Supplemental Text

Constructing a computational insertion energy scale. A scale of amino acid insertion energies is only meaningful if the energy values are largely independent of the protein complex to which they belong. In the present case, we wish to extract individual insertion energies for particular amino acids that reside in test α-helices. If we compute the total insertion energy of two α-helices that are identical except at the central position, then we can ascribe the total energy difference between the two helices to the energy difference between the two central residues. However, this only gives us the *relative* energy difference between the two residues of interest. Continuing along this line of reasoning, the absolute insertion energy of residue X, ΔG^X can be determined by comparing the insertion energy of the appropriate helices as shown next.

(a) Determination of the absolute insertion energies for leucine and alanine.

We started by determining the insertion energy for alanine, ΔG^A, and leucine, ΔG^L, using a fitting protocol similar to the one used by Hessa et al. [1]. First, we used the program MOLDA to design helical peptides of the form GGGG-(L_nA_{19-n})-GGGG that contain 19 central amino acids and n leucine residues, where n ranged from 0 to 7 [2]. Below, we refer to a helix containing n leucines as n-L. Experimentally, it is believed that the helix inserts with its long axis perpendicular to the plane of the membrane with the alanine residues spanning the transmembrane (TM) region, so we carried out calculations with the same assumption. The energy for a test peptide is:

$$\Delta G_{tot} = n\Delta G^L + (19-n)\Delta G^A + \Delta G^{flank}, \tag{1}$$

where ΔG^{flank} is the contribution to the insertion energy from the flanking glycine residues. We assumed that the flanking residues are sufficiently solvated such that they do not contribute to the insertion energy, thus we set $\Delta G^{flank} = 0$ as assumed in Ref. [1]. Rewriting Eq. 1 we arrive at

$$\Delta G_{tot} = n\underbrace{(\Delta G^L - \Delta G^A)}_{slope} + \underbrace{19\Delta G^A}_{intercept}. \tag{2}$$

Next, we systematically calculated ΔG_{tot} for peptides harboring 2 to 4 leucine residues, and we constructed a graph of ΔG_{tot} against n (Fig. S1). We carried out a linear regression on this data to determine the slope, m, and the intercept, a, in Eq. 2. Finally, we have $\Delta G^A = a/19$ and $\Delta G^L = m + a/19$. For these calculations, we only considered the electrostatic and non-polar components of the free energy. Since the TM domain is 20 residues long, there are many possible positions for the placement of the leucine residues. We considered the following symmetric sequences:

$$\begin{array}{ll} 2-L & \texttt{GGGGLAAAAAAAAAAAAAAAALGGGG} \\ & \texttt{GGGGALAAAAAAAAAAAAAAALAGGGG} \\ & \texttt{GGGGAALAAAAAAAAAAAAALAAGGGG} \end{array} \tag{3}$$

GGGGAAALAAAAAAAAAAALAAAGGGG
GGGGAAAALAAAAAAAAALAAAAGGGG
GGGGAAAAALAAAAAAALAAAAAGGGG
GGGGAAAAAALAAAAALAAAAAAGGGG
GGGGAAAAAAALAAALAAAAAAAGGGG
GGGGAAAAAAAALALAAAAAAAAGGGG

$3-L$ GGGGAAAAAAALALALAAAAAAAGGGG (4)
GGGGAAAAAALAALAALAAAAAAGGGG
GGGGAAAAALAAALAAALAAAAAGGGG
GGGGAAAALAAAALAAAALAAAAGGGG
GGGGAAALAAAAALAAAAALAAAGGGG
GGGGAALAAAAAALAAAAAALAAGGGG
GGGGALAAAAAAALAAAAAAALAGGGG
GGGGLAAAAAAAALAAAAAAAALGGGG

$4-L$ GGGGLAAALAAAAAAAALAAALGGGG (5)
GGGGALAALAAAAAAAALAALAGGGG
GGGGAALALAAAAAAAALALAAGGGG
GGGGAAALLAAAAAAAALLAAAGGGG
GGGGAAAALLAAAAAAALLAAAAGGGG
GGGGAAAALALAAAAALALAAAAGGGG
GGGGAAAALAALAAALAALAAAAGGGG
GGGGAAAALAAALALAAALAAAAGGGG

For unknown reasons, MOLDA produces improper helices for the first and sixth sequences in Eq. 4, so we discarded these helices in our analysis. For each value of n, there are many energy values corresponding to each of the sequences above. A linear regression of this data revealed that ΔG^A = –1.97 and ΔG^L = –3.49 kcal/mol. At each value of n, the spread in the computational energies is roughly 0.75-1.5 kcal/mol, and the translocon data exhibits a similar spread of about 0.5 kcal/mol [1].

(b) Determination of the remaining insertion energies.

Based on the values of ΔG^A and ΔG^L, we calculated the free energy cost for inserting all other residues into the membrane using the following equation:

$$\begin{aligned} \Delta G^X &= \Delta\Delta G^{L\to X} + \Delta G^L \\ &= (\Delta G^X - \Delta G^L)_{elec} + (\Delta G^X - \Delta G^L)_{np} + (\Delta G^X - \Delta G^L)_{mem} \\ &+ (\Delta G^X - \Delta G^L)_{dipole} + \Delta G^L, \end{aligned} \tag{6}$$

where X denotes the central amino acid in the helix. The electrostatic, non-polar, membrane deformation, and dipole energy terms can all be calculated by comparing the insertion energy of a reference peptide containing a leucine at the central position with a helix in which the central residue was mutated to X, as in Fig. 1 of the main text. For example, to estimate ΔG^I the following peptides were constructed and their electrostatic and non-polar energies were computed:

$$\begin{aligned} Ile(I) \quad &: \quad \texttt{GGGGAAAALAAAALAAAALAAAAGGGG} \\ &\quad\quad\ \texttt{GGGGAAAALAAAAIAAAALAAAAGGGG} \end{aligned} \tag{7}$$

The difference in electrostatic energies between these two configurations is –0.47 kcal/mol. The non-polar component of the energy was calculated as detailed in the main text, and it is 0.57 kcal/mol. In this case, there is no contribution from the membrane deformation and the dipole energy. Thus, the total free energy for I is:

$$\Delta G^I = -0.47 + 0.57 - 3.49 = -3.39 \text{ kcal/mol}. \tag{8}$$

The free energies of other residues were obtained in a similar way. Here we provide all the sequences used in this work, which are similar to those used in Ref. [1]:

$$\begin{aligned} Phe(F) \quad &: \quad \texttt{GGGGAAAALAAAALAAAALAAAAGGGG} \\ &\quad\quad\ \texttt{GGGGAAAALAAAAFAAAALAAAAGGGG} \\ Val(V) \quad &: \quad \texttt{GGGGAAAALAAAALAAAALAAAAGGGG} \\ &\quad\quad\ \texttt{GGGGAAAALAAAAVAAAALAAAAGGGG} \\ Cys(C) \quad &: \quad \texttt{GGGGAAAALAAAALAAAALAAAAGGGG} \\ &\quad\quad\ \texttt{GGGGAAAALAAAACAAAALAAAAGGGG} \\ &\quad\quad\ \texttt{GGGGAAAALALAALAALALAAAAGGGG} \\ &\quad\quad\ \texttt{GGGGAAAALALAACAALALAAAAGGGG} \\ Met(M) \quad &: \quad \texttt{GGGGAAAALAAAALAAAALAAAAGGGG} \\ &\quad\quad\ \texttt{GGGGAAAALAAAAMAAAALAAAAGGGG} \end{aligned} \tag{9}$$

$Trp(W)$	:	GGGGAAAALAAAALAAAALAAAAGGGG
		GGGGAAAALAAAAWAAAALAAAAGGGG
		GGGGAAAALALAALAALALAAAAGGGG
		GGGGAAAALALAAWAALALAAAAGGGG
$Thr(T)$	:	GGGGAAAALALAALAALALAAAAGGGG
		GGGGAAAALALAATAALALAAAAGGGG
$Tyr(Y)$	:	GGGGAAAALALAALAALALAAAAGGGG
		GGGGAAAALALAAYAALALAAAAGGGG
$Gly(G)$	:	GGGGAAAALALAALAALALAAAAGGGG
		GGGGAAAALALAAGAALALAAAAGGGG
$Ser(S)$	:	GGGGAAAALALAALAALALAAAAGGGG
		GGGGAAAALALAASAALALAAAAGGGG
$Asn(N)$	:	GGGGAAAALALALLLALALAAAAGGGG
		GGGGAAAALALALNLALALAAAAGGGG
$His(H)$	:	GGGGAAAALALALLLALALAAAAGGGG
		GGGGAAAALALALHLALALAAAAGGGG
$Gln(Q)$	:	GGGGAAAALALALLLALALAAAAGGGG
		GGGGAAAALALALQLALALAAAAGGGG
$Arg(R)$	:	GGGGAAAALALALLLALALAAAAGGGG
		GGGGAAAALALALRLALALAAAAGGGG
$Glu(E)$	:	GGGGAAAALALALLLALALAAAAGGGG
		GGGGAAAALALALELALALAAAAGGGG
$Lys(K)$	:	GGGGAAAALALALLLALALAAAAGGGG
		GGGGAAAALALALKLALALAAAAGGGG
$Asp(D)$	:	GGGGAAAALALALLLALALAAAAGGGG
		GGGGAAAALALALDLALALAAAAGGGG

For cysteine and tryptophan, we used the two configurations shown in Ref. [1] to determine the electrostatic and non-polar energies, and we presented the average of both values. Note that for asparagine we used a different sequence from that in Ref. [1]. In Table S1, we summarize the insertion energies for each residue. In the case of charged residues (R,E,K, and D) we show two values: the first is the insertion energy when the membrane is allowed to deform, and the value in parentheses is the classic calculation assuming a flat membrane.

Electrostatic calculations. We used the APBS package to carry out all electrostatic calculations along with the PARSE parameter set for the atomic partial charges and radii [3],[4]. We solved the linearized form of the Poisson-Boltzmann equation using 3 levels of focusing starting with a system size of 300 Å and ending at a size of 50 Å. 161 grid points were used in each direction resulting in a final grid spacing of 0.31 Å. All calculations were carried out in the presence of a 100 mM symmetric salt concentration in water accessible regions.

To include the effect of the membrane, we wrote scripts to manipulate the local environment around the helix before carrying out electrostatic calculations. Briefly, we carried out dummy runs in APBS to create 3D maps representing the system's dielectric and solvent accessibility properties using the molecular surface representation to define the water-protein interface. Next, the scripts examined the spatial position of each grid point and determined if the point fell into a predefined region corresponding to either the hydrocarbon-core or the polar head group. Points in either of these regions were labeled solvent inaccessible and their dielectric value was changed to either 2.0 or 80.0, respectively. The shapes of the predefined regions were determined from the elastostatic calculations in the next section. We fit the numerical solution for the membrane height to a high-order polynomial in r and θ. For the class of distortions that we considered, the membrane surface is very well described by the following equation:

$$h(r,\theta) = \left(\frac{L_0 - u_0}{2} \cdot \frac{1}{2}(1-\cos\theta) + \frac{u_0}{2}\right)(1 - A(r)) + \frac{L_0}{2}A(r) \tag{10}$$

where

$$A(r) = \frac{\frac{1}{5}(r-r_0) + \frac{1}{5000}(r-r_0)^4}{1 + \frac{1}{5}(r-r_0) + \frac{1}{5000}(r-r_0)^4}, \tag{11}$$

r_0 is the effective radius of the helical peptide, L_0 is the equilibrium length of the membrane, and $u_0/2$ is the height of the membrane at the site of maximum compression when it meets the helix (see Fig. 3 of the main text). When we examined leaflet thickness changes, we varied the value of u_0 in the boundary conditions (BCs) for the elasticity calculations and in Eq. 10 when carrying out the electrostatic calculations. For instance, when $u_0 = L_0$, the membrane is completely flat, and when $u_0 = 0$, the membrane is compressed to half of its width as in Fig. 4A and B of the main text. Finally, we assumed that the height of the lipid head group region, L_{hg}, was constant, so that the head group region falls between h and $(h - L_{hg})$. Thus, given a value of u_0, the scripts use Eq. 10 to unambiguously define the hydrocarbon-core of the membrane and the head group region.

The membrane dipole potential was modeled by adding a thin layer of dipole charges at both the upper and lower membrane core-lipid head group interface. This approximates a sheet of dipole charges, and when the sheets are flat, electric fields only exist between the dipole charges. Positive charges were placed closer to the core; thus, the membrane potential sharply rises upon entering the membrane core perpendicular to the plane of the membrane, and it drops to zero upon exiting the core. Charges were only added in pairs, so that no net charge was added to the system. They were placed at adjacent grid points along the z-axis, so that their separation was Δz. Charges were not added to grid points within the molecular surface of the TM protein. In a uniform dielectric material, the potential gained by crossing the dipole sheet is:

$$\psi = \frac{\rho \Delta z}{\epsilon}, \tag{12}$$

where ρ is a 2D charge density corresponding to one-half of the dipole sheet, either the plus or minus

sheet, since both have the same density, and ψ is the membrane dipole potential. We can solve for ρ in Eq. 12 for a desired value of ψ, and then use this value when editing the charge maps. During focusing, Δz gets smaller, so to match the desired ψ one must increase the effective charge on the dipoles. The dielectric discontinuity at the membrane interface invalidates Eq. 12. Therefore, the value of the dipole charge added to each grid point was manually adjusted to reach the desired inner membrane potential of +300 mV. After this was done at the coarsest level, scaling this charge by $1/\Delta z$ produced the desired membrane potential upon focusing. At the coarsest level, we imposed Dirichlet boundary conditions on the system, setting the potential to zero. Technically, we should have imposed boundary conditions consistent with the dipole membrane potential; however, this was technically difficult. Nonetheless, at the first level of focusing the desired value of the dipole membrane potential everywhere along the boundary was achieved. Therefore, we feel comfortable that the desired physical model was achieved. Finally, we calculated the energy of interaction of the dipole charges on the TM helix by carrying out a calculation with the dipole charges on and then one with the charges off. In each case, we had APBS report the total energy for each atom on the TM helix. Twice the difference in this value between the two cases is the interaction energy between the helix charges and the dipole membrane potential. This is what is reported throughout the manuscript.

For more explicit details on editing maps please see Ref. [5].

Determining the membrane deformation energy. We started by writing down the equation that describes how the membrane deformation energy depends on the shape of the membrane (see Ref. [6]):

$$\Delta G_{mem} = \int_{\Omega} \frac{1}{2} \left(K_a \left(\frac{2u}{L_0} \right)^2 + K_c \left(\vec{\nabla}^2 u \right)^2 + \alpha \left(\vec{\nabla} u \right)^2 \right) d\Omega, \tag{13}$$

where K_a is the bilayer compression modulus, K_c is the bending modulus, α is the surface tension, and u is the deviation of the leaflet height from its equilibrium value: $u(x, y) = h(x, y) - L_0/2$. Physically, the first term in Eq. 13 corresponds to the compression of the bilayer, the second term refers to the curvature deformation energy of the individual monolayer leaflets, and the last term accounts for the energy associated with area expansion of each leaflet of the bilayer. Moreover, the equation above assumes that the profile of the upper and lower leaflets are mirror images of each other through the x-y plane. In our case, we considered only distortions in the upper leaflet, so we would like to rewrite Eq. 13 to reflect this situation. Since we assumed no deformation in the lower monolayer, the curvature and area expansions terms are exactly 1/2 of the values in Eq. 13. The compression term depends on the total length of the bilayer so it scales differently. Since the distortions in the upper and lower leaflet are assumed to be equal and opposite in Eq. 13, the width of the bilayer is $2h(x, y) = 2(u(x, y) + L_0/2) = 2u(x, y) + L_0$. In a compression spring model, the energy is assumed to be zero when the spring adopts its equilibrium length, and it increases quadratically as the relative length of the spring increases or decreases. Therefore, we have:

$$\Delta G_{compression} = \frac{1}{2} K_a \left(\frac{(2u(x, y) + L_0) - L_0}{L_0} \right)^2 = \frac{1}{2} K_a \left(\frac{2u}{L_0} \right)^2. \tag{14}$$

In our case, the deformation in the upper leaflet is not reflected in the lower leaflet, and we have the following modification:

$$\Delta G_{compression} = \frac{1}{2} K_a \left(\frac{(u(x,y) + L_0) - L_0}{L_0} \right)^2 = \frac{1}{2} K_a \left(\frac{u}{L_0} \right)^2. \tag{15}$$

Making these changes to Eq. 13 we arrive at the following energy functional:

$$\Delta G_{mem} = \int_\Omega \frac{1}{2} \left(\frac{K_a}{L_0^2} u^2 + \frac{K_c}{2} \left(\vec{\nabla}^2 u \right)^2 + \frac{\alpha}{2} \left(\vec{\nabla} u \right)^2 \right) d\Omega. \tag{16}$$

At this point, one can determine the underlying partial differential equation that u satisfies by finding the solution that minimizes the energy of Eq. 16:

$$\delta \Delta G_{mem} = \int_\Omega \frac{1}{2} \left(2 \frac{K_a}{L_0^2} u \delta u + K_c \left(\vec{\nabla}^2 u \right) \delta \left(\vec{\nabla}^2 u \right) + \alpha \left(\vec{\nabla} u \right) \delta \left(\vec{\nabla} u \right) \right) d\Omega = 0. \tag{17}$$

After performing integration by parts twice and discarding boundary terms, we arrive at the following:

$$\int_\Omega \left(\left(\frac{2K_a}{L_0^2} u + K_c \vec{\nabla}^4 u - \alpha \vec{\nabla}^2 u \right) \delta u \right) d\Omega = 0. \tag{18}$$

For arbitrary variations δu the term in parenthesis must vanish, so that u solves the following 2D Euler-Lagrange equation:

$$\nabla^4 u - \frac{\alpha}{K_c} \nabla^2 u + \frac{2K_a}{L_0^2 K_c} u = 0, \tag{19}$$

where in the main text we let $\gamma = \alpha / K_c$ and $\beta = 2K_a / (L_0^2 K_c)$. This equation is the same as the one derived by Nielsen and colleagues for symmetric deviations up to a factor of 2 multiplying the last term [6]. Note that our definition of β here deviates from theirs by this factor.

Finally, we rewrite Eq. 16 in polar coordinates to facilitate the determination of the membrane bending energy once we have determined the deformation profile for a given set of BCs:

$$\Delta G_{mem} = \int_\Omega \frac{1}{2} \left(\frac{K_a}{L_0^2} u^2 + \frac{K_c}{2} \left(\frac{1}{r} u_r + u_{rr} + \frac{1}{r^2} u_{\theta\theta} \right)^2 + \frac{\alpha}{2} \left(u_r^2 + \frac{1}{r^2} u_\theta^2 \right) \right) d\Omega. \tag{20}$$

For all the calculations in the main text, we first solved Eq. 19 for a given set of BCs, and then we used the solution along with Eq. 20 to determine the total membrane deformation energy. We accomplished this task by first computing all of the partial derivatives in Eq. 20 via a finite difference scheme, and then we used a 2D trapazoid rule to perform the integration.

Solving the elastostatic equation. Here we discuss the finite difference solution to the 2D Euler-Lagrange equation used to determine the shape of the membrane. It is useful to discuss membrane bending in terms of the deviation of the membrane height from its equilibrium value. Calling this value u, we have $u(x, y) = h(x, y) - L_0/2$, where L_0 is the membrane equilibrium thickness and $h(x, y)$ is the local *monolayer* height [6]. The Euler-Lagrange equation has the following form:

$$\nabla^4 u - \gamma \nabla^2 u + \beta u = 0, \tag{21}$$

where the first term is biharmonic, the second is harmonic, and the third is proportional to the surface deformation, u. The constant prefactors are related to the material properties of the bilayer. While analytic solutions to these equations are readily produced for cylindrically symmetric problems, we were interested in problems that lack this symmetry. Therefore, we numerically solved this equation. We first rewrite Eq. 21 in polar coordinates:

$$\begin{aligned} &u_{rrrr} + \frac{2}{r^2} u_{rr\theta\theta} + \frac{1}{r^4} u_{\theta\theta\theta\theta} + \frac{2}{r} u_{rrr} - \frac{2}{r^3} u_{r\theta\theta} - \frac{1}{r^2} u_{rr} + \frac{4}{r^4} u_{\theta\theta} \\ &+ \frac{1}{r^3} u_r - \gamma \left(u_{rr} + \frac{1}{r} u_r + \frac{1}{r^2} u_{\theta\theta} \right) + \beta u = 0. \end{aligned} \tag{22}$$

We discretized each term in the above equation using a symmetric finite differencing scheme, and then we assembled all of the terms into the following linear form $Ax = b$. After discretization, the elements of A can be read off from the coefficients of the terms above, the constant vector b is mostly zero except for terms involving the boundary, and the solution vector x is the numeric solution to Eq. 21 in column form.

We imposed periodic BCs on θ, and we enforced clamped BCs on r at the inner radial boundary, r_0, and the outer boundary, r_f. At the outer boundary u and $\partial u/\partial r$ were both set to zero. The inner boundary is at a radius r_0, and we fixed the height and slope of the surface as follows:

$$u(r_0, \theta) = \frac{L_0 - u_0}{2} \cdot \frac{1}{2} \left(\cos\theta - 1 \right), \tag{23}$$

and

$$\frac{\partial u(r_0, \theta)}{\partial r} = -\frac{L_0 - u_0}{2 r_0} \cdot \frac{1}{2} \left(\cos\theta - 1 \right), \tag{24}$$

where u_0 is the same quantity shown in Fig. 3 of the main text. Also, notice that $u(r = r_0, \theta)$ is not u_0. The choice of boundary conditions is critical in these problems [6], and our choice was motivated by two observations. First, the values of $u(r_0, \theta)$ and $u_r(r_0, \theta)$ shown above give rise to membrane shapes that look quantitatively similar to those from MD simulations [7]. Second, for the contact angles we explicitly examined, the form shown above minimized the membrane deformation energy. The equations were solved in Matlab, and the accuracy of the code was tested against the analytic solutions of Nielsen et al. for radially symmetric test cases [6]. Our solver converges quickly to the analytic values. Finally, for the graphs presented in Fig. 4 in the main text, rigorous convergence testing was carried out on the solutions and energy values. The final energies exhibited uniform convergence to within 0.01 kcal/mol.

PMF calculation of arginine insertion and the pK_a of the arginine side chain. We designed a 61 residue polyleucine peptide with an arginine at the central position, $L_{30}RL_{30}$. To calculate the potential of mean force (PMF), we positioned the helix with the C_α atom of the arginine at the upper leaflet (z = 21 Å), and we translated the helix in 1 Å increments until the C_α atom reached the membrane core (z = 0). At each step, we calculated the total free energy (electrostatic + non-polar + membrane deformation energy). We also explored the dependence of the PMF profile on the choice of rotamer libraries by carrying out these calculations with SCCOMP and SCWRL. While the arginine conformations are not the same, the profiles are nearly identical.

Since the energy of inserting a charged residue into a flat membrane is large compared to polar residues, it is reasonable to ask if arginine is protonated or neutral in the membrane. Several research groups have used detailed MD simulations to ask this same question [8, 9, 10], and we used our continuum method to calculate the pK_a of arginine at different depths in the membrane. We did this by carrying out the PMF calculation in Fig. 5D with a neutralized arginine side chain, mimicking the deprotonated state. We did this by changing the PARSE charges on the following arginine atoms: HE ($+0.45 \rightarrow -0.35$), NH1 ($-0.7 \rightarrow -0.8$), and NH2 ($-0.7 \rightarrow -0.8$). These changes resulted in an overall neutral molecule. The two curves are superposed in Fig. S2A (red line, charged; blue line, neutral). Both PMF curves are set equal to zero at z = 21 Å when the arginine is at the equilibrium height of the unstrained upper leaflet. The energies are nearly identical for z values from 21 to 10; however, from z = 10-6 Å the neutral form takes on a higher value. When the arginines penetrate the core of the membrane the roles reverse, and the charged form is about 7-8 kcal/mol higher in energy. We used the energy difference between the neutral and charged forms to calculate the pK_a of arginine as a function of membrane depth in Fig. S2B using the following equation:

$$pK_a(z) = pK_a^{bulk} + \left(\frac{\Delta G_{tot}^{neutral}(z) - \Delta G_{tot}^{charged}(z)}{2.3\, k_B T} \right), \tag{25}$$

where pK_a^{bulk} is the pK_a of arginine in solution, which we take as 12.5. There are two features of this curve that are in good agreement with MD simulations. First, the pK_a of the side chain drops below 7 at the center of the membrane, suggesting that the likelihood of being protonated is about 50%. Second, the pK_a rises as the side chain begins to enter the leaflet around z = 10 Å. This general agreement with simulation again supports the accuracy of our present method.

Exploring the model's dependence on key parameters.

(a) Influence of rotamer conformation.

In addition to SCCOMP, we also used the program SCRWL to generate side-chain rotamer conformations for the peptides constructed using MOLDA [11],[12]. We then recalculated all of the insertion energy values examined in section (a) of Constructing a computational insertion energy scale. ΔG^A and ΔG^L were –1.95 and –3.93 kcal/mol, respectively. Each of these values differ by less than 0.5 kcal/mol from the values determined using SCCOMP. We then recalculated the insertion free energy of each residue ΔG^X (Table S2). Most values between the two scales are in good agreement, but noticeably, the lysine values differ by 5.0 kcal/mol. The reason for this is that SCWRL produces a side-chain conformation in which the residue extends away from the protein, perpendicular to the helical axis, while SCCOMP produces a conformation in which the side chain "reaches up" out of the membrane toward the extracellular space. This is termed "snorkeling", and it has been documented in bilayer MD simulations [13]. The SCCOMP conformation minimizes the charged residue's exposure to the low-dielectric membrane core, which is why its value is lower.

(b) Dependence on head group dielectric value.

The dielectric value of the lipid head group region is not well characterized. For the calculations in the main text we chose a relatively high value of 80. To test the effect of varying the head group dielectric constant, we recalculated the two main graphs (Fig. 5D and Fig. 7A) using a lower value of 40 to see how it affected our main conclusions. Changing the head group dielectric value from 80 to 40 has very little influence on the PMF for inserting a charged arginine (Fig. S3A compared to Fig. 5D), and the computational biological hydrophobicity scale is also robust to this change (Fig. S3B compared to Fig. 7A). The membrane is stable and flat for all of the non-charged residues. So, it is easy to see why the scale is insensitive to changes in ϵ_{hg} since these residues reside in the low-dielectric of the unaltered hydrocarbon core. However, the charged residues reside at the membrane-water interface, which should be affected by changing the lipid head group dielectric value. Nonetheless, the insertion energies of the charged residues are only marginally increased: R (+1.6 kcal/mol), E (+0.8 kcal/mol), K (+1.4 kcal/mol), and D (+0.7 kcal/mol), demonstrating that the proximity to the high-dielectric water region mitigates changes to the solvation energy.

References

[1] Hessa, T. et al. 2005. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature.* 433:377-381.

[2] Yoshida, H., and H. Matsuura. 1997. MOLDA for Windows - A Molecular Modeling and Molecular Graphics Program Using a VRML Viewer on Personal Computers. *J. Comput. Chem. Software.* 3:147-156.

[3] Baker, N.A., D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.* 98:10037-10041.

[4] Sitkoff, D., K.A. Sharp, and B. Honig. 1994. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* 98:1978-1988.

[5] Grabe, M., H. Lecar, Y.N. Jan, and L.Y Jan. 2004. A quantitative assessment of models for voltage-dependent gating of ion channels. *Proc. Natl. Acad. Sci. USA.* 101:17640-17645.

[6] Nielsen, C., M. Goulian, and O.S. Andersen. 1998. Energetics of inclusion-induced bilayer deformations. *Biophys. J.* 74:1966-1983.

[7] Dorairaj, S., and T.W. Allen. 2007. On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *Proc. Natl. Acad. Sci. USA.* 104:4943-4948.

[8] MacCallum, J.L., W.F. Bennett, and D.P. Tieleman. 2007. Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment. *J. Gen. Physiol.* 129:371-377.

[9] Li, L., I. Vorobyov, A.D. MacKerell, and T.W. Allen. 2008. Is arginine charged in a membrane? *Biophys. J.* 94:L11-L13.

[10] Yoo, J., and Q. Cui. 2008. Does arginine remain protonated in the lipid membrane? Insight from microscopic pKa calculations. *Biophys. J.* doi:10.1529/biophysj.107.122945.

[11] Eyal, E., R. Najmanovich, B.J. McConkey, M. Edelman, and V. Sobolev. 2004. Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. *J. Comput. Chem.* 25:712-724.

[12] Canutescu, A.A., A.A. Shelenkov, and R.L. Dunbrack, Jr. 2003. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* 12:2001-2014.

[13] Mishra, V.K., M.N. Palgunachari, J.P. Segrest, and G.M. Anantharamaiah. 1994. Interactions of synthetic peptide analogs of the class A amphipathic helix with lipids. Evidence for the snorkel hypothesis. *J. Biol. Chem.* 269:7185-7191.

Table S1: Free energy cost for insertion, ΔG^X, where X denotes the central amino acid in the helix. In the case of charged residues we show two values: the first one is the insertion energy when the membrane is allowed to deform while the value in parentheses is the classic calculation assuming a flat membrane.

amino acid	ΔG^X (kcal/mol)
Ile(I)	– 3.39
Leu(L)	– 3.49
Phe(F)	– 2.29
Val(V)	– 2.65
Cys(C)	– 2.12
Met(M)	– 0.56
Ala(A)	– 1.97
Trp(W)	0.27
Thr(T)	– 1.96
Tyr(Y)	0.98
Gly(G)	– 1.77
Ser(S)	– 0.97
Asn(N)	3.95
His(H)	3.07
Gln(Q)	2.09
Arg(R)	8.07 (29.63)
Glu(E)	6.03 (35.67)
Lys(K)	5.58 (32.26)
Asp(D)	7.18 (37.80)

Table S2: Same as Table S1, but using the SCWRL rotamer library.

amino acid	ΔG^X (kcal/mol)
Ile(I)	– 3.78
Leu(L)	– 3.93
Phe(F)	– 2.45
Val(V)	– 3.59
Cys(C)	– 1.53
Met(M)	– 1.62
Ala(A)	– 1.95
Trp(W)	0.02
Thr(T)	– 1.14
Tyr(Y)	1.12
Gly(G)	– 2.09
Ser(S)	– 0.35
Asn(N)	3.74
His(H)	2.86
Gln(Q)	1.66
Arg(R)	4.19 (31.30)
Glu(E)	5.33 (36.20)
Lys(K)	11.40 (32.53)
Asp(D)	6.30 (36.17)

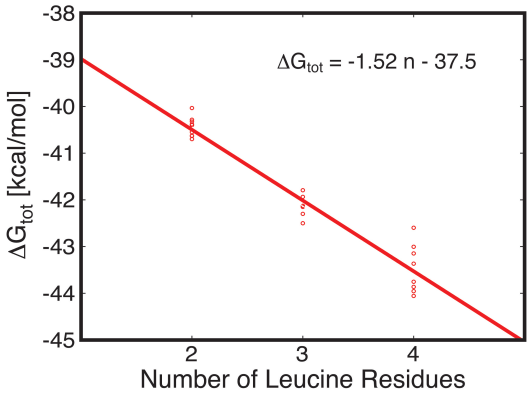

Figure S1: Insertion energies for test peptides containing 2-4 leucine residues. Red circles represent the calculated values for different helical segments containing a given number of leucine and alanine residues. The exact sequences are given in the Supplemental Text. The red line is the linear regression to the data.

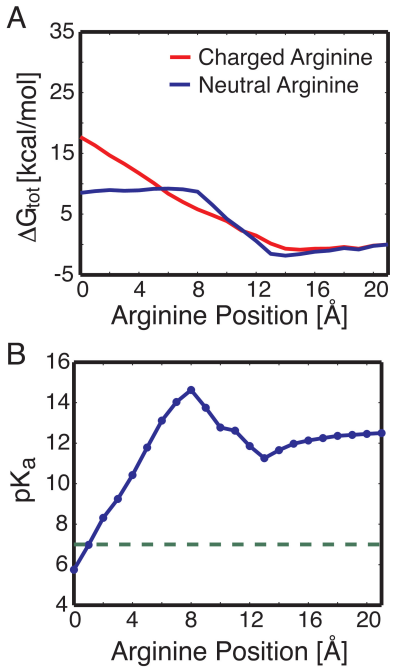

Figure S2: pK_a calculation of arginine side chain as it enters the membrane. (A) PMF profile for the charged side chain (red line) and the neutral side chain (blue line). (B) The pK_a of the arginine side chain can be determined from the two curves in panel A as discussed in the Supplemental Text. The value at when the arginine position is 21 Å is the bulk value, which we take as 12.5.

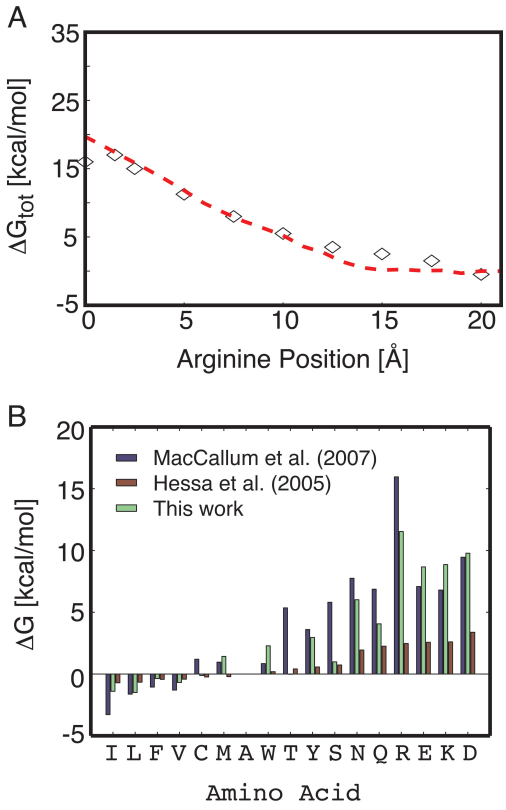

Figure S3: Calculation of leucine-arginine PMF and the biological hydrophobicity scale with a head group dielectric value of 40. (A) The total PMF for inserting a polyleucine α-helix harboring a central arginine was recomputed with $\epsilon_{hg} = 40$ (dashed red line) and compared to the MD data from Dorairaj and Allen (Diamonds). There is very little difference between this curve and the one computed using $\epsilon_{hg} = 80$ in Fig. 5D of the main text. (B) The biological hydrophobicity scale was also recomputed using $\epsilon_{hg} = 40$ (green bars) and compared against the values from the translocon experiments (red bars) and the lone amino acid scale calculated from MD simulations (blue bars). Comparing these values with those in Fig. 7A of the main text, there is very little difference.