group, which has a larger ASA than NH of the Trp indole.

Any molecule that combines a hydrophobic moiety with non-zero dipole moment is expected to be preferentially distributed in the "mid-polar" region. Indeed, we found that His, Cys and Thr residues have shallow energy minima in this area. The minima for Tyr, Thr and Ser residues are enforced by their hydrogen bonds with lipid carbonyls and phosphates. Amphiphilic drugs and drug-like compounds also tend to accumulate in the "mid-polar" region (Figure 7). This may explain the observed correlation between water-liposome partition coefficients of drugs and the corresponding water-octanol logP values[48].

Intramembrane water distribution

Comparison of transfer energies calculated for residues in α-helix with statistical energies derived from analysis of protein structures (Figures 5, S5) allowed us to refine the distribution of water along the membrane normal. We assessed the water concentrations in the "mid-polar" and the central regions of different biological membranes by fitting the theoretically predicted and statistical energy profiles for Trp and Tyr residues in α-helical or β-barrel proteins. This estimate is rather approximate because we combined α-helical proteins from different types of membranes (archaeal, inner bacterial and eukaryotic) to increase the statistics, whereas β-barrel proteins were taken from outer membranes of different gram-negative bacteria.

The best fit for α-helical proteins was obtained with the average water concentration in the "mid-polar" region of 3.66 M, i.e. almost twice as much as the corresponding values obtained by spin probes for DOPC membrane (2.05 M)[20]. The increased water concentration in biological membranes may be attributed to the higher lipid heterogeneity and to the presence of proteins. Indeed, it was shown by spin probes that modifications of lipid composition or the addition of cholesterol increased the hydration level of spin probes in the region adjacent to the lipid head groups[39, 72]. The presence of peptides or proteins may also induce water permeation into the lipid bilayer, similar to the "dragging" of water by small polar molecules to nonpolar solvents[78]. Water distribution parameters, which we obtained

for membranes bearing β-barrel proteins, indicated even higher water concentration inside outer bacterial membranes (5.33 M in the "mid-polar region and 0.55 M in the central region) (Table S4). This is consistent with neutron scattering studies of outer bacterial membranes of *Pseudomonas aeruginosa*, which found the presence of water in the entire length of the lipopolysaccharide bilayer, including the central core region, unlike that in DOPC bilayer[79].

Whole-residue hydrophobicity scales

The physical and statistical energy profiles obtained in this work (Figures 5 and S5) can be regarded as whole-residue hydrophobicity scales of α-helical or β-barrel proteins defined for any location along the bilayer normal. Similar scales for residues in α-helical proteins have been developed using statistical analysis, MD simulations or *in vivo* experimental studies[45, 80–83]. In order to compare our energy profiles to experimentally-derived hydrophobicity scales[45, 57, 84–86], we used transfer energy values for residues located in the membrane center. There is a good correlation between our physical and statistical scales for both α-helical and β-barrel proteins (R^2 are equal to 0.92 and 0.91, respectively, Figure 6A). The physical scale correlates remarkably well with water/cyclohexane scale for amino acid side-chain analogues[57] ($R^2 = 0.93$), but shows only moderate correlation with Wimley-White (WW) water/octanol scale[84] ($R^2 = 0.73$, Figure 11A). This is expected because the central part of the lipid bilayer can be approximated well by a nonpolar aliphatic solvent, rather than by wet *n*-octanol solution[87, 88], which dissolves up to 2.3 M of water[89]. On the other hand, *n*-octanol can mimic properties of the "mid-polar" region of the lipid bilayer. Indeed, the correlation between WW water/octanol scale[84] and physical scale obtained for the "mid-polar" region (at 12 Å distance from the interior) was much better ($R^2 = 0.82$, slope 0.81; not shown).

A plot of the physical *versus* the statistical scale shows a slop factor of 0.48, while a plot of the physical scale *versus* the water-cyclohexane scale has slop factor of 1.42 (Figures 6A and 11A). The observed difference in slopes may reflect different accessibility of side chains in various systems: ~2 fold smaller ASA of residues in multi-spanning α-helical proteins than in an isolated α-helix and ~1.4 fold larger ASA of side chain analogues than of residues incorporated in an α-helix. Such dependence of transfer free energy of residues on their solvent exposure in different model systems was mentioned above (Figure 4).

We also compared transfer energies obtained in the present work with several biological scales recently developed for different *in-vivo* helix translocation systems: the translation of model proteins with a hydrophobic peptide by Sec61 machinery in mammalian[45, 81] or yeast[85] endoplasmic reticulum membranes, or insertion of an α-helical hairpin in inner membrane of *E. coli* by YidC machinery[86] (Figure 11B). Correlation of the physical scale with different biological scales is reasonably good (R^2 from 0.79 to 0.86), but the slopes vary significantly from 0.1 to 0.68, depending of the *in-vivo* translocation machinery. These differences in slopes may be related to the distinct local environment of the tested hydrophobic peptides inside the translocation systems and to the different interactions of the nascent helix with other transmembrane helices of the growing protein. However, the observed variations may also be related to the differences in data analysis[45, 81].

The physical energy profiles obtained in the present work may be used for prediction of single-spanning (bitopic) α-helical proteins whose helices must be individually stable in the lipid bilayer without interactions with neighboring helices. Furthermore, the normalized physical scales or statistical scales from this work may also be used for prediction of transmembrane helices in polytopic proteins, just as the biological hydrophobicity scale derived from *in vivo* insertion of hydrophobic helixes mediated by translocon machinery[45].

Although the whole-residue scales may be applied for the genome-wide prediction of transmembrane α-helices, the whole-residue approach is inadequate for the correct evaluation of lipid-protein interactions, because it does not account for significant variations of residue accessibilities on the surface of individual protein structures (Figure 4). In contrast, only all-atom representation of proteins and small molecules, as developed in this work, can accurately reproduce their interactions with surrounding water and anisotropic lipid environment.

Computational implementation and validation

The developed theoretical method was integrated in a new version of our program (PPM 2.0) that computes the energetically preferred spatial arrangements of molecules with respect to the lipid bilayer (see Methods). The new method was validated for peripheral proteins, peptides and small molecules whose membrane penetration depth, spatial orientation with respect to the lipid bilayer or membrane binding affinities have been experimentally studied. We found a much better accuracy of PPM 2.0 than PPM 1.0 in prediction of membrane binding energies of small molecules and peripheral proteins (R^2 increased from 0.47 to 0.78 and rmse improved from 2.73 to 1.13 kcal/mol). An especially challenging test was prediction of the membrane binding mode for a short Arg-rich peptide, which requires a precise balance of the hydrophobic, electrostatic and other interactions in the "mid-polar" region of the bilayer that was provided only by the improved method. It is noteworthy that none of other published theoretical methods was validated against experimental membrane binding affinities of small molecules, peptides and proteins or for a large set of peripheral proteins with known spatial positions in membranes.

CONCLUSION

We improved our computational method (PPM 2.0) by developing a significantly more advanced anisotropic solvent model of the lipid bilayer. High efficiency, accuracy, and reliability of this method makes it a unique tool for large-scale computational analysis of binding and preferred spatial arrangement of integral and peripheral proteins, as well as small molecules, including peptides, organic compounds and drugs. Therefore, PPM 2.0 was applied for recalculation of spatial arrangement in membranes of ~1000 membrane-associated proteins and peptides from the OPM database[23].

The current method operates with transverse properties of the lipid bialyer by calculating complex polarity profiles from the known distributions of lipid fragments and water along the membrane normal. However, increasing attention has recently been paid to the lateral structure of lipid bilayer: formation of protein clusters and lipid rafts. The impact of lateral pressure, curvature, charge, lipid order and the lipid domains on the insertion and orientation of proteins in membranes can be also included in the implicit solvent model of the lipid bilayer, but this will be the subject of future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ABBREVIATIONS

DOPC	dioleoylphosphatidylcholine
DOPS	dioleoylphosphatidylserine
ASA	solvent accessible surface area

| **rmse** | root mean square error |
| **PPM** | Positioning of Proteins in Membranes computational method |
| **OPM** | Orientations of Proteins in Membranes database |

Acknowledgments

REFERENCES AND NOTES

1. Grossfield, A. Implicit modeling of membranes. In: Feller, SE., editor. Curents Topics in Membranes. Computational Modeling of Membrane Bilayers. Vol. Vol. 60. Academic Press; 2008. p. 131-157.

2. Marsh D. Protein modulation of lipids, and vice-versa, in membranes. BBA-Biomembranes. 2008; 1778:1545–1575. [PubMed: 18294954]

3. McMahon HT, Gallop JL. Membrane curvature and mechanisms of dynamic cell membrane remodelling. Nature (London). 2005; 438:590–596. [PubMed: 16319878]

4. Zimmerberg J, Kozlov MM. How proteins produce cellular membrane curvature. Nat. Rev. Mol. Cell Bio. 2006; 7:9–19. [PubMed: 16365634]

5. Bogdanov M, Dowhan W. Lipid-assisted protein folding. J. Biol. Chem. 1999; 274:36827–36830. [PubMed: 10601231]

6. Bogdanov M, Xie J, Dowhan W. Lipid-Protein Interactions Drive Membrane Protein Topogenesis in Accordance with the Positive Inside Rule. J. Biol. Chem. 2009; 284:9637–9641. [PubMed: 19074771]

7. White SH, von Heijne G. The machinery of membrane protein assembly. Curr. Opin. Struct. Biol. 2004; 14:397–404. [PubMed: 15313232]

8. Dowhan W, Bogdanov M. Lipid-Dependent Membrane Protein Topogenesis. Ann. Rev. Biochem. 2009; 78:515–540. [PubMed: 19489728]

9. Bos MP, Robert V, Tommassen J. Biogenesis of the gram-negative bacterial outer membrane. Ann. Rev. Microbiol. 2007; 61:191–214. [PubMed: 17506684]

10. Hanshaw RG, Stahelin RV, Smith BD. Noncovalent keystone interactions controlling biomembrane structure. Chem.-Eur. J. 2008; 14:1690–1697.

11. van Meer G, Voelker DR, Feigenson GW. Membrane lipids: where they are and how they behave. Nat. Rev. Mol. Cell Biol. 2008; 9:112–124. [PubMed: 18216768]

12. MacCallum JL, Tieleman DP. Interactions between small molecules and lipid bilayers. Computational Modeling of Membrane Bilayers. 2008; Vol. 60:227–256.

13. Im W, Feig M, Brooks CL. An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins. Biophys. J. 2003; 85:2900–2918. [PubMed: 14581194]

14. Feig M, Brooks CL. Recent advances in the development and application of implicit solvent models in biomolecule simulations. Curr. Opin. Struct. Biol. 2004; 14:217–224. [PubMed: 15093837]

15. Cramer CJ, Truhlar DG. A universal approach to solvation modeling. Acc. Chem. Res. 2008; 41:760–768.

16. Marcus Y. Some thermodynamic aspects of ion transfer. Electrochim. Acta. 1998; 44:91–98.

17. Abraham MH, Zhao YH. Determination of solvation descriptors for ionic species: Hydrogen bond acidity and basicity. J. Org. Chem. 2004; 69:4677–4685. [PubMed: 15230589]

18. Reichardt C. Solvents and solvent effects: An introduction. Org. Process Res. Dev. 2007; 11:105–113.

19. Kucerka N, Nagle JF, Sachs JN, Feller SE, Pencer J, Jackson A, Katsaras J. Lipid bilayer structure determined by the simultaneous analysis of neutron and x-ray scattering data. Biophys. J. 2008; 95:2356–2367. [PubMed: 18502796]

20. Marsh D. Membrane water-penetration profiles from spin labels. Eur. Biophys. J. 2002; 31:559–562. [PubMed: 12602343]

21. Marsh D. Energetics of hydrophobic matching in lipid-protein interactions. Biophys. J. 2008; 94:3996–4013. [PubMed: 18234817]

22. Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI. Positioning of proteins in membranes: A computational approach. Protein Sci. 2006; 15:1318–1333. [PubMed: 16731967]

23. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI. OPM: Orientations of proteins in membranes database. Bioinformatics. 2006; 22:623–625. [PubMed: 16397007]

24. Marsh D, Toniolo C. Polarity dependence of EPR parameters for TOAC and MTSSL spin labels: Correlation with DOXYL spin labels for membrane studies. J. Magn. Reson. 2008; 190:211–221. [PubMed: 18042415]

25. Zhuo KL, Chen YJ, Kang L, Xu SJ, Wang JJ. Dielectric Constants for Binary Amino Acid-Water Solutions from (278.15 to 313.15) K. J. Chem. Eng. Data. 2009; 54:137–141.

26. Marcus Y. The use of chemical probes for the characterization of solvent mixtures.2. Aqueous mixtures. J. Chem. Soc.-Perk. T. 2. 1994; 8:1751–1758.

27. Marcus, Y. Solvent Mixtures. Properties and Selective Solvation. New York, Basel: Marcel Dekker, Inc.; 2002. p. 259

28. Abe T. A modification of the Born equation. J. Phys. Chem. 1986; 90:713–715.

29. Block H, Walker SM. Modification of Onsager theory for a dielectric. Chem. Phys. Lett. 1973; 19:363–364.

30. Marcus Y. On the preferential solvation of drugs and PAHs in binary solvent mixtures. J. Mol. Liq. 2008; 140:61–67.

31. Jorgensen WL, Gao J. Monte-Carlo simulations of the hydration of ammonium and carboxylate ions. J. Phys. Chem. 1986; 90:2174–2182.

32. Leung K, Rempe SB. *Ab initio* molecular dynamics study of formate ion hydration. J. Am. Chem. Soc. 2004; 126:344–351. [PubMed: 14709101]

33. Mason PE, Neilson GW, Enderby JE, Saboungi ML, Dempsey CE, MacKerell AD, Brady JW. The structure of aqueous guanidinium chloride solutions. J. Am. Chem. Soc. 2004; 126:11462–11470. [PubMed: 15366892]

34. Lien EJ, Guo ZR, Li RL, Su CT. Use of dipole-moment as a parameter in drug receptor interaction and quantitative structure activity relationship studies. J. Pharm. Sci. 1982; 71:641–655. [PubMed: 7097526]

35. Li WY, Guo ZR, Lien EJ. Examination of the interrelationship between aliphatic group dipole-moment and polar substituent constants. J. Pharm. Sci. 1984; 73:553–558. [PubMed: 6726644]

36. Spackman MA, Munshi P, Dittrich B. Dipole moment enhancement in molecular crystals from X-ray diffraction data. ChemPhysChem. 2007; 8:2051–2063. [PubMed: 17676648]

37. Israelachvili, JN. Intermolecular and surface forces. London; Orlando, USA: Academic Press; 1992. p. 296

38. Lomize AL, Reibarkh MY, Pogozheva ID. Interatomic potentials and solvation parameters from protein engineering data for buried residues. Protein Sci. 2002; 11:1984–2000. [PubMed: 12142453]

39. Erilov DA, Bartucci R, Guzzi R, Shubin AA, Maryasov AG, Marsh D, Dzuba SA, Sportelli L. Water concentration profiles in membranes measured by ESEEM of spin-labeled lipids. J. Phys. Chem. B. 2005; 109:12003–12013. [PubMed: 16852481]

40. Petrache HI, Tristram-Nagle S, Gawrisch K, Harries D, Parsegian VA, Nagle JF. Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt. Biophys. J. 2004; 86:1574–1586. [PubMed: 14990484]

41. Xiang TX, Xu YH, Anderson BD. The barrier domain for solute permeation varies with lipid bilayer phase structure. J. Membrane Biol. 1998; 165:77–90. [PubMed: 9705984]

42. Yau WM, Wimley WC, Gawrisch K, White SH. The preference of tryptophan for membrane interfaces. Biochemistry. 1998; 37:14713–14718. [PubMed: 9778346]

43. Liu W, Caffrey M. Interactions of tryptophan, tryptophan peptides, and tryptophan alkyl esters at curved membrane interfaces. Biochemistry. 2006; 45:11713–11726. [PubMed: 17002272]

44. Killian JA, von Heijne G. How proteins adapt to a membrane-water interface. Trends Biochem. Sci. 2000; 25:429–434. [PubMed: 10973056]

45. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH, von Heijne G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature (London). 2005; 433:377–381. [PubMed: 15674282]

46. Goder V, Junne T, Spiess M. Sec61p contributes to signal sequence orientation according to the positive-inside rule. Mol. Biol. Cell. 2004; 15:1470–1478. [PubMed: 14668483]

47. Jackups R, Cheng S, Liang J. Sequence motifs and antimotifs in beta-barrel membrane proteins from a genome-wide analysis: The Ala-Tyr dichotomy and chaperone binding motifs. J. Mol. Biol. 2006; 363:611–623. [PubMed: 16973175]

48. Avdeef, A. Absorption and Drug Development: Solubility, Permeability, and Charge State. NJ, USA: John Wiley & Sons; 2003. Partitioning into Liposomes; p. 67-90.

49. Gaede HC, Yau WM, Gawrisch K. Electrostatic contributions to indole-lipid interactions. J. Phys. Chem. B. 2005; 109:13014–13023. [PubMed: 16852615]

50. Evanics F, Prosser RS. Discriminating binding and positioning of amphiphiles to lipid bilayers by H-1 NMR. Anal. Chim. Acta. 2005; 534:21–29.

51. Al-Abdul-Wahid MS, Neale C, Pomes R, Prosser RS. A Solution NMR Approach to the Measurement of Amphiphile Immersion Depth and Orientation in Membrane Model Systems. J. Am. Chem. Soc. 2009; 131:6452–6459. [PubMed: 19415935]

52. Petersen FNR, Jensen MO, Nielsen CH. Interfacial tryptophan residues: A role for the cation-pi effect? Biophys. J. 2005; 89:3985–3996. [PubMed: 16150973]

53. Wimley WC, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat. Struct. Biol. 1996; 3:842–848. [PubMed: 8836100]

54. Jacobs RE, White SH. The nature of the hydrophobic binding of small peptides at the bilayer interface - implications for the insertion of transbilayer helices. Biochemistry. 1989; 28:3421–3437. [PubMed: 2742845]

55. Brown JW, Huestis WH. Structure and orientation of a bilayer-bound model tripeptide - a H-1-NMR study. J. Phys. Chem. 1993; 97:2967–2973.

56. Zhang WX, Thakur V, Lomize A, Pogozheva I, Panagabko C, Cecchini M, Baptist M, Morley S, Manor D, Atkinson J. The Contribution of Surface Residues to Membrane Binding and Ligand Transfer by the α-Tocopherol Transfer Protein (alpha-TTP). J. Mol. Biol. 2011; 405:972–988. [PubMed: 21110980]

57. Radzicka A, Wolfenden R. Comparing the polarities of the amino-acids - side-chain distribution coefficients between the vapor-phase, cyclohexane, 1-octanol, and neutral aqueous-solution. Biochemistry. 1988; 27:1664–1670.

58. Doherty T, Su YC, Hong M. High-Resolution Orientation and Depth of Insertion of the Voltage-Sensing S4 Helix of a Potassium Channel in Lipid Bilayers. J. Mol. Biol. 2010; 401:642–652. [PubMed: 20600109]

59. Hessa T, White SH, von Heijne G. Membrane insertion of a potassium-channel voltage sensor. Science. 2005; 307:1427–1427. [PubMed: 15681341]

60. Kucerka N, Liu YF, Chu NJ, Petrache HI, Tristram-Nagle S, Nagle JF. Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using X-ray scattering from oriented multilamellar arrays and from large unilamellar vesicles. Biophys. J. 2005; 88:245A–245A.

61. Mitra K, Ubarretxena-Belandia I, Taguchi T, Warren G, Engelman DM. Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. Proc. Natl. Acad. Sci. U.S.A. 2004; 101:4083–4088. [PubMed: 15016920]

62. Schow EV, Freites JA, Cheng P, Bernsel A, von Heijne G, White SH, Tobias DJ. Arginine in membranes: the connection between molecular dynamics simulations and translocon-mediated insertion experiments. J. Membr. Biol. 2010; 239:35–48. [PubMed: 21127848]

63. Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI. The role of hydrophobic interactions in positioning of peripheral proteins in membranes. BMC Struct. Biol. 2007; 7:44. [PubMed: 17603894]

64. Lomize AL, Pogozheva ID, Mosberg HI. Quantification of helix-helix binding affinities in micelles and lipid bilayers. Prot. Sci. 2004; 13:2600–2612.

65. Fuguet E, Rafols C, Roses M. Characterization of the solvation properties of surfactants by solvatochromic indicators. Langmuir. 2003; 19:6685–6692.

66. Carrozzino JM, Fuguet E, Helburn R, Khaledi MG. Characterization of small unilamellar vesicles using solvatochromic pi* indicators and particle sizing. J. Biochem. Bioph. Meth. 2004; 60:97–115.

67. Efremov RG, Nolde DE, Vergoten G, Arseniev AS. A solvent model for simulations of peptides in bilayers. I. Membrane-promoting alpha-helix formation. Biophys. J. 1999; 76:2448–2459. [PubMed: 10233062]

68. Lazaridis T, Mallik B, Chen Y. Implicit solvent simulations of DPC micelle formation. Journal of Physical Chemistry B. 2005; 109:15098–15106.

69. Ulmschneider MB, Ulmschneider JP, Sansom MSP, Di Nola A. A generalized Born implicit-membrane representation compared to experimental insertion free energies. Biophys. J. 2007; 92:2338–2349. [PubMed: 17218457]

70. Kurad D, Jeschke G, Marsh D. Lipid membrane polarity profiles by high-field EPR. Biophys. J. 2003; 85:1025–1033. [PubMed: 12885649]

71. Koehorst RBM, Spruijt RB, Vergeldt FJ, Hemminga MA. Lipid bilayer topology of the transmembrane α-helix of M13 major coat protein and bilayer polarity profile by site-directed fluorescence spectroscopy. Biophys. J. 2004; 87:1445–1455. [PubMed: 15345527]

72. Subczynski WK, Widomska J, Feix JB. Physical properties of lipid bilayers from EPR spin labeling and their influence on chemical reactions in a membrane environment. Free Radical Bio. Med. 2009; 46:707–718. [PubMed: 19111611]

73. Marrink SJ, Berendsen HJC. Simulation of water transport through a lipid-membrane. J. Phys. Chem. B. 1994; 98:4155–4168.

74. Chamberlain AK, Lee Y, Kim S, Bowie JU. Snorkeling preferences foster an amino acid composition bias in transmembrane helices. J. Mol. Biol. 2004; 339:471–479. [PubMed: 15136048]

75. Voinov MA, Kirilyuk IA, Smirnov AI. Spin-Labeled pH-Sensitive Phospholipids for Interfacial pK(a) Determination: Synthesis and Characterization in Aqueous and Micellar Solutions. J. Phys. Chem. B. 2009; 113:3453–3460. [PubMed: 19235992]

76. Wang PM, Anderko A. Computation of dielectric constants of solvent mixtures and electrolyte solutions. Fluid Phase Equilibr. 2001; 186:103–122.

77. Hong HD, Park S, Jimenez RHF, Rinehart D, Tamm LK. Role of aromatic side chains in the folding and thermodynamic stability of integral membrane proteins. J. Am. Chem. Soc. 2007; 129:8320–8327. [PubMed: 17564441]

78. Tsai RS, Fan WZ, Eltayar N, Carrupt PA, Testa B, Kier LB. Solute water interactions in the organic-phase of a biphasic system.1. Structural influence of organic solutes on the water-dragging effect. J. Am. Chem. Soc. 1993; 115:9632–9639.

79. Abraham T, Schooling SR, Nieh MP, Kucerka N, Beveridge TJ, Katsaras J. Neutron diffraction study of Pseudomonas aeruginosa lipopolysaccharide bilayers. J. Phys. Chem. B. 2007; 111:2477–2483. [PubMed: 17305384]

80. Ulmschneider MB, Sansom MSP, Di Nola A. Properties of integral membrane protein structures: Derivation of an implicit membrane potential. Proteins Struc. Func.Bioinf. 2005; 59:252–265.

81. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH, von Heijne G. Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature (London). 2007; 450:1026-U2. [PubMed: 18075582]

82. 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. J. Mol. Biol. 2007; 366:436–448. [PubMed: 17174324]

83. MacCallum JL, Bennett WFD, Tieleman DP. Partitioning of amino acid side chains into lipid bilayers: Results from computer simulations and comparison to experiment. J. Gen. Physiol. 2007; 129:371–377. [PubMed: 17438118]

84. Wimley WC, Creamer TP, White SH. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry. 1996; 35:5109–5124. [PubMed: 8611495]

85. Hessa T, Reithinger JH, von Heijne G, Kim H. Analysis of Transmembrane Helix Integration in the Endoplasmic Reticulum in S. cerevisiae. J. Mol. Biol. 2009; 386:1222–1228. [PubMed: 19452628]

86. Xie K, Hessa T, Seppala S, Rapp M, von Heijne G, Dalbey RE. Features of transmembrane segments that promote the lateral release from the translocase into the lipid phase. Biochemistry. 2007; 46:15153–15161. [PubMed: 18052199]

87. Walter A, Gutknecht J. Permeability of small nonelectrolytes through lipid bilayer-membranes. J. Membane Biol. 1986; 90:207–217.

88. Xiang TX, Anderson BD. The relationship between permeant size and permeability in lipid bilayer-membranes. J. Membrane Biol. 1994; 140:111–122. [PubMed: 7932645]

89. Leo A, Hansch C, Elkins D. Partition coefficients and their uses. Chem. Rev. 1971; 71:525–616.
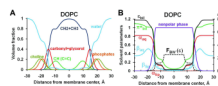
Figure 1.
(**A**) Volume fractions of lipid segments determined by X-ray and neutron scattering for fluid DOPC bilayer[19]. The volume fractions of individual lipid components are shown as colored solid lines for alkyl groups (CH_2+CH_3, purple), double bonds (CH, green), water (blue), carbonyls with glycerols (CG, red), phosphate (orange), and choline (dark green) groups.
(**B**) Profiles of hydrogen bonding donor (α_{aq} and α_{lip}) and acceptor (β_{aq} and β_{lip}) capacities, solvatochromic dipolarity parameter (π^*_{lip}), dielectric constant (ε_{bil}), Block-Walker dielectric function, $F_{BW}(\varepsilon)$, and volume fraction of nonpolar phase ($CH+CH_2+CH_3+CG$) calculated for DOPC bilayer.
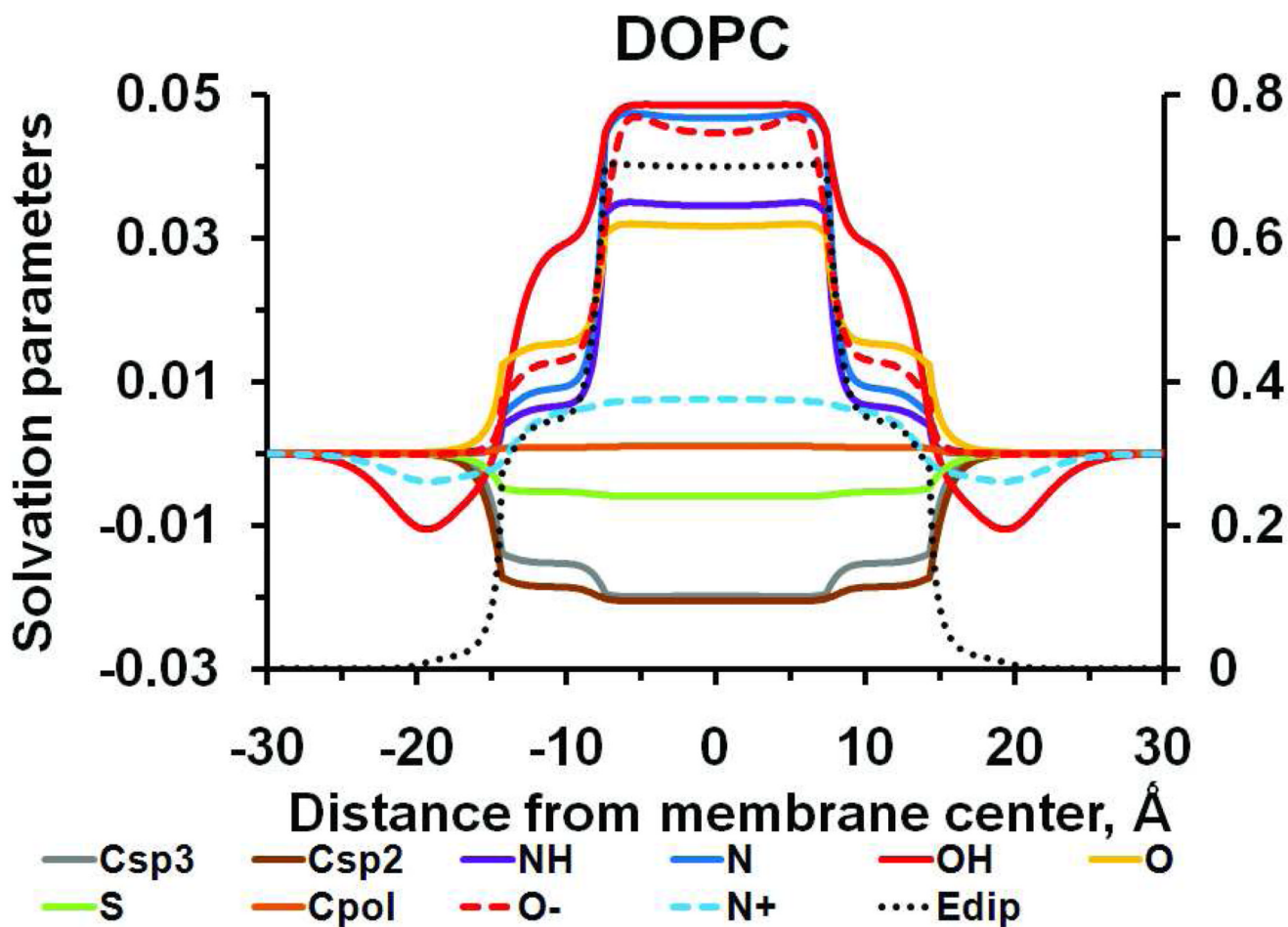
Figure 2.
Transbilayer profiles of atomic solvation parameters, σ_{Csp3} (gray solid line), σ_{Csp2} (brown solid line), σ_{NH} (purple solid), σ_{OH} (red solid line), σ_{O}, (yellow solid line), σ_{S}, (green solid line), σ_{pol}, (orange solid line), σ_{COO-}, (red dashed line), σ_{NH3+} (blue dashed line), and dipolar solvation parameter η for 1D dipole (black dotted line) calculated for DOPC bilayer.
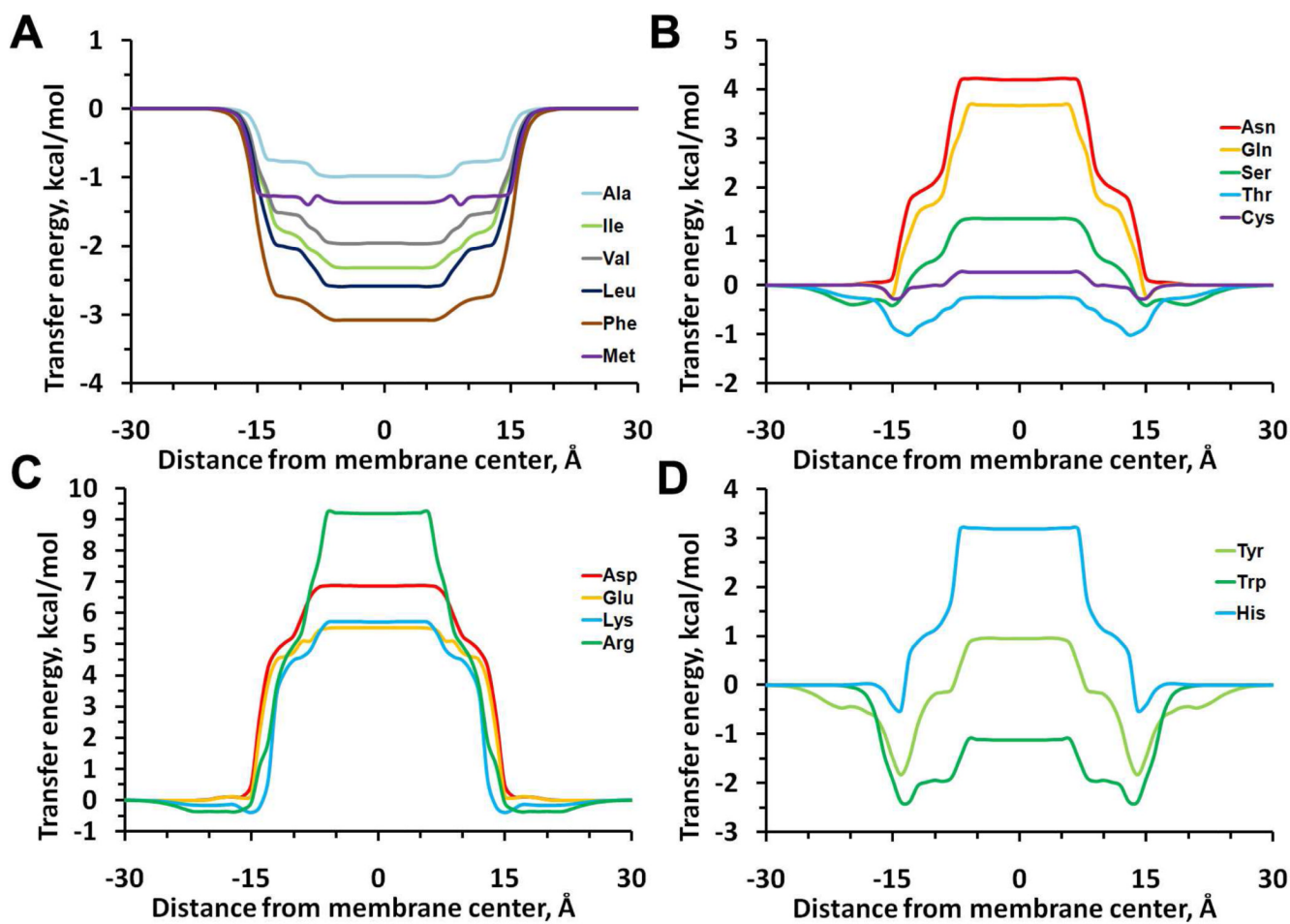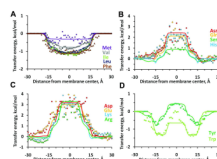
Figure 3.
Transfer energy profiles of 18 amino-acid residues incorporated into an α-helix moving across the DOPC bilayer, as calculated with PPM 2.0.
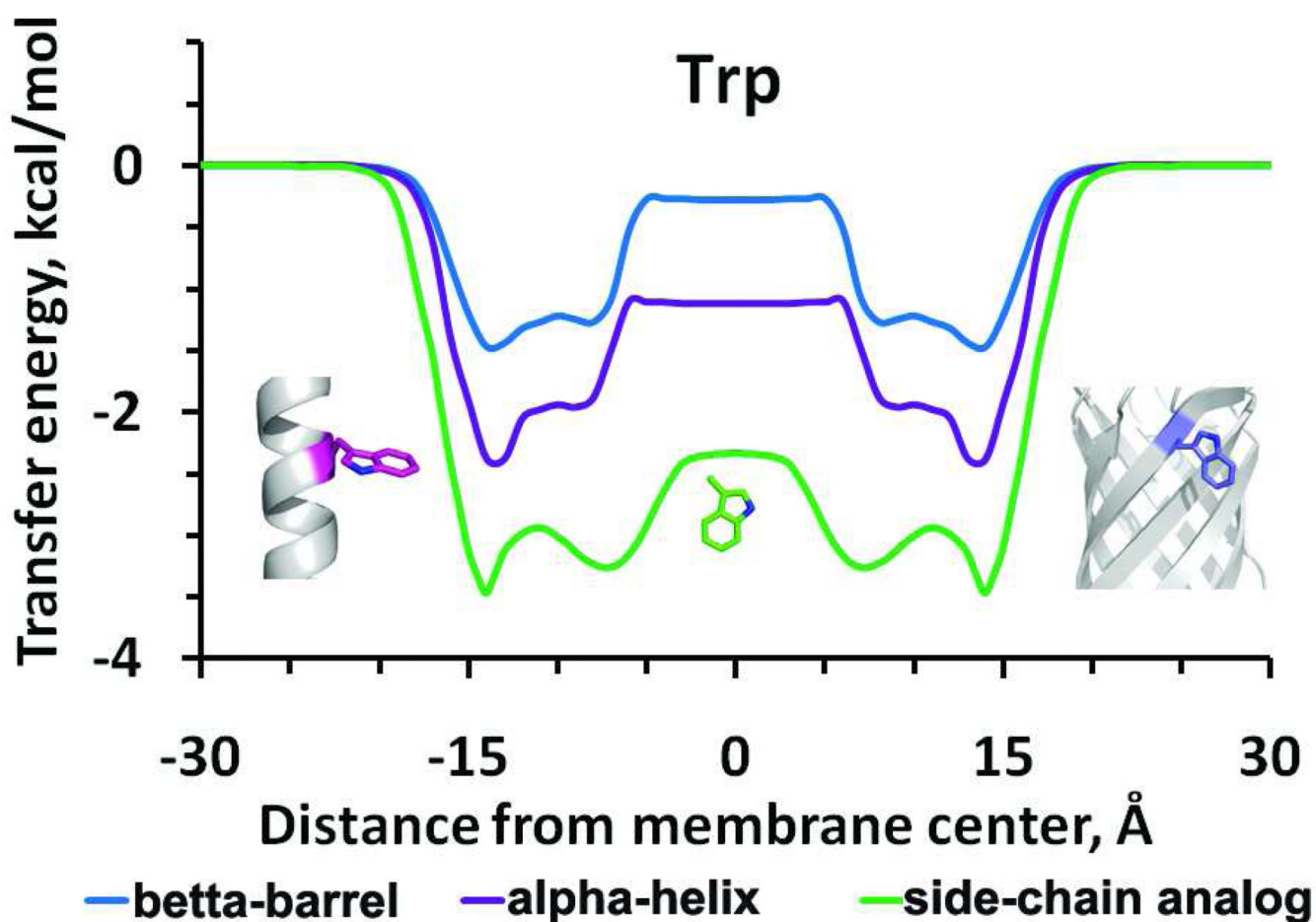
Figure 4.
Transfer energy profiles of Trp in the DOPC bilayer. Energy was calculated for Trp side chain analog, 3-methylindole (green), for Trp in a poly-Ala α-helix (purple), and Trp in a β-barrel (blue). The energies are roughly proportional to solvent-accessible surface area (ASA) of Trp in each environment: ASA = 293 $Å^2$ for Trp side chain analogue, ASA = 176 $Å^2$ for Trp in α-helix and ASA = 130 $Å^2$ for Trp in β-barrel.

Figure 5.
Comparison of transfer energies calculated for 15 amino acid residues in α-helix after optimization of parameters for water (solid lines) and statistical energies (squares, circles, triangles and diamonds) obtained for solvent-exposed residues in a set of 119 α-helical transmembrane proteins from bacterial, archaeal, and eukaryotic membranes from the OPM database. The calculated energies were normalized to account for the difference of ASA for the residue in the set of proteins and in isolated α-helix.
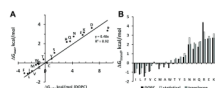
Figure 6.
Comparison of physical, statistical, and biological hydrophobicity scales of amino acid residues from α-helices in the middle of membrane. **A.** Correlation ($R^2=0.92$) between statistical energies derived from distributions of surface residues in 119 α-helical transmembrane proteins (ΔG_{stat}) and transfer energies calculated for residues from an α-helix in DOPC (ΔG_{calc}). The obtained slope (k=0.48) is likely related to the decreased accessibility of residues to lipid in polytopic membrane proteins in comparison to that in a single α-helix. **B**. Comparison of transfer energies: (a) calculated for the residues in an α-helix in DOPC after normalization by k=0.48 (black bars); (b) derived from statistical distributions of residues in transmembrane proteins (light gray bars); and (c) determined from translocon-mediated insertion experiments[45] (dark gray bars).
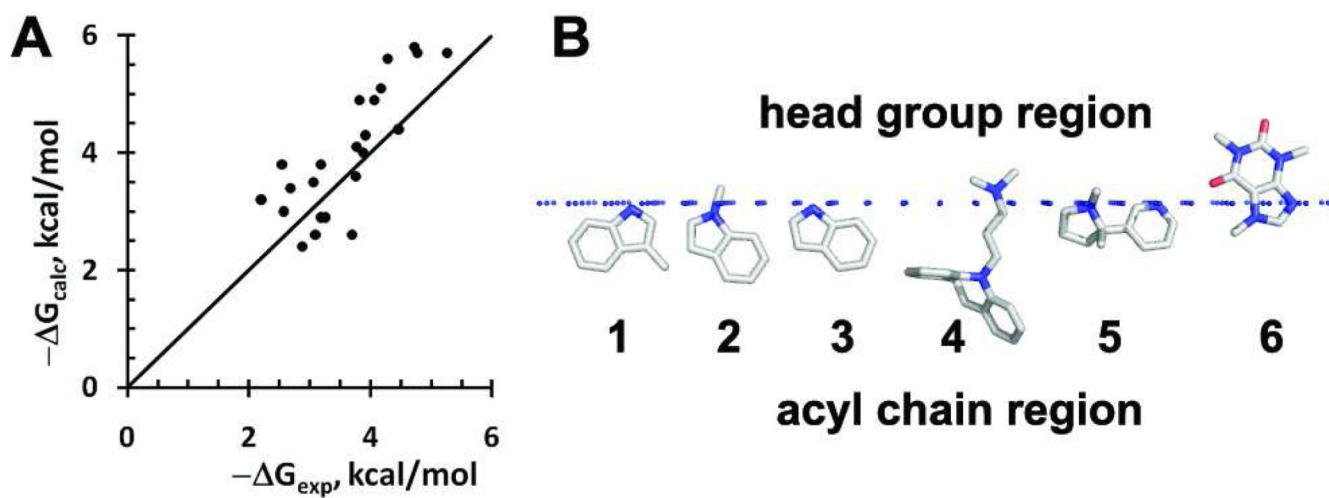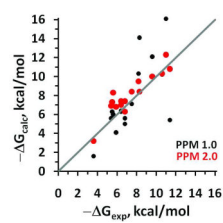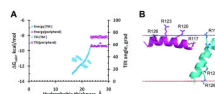
Figure 7.
Membrane binding of small drug-like molecules. **A**. Comparison of experimental and calculated transfer energies of small molecules from water to DOPC bilayer (data from Table S5; R^2=0.65). **B**. Membrane-bound poses of small-molecules: 1: 3-methylindole; 2: N-methylindole; 3: indole; 4: imipramine; 5: nicotine; 6: caffeine. The hydrophobic boundary of the lipid bilayer (corresponding to location of lipid carbonyl groups) is shown by blue dots.
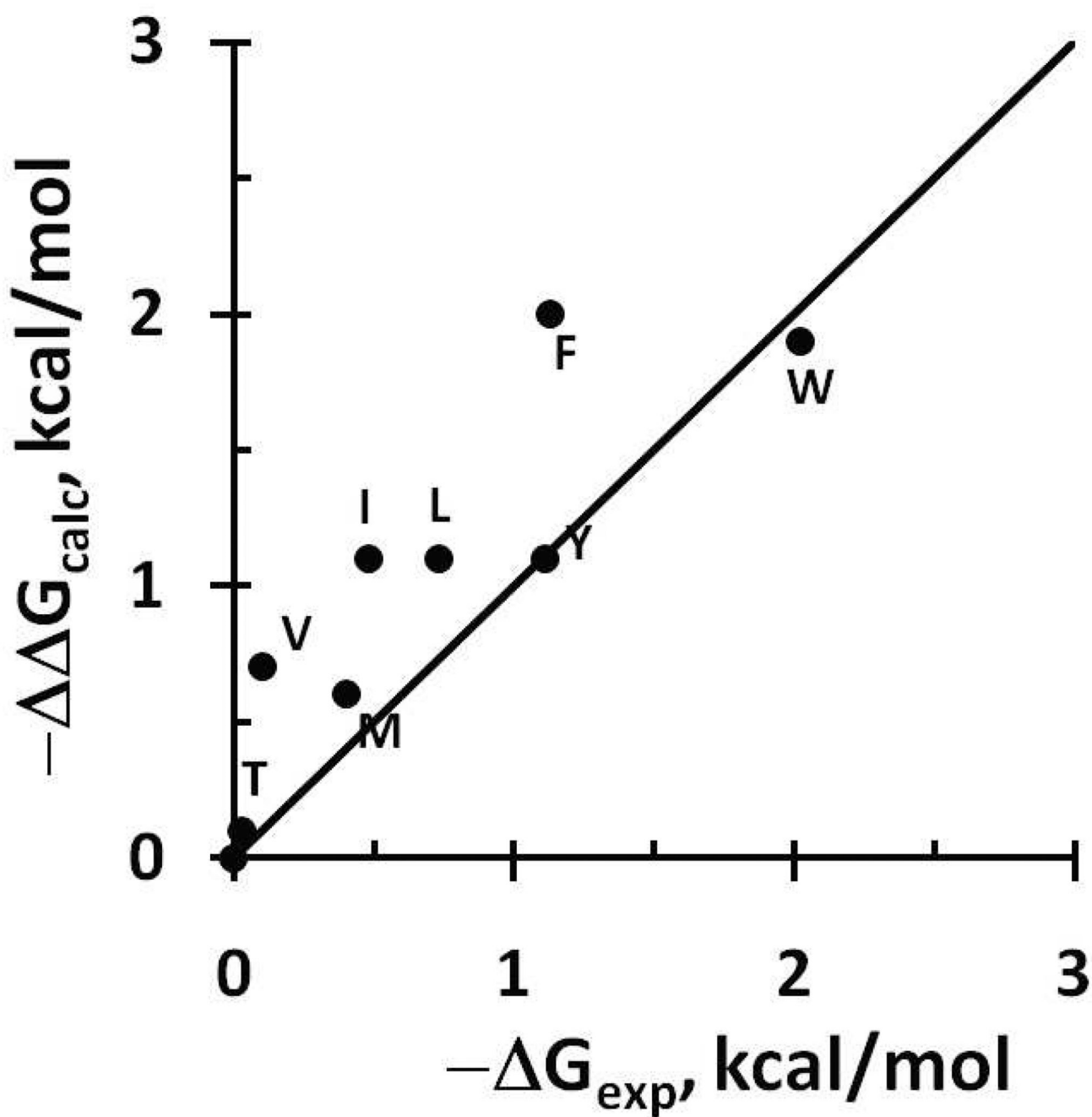
Figure 8.
Comparison of calculated transfer energies of N-methyl acetamides of nonpolar residues and membrane binding energies of the corresponding residues included in host Ac-WL-X-LL pentapeptides[53] (black circles; data from Table S6; R^2=0.63). Energy of the reference peptide with Ala residue was subtracted.
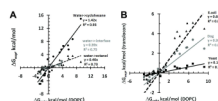
Figure 9.
Correlations of experimental (ΔG_{exp}) and calculated (ΔG_{calc}) free energies of membrane insertion of 17 peripheral proteins (data from Table S7). Transfer energies were calculated by PPM 1.0 (black circles; R^2=0.47) or PPM 2.0 (red circles; R^2=0.78).

Figure 10.
Transfer energies and tilt angles of S4-helix immersed in the lipid bilayer. **A**. Energies (diamonds) and tilt angles (open circles) calculated for different hydrophobic thicknesses of the bilayer for transmembrane (colored blue) and peripheral (colored purple) orientations of the S4 helix. ΔG_{calc} does not include the penalty for hydrophobic mismatch. **B**. Alternative orientations in the lipid bilayer of S4 peptide.

Figure 11.
Correlations between theoretically predicted transfer energies of residues in α-helix from water to the middle of DOPC bilayer (ΔG_{calc}) and published whole-residue hydrophobicity scales. **A**. Comparison of ΔG_{calc} with experimental water/cyclohexane transfer energies of side chain analogues[57] (black circles, black solid line, R^2=0.92) and WW water/octanol scale[84] (grey circles, gray solid line, R^2=0.73). **B**. Correlations between physical scale and different biological scales obtained for mammalian[45] (grey circles, grey solid line, R^2=0.83), yeast[85] (black squares, black solid line, R^2=0.79) or bacterial[86] (black triangles, black dashed line, R^2=0.82) translocation systems.