SI Materials and Methods

Energy Function. The effective energy of a membrane protein is calculated as follows.

$$E_{\text{prot}} = W_{\text{rot}}E_{\text{rot}} + W_{\text{atr}}E_{\text{atr}} + W_{\text{rep}}E_{\text{rep}} + W_{\text{solv}}E_{\text{solv}} + W_{\text{pair}}E_{\text{pair}} + W_{\text{mbenv}}E_{\text{mbenv}} + W_{\text{hbond}}E_{\text{hbond}} - E_{\text{ref}} \quad \textbf{[1]}$$

As described previously (1), the W, weights that scale each energy term, were determined by optimizing the recovery of the native amino acid identities of membrane proteins in a set of 18 high-resolution membrane protein crystal structures (see SI Table 5).

The functional forms of the potential energy terms E_{rot}, E_{atr}, E_{rep}, and E_{pair} describing rotamer self-energies, attractive, repulsive portions of the Lennard–Jones potential energies and knowledge-based electrostatic pair energies, respectively, are identical to those used in the Rosetta full-atom energy function for water-soluble proteins (1).

E_{solv} is the change in solvation energy of an atom upon burial in the protein and is calculated by using the implicit solvent model IMM1, an extension of EEF1 for membrane proteins developed by Lazaridis (2).

$$E_{\text{solv}} = -\sum_{i}^{\text{natom}} \sum_{j>i}^{\text{natom}} \left\{ \frac{2\Delta G_i^{\text{free}}}{4\pi\sqrt{\pi}\lambda_i r_{ij}^2} \exp\left(-d_{ij}^2\right) V_j + \frac{2\Delta G_j^{\text{free}}}{4\pi\sqrt{\pi}\lambda_j r_{ij}^2} \exp\left(-d_{ji}^2\right) V_i \right\} \quad \textbf{[2]}$$

d_{ij} is the distance between two atoms i and j, r_{ij} is the sum of their van der Waals radii, ΔG^{free} is the solvation free energy of the isolated atom i and λ the correlation length. As described below, ΔG^{free} is a function of the position of each atom in the membrane.

E_{mbenv} represents the change in solvation free energy of isolated atoms when transferred from pure water (which is the reference state in our solvation model) to the membrane bilayer at position z along the membrane normal (see description below).

$$E_{\text{mbenv}} = \sum_{i}^{\text{natom}} \Delta G_i^{\text{ref}}(z') \qquad \textbf{[3]}$$

Our membrane model is very similar to that described in IMM1 (2). It represents the membrane bilayer with three different planar phases: two isotropic phases (pure water and pure hydrocarbon for the nonpolar core of the membrane) and one anisotropic phase for the hydrocarbon-polar headgroup interface region. The membrane bilayer is considered to be parallel to the *xy* plane and centered at $z = 0$. The solvation energies of each atom i (ΔG^{free} and ΔG^{ref}) depends on its position z along the membrane normal (vertical z axis) as follows:

$$\begin{aligned} \Delta G_i^{\text{free}}(z') &= f(z')\Delta G_i^{\text{free,water}} + (1 - f(z'))\Delta G_i^{\text{free,chex}} \\ \Delta G_i^{\text{ref}}(z') &= (1 - f(z')) * \left(\Delta G_i^{\text{ref,chex}} - \Delta G_i^{\text{ref,water}}\right) \end{aligned} \qquad \textbf{[4]}$$

As described in ref. 2, the atomic solvation energies for the water phase ($\Delta G^{\text{free,water}}$ and $\Delta G^{\text{ref,water}}$) are derived from transfer free energies of amino acid side-chain analogs from vacuum to water. The atomic solvation energies ($\Delta G^{\text{free,chex}}$ and $\Delta G^{\text{ref,chex}}$) for the membrane core are derived from transfer free energies of amino acid side-chain analogs from vacuum to cyclohexane (chex).

$z' = |z|/(T/2)$ with T corresponding roughly to the thickness of the nonpolar core of the membrane. The function f(z') describes the transition between the two isotropic phases, i.e., the anisotropic phase:

$$f(z') = \frac{z'^{n}}{1 + z'^{n}} \qquad \textbf{[5]}$$

n controls the steepness of the transition and therefore the thickness of the anisotropic phase. Based on native amino acid sequence recovery tests, optimal values for T and n were found to range from 12.5 to 15 Å and from 9 to 11, respectively.

The previously developed ROSETTA hydrogen-bond potential (3) was modified to model the effect of the membrane environment on the strength of the hydrogen bonds. Liquid water has high-dielectric properties. Water molecules are strong hydrogen bond donors and acceptors and can compete with solvent-exposed hydrogen bonds between protein atoms. Therefore, the strength of hydrogen bonds between protein atoms in the water phase depends on the atomic burial in the protein (see Eq. **8**). Unlike water, the acyl chains of the lipids have no polar groups to compete with solvent-exposed hydrogen bonds and the dielectric properties of the solvent in the hydrophobic core of the membrane are close to those of protein interiors. At the center of the membrane, the effect of the solvent on the strength of the hydrogen bonds is therefore negligible and buried or lipid-exposed hydrogen bonds have the same energies in our model. As described by Eq. **7**, the effect of the solvent on the strength of hydrogen bonds in the anisotropic phase is interpolated from the values in water and at the membrane center based on the depth of the donor and acceptor atoms in the membrane.

Equation **6** describes the effective hydrogen-bonding energy $E_{\text{effective}}^{\text{hbond}}$ between a donor *D* and acceptor *A* is:

$$E_{\text{effective}}^{\text{hbond}} = W_{\text{hbond}} * W_{\text{burial}}^{\text{hbond}}\left(z_{DA}^{'}, nb\right) * E_{\text{geo}}^{\text{hbond}}\left(d, \Psi, \Theta\right) \quad \textbf{[6]}$$

W_{hbond} is the global weight optimized by amino acid sequence recovery and $E_{\text{geo}}^{\text{hbond}}\left(d, \Psi, \Theta\right)$ is the geometrical dependent hydrogen-bond energy (see below for further description)

$$W_{\text{burial}}^{\text{hbond}}(z_{DA}^{'}, nb) = f(z_{DA}^{'}) W_{\text{burial,water}}^{\text{hbond}}\left(nb\right) + (1 - f(z_{DA}^{'})) W_{\text{burial,membrane}}^{\text{hbond}} \quad \textbf{[7]}$$

$W_{\text{burial}}^{\text{hbond}}(z_{DA}^{'}, nb)$ represents the weight that scales the hydrogen bond energy based on the atomic burial in the protein [measured by the number of neighbors (*nb*) around the donor (*D*) and acceptor (*A*) atoms] and on the averaged atomic depth of the donor (*D*) and acceptor (*A*) atoms in the membrane z'_{DA}. $W_{\text{burial,water}}^{\text{hbond}}(nb)$ and $W_{\text{burial,membrane}}^{\text{hbond}}$ are the weights

in pure water and at the membrane center, respectively. In our model, $W^{\text{hbond}}_{\text{burial,membrane}}$ is considered to be independent of the atomic burial in the protein and set to its maximal value (0.5). $W^{\text{hbond}}_{\text{burial,water}}$ is a simple function of the number of atom neighbors (*nb*) within 10 Å from the beta carbon of the atom considered and has the following functional form:

$$\Delta G^{\text{hbond}}_{\text{burial,water}} = 0.1 \quad for \quad nb < 7; \quad \Delta G^{\text{hbond}}_{\text{burial,water}} = \frac{(nb - 2.75)}{42.5} \quad for \quad 24 <= nb <= 7;$$
$$\Delta G^{\text{hbond}}_{\text{burial,water}} = 0.5 \quad for \quad 24 < nb \qquad \textbf{[8]}$$

The hydrogen-bond potential was also further developed to explicitly model weak CH—O and bifurcated side-chain/backbone hydrogen bonds that play important roles in inducing helical distortions and stabilizing polar residues in membrane proteins. Bifurcated hydrogen-bonds are non-pairwise factorable interactions (SI Fig. 3, where one oxygen accepts two hydrogens) and are generally weaker by 15–20% than normal hydrogen-bonds (4). An efficient pairwise approximation to these interactions was defined by Eq. **9**:

$$E^{\text{bif}}_{\text{total}} = 0.85 * \left(E^{\text{normal}}_{\text{bb/bb}} + E^{\text{normal}}_{\text{bb/sc}} \right); \quad E^{\text{bif}}_{\text{bb/bb}} = 0.85 * E^{\text{normal}}_{\text{bb/bb}}; \quad E^{\text{bif}}_{\text{sc/bb}} = 0.85 * E^{\text{normal}}_{\text{bb/sc}} \qquad \textbf{[9]}$$

A weak hydrogen-bond potential was developed based on *ab initio* calculation of the interaction energies between aliphatic protons covalently bound to polarized carbons and oxygens in model compounds (5). Distance and angular dependencies of these interactions were fitted with polynomials calibrated so that the energy of an optimal weak hydrogen bond would equate one half of an optimal normal hydrogen bond (5-7):

$$E^{\text{weakhbond}}_{\text{optimal}} = 0.5 * E^{\text{regularhbond}}_{\text{optimal}} \qquad \textbf{[10]}$$

with $$E^{\text{regularhbond}}_{\text{optimal}} = E^{dis\tan ce}_{\text{optimal}}(d_{HA} = 1.9) + E^{\Psi}_{\text{optimal}}(\Psi = 115) + E^{\Theta}_{\text{optimal}}(\Theta = 175)$$

and $$E^{\text{weakhbond}}_{\text{optimal}} = E^{dis\tan ce}_{\text{optimal}}(d_{HA} = 2.3) + E^{\Psi}_{\text{optimal}}(\Psi = 135) + E^{\Theta}_{\text{optimal}}(\Theta = 170)$$

where d_{HA} is the distance in angstroms between hydrogen and acceptor atoms, Ψ is the angle between the acceptor base, the acceptor atom and the hydrogen atom and Θ is the angle between the acceptor atom, the hydrogen atom and the donor atom. Optimal values for these parameters are given in brackets. SI Fig. 9 represents the variations of the energy of a weak hydrogen bond as a function of the distance d_{HA} and angles Ψ and Θ.

Side-Chain Conformation and Amino Acid Sequence Recovery. From the database of 18 membrane protein crystal structures (resolution better than 3.5 Å), two data sets of nine structures were created. Each data set was used to optimize the weights of the energy function and, then, to cross-validate the energy function optimized with the other data set. Most membrane proteins considered in our recovery tests were too large to design all positions simultaneously with our expanded side-chain rotamer library. For each membrane protein structure, two regions were defined, i.e., the "core" region facing the membrane hydrophobic core and the "interface" region facing the membrane interface. In conformation recovery experiments, the backbone structure was kept fixed to the crystallographic coordinates and all side-chains belonging either to the "core" or to the "interface" region were repacked simultaneously. Side-chain dihedral angles were considered correctly predicted if they were within 40° of the crystallographically determined values. In sequence recovery experiments, the backbone structure was also held constant and sequence space was searched simultaneously at all positions belonging either to the "core" or the "interface" region for the combination of amino-acids that minimizes the free energy of the system. Residues lining channels or pore regions as well as residues binding cofactors were excluded from these experiments.

1. Kuhlman B, Baker D (2000) *Proc Natl Acad Sci USA* 97:10383-10388.

2. Lazaridis T (2003) *Proteins* 52:176-192.

3. Kortemme T, Morozov AV, Baker D (2003) *J Mol Biol* 326:1239-1259.

4. Keutsch FN, Saykally RJ (2001) *Proc Natl Acad Sci USA* 98:10533-10540.

5. Gu Y, Kar T, Scheiner S (1999) *J Am Chem Soc* 121:9411-9422.

6. Vargas R, Garza J, Dixon DA, Hay BP (2000) *J Am Chem Soc* 122:4750-4755.

7. Scheiner S, Kar T, Gu Y (2001) *J Biol Chem* 276:9832-9837.