Resource

Structure

Membrane Protein Properties Revealed through Data-Rich Electrostatics Calculations

Highlights

- Incorporation of PDB2PQR for automated protein preparation
- A new model for determining membrane-induced pK_a shifts
- Detailed electrostatic analysis on the TRPV1 ion channel
- A survey of the electrostatic properties of 1,614 membrane protein structures

Authors

Frank V. Marcoline, Neville Bethel, Christopher J. Guerriero, Jeffrey L. Brodsky, Michael Grabe

Correspondence

michael.grabe@ucsf.edu

In Brief

Marcoline et al. highlight major new features of the software, APBSmem, which allow users to carry out electrostatics calculations on membrane proteins and membrane-associated proteins. The authors use APBSmem to reveal new biological insight for several specific membrane proteins.

Marcoline et al., 2015, Structure *23*, 1526–1537
August 4, 2015
http://dx.doi.org/10.1016/j.str.2015.05.014

CellPress

CellPress

Structure
Resource

Membrane Protein Properties Revealed through Data-Rich Electrostatics Calculations

Frank V. Marcoline,[1,2,5] Neville Bethel,[1,2,3,5] Christopher J. Guerriero,[4] Jeffrey L. Brodsky,[4] and Michael Grabe[1,2,*]
[1]Cardiovascular Research Institute, University of California, San Francisco, CA 94158, USA
[2]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158, USA
[3]Integrative Program in Quantitative Biology, University of California, San Francisco, CA 94158, USA
[4]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA
[5]Co-first author
*Correspondence: michael.grabe@ucsf.edu
http://dx.doi.org/10.1016/j.str.2015.05.014

SUMMARY

The electrostatic properties of membrane proteins often reveal many of their key biophysical characteristics, such as ion channel selectivity and the stability of charged membrane-spanning segments. The Poisson-Boltzmann (PB) equation is the gold standard for calculating protein electrostatics, and the software APBSmem enables the solution of the PB equation in the presence of a membrane. Here, we describe significant advances to APBSmem, including full automation of system setup, per-residue energy decomposition, incorporation of PDB2PQR, calculation of membrane-induced pK_a shifts, calculation of non-polar energies, and command-line scripting for large-scale calculations. We highlight these new features with calculations carried out on a number of membrane proteins, including the recently solved structure of the ion channel TRPV1 and a large survey of 1,614 membrane proteins of known structure. This survey provides a comprehensive list of residues with large electrostatic penalties for being embedded in the membrane, potentially revealing interesting functional information.

INTRODUCTION

The Poisson-Boltzmann (PB) equation is a popular method for calculating the electrostatic properties of proteins (Baker et al., 2001; Brooks et al., 2009; Gilson and Honig, 1987; Grant et al., 2001; Zhou et al., 2008). The equation relates the fixed charges on a protein of known structure to the electrostatic potential from which electrostatic energies can be determined (Fogolari et al., 2002). Formally, the PB equation is a second-order partial differential equation,

$$-\nabla\cdot[\varepsilon(r)\nabla\phi(r)]+\varepsilon(r)\kappa^2(r)\sinh[\phi(r)]=\frac{e}{k_BT}4\pi\rho(r), \quad \text{(Equation 1)}$$

where $\phi = e\Phi/k_BT$ is the reduced electrostatic potential, ε is the dielectric value of the different spatial regions (water, membrane, protein), κ is the Debye-Hückel screening parameter related to the ionic conditions of the solvent, ρ is the spatial distribution of the fixed charges on the protein, and r is the position in three-dimensional space. This theory has been applied widely to study ligand binding, protein-protein interactions, and conformational change, with the majority of the studies aimed at soluble proteins.

Electrostatics play also an intimate role in the function of membrane proteins, and the low-dielectric nature of the membrane has a large influence on the electric fields and energetics of proteins and small molecules at or near the lipid bilayer. Key studies have used PB theory to determine the protonation state of residues in membrane-spanning regions (Bashford and Gerwert, 1992; Karshikoff et al., 1994), the insertion energetics of hydrophobic helices (Ben-Tal et al., 1996), the influence of the membrane potential on transmembrane proteins (Roux, 1997), and how the membrane alters the electrostatic potential experienced by ions passing through channels (Roux and MacKinnon, 1999). While there are several PB solvers available for studying soluble proteins, few have been adapted to explore the influence of the membrane. Previously, we developed the APBSmem software to enable users to carry out a number of calculations relevant to specific membrane processes (Callenberg et al., 2010). APBSmem uses the Adaptive Poisson-Boltzmann Solver (APBS), an open-source finite difference PB solver, as the back-end for its electrostatics calculations (Baker et al., 2001).

Here, we report several significant advances to APBSmem that make it more versatile, providing additional energetics information for users, increased protein and membrane manipulation, bundling with PDB2PQR for pK_a calculations, and the ability to report non-polar energy values, which are needed to better model membrane protein stability. Several of these additions are shown in Figure 1 and are discussed in detail in the Experimental Procedures and Supplemental Information. We demonstrate the new features of APBSmem through five case studies. The first two cases explore permeation of cations through the recently solved structure of the thermosensitive channel TRPV1. APBSmem automatically identifies residues known to influence conduction, and provides a rationale for pH-dependent changes in ion selectivity. Case III shows how the software can be used to quickly identify residues whose protonation states are altered by the membrane, and case IV explores properties revealed from a scan of electrostatic insertion energies for all multipass membrane proteins of known structure. Finally, case V shows how protein stability and orientation in the membrane

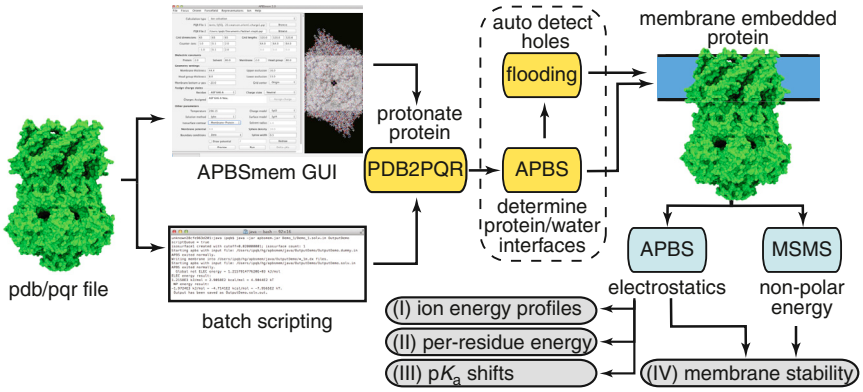

Figure 1. Workflow for APBSmem

PDB or PQR files (green) can be loaded into APBSmem through the GUI or via input files executed from a command line. The bundled PDB2PQR program will protonate PDB files followed by initial protein surface determination with APBS. A membrane-flooding algorithm (see Figure 2) will add the presence of a low-dielectric membrane. APBS and MSMS can then be used to determine the energies for a number of situations outlined in cases I–V.

can be predicted with a simple model based on non-polar energetics coupled with electrostatics.

RESULTS

Case I: Ion and Small-Molecule Placement and Manipulation for Computing Electrostatic Energy Profiles

Due to the low-dielectric nature of the lipid bilayer, ions and small charged molecules cannot readily cross the membranes of cells and organelles. Instead, ion channels and transporters span membranes to facilitate movement. Since ions and many small molecules are electrically charged, electrostatic interactions with the channel or transporter are key determinants of the magnitude of the flux and substrate selectivity. Previously, we demonstrated the ease with which APBSmem can be used to calculate the electrostatic solvation free energy of potassium ions in the pore of the membrane-embedded KcsA potassium channel (Callenberg et al., 2010) by revisiting the seminal study on this topic by Roux and MacKinnon (1999). The ion transfer free energy, ΔG_{elec}, is calculated as

$$\Delta G_{elec} = E_{P,I} - E_P - E_I, \qquad \text{(Equation 2)}$$

where $E_{P,I}$ is the electrostatic energy of the protein plus ion embedded in the membrane, E_P is the energy of the protein in the membrane, and E_I is the energy of the ion in solution. A description of how total electrostatic energies (as in Equation 2) are computed from Equation 1 is presented in the Supplemental Information. Unfortunately, there are still several major hurdles to carrying out these calculations that make them difficult for non-experts, including charge assignments, orienting the protein in the membrane, editing the dielectric around the protein to include the influence of the membrane, and then placing and moving ions through pathways of interest. We have added features to APBSmem that streamline these steps (Figure 1).

We examine ion movement through TRPV1, a narrow channel that is selective for Ca^{2+} and to a lesser degree for Na^+ (Caterina et al., 1997). TRPV1 is thermosensitive, mildly voltage dependent, and sensitive to several toxins and irritants such as capsaicin, which is the active ingredient in chili peppers (Caterina et al., 1997). The channel is a tetramer with each subunit having six transmembrane (TM) segments, and the last two TMs form the central pore domain through which ions flow (A). The pore has two constriction zones: one at the selectivity filter composed of residues GMGD, and a second hydrophobic gate near the cytoplasmic side of the membrane where the TM6 helices cross (Liao et al., 2013). TRPV1 agonists can induce large conformal changes, opening one or both gates (Cao et al., 2013). The most open conformation of the channel was recently determined via electron cryomicroscopy in the presence of a vanilloid agonist resiniferatoxin (RTX) and double-knot toxin (DkTx) (Cao et al., 2013; Liao et al., 2013).

The channel structure (PDB: 3j5q) was loaded into APBSmem and then new features in the *Orient* menu were used to translate the channel −20 Å along the z axis and rotate by 180° about the x axis, followed by a 45° rotation about the z axis. Next, we chose a smoothed molecular surface representation for the protein (Nina et al., 1997), and the SWANSON parameter set for the atomic radii and charges (Swanson et al., 2007), since the dielectric smoothing inherent in this method generally gives rise to non-rugged ion energy profiles. Parameterizing the PDB file to create what is known as a PQR file is quite easy now that we have bundled PDB2PQR into the APBSmem distribution (Dolinsky et al., 2004). For cases III–V, we use the PARSE charge and radii set to parameterize the proteins, since that model was specifically developed to explore the free energy of partitioning between aqueous and non-polar environments (Sitkoff et al., 1994). The upper and lower boundaries of the membrane must be set by hand in the graphical user interface (GUI) for the protein of interest (Figure 1), after which APBSmem edits the local dielectric, charge, and ion accessibility around the protein to include the presence of the membrane for electrostatics calculations. The presence of aqueous cavities makes it difficult to unambiguously identify the membrane protein boundaries when adding the membrane. This task is particularly difficult for channels containing fenestrations that connect the inner pore directly to the mid-plane of the bilayer, such as the voltage-gated sodium channels (Payandeh et al., 2012; Shaya et al., 2014; Zhang et al., 2012) (Figure 2B). Programs exist for detecting cavities in proteins (Smart et al., 1996; Voss and Gerstein, 2010); however, detection can also be difficult when the water pathways are convoluted and the protein lacks symmetry, in which case more computationally demanding methods are needed (Adelman et al., 2014). To this end, we use a six-way flood-filling method illustrated in Figure 2A, which starts from a "seed" point known to be within the membrane and then tests surrounding regions to determine whether they are within the

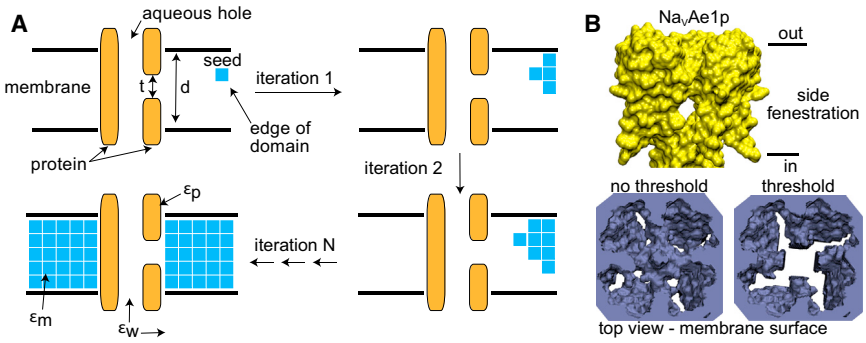

Figure 2. Automatic Detection of Aqueous Channels

(A) Membrane-flooding method. Protein is orange, solution white, initial membrane boundaries black lines, and membrane blue. The membrane is added iteratively in small units starting from an initial seed at the outer boundary of the system. Membrane will not flood protein fenestrations with a vertical dimension less than *t*.

(B) Surface of the Na_VAe1p sodium channel showing large fenestrations in the hydrophobic core of the membrane (top). If no threshold of lipid penetration is set, membrane fills the central aqueous cavity of the channel (blue surface, bottom left); however, for t = 8 Å, the membrane will not penetrate into aqueous cavities (bottom right).

membrane boundaries and external to the protein. The entire membrane is drawn by expanding from this seed in an iterative manner. An additional threshold can be set that prevents expansion into holes smaller than a vertical thickness of *t* (Figure 2B). Here, we use a value of t = 8 Å, which roughly approximates the size of a lipid molecule, and this value successfully allows for the proper identification of aqueous cavities.

The final membrane-embedded protein is shown in Figure 3A with the corresponding membrane boundaries. We then used the new *Ion/Step ion* function to create a Ca^{2+} ion and move it along the z axis through the center of the pore from −40 to +80 Å. A series of calculations were initiated along the path to determine the electrostatic component of the free energy, ΔG_{elec} in Equation 2, for each position (Figure 3B). Parameter values for all calculations are listed in Table 1. The energy profile is marked by asterisks corresponding to locations of interest, and the positions are also identified on the structure (from top to bottom): two minima in the selectivity filter, one minimum in the central cavity between both gates, and one location near the lower gate. These positions most likely reveal regions of the channel involved in selectivity or function, which we explore in more detail in case II. Additional technical aspects of the calculation concerning timings, the choice of molecular surfaces, linear versus non-linear solutions, and grid spacing are discussed in the Supplemental Information.

Case II: Contribution of Individual Residues to the Electrostatic Interaction

Specific residues often play a crucial role in determining protein function by stabilizing bound ligands, facilitating ion permeation, or providing structural integrity through salt bridge interactions. While several phenomena contribute to stabilization, electrostatics is often a key factor, and in some cases it can be the dominant term. With this in mind, it is useful to determine the contribution of a particular residue to an electrostatic interaction, and this information can help interpret structural information to guide future experiments, as reported by Robertson et al. (2008) in their work on inward rectifier channels. The technical details for isolating electrostatic interaction energies between ions or small molecules with specific residues in a protein can be found in the Experimental Procedures and Supplemental Information.

Here, we highlight the utility of APBSmem's ability to dissect the electrostatic contribution of each residue by re-examining Ca^{2+} permeation through TRPV1. As described in case I, the ion experiences energy minima at four locations in the channel (asterisks in Figure 3B), and we identified the top five amino acids that interact most strongly with the ion at each of these positions (vertical dashed lines in Figures 4A–4D). The interaction energy of these residues with the ion is plotted throughout the channel to reveal the spatial extent of their influence. The channel with the ions (red) at each of the four positions is shown to the right of each profile. Not surprisingly, charged residues contribute the most to the electrostatic interaction energy. However, this is not always the case, as the carbonyl group of G643 plays an important role in stabilizing the cation in the narrow portion of the filter (Figure 4B), which is observed for potassium channels. The ease with which APBSmem identifies these crucial residues through these calculations provides a convenient and rational means to select targets for mutational and functional studies.

Many of the residues in Figure 4 were previously shown to play a role in conduction. Counterintuitively, Liu et al. (2009a) demonstrated that a basic residue, K639, is essential for cation conduction, and that neutralization (i.e., K639Q) reduces current. Our calculations reveal that K639 destabilizes permeating ions more than any other residue in the channel, offsetting the deep

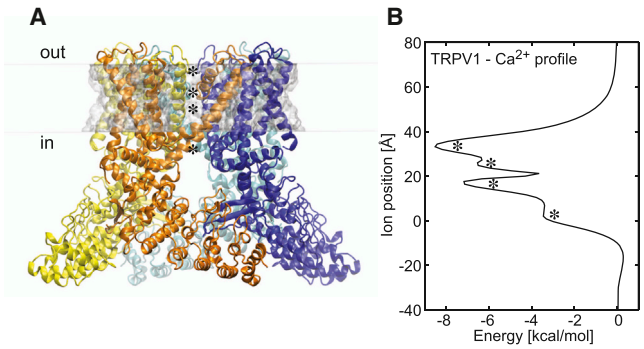

Figure 3. Ion Stepping for Potential Energy Profiles through TRPV1

(A) Molecular image of the fully open TRPV1 tetramer bound to RTX and DkTx (PDB: 3j5q) colored by chain and embedded in a low-dielectric, ion-impermeable membrane. The upper and lower leaflets of the membrane are pale gray surfaces. Asterisks highlight positions along the channel that correspond to energy minima in (B).

(B) Calcium ion solvation energy through TRPV1. The upper two minima are in the selectivity filter, the third from the top is in the cavity, and the bottom position is near the inner gate.

Table 1. Input Parameters

Molecule	TRPV1 (3j5q)	VDAC (3emn)	LeuT (2a65)	Ste6p*
Force field	SWANSON	PARSE	PARSE	PARSE
Counterions	±1|e|, 0.1 M, 2.0 Å	±1|e|, 0.1 M, 2.0 Å	±1|e|, 0.1 M, 2.0 Å	±1|e|, 0.1 M, 2.0 Å
Temperature (K)	298.15	298.15	298.15	298.15
Grid dimensions	161 × 161 × 161	161 × 161 × 161	161 × 161 × 161	161 × 161 × 161
Coarse grid size ($Å^3$)	320 × 320 × 320	300 × 300 × 300	300 × 300 × 300	300 × 300 × 300
Medium grid size ($Å^3$)	160 × 160 × 160	120 × 120 × 120	120 × 120 × 120	120 × 120 × 120
Fine grid size ($Å^3$)	64 × 64 × 64	60 × 60 × 60	60 × 60 × 80	60 × 60 × 60
Protein dielectric	2	2 or 8	2 or 8	2
Membrane dielectric	2	2.0	2.0	2.0
Headgroup dielectric	80	80	80	80
Solvent dielectric	80	80	80	80
Membrane thickness (Å)	42.5	39.9	42.0	39.9
Membrane bottom (Å)	−19.0	−19.95	−21.0	−19.95
Headgroup thickness (Å)	8.0	8.0	9.0	8.0
Upper/lower exclusion radii (Å)	16/13	18.5/18.5	12/12	12/12
Grid center	Origin	Origin	Origin	Origin
Solution method	Npbe	Npbe	Npbe	Npbe
Boundary condition	Zero	Zero	Zero	Zero
Membrane potential (mV)	0	0	0	0
Charge model	Spl2	Spl2	Spl2	Spl2
Surface model	Spl4	Mol	Mol	Mol
Surface spline width (Å)	0.3	NA	NA	NA
Solvent probe radius (Å)	NA	1.4	1.4	1.4
Surface sphere density ($Å^{-2}$)	NA	10	10	10
Ion initial position (Å)	(0, 0, −60)	NA	NA	NA
Ion final position (Å)	(0, 0, 60)	NA	NA	NA
No. of ion steps	100	NA	NA	NA

NA, not applicable.

energy well created by acidic amino acids that would otherwise trap Ca^{2+}. This result supports the hypothesis that decreased single-channel conductance in the K639Q channel results from longer Ca^{2+} dwell times in the primary binding sites. Stabilization in the selectivity filter is dominated by D646, E636, and to a lesser extent E648 (Figures 4A and 4B). Indeed, mutations that neutralize D646, E648, and E651 reduce Ca^{2+} permeability (Chung et al., 2008; Garcia-Martinez et al., 2000; Samways et al., 2008; Welch et al., 2000), and result in a loss of Ca^{2+} selectivity with respect to Na^+ (Samways et al., 2008). A more recent study showed that D646, E648, and E651 provide a strong Ca^{2+} binding site (Samways and Egan, 2011), which is in agreement with the deep electrostatic well shown in Figure 3B, but APBSmem does not reveal an electrostatic role for E651. Unlike the other acidic residues, E636Q causes a significant increase in the agonist induced fraction of total current carried by Ca^{2+} (Samways and Egan, 2011). It is often difficult to determine the kinetic properties of a channel, such as conduction rate and non-equilibrium selectivity, from equilibrium free energy profiles. However, these profiles can be coupled with simple kinetics models to reveal estimates of single-channel flux and differential flux for different ions, thus providing deeper mechanistic insight into how residues control channel properties.

Data suggest the existence of an ion binding site deeper in the pore than the ones stabilized by D646/E648/E651 in the filter (Chung et al., 2008). Our calculations show that the minimum in the central cavity at 17.6 Å is slightly more stable than the other sites, and the top contributor at this position is D576 (Figure 4C). Previously, the charge at this position was recognized as being important for capsaicin-dependent activation (Boukalova et al., 2010), but to our knowledge its importance in ion stabilization had gone unnoticed. We suggest that mutations at D576 may change channel conduction properties. Together, APBSmem provides an automated pipeline to gain inferences about new protein structures in which critical residues may not yet have been identified.

Next, we explored the influence that the protonation state of particular acidic residues had on the very stable energy profiles in Figure 3B. Using patch-clamp photometry and site-directed mutagenesis, Samways et al. (2008) demonstrated that protonation of residues D646, E648, and E651 significantly reduces the fraction of total current carried by Ca^{2+} in a manner indistinguishable from pH-dependent loss of selectivity, hinting at a mechanism for pH-dependent loss of Ca^{2+} selectivity in which these residues become protonated. They later estimated that the D646N/E648Q/E651Q triple mutant lacked any selectivity

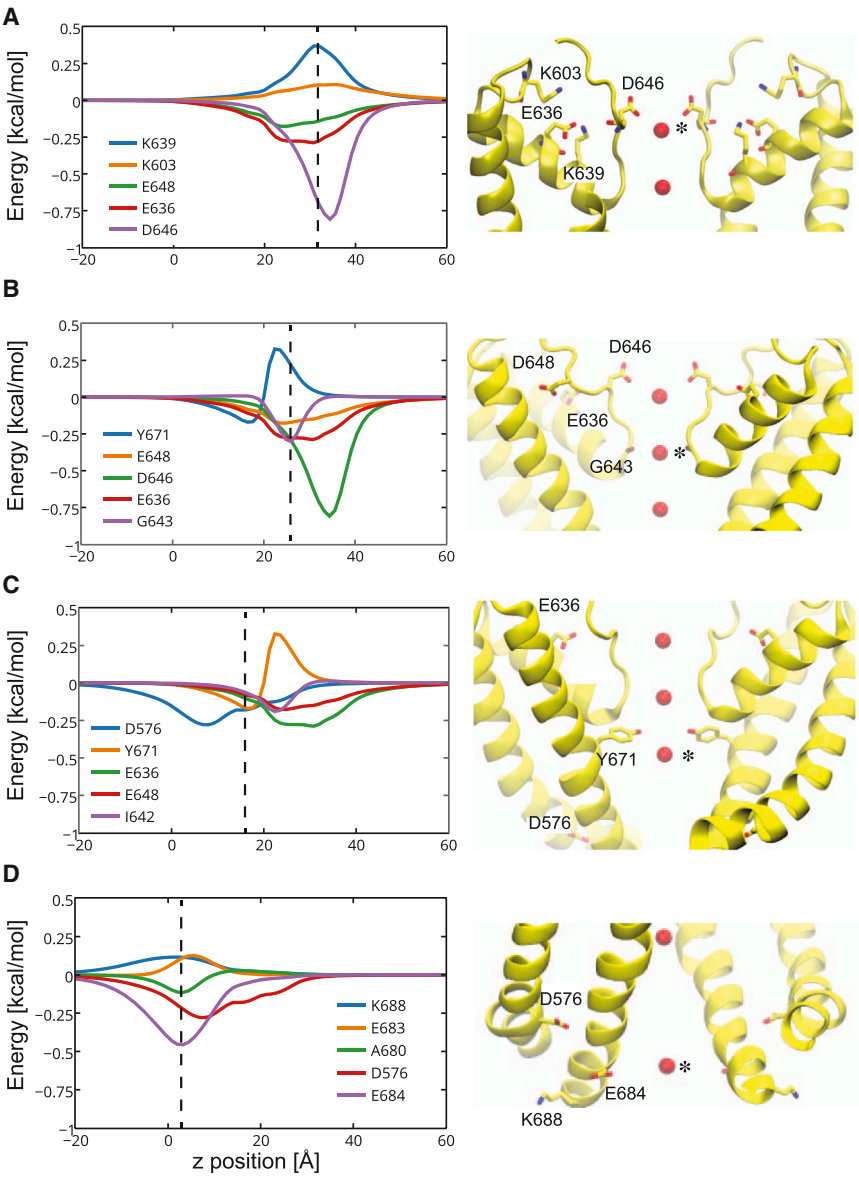

Figure 4. Critical Residues Contributing to the Electrostatics of Ion Permeation through TRPV1

The top five residues interacting with the permeating ion (by absolute value) at each of the positions (A, 32 Å; B, 24.8 Å; C, 17.6 Å; and D, 2 Å) identified in Figure 3. The vertical dashed line is the ion position of interest at which the rank order was compiled, but interaction strengths are plotted through the entire channel. The molecular image to the right of each graph shows the ion (red sphere with asterisk) at the z position corresponding to the dashed line. TRPV1 is yellow, with impactful residues rendered in stick mode. For clarity, only two subunits of the channel are represented, and a Ca^{2+} ion is pictured at all four positions for perspective, but calculations are performed with only a single ion in the channel.

for Ca^{2+} over Na^{+} (Samways and Egan, 2011). Our initial calculation showed that D646 is most important for ion stabilization, and since TRPV1 is a tetramer, there are four copies of this residue. We protonated each in turn and recomputed the corresponding Ca^{2+} and Na^{+} profiles. To facilitate such calculations, we created a dialog box so that charge states for individual residues can be set prior to each calculation using the integrated PROPKA plug-in (Li et al., 2005; Olsson et al., 2011). In the *Assign charge states* section of the GUI a button exists to select any amino acid in the protein, identifying them by their chemical name, residue number, and chain ID.

Neutralization of all four D646 residues results in a decrease in Ca^{2+} binding energy of 6.8 kcal/mol at the most extracellular site (32.0 Å) and an extracellular shift in the position (Figure 5A). The sites at 24.8 and 17.6 Å are also destabilized, but by a much smaller amount: 4 and 0.4 kcal/mol, respectively. Protonation changes in the protein can therefore dramatically influence binding energy and ion dwell times at positions along the pore. In addition, we carried out the electrostatic free energy calculations on a sodium-like cation (Figure 5B). Even when all four D646 residues are charged, the binding energy is much smaller due to the reduced valency of sodium, and the most extracellular site is barely present. Correspondingly, protonating D646 has less of an impact on the energy profile, and a rough comparison of the binding energy changes at each site reveal an energy difference between Ca^{2+} and Na^{+} of −4.5, −2.3, and −3.7 kcal/mol prior to protonation and 1.7, 0.4, and −3.4 kcal/mol after full protonation at the extracellular filter site, intracellular filter site, and the cavity site, respectively (Figure 5C). Thus, the electrostatic changes alone show a dramatic loss of binding energy for Ca^{2+} compared with Na^{+} when the filter sites are protonated in response to a drop in pH. These calculations corroborate the loss of selectivity as the pH decreases, as reported previously (Chung et al., 2008; Samways and Egan, 2011; Samways et al., 2008).

Case III: Determination of Membrane-Induced pK_a Shifts

The protonation state of a residue can be influenced by the local electrical environment, changes in pH (as discussed in case II), and the dielectric environment. There is significant literature centered on the use of continuum electrostatics for predicting pK_a shifts of residues, since charge changes can impact protein structure, ligand binding, and protein-protein interactions (see Alexov et al., 2011), and earlier studies have used solutions to the PB equation to explore pK_a shifts in the presence of the membrane for membrane proteins such as bacteriorhodopsin (Bashford and Gerwert, 1992) and outer membrane porins (Karshikoff et al., 1994). When a charged group moves from a high dielectric environment, like water, into a low-dielectric medium,

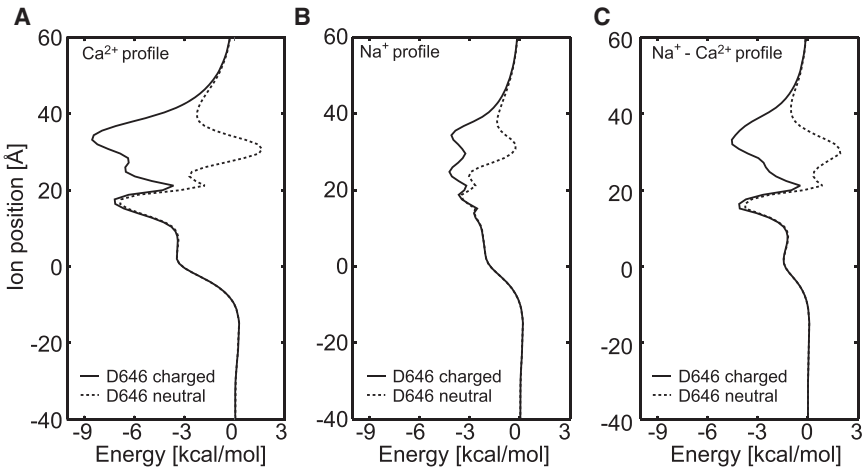

Figure 5. Selectivity Is Influenced by Protonation State

(A and B) Ion stepping profile with zero to four D646 residues neutralized for Ca^{2+} and Na^{+}, respectively. Note that the energies near 32 Å for both ions become comparable once all D646 residues are neutralized.

(C) Energy difference between profiles shown in (A) and (B). This energy is the Ca^{2+} energy minus the Na^{+} energy.

such as the bilayer core, there is an electrostatic penalty, which can be thought of as the energy associated with stripping away polar water molecules from the protein. Neutralizing the residue can mitigate this energetic cost, and if the resulting energy decrease is greater than the free energy of ionization, the group will likely be neutral in the membrane. As shown in Figure S1, we calculate these shifts using two thermodynamic cycles for (de)protonation of a residue of interest in solution (ΔpK_a^1, cycle 1) and a cycle corresponding to a change in charge state in the membrane (ΔpK_a^2, cycle 2). PROPKA is used to compute the shift along cycle 1 (Li et al., 2005; Olsson et al., 2011), and APBSmem is used to estimate the shift due to the membrane using the PB approach developed by Honig and co-workers (Yang et al., 1993) as described in the Supplemental Information. Thus, the pK_a of a residue is given by

$$pK_a = pK_a^0 + \Delta pK_a^1 + pK_a^2, \qquad \text{(Equation 3)}$$

where pK_a^0 is the experimentally determined pK_a of the isolated residue. APBSmem will calculate the ΔpK_a of a single residue, or will provide a rank-ordered list of residues most likely to be shifted based on a single, heuristic approximation presented next.

Here, we demonstrate APBSmem's ΔpK_a calculator with mVDAC1 and LeuT, for which timings and memory usage can be found in Table S1. As an initial evaluation, we ran a single solvation energy calculation for mVDAC1 with all residues set to standard protonation states at pH 7 to obtain the per-residue components of the solvation energy (Figure 6A). For these calculations, we subtract the total fixed-charge energy of mVDAC1 in solution from the membrane-embedded value and report the per-residue contribution. While not a direct indicator of ΔpK_a, large energy values might be reduced if the residue is neutralized. This quick calculation singled out E73 and K110 as having the largest solvation energies. The side chain of E73 is oriented toward the hydrophobic core of the membrane, as suggested earlier by De Pinto et al. (1993), while K110 points toward the bilayer at the headgroup-core interface (Figure 6C). Figure 6B shows membrane-induced residue pK_a shifts for all R, K, D, E, Y, and C residues in mVDAC1 (ΔpK_a^2), assuming a protein dielectric of 2 and 8, since it is often debated which dielectric value is most appropriate (Kukic et al., 2013). As predicted by the fixed energy solvation values, both E73 and K110 have the most profound shifts, regardless of the protein dielectric constant employed. The modified pK_a of E73 and K110 are 25.37 and −0.87, respectively (Table 2). Values above 7 indicate protonation at neutral pH while values below 7 indicate a lack of protonation; thus, we predict that K110 and E73 are both neutral. Our findings corroborate recent molecular dynamics (MD) simulations showing that E73 causes bending and water penetration into the membrane when charged (Villinger et al., 2010). However, simulations have not reported deprotonation of K110 (Choudhary et al., 2014; Noskov et al., 2013; Rui et al., 2011; Villinger et al., 2010), potentially due to electrostatic compensation caused by snorkeling into the headgroups.

The per-residue solvation energies for the LeuT transporter reveal that K288 is an outlier (Figure 6D). The membrane-induced pK_a shift is −12.64 (Figure 6E), indicating that it is most likely neutral, which is consistent with its position in the bilayer core (Figure 6F). Not surprisingly, MD simulations with K288 charged result in membrane deformations, water penetration, and membrane thinning at the site (Mondal et al., 2013, 2014), and these deformations could help keep the residue charged (Callenberg et al., 2012; Li et al., 2008; Mondal et al., 2013, 2014; Yoo and Cui, 2008).

We wish to emphasize that the membrane-induced pK_a shifts described here are a first-order approximation of a more complete statistical-mechanical treatment, which accounts for the interaction of multiple charged sites in all possible ionization configurations (Yang et al., 1993). This full treatment can be done using the shell scripting feature with APBSmem, but except for the smallest proteins this calculation can be extremely time consuming; thus, it is useful to consider solving this problem with approximate methods (Bashford and Karplus, 1991; Tanford and Roxby, 1972) or Monte Carlo-based approaches (Beroza et al., 1991). Furthermore, charged side chains are often able to reduce their electrostatic penalty either by snorkeling into the headgroup region or by forming hydrogen bonds with nearby residues that would otherwise not form if the protein were in aqueous solution. Thus, in addition to exploring the full ensemble of possible titration states, one would also need to optimize side-chain conformations to obtain a more accurate pK_a.

Case IV: Electrostatic Survey of Membrane Proteins of Known Structure

We combined the automatic membrane detection algorithm with command-line scripting to expand our per-residue solvation

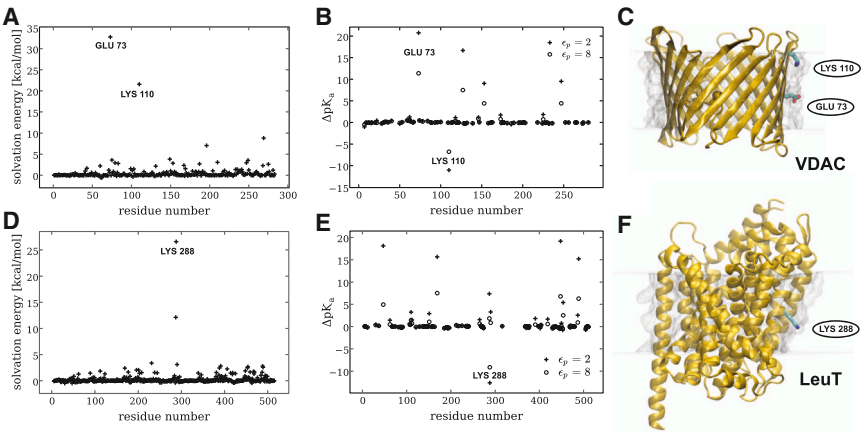

Figure 6. Membrane-Induced pK_a Shifts

(A) mVDAC1 (PDB: 3emn) per-residue solvation energy contributions.

(B) Membrane-induced pK_a (ΔpK_a^2) shifts for mVDAC1 calculated with the protein dielectric set to 2 (plus signs) or 8 (circles).

(C) Image of mVDAC1 showing E73 poking into the membrane and K110 in the headgroup.

(D) LeuT (PDB: 2a65) per-residue solvation energy contributions.

(E) Membrane-induced pK_a shifts for LeuT with the protein dielectric set to 2 (plus signs) or 8 (circles).

(F) LeuT showing K288 poking out into the membrane.

energy analysis to 1,614 multi-pass membrane proteins available in the OPM (Orientations of Proteins in Membranes) database (Lomize et al., 2006). For each protein, we identified residues that incurred a 10-kcal/mol or greater electrostatic penalty for residing in the membrane, assuming standard protonation states at pH 7. We used the membrane thicknesses determined by OPM for each protein. While most membrane proteins have no residues (41%) or one residue (15%) in violation, APBSmem identified 707 proteins that have five or more residues in violation (44%). For instance, the method correctly identified all 12 copies of the titratable rotor site (D61) on the rotor domain of the F_0-ATPase (PDB: 1c17), whose protonation is crucial for ion transport (Rastogi and Girvin, 1999). In Figure 7A, we picture three of the top five structures with the most violations: the KvAP voltage-gated potassium channel (PDB: 2a0l; Jiang et al., 2003) with 52 penalties, the capsaicin-bound TRPV1 channel (PDB: 3j5r; Liao et al., 2013) with 25 penalties, and the mechanosensitive channel of large conductance MscL (PDB: 3hzq; Liu et al., 2009b), with 24 penalties. The physiological relevance of the KvAP structure is not clear (Cohen et al., 2003) and many of the high-energy residues are buried deep in what would be the core of the membrane, but offending residues are located near the headgroup interface for both MscL and TRPV1. It is possible that minor membrane bending could accommodate the residual hydrophobic mismatch for these proteins. In the future, we will explore membrane-bending effects with our implicit membrane-bending model (Callenberg et al., 2012).

We next categorized the proteins by family using the Mpstruct database (http://blanco.biomol.uci.edu/mpstruc). Figure 7B shows the proportion of α-helical structures evaluated that had at least five residues with electrostatic insertion penalties greater than 10 kcal/mol for the top 20 families. The majority of these groups are transporters and ion channels, but we also identified families involved in the electron transport chain and light harvesting. For example, APBSmem indicates large penalties for E78 and R207 in cytochrome b6f, which line the proton transfer pathway (Hasan et al., 2013). Thus, α-helical proteins that move charge across membranes appear to have an increased number of charged residues in the transmembrane region that are energetically costly. In contrast, β-barrel proteins generally have fewer residues with large electrostatic penalties. However, many of these proteins also facilitate charge movement, and we hypothesize that the large water-filled cavities found in porins reduce the electrostatic fields and corresponding energy penalties. Only the PagP outer membrane palimitoyl transferase (PDB: 1mm4; Hwang et al., 2002) returned five or more high-energy residues, and these residues are located on the loops of the barrel and at the headgroup-core interfaces. We found only a very weak correlation ($R^2 < 0.1$) between the resolution of the structures and the number of electrostatically unfavorable residues for α-helical proteins, and there was no correlation for β barrels (Figures S2A–S2E). That said, the five structures with the highest number of reported residues have resolutions greater than 3.2 Å.

Finally, we categorized the total number of residues with large electrostatic insertion penalties for both α-helical (Figure 7C) and

Table 2. Predicted pK_a Shifts for Select Charged Groups

Protein	Residue	Calculation	Method	pK_a
mVDAC1	Glu73	pK_a^0 isolated amino acid in solution	Experiment	4.50
		ΔpK_a^1 in protein in solution	PROPKA	0.31
		ΔpK_a^2 solution to membrane	APBSmem	20.74
		modified pK_a	Equation 3	25.55
mVDAC1	Lys110	pK_a^0 isolated amino acid in solution	Experiment	10.50
		ΔpK_a^1 in protein in solution	PROPKA	−0.26
		ΔpK_a^2 solution to membrane	APBSmem	−11.03
		modified pK_a	Equation 3	−0.79
LeuT	Lys288	pK_a^0 isolated amino acid in solution	Experiment	10.50
		ΔpK_a^1 in protein in solution	PROPKA	−0.20
		ΔpK_a^2 solution to membrane	APBSmem	−12.64
		modified pK_a	Equation 3	−2.34

Membrane dielectric is 2 for all calculations.

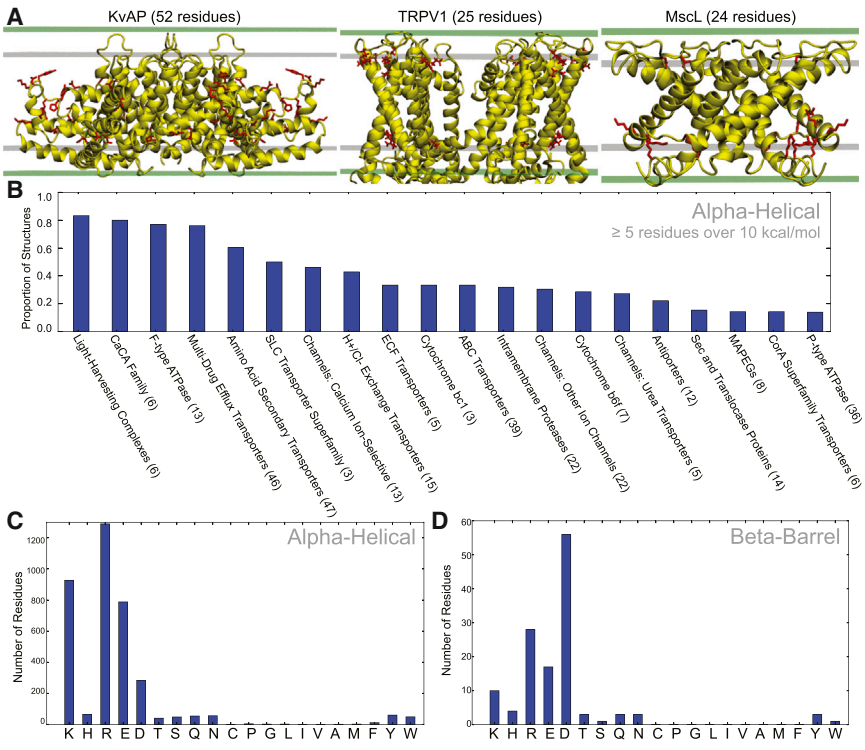

Figure 7. Electrostatic Scan of Multi-pass Membrane Proteins

(A) Three of the structures with the largest number of electrostatically unfavorable residues: KvAP, TRPV1, and MscL (PDB: 2a0l, 3j5r, and 3hzq, respectively). Residues with greater than 10 kcal/mol electrostatic insertion penalty are shown in red licorice, with the total number of offending residues given in parentheses. The water-headgroup and headgroup-tail interfaces are shown as green and white surfaces, respectively.

(B) Proportion of α-helical structures with five or more residues characterized as electrostatically unfavorable. Families are defined according to the Mpstruct database, and we excluded families with fewer than three structures used in the final calculations. The total number of structures analyzed for each family is given in parentheses.

(C and D) Total number of electrostatically unfavorable residues for α-helical and β-barrel proteins, respectively. In total, calculations involved 794 α-helical proteins and 215 β-barrel proteins.

β-barrel (Figure 7D) proteins. As expected, the majority of identified residues are charged, and α-helical proteins contain more basic residues while the most prevalent residue in β barrels is aspartate. Clearly many of these residues will be neutralized in the membrane, but we note that some residues we identified are not titratable. For example, the backbone of residue S2 in mitochondrial cytochrome *c* oxidase (PDB: 2zxw; Aoyama et al., 2009) is exposed to the membrane core, giving the residue an electrostatic penalty of 11.0 kcal/mol. Table S2 contains the full list of residues with large solvation energy penalties greater than or equal to 10 kcal/mol, and details on the calculations can be found in the Supplemental Information.

Case V: Prediction of Membrane Protein Insertion Energies

Next, we use APBSmem to explore the stability of proteins in the membrane. There are several computational methods available to quantitatively assess the energy of partitioning from water into the membrane, including computationally expensive fully atomistic MD simulations (Dorairaj and Allen, 2007; MacCallum et al., 2007), more tractable physics-based approaches (Ben-Tal et al., 1996; Lomize et al., 2006), and statistical potentials (Bernsel et al., 2008; Schramm et al., 2012). We employ the method outlined by Honig and coworkers, which assumes that the non-polar energy of insertion (ΔE_{np}) is proportional to the surface area of the molecule (Sitkoff et al., 1994):

$$\Delta E_{np} = a \cdot (A_{mem} - A_{sol}), \qquad \text{(Equation 4)}$$

where A_{mem} is the solvent accessible surface area (SASA) of the protein or molecule in the membrane and A_{sol} is the SASA in solution, and a = 0.028 kcal/mol/Å^2 is a constant of proportionality. Both areas are calculated by APBSmem using a call to the external program MSMS (Sanner et al., 1996), which must be separately downloaded and installed locally. APBSmem includes three models for calculating ΔE_{np}: (1) all surface atoms between the upper and lower leaflets are included in the calculation of A_{mem}; (2) only atoms in the hydrophobic core of the membrane are included; or (3) the constant a in the headgroup regions is linearly scaled from 0.28 kcal/mol/Å^2 to zero as the z position of a buried atom approaches the membrane-water interface from the headgroup-core interface. The latter model is consistent with the observation that water penetration falls off linearly in the headgroup region (Nagle and Tristram-Nagle, 2000), and this is the model we employ here.

Predicting the relative stability of membrane-spanning domains is critical in understanding the balance between membrane protein biosynthesis and quality control. As integral membrane proteins are translated and inserted into the ER, their TMs must adopt the proper topology and interact correctly with subsequent TMs to ensure native folding and function (Skach, 2009). Failure to fold can result in protein destruction by ER-associated degradation (ERAD), a quality control pathway that triages misfolded proteins (Needham and Brodsky, 2013). Folding of multi-pass membrane proteins is highly problematic due to the number of membrane-spanning domains and the complexity of intermembrane interactions, especially for ATP binding cassette (ABC) transporters, which possess 12 TMs and two large cytoplasmic nucleotide binding domains (NBDs). Indeed, destabilizing mutations within ABC transporters lead to a number of diseases (Guerriero and Brodsky, 2012).

A model misfolded ABC transporter is a truncated form of the yeast mating pheromone transporter Sterile 6 (Ste6p*). Following translation, wild-type Ste6p* traffics to the plasma membrane; however, a 42-amino-acid truncation in the second NBD (NBD2) results in ER retention and destruction by the ERAD

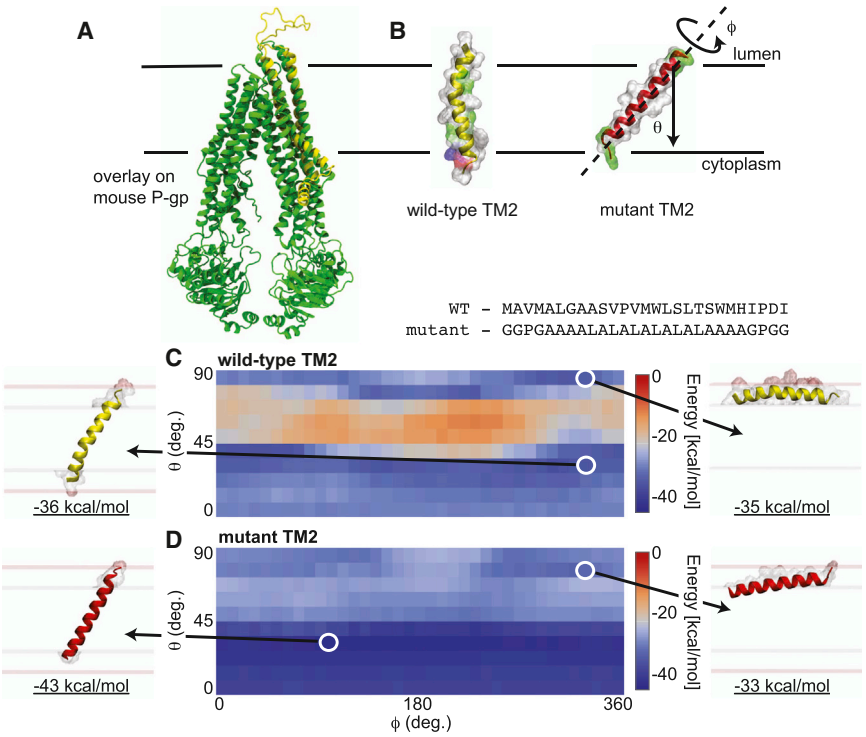

Figure 8. Membrane Protein Insertion Energies

(A) Superposition of homology model of Ste6p* 2TM construct (yellow) on the P-gp template structure (green).

(B) Model of isolated TM2 from wild-type (yellow on left) and poly-alanine/leucine construct (red on right) with molecular surface showing amino acid chemistry (white, hydrophobic; green, polar; blue, basic). Coordinate system corresponds to (C) and (D). The primary sequence of each construct is shown at the bottom.

(C and D) Insertion energy heatmap for wild-type TM2 (C) and the artificial hydrophobic TM2 (D). The total energy consists of electrostatic and non-polar terms. The most stable configurations are fully inserted ($\theta < 30^\circ$) or interfacial (θ near 90°). Minimum energy configurations of each orientation are depicted to the left and right, respectively. The headgroup-water interface is pink and the headgroup-core interface is white.

pathway (Loayza et al., 1998). To model ABC transporter TM insertion, folding, and quality control, we created an internal deletion of Ste6p* to remove all but the first two TMs, which were then appended to the truncated NBD2. This species was termed Chimera N*. When expressed in yeast cells, the native TM2 in Chimera N* fails to partition into the membrane (data not shown). However, proper topology is corrected by substitution of TM2 with an artificial poly-A/L hydrophobic stretch offset by helix-terminating linkers (Hessa et al., 2007) (Figure 8).

Because a high-resolution structure of Ste6p* is lacking, we created a homology model of the wild-type and mutant constructs based on the related P-glycoprotein (P-gp) ABC transporter (PDB: 3g5u, 26% identity; Aller et al., 2009) (see Supplemental Information and Figure S3 for details on model construction). Superposing the 2TM model of Ste6p* (yellow) on the P-gp transporter (green) reveals that both helices have extensive interactions with other TM segments (Figure 8A), which may explain why Chimera N* fails to adopt the correct topology. To explore the energetic stability of the isolated wild-type and artificial hydrophobic TM2 segments, we calculated the sum of the non-polar (Equation 4) and electrostatic energies (from Equation 1) for each segment at different positions in the membrane, with each compared with their respective value in solution. We used command-line scripting to rotate each segment through a wide range of positions, including fully transmembrane, titled, and interfacial configurations, with the hypothesis that the native TM2 sequence may be predisposed to adopt an interfacial configuration, while the engineered TM2 may be more stable in the fully inserted state. An energetic heatmap was created for each helix by rotating 360° along the long axis (ϕ) and then pivoting the helix from 0° to 90° with respect to the membrane normal (θ) while pinning the N terminus as shown in Figure 8B. See the Supplemental Information for a complete description of the electrostatic energy and scripting.

As hypothesized, the native TM2 is significantly more stably bound to the membrane than in aqueous solution, with nearly equal values in interfacial (−35 kcal/mol) and TM configurations (−36 kcal/mol). The overall stability in or near the membrane is not surprising given the hydrophobic character of the primary sequence, and the similarity between these two energetic values may explain why this segment fails to insert. Meanwhile, the artificial hydrophobic TM2 segment is more stable in the TM (−43 kcal/mol) than the interfacial configuration (−33 kcal/mol) due to the poly-A/L sequence. The 10-kcal/mol increased stability of the TM configuration likely explains its correct topological insertion.

DISCUSSION

We have explored a diverse set of biological problems related to the electrostatics and stability of membrane proteins. In doing so, we used several new features of the APBSmem software that significantly enhance its usability and power. The ability to manipulate protein orientation and position coupled with the bundling of PDB2PQR now makes it possible to initiate calculations entirely within APBSmem directly from a PDB file without the use of external software packages. Added functionality for ion placement and movement allows users to easily explore the electrostatics of ion movement through channels. Moreover, increased handling of energetic terms makes it possible to extract specific interactions between protein residues and ions or small molecules in the system, which helps identify amino acids critical to permeation or binding. We also developed a scheme to determine the membrane-induced shifts in residue pK_a values in combination with existing methods in the incorporated PROPKA software. Using newly added command-line scripting, we compiled a comprehensive list of residues likely to have altered protonation states for all integral membrane

proteins in the OPM database. Finally, a more complete model of membrane protein stability that includes non-polar energies can now be calculated with APBSmem, since it interfaces with the program MSMS. While other energetic terms such as protein conformational change, entropy, and membrane distortions are ignored, non-polar and electrostatic energies alone can provide a first approximation of protein stability. Thus, the ease and speed of APBSmem coupled with its ability to predict changes at the single amino acid level make it a first-line approach for exploring the stability of membrane and membrane-associated proteins.

Throughout this study we have assumed that the membrane remains flat and undeformed, and there are instances where this will not be true. An ongoing effort in our laboratory is to incorporate membrane deformations into a continuum framework consistent with APBSmem (Callenberg et al., 2012), and future releases will include this feature. In addition, the parameters listed in Table 1 are based on typical continuum electrostatics calculations. For example, the high dielectric value of 80 used in the headgroups is inspired by fully atomistic simulations carried out in the Feller laboratory (Stern and Feller, 2003), but they are not intended to be applicable to all situations. The APBSmem GUI makes it easy for researchers to explore different values. In particular, it will be interesting to use APBSmem with the flexibility provided through batch scripting to benchmark known pK_a shifts for membrane protein residues to identify optimal model parameters and test quantitative aspects of our method.

The latest free software version of APBSmem can be downloaded from http://apbsmem.sourceforge.net.

EXPERIMENTAL PROCEDURES

All calculations were carried out with APBSmem, which is a Java-based program that aids in solving the PB equation in the presence of membrane-like environments. It can be run from a GUI or from the command line using prespecified input files. PDB files were loaded into the software, and charge models were set using the bundled PDB2PQR package. Non-standard protonation states of specific residues were assigned with PROPKA, which is also now bundled with APBSmem. Through the GUI interface calculation parameters were set, including calculation type (ion solvation energy, gating charge/voltage dependence calculations, membrane insertion energy, pK_a shifts, etc.), spatial dimensions and grid discretization, membrane/protein/solution dielectric values, protein surface representation, and far-field boundary conditions, as well as other parameters typical of molecular PB calculations. Once all parameters are set, APBSmem calls APBS to generate an initial dielectric environment map (ε), ion accessibility map (κ), and explicit charge map (ρ), based on the molecular coordinates and any explicit ions in solution. These maps are then manipulated by APBSmem to include the presence of an implicit membrane in the dielectric and ion accessibility maps, as well as the charge maps if a membrane potential is imposed on the system. The *Preview* button allows the user to quickly visualize the protein's placement in the membrane, which is crucial at this stage to ensure proper orientation in the membrane with the desired membrane protein boundaries. APBSmem next calls on APBS, again using the altered maps to numerically solve the PB equation in Equation 1. This flow makes APBSmem operation transparent: after an APBSmem calculation, calculations can be repeated without using APBSmem by simply running APBS on the generated input files. An in-depth discussion of the basic features of the software is provided in our original paper (Callenberg et al., 2010).

Per-residue solvation energies in Figures 6 and 7 are calculated as the sum of the fixed-charge energies for all residue atoms computed in the presence of the membrane subtracted from the fixed-charge energy when the protein is in solution. Per-residue ion interaction energies are calculated by summing the per-atom fixed-charge energies over each atom in a residue. This energy value is formally divergent because it evaluates the potential at the position of the atoms, so APBSmem isolates the residue ion interaction energy by subtracting the protein-protein energies. These per-residue interaction energies are saved to a log file in the output directory and used to produce the curves in Figure 4.

pK_a shifts are calculated from a set of thermodynamic cycles. We consider two cycles: first, the energy required to protonate/deprotonate the residue of interest in solution (cycle 1), and second, the energy required to protonate/deprotonate the residue in the presence of the membrane (cycle 2). We use PROPKA to compute the values along cycle 1. APBSmem solves the PB equation to determine the change in total electrostatic energy for inserting a protein into the membrane with the residue deprotonated and the change in energy with the residue protonated. This difference between these energies is used to calculate the shift in pK_a along cycle 2. Details of these cycles are further described in Figure S1 and the Supplemental Information.

Homology models of the TM domains (TM1 and TM2) of the ABC transporter Ste6p* described in case V were constructed with Modeller9v13 (Sali and Blundell, 1993) using the P-gp transporter (PDB: 3g5u) as a template structure and the alignment provided in Figure S3. Additional hand adjustments were then carried out to close up gaps and maximally align the second TM regions. The final alignment used for construction of the first two TM segments of Ste6p* and the wild-type TM2 segment is shown in Figure S3. The mutant TM2 segment was then constructed using the wild-type TM2 model as a template. Please note that for the TM insertion energy calculations these alignments are not crucial, since we simply assume that both TM2 segments take on roughly straight helical configurations.

The new features added to APBSmem that are used throughout this study are more fully described in the Supplemental Information and include enhanced PDB/PQR file processing, greater geometric control over proteins, ions, and small molecules, automatic identification of the membrane, non-polar energy calculations with MSMS, per-residue contributions to energies, command-line scripting, and ligand solvation energy calculations.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, Supplemental Calculation Methodology, three figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2015.05.014.

AUTHOR CONTRIBUTIONS

F.V.M., N.B., and M.G. designed all aspects of the research project, while C.J.G. and J.L.B. aided in the design of case V. F.V.M. and N.B. programmed the new functionality into APBSmem code and carried out all calculations. All authors contributed to writing and editing the manuscript.

ACKNOWLEDGMENTS

We thank Keith Callenberg and Naomi Latorraca for their help in this project, and Marco Lolicato for his help in testing the software. We also thank Nathan Baker and Keith Star for continued support concerning the integration of APBSmem and APBS. Finally, we thank Yifan Cheng for his helpful discussions concerning the electrostatic calculations on TRPV1. This work was funded by NSF CAREER award MCB-0845286 (M.G.), NIH grant DK101584 (C.J.G.), and NIH grants GM75061 and DK79307 (J.L.B.).

Received: January 16, 2015
Revised: April 9, 2015
Accepted: May 2, 2015
Published: June 25, 2015

REFERENCES

Adelman, J.L., Sheng, Y., Choe, S., Abramson, J., Wright, E.M., Rosenberg, J.M., and Grabe, M. (2014). Structural determinants of water permeation

through the sodium-galactose transporter vSGLT. Biophys. J. *106*, 1280–1289.

Alexov, E., Mehler, E.L., Baker, N., Baptista, A.M., Huang, Y., Milletti, F., Nielsen, J.E., Farrell, D., Carstensen, T., Olsson, M.H., et al. (2011). Progress in the prediction of pKa values in proteins. Proteins *79*, 3260–3275.

Aller, S.G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., Harrell, P.M., Trinh, Y.T., Zhang, Q., Urbatsch, I.L., et al. (2009). Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science *323*, 1718–1722.

Aoyama, H., Muramoto, K., Shinzawa-Itoh, K., Hirata, K., Yamashita, E., Tsukihara, T., Ogura, T., and Yoshikawa, S. (2009). A peroxide bridge between Fe and Cu ions in the O_2 reduction site of fully oxidized cytochrome c oxidase could suppress the proton pump. Proc. Natl. Acad. Sci. USA *106*, 2165–2169.

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

Bashford, D., and Karplus, M. (1991). Multiple-site titration curves of proteins—an analysis of exact and approximate methods for their calculation. J. Phys. Chem. *95*, 9556–9561.

Bashford, D., and Gerwert, K. (1992). Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. J. Mol. Biol. *224*, 473–486.

Ben-Tal, N., Ben-Shaul, A., Nicholls, A., and Honig, B. (1996). Free-energy determinants of alpha-helix insertion into lipid bilayers. Biophys. J. *70*, 1803–1812.

Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G., and Elofsson, A. (2008). Prediction of membrane-protein topology from first principles. Proc. Natl. Acad. Sci. USA *105*, 7177–7181.

Beroza, P., Fredkin, D.R., Okamura, M.Y., and Feher, G. (1991). Protonation of interacting residues in a protein by a Monte Carlo method: application to lysozyme and the photosynthetic reaction center of *Rhodobacter sphaeroides*. Proc. Natl. Acad. Sci. USA *88*, 5804–5808.

Boukalova, S., Marsakova, L., Teisinger, J., and Vlachova, V. (2010). Conserved residues within the putative S4-S5 region serve distinct functions among thermosensitive vanilloid transient receptor potential (TRPV) channels. J. Biol. Chem. *285*, 41455–41462.

Brooks, B.R., Brooks, C.L., 3rd, Mackerell, A.D., Jr., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al. (2009). CHARMM: the biomolecular simulation program. J. Comput. Chem. *30*, 1545–1614.

Callenberg, K.M., Choudhary, O.P., de Forest, G.L., Gohara, D.W., Baker, N.A., and Grabe, M. (2010). APBSmem: a graphical interface for electrostatic calculations at the membrane. PLoS One *5*, http://dx.doi.org/10.1371/journal.pone.0012722.

Callenberg, K.M., Latorraca, N.R., and Grabe, M. (2012). Membrane bending is critical for the stability of voltage sensor segments in the membrane. J. Gen. Physiol. *140*, 55–68.

Cao, E., Liao, M., Cheng, Y., and Julius, D. (2013). TRPV1 structures in distinct conformations reveal activation mechanisms. Nature *504*, 113–118.

Caterina, M.J., Schumacher, M.A., Tominaga, M., Rosen, T.A., Levine, J.D., and Julius, D. (1997). The capsaicin receptor: a heat-activated ion channel in the pain pathway. Nature *389*, 816–824.

Choudhary, O.P., Paz, A., Adelman, J.L., Colletier, J.P., Abramson, J., and Grabe, M. (2014). Structure-guided simulations illuminate the mechanism of ATP transport through VDAC1. Nat. Struct. Mol. Biol. *21*, 626–632.

Chung, M.K., Guler, A.D., and Caterina, M.J. (2008). TRPV1 shows dynamic ionic selectivity during agonist stimulation. Nat. Neurosci. *11*, 555–564.

Cohen, B.E., Grabe, M., and Jan, L.Y. (2003). Answers and questions from the KvAP structures. Neuron *39*, 395–400.

De Pinto, V., al Jamal, J.A., and Palmieri, F. (1993). Location of the dicyclohexylcarbodiimide-reactive glutamate residue in the bovine heart mitochondrial porin. J. Biol. Chem. *268*, 12977–12982.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., and Baker, N.A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. *32*, W665–W667.

Dorairaj, S., and Allen, T.W. (2007). On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. Proc. Natl. Acad. Sci. USA *104*, 4943–4948.

Fogolari, F., Brigo, A., and Molinari, H. (2002). The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. J. Mol. Recognit. *15*, 377–392.

Garcia-Martinez, C., Morenilla-Palao, C., Planells-Cases, R., Merino, J.M., and Ferrer-Montiel, A. (2000). Identification of an aspartic residue in the P-loop of the vanilloid receptor that modulates pore properties. J. Biol. Chem. *275*, 32552–32558.

Gilson, M.K., and Honig, B.H. (1987). Calculation of electrostatic potentials in an enzyme active site. Nature *330*, 84–86.

Grant, J.A., Pickup, B.T., and Nicholls, A. (2001). A smooth permittivity function for Poisson-Boltzmann solvation methods. J. Comput. Chem. *22*, 608–640.

Guerriero, C.J., and Brodsky, J.L. (2012). The delicate balance between secreted protein folding and endoplasmic reticulum-associated degradation in human physiology. Physiol. Rev. *92*, 537–576.

Hasan, S.S., Yamashita, E., Baniulis, D., and Cramer, W.A. (2013). Quinone-dependent proton transfer pathways in the photosynthetic cytochrome b6f complex. Proc. Natl. Acad. Sci. USA *110*, 4297–4302.

Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S.H., and von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature *450*, 1026–1032.

Hwang, P.M., Choy, W.Y., Lo, E.I., Chen, L., Forman-Kay, J.D., Raetz, C.R., Prive, G.G., Bishop, R.E., and Kay, L.E. (2002). Solution structure and dynamics of the outer membrane enzyme PagP by NMR. Proc. Natl. Acad. Sci. USA *99*, 13560–13565.

Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B.T., and MacKinnon, R. (2003). X-ray structure of a voltage-dependent K^+ channel. Nature *423*, 33–41.

Karshikoff, A., Spassov, V., Cowan, S.W., Ladenstein, R., and Schirmer, T. (1994). Electrostatic properties of two porin channels from *Escherichia coli*. J. Mol. Biol. *240*, 372–384.

Kukic, P., Farrell, D., McIntosh, L.P., Garcia-Moreno, E.B., Jensen, K.S., Toleikis, Z., Teilum, K., and Nielsen, J.E. (2013). Protein dielectric constants determined from NMR chemical shift perturbations. J. Am. Chem. Soc. *135*, 16968–16976.

Li, H., Robertson, A.D., and Jensen, J.H. (2005). Very fast empirical prediction and rationalization of protein pKa values. Proteins *61*, 704–721.

Li, L., Vorobyov, I., MacKerell, A.D., Jr., and Allen, T.W. (2008). Is arginine charged in a membrane? Biophys. J. *94*, L11–L13.

Liao, M., Cao, E., Julius, D., and Cheng, Y. (2013). Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature *504*, 107–112.

Liu, B., Yao, J., Wang, Y., Li, H., and Qin, F. (2009a). Proton inhibition of unitary currents of vanilloid receptors. J. Gen. Physiol. *134*, 243–258.

Liu, Z., Gandhi, C.S., and Rees, D.C. (2009b). Structure of a tetrameric MscL in an expanded intermediate state. Nature *461*, 120–124.

Loayza, D., Tam, A., Schmidt, W.K., and Michaelis, S. (1998). Ste6p mutants defective in exit from the endoplasmic reticulum (ER) reveal aspects of an ER quality control pathway in *Saccharomyces cerevisiae*. Mol. Biol. Cell *9*, 2767–2784.

Lomize, M.A., Lomize, A.L., Pogozheva, I.D., and Mosberg, H.I. (2006). OPM: orientations of proteins in membranes database. Bioinformatics *22*, 623–625.

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

Mondal, S., Khelashvili, G., Shi, L., and Weinstein, H. (2013). The cost of living in the membrane: a case study of hydrophobic mismatch for the multi-segment protein LeuT. Chem. Phys. Lipids *169*, 27–38.

Mondal, S., Khelashvili, G., and Weinstein, H. (2014). Not just an oil slick: how the energetics of protein-membrane interactions impacts the function and organization of transmembrane proteins. Biophys. J. *106*, 2305–2316.

Nagle, J.F., and Tristram-Nagle, S. (2000). Structure of lipid bilayers. Biochim. Biophys. Acta *1469*, 159–195.

Needham, P.G., and Brodsky, J.L. (2013). How early studies on secreted and membrane protein quality control gave rise to the ER associated degradation (ERAD) pathway: the early history of ERAD. Biochim. Biophys. Acta *1833*, 2447–2457.

Nina, M., Beglov, D., and Roux, B. (1997). Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. J. Phys. Chem. B *101*, 5239–5248.

Noskov, S.Y., Rostovtseva, T.K., and Bezrukov, S.M. (2013). ATP transport through VDAC and the VDAC-tubulin complex probed by equilibrium and nonequilibrium MD simulations. Biochemistry *52*, 9246–9256.

Olsson, M.H.M., Sondergaard, C.R., Rostkowski, M., and Jensen, J.H. (2011). PROPKA3: consistent treatment of internal and surface residues in empirical pK(a) predictions. J. Chem. Theor. Comput. *7*, 525–537.

Payandeh, J., Gamal El-Din, T.M., Scheuer, T., Zheng, N., and Catterall, W.A. (2012). Crystal structure of a voltage-gated sodium channel in two potentially inactivated states. Nature *486*, 135–139.

Rastogi, V.K., and Girvin, M.E. (1999). Structural changes linked to proton translocation by subunit c of the ATP synthase. Nature *402*, 263–268.

Robertson, J.L., Palmer, L.G., and Roux, B. (2008). Long-pore electrostatics in inward-rectifier potassium channels. J. Gen. Physiol. *132*, 613–632.

Roux, B. (1997). Influence of the membrane potential on the free energy of an intrinsic protein. Biophys. J. *73*, 2980–2989.

Roux, B., and MacKinnon, R. (1999). The cavity and pore helices in the KcsA K^+ channel: electrostatic stabilization of monovalent cations. Science *285*, 100–102.

Rui, H., Lee, K.I., Pastor, R.W., and Im, W. (2011). Molecular dynamics studies of ion permeation in VDAC. Biophys. J. *100*, 602–610.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779–815.

Samways, D.S., and Egan, T.M. (2011). Calcium-dependent decrease in the single-channel conductance of TRPV1. Pflugers Arch. *462*, 681–691.

Samways, D.S., Khakh, B.S., and Egan, T.M. (2008). Tunable calcium current through TRPV1 receptor channels. J. Biol. Chem. *283*, 31274–31278.

Sanner, M.F., Olson, A.J., and Spehner, J.C. (1996). Reduced surface: an efficient way to compute molecular surfaces. Biopolymers *38*, 305–320.

Schramm, C.A., Hannigan, B.T., Donald, J.E., Keasar, C., Saven, J.G., Degrado, W.F., and Samish, I. (2012). Knowledge-based potential for positioning membrane-associated structures and assessing residue-specific energetic contributions. Structure *20*, 924–935.

Shaya, D., Findeisen, F., Abderemane-Ali, F., Arrigoni, C., Wong, S., Nurva, S.R., Loussouarn, G., and Minor, D.L., Jr. (2014). Structure of a prokaryotic sodium channel pore reveals essential gating elements and an outer ion binding site common to eukaryotic channels. J. Mol. Biol. *426*, 467–483.

Sitkoff, D., Sharp, K.A., and Honig, B. (1994). Accurate calculation of hydration free-energies using macroscopic solvent models. J. Phys. Chem. *98*, 1978–1988.

Skach, W.R. (2009). Cellular mechanisms of membrane protein folding. Nat. Struct. Mol. Biol. *16*, 606–612.

Smart, O.S., Neduvelil, J.G., Wang, X., Wallace, B.A., and Sansom, M.S. (1996). HOLE: a program for the analysis of the pore dimensions of ion channel structural models. J. Mol. Graph. *14*, 354–360, 376.

Stern, H.A., and Feller, S.E. (2003). Calculation of the dielectric permittivity profile for a nonuniform system: application to a lipid bilayer simulation. J. Chem. Phys. *118*, 3401–3412.

Swanson, J.M.J., Wagoner, J.A., Baker, N.A., and McCammon, J.A. (2007). Optimizing the Poisson dielectric boundary with explicit solvent forces and energies: lessons learned with atom-centered dielectric functions. J. Chem. Theor. Comput. *3*, 170–183.

Tanford, C., and Roxby, R. (1972). Interpretation of protein titration curves. Application to lysozyme. Biochemistry *11*, 2192–2198.

Villinger, S., Briones, R., Giller, K., Zachariae, U., Lange, A., de Groot, B.L., Griesinger, C., Becker, S., and Zweckstetter, M. (2010). Functional dynamics in the voltage-dependent anion channel. Proc. Natl. Acad. Sci. USA *107*, 22546–22551.

Voss, N.R., and Gerstein, M. (2010). 3V: cavity, channel and cleft volume calculator and extractor. Nucleic Acids Res. *38*, W555–W562.

Welch, J.M., Simon, S.A., and Reinhart, P.H. (2000). The activation mechanism of rat vanilloid receptor 1 by capsaicin involves the pore domain and differs from the activation by either acid or heat. Proc. Natl. Acad. Sci. USA *97*, 13889–13894.

Yang, A.S., Gunner, M.R., Sampogna, R., Sharp, K., and Honig, B. (1993). On the calculation of pKas in proteins. Proteins *15*, 252–265.

Yoo, J., and Cui, Q. (2008). Does arginine remain protonated in the lipid membrane? Insights from microscopic pKa calculations. Biophys. J. *94*, L61–L63.

Zhang, X., Ren, W., DeCaen, P., Yan, C., Tao, X., Tang, L., Wang, J., Hasegawa, K., Kumasaka, T., He, J., et al. (2012). Crystal structure of an orthologue of the NaChBac voltage-gated sodium channel. Nature *486*, 130–134.

Zhou, Y.C., Feig, M., and Wei, G.W. (2008). Highly accurate biomolecular electrostatics in continuum dielectric environments. J. Comput. Chem. *29*, 87–97.

Structure, Volume *23*

Supplemental Information

Membrane Protein Properties Revealed through Data-Rich Electrostatics Calculations

Frank V. Marcoline, Neville Bethel, Christopher J. Guerriero, Jeffrey L. Brodsky, and Michael Grabe

1 Supplemental Figures

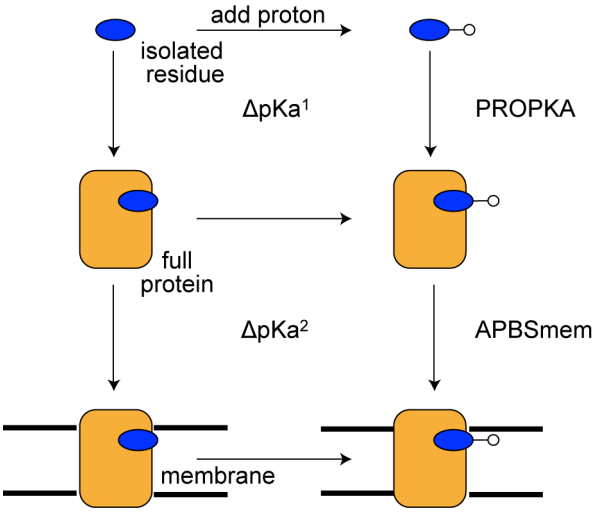

Figure S1: Related to Figure 6. Thermodynamic cycle of residue protonation and membrane insertion. The protonation state of a residue (blue) is affected by its incorporation into a folded protein (gold) as well as the embedding of the protein into the low-dielectric environment of the membrane. The residue's pK_a change for the protein in solution (ΔpK_a^1) is calculated using PROPKA. The second pK_a shift resulting from insertion of the protein into the membrane (ΔpK_a^2) is calculated using APBSmem. These two pK_a changes are summed with the isolated pK_a^0 to yield the final pK_a.

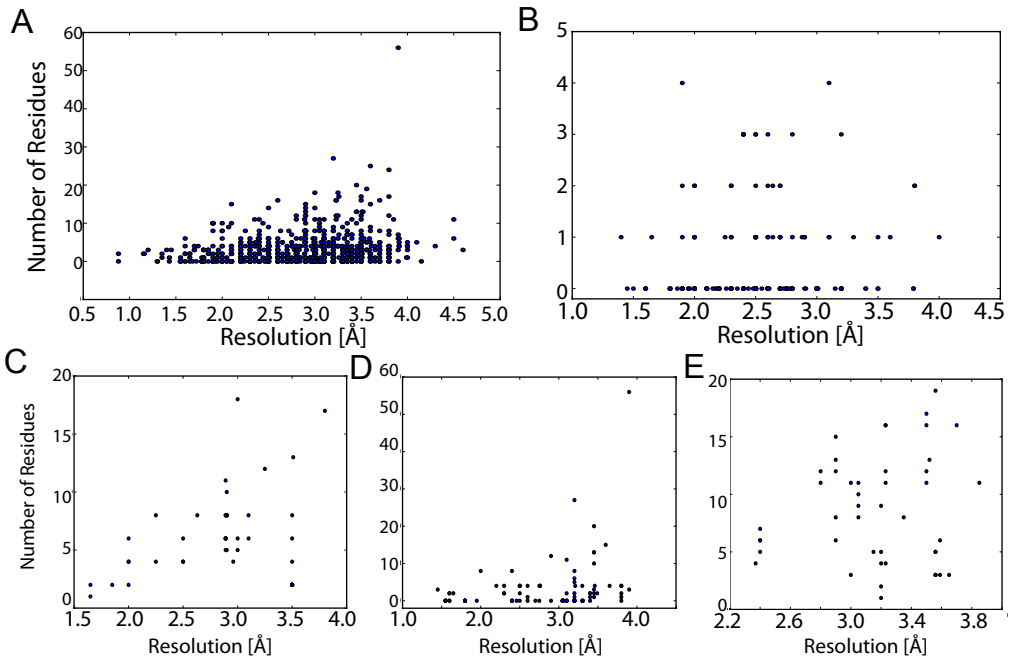

Figure S2: Related to Figure 7. Summary data from the large survey of membrane proteins. **A,B,** Number of electrostatically unfavorable residues over resolution for alpha-helical and beta-barrel proteins, respectively. **C,D,E,** Number of electrostatically unfavorable residues over resolution for amino acid secondary transporters, monovalent cation selective channels, and multi drug efflux transporters, respectively.

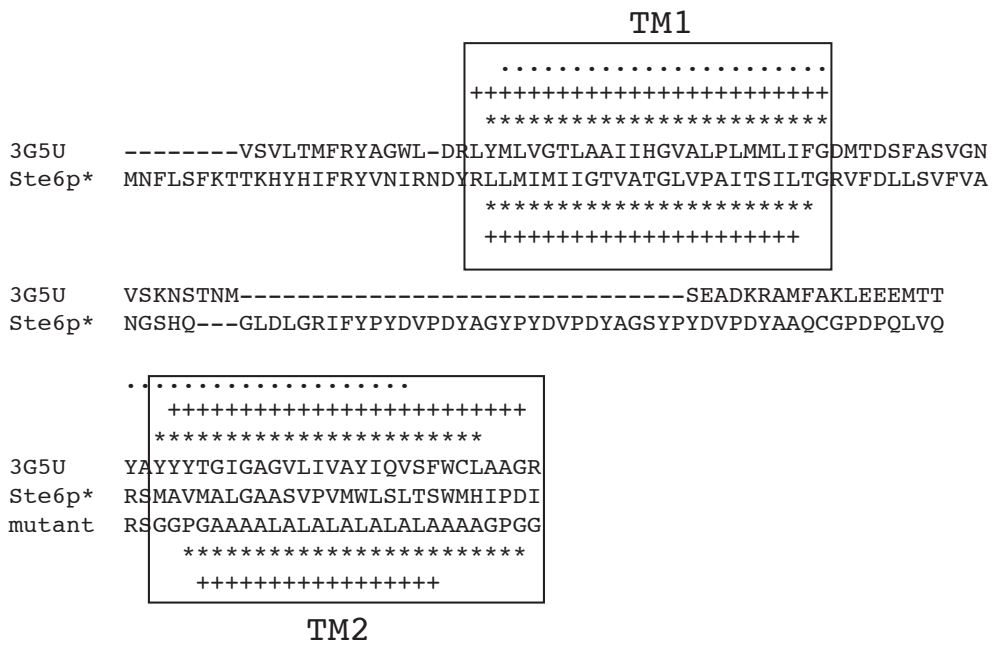

Figure S3: Related to Figure 8. Alignment between Ste6p* and P-gp transporter (PDB: 3g5u). The alignment of the first two transmembrane segments (boxed regions) was carried out using ClustalW followed by adjustments by hand. The hydrophobic mutant protein differs from Ste6p* only in TM2, and its sequence was therefore omitted from the first two lines. Symbols above the sequences refer to the P-gp transporter, while symbols below refer to Ste6p*. Dots are TM stretches of the 3g5u structure identified by inspection, * indicates transmembrane segments identified with the program TMHMM, and + indicates transmembrane segments identified with TMPRED.

Protein	Atoms	Grid dimensions	Focus levels	Grid spacing (Å)	Memory	Time
LeuT	4366	161^3	2	0.375	2.27 GB	4 m 25 s
LeuT	4366	65^3	0	0.9375	363 MB	18 s
VDAC	2250	161^3	2	0.375	2.26 GB	3 m 50 s
VDAC	2250	65^3	0	0.9375	347 MB	15 s

Table S1: Related to Figure 6. Timings for per-residue solvation energy calculations.

Table S2, related to Figure 7, can be found online in a separate spreadsheet as part of the Supplemental Information.

2 Additional APBSmem Features

PDB file processing. Membrane protein PDB files are required for electrostatics calculations, and they can be obtained from many sources such as Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (Berman et al., 2000), PDBe in Europe (Velankar et al., 2010), PDBj in Japan (Kinjo et al., 2012), Orientations of Proteins in Membranes (OPM) database (Lomize et al., 2006), and many other sources, including locally refined structures. PDB files lack atomic radii and charge state information and also lack hydrogen atoms, all of which are crucial for electrostatics calculations. Previously, we required users to first convert PDBs to a PQR format. Now, we have integrated the open source PDB2PQR software (Dolinsky et al., 2004) with APBSmem so that PDB files can be read in, protonated, and assigned atomic radii and charges from one of several different charge models. If a PDB file is loaded, APBSmem will present a pop up window that will ask the user to chose a charge model (PARSE, SWANSON, CHARMM, AMBER, PEOEPB, or TYL06), whether waters should be stripped from the PDB file, and which molecular chains are to be used since many PDB files contain more than one protein. Additional options can be selected such as whether protons should be moved (***nodebump***), whether to perform optimization on hydrogen placement (***noopt***), whether to add a ligand to the PQR (***ligand***), and whether pK_a states should be determined (***with-ph=7.0***). The last option will be discussed more fully in Case III. More information on these options can be found PDB2PQR website: www.poissonboltzmann.org. A PQR file will be generated and saved with the force field indicated in the filename. If a PQR file is loaded from the beginning, APBSmem will not attempt to set the charge model. In our previous calculations (Callenberg et al., 2010), we used the PARSE parameter set with a molecular representation of the protein surface. This was due to the fundamental work by the Honig lab on the role of electrostatics and non-polar interactions when small molecules partition from water to lipid-like phases (Sitkoff et al., 1994). For Case I, we use a spline-based representation of the protein surface invented by the Roux lab since it smooths dielectric boundaries, producing more slowly varying energy profiles that are not highly dependent on grid spacing (Nina et al., 1997). In particular, we use the SWANSON spline-based parameters (Swanson et al., 2007).

Greater geometric control over proteins and ions. Proteins deposited in the OPM database are oriented with respect to the z-axis and centered at the origin (Lomize et al., 2006), but this is not the case for structures stored in most databases. APBSmem requires that the membrane lies in the x-y plane, so we added a feature to automatically align proteins to the z-axis under the menu item ***Orient***. The ***Auto-orient*** function under ***Orient*** will pick one of the principal axes of the molecule to align with the z-axis in an attempt to correctly orient the protein in the membrane. The protein will also be translated to the origin, and the transformed structure will be saved to a new file. The details of the transformation are stored in the file header. In all cases, the ***Preview*** button must be used after setting the desired membrane parameters to visualize the dielectric boundaries of the membrane and how the protein is oriented in the membrane. Please see our previous paper for more detail on this step and other GUI parameters (Callenberg et al., 2010). If the desired position is not achieved, the ***Orient*** menu item provides options for arbitrary rotations and translations of the protein for fine-tuning.

Ligand solvation energy calculation. APBSmem can now calculate the electrostatic transfer energy of a small ligand molecule. Addition of a ligand molecule is specified as an option when converting a PDB to PQR file in APBSmem (***ligand*** option). If this option is selected, the user is prompted to provide a MOL2 file for the ligand. This file provides the partial charges and bond connectivity of the ligand, and can be obtained from

the PRODRG server (Schüttelkopf and van Aalten, 2004). Once the PQR is generated, the electrostatic transfer energy can be calculated by selecting ***Ligand solvation*** as the calculation type and clicking ***Run***.

Automatic identification of the membrane. APBSmem edits the local dielectric environment around a protein to include the presence of the membrane for electrostatics calculations. To do this, the dielectric, ion accessibility, and charge maps produced by APBS are redrawn with membrane parameters provided by the user. Identifying where the membrane lies is difficult since many proteins contain aqueous cavities that must be ignored when adding the lipid bilayer. We added an automatic detection method that requires minimal user input. Specifically, we use a six-way flood-filling method illustrated in Figure 2A. Briefly, a three dimensional logical array specifies where the membrane should be drawn, and the user provides only the upper and lower membrane boundaries in the z-direction. An initial seed element within the membrane at the edge of the x-y plane, far from the protein, is used as the starting point for the flood fill method (Figure 2A). APBSmem expands the seed north, south, east, west, forward and backward, creating new membrane elements of the array for any element inside the membrane defined by thickness, d, and outside the protein. Thus, the provided membrane boundaries and dielectric protein surface determined via APBS serve as the boundaries for the flood fill. The procedure is iterated for every new element of the array that becomes labeled as membrane until no more elements are added. This array is then used to redraw the initial dielectric maps produced by APBS to be used in the electrostatics calculations in the presence of a membrane. However, this method may be confounded by channels that have fenestrations that connect to the inner pore, such as the voltage-gated sodium channels (Payandeh et al., 2012; Shaya et al., 2014; Zhang et al., 2012) (Figure 2B). Since flooding fills the pore with membrane through these fenestrations (bottom left), we apply an additional threshold that prevents expansion into holes smaller than a vertical thickness of t (8 Å). This threshold allows for the proper identification of aqueous cavities, as shown in the bottom right panel. The automatic membrane detection also makes it possible to systematically carry out electrostatic calculations in an automated manner on large sets of proteins, as discussed in Case III.

Non-polar energy calculations and per-residue energies A major energetic consideration with respect to the stability of membrane proteins is the non-polar energy arising from solvent reorganization when portions of the protein are buried in the bilayer. Simple non-polar energy models assume a linear dependence on the solvent exposed surface area (Hermann, 1972; Lee and Richards, 1971). APBSmem now calculates the non-polar energy (Enp) for all buried residues using the software MSMS, which is freely available for academic use (http://www.scripps.edu/sanner) (Sanner et al., 1996). Upon starting APBSmem, the user will be prompted to set the path to MSMS. [[2.4]] APBS can also be used to calculate the nonpolar energy of a protein, and the user has the choice to use APBS for this calculation instead. Additionally, APBSmem now provides the non-polar and electrostatic energies for each residue in the protein, which aids in designing mutational analyses.

Command line scripting. When coupled with a simple shell script to modify parameters in an input file – such as which protein to use for each calculation – APBSmem can readily perform many calculations in a fully automated manner. The initial APBSmem input file (*file.in*) can be generated using the GUI once parameters have been set. This happens automatically during a run, or when the user selects the menu item ***File/Save***. This input file is a simple text file that can be opened up and edited for use in subsequent command line calls to APBSmem, and is a valid ABBS input file. To run a calculation, the user provides the input file and output directory (*out-dir*) on the command line: `java`

`-jar apbsmem.jar` *file.in out-dir*. Additionally, APBSmem can be run from the command line without doing a full electrostatics calculation to: add an implicit membrane to a set of APBS .dx files (`java -cp apbsmem.jar apbsmem.DrawMembrane`); modify the orientation of a PQR/PDB file (`java -cp apbsmem.jar apbsmem.orient.Orient`), or change the charge state of a residue (`java -cp apbsmem.jar apbsmem.AssignCharge`). For the later three commands, omission of further command line arguments causes APBSmem to print usage information. Between calculations it may be advisable to delete the APBS map files, as they can be quite large. Alternatively, if the map file type is changed from dx to gz in the input file, APBS and APBSmem will work with compressed maps, reducing disk usage significantly.

Other enhancements. APBSmem now supports additional APBS keywords to choose the charge model (***chgm***), the surface model (***srfm***), and the spline width (***swin***), and provides improved support for setting the geometric center (***gcent***) of the grid volume. These keywords can be read from an APBS input file or set directly using the APBSmem interface. For a typical APBS calculation, the charge density, dielectric constant and ion accessibility maps are computed from a molecule in PQR format and from keywords describing the solution dielectric (***sdie***), the molecular dielectric (***pdie***), discretization of the charge distribution (***chgm***), charge, concentration, and radius for each ion species (***ion***), and definition and parameterization of the solvent accessible surface (***srfm***, ***srad***, ***sdens*** and ***swin***). For example, the flags to the keyword chgm specify how the idealized point charge of an atom in a PQR file is partitioned to nearby grid points in the charge density map.

3 APBSmem Electrostatic Energy Calculations with APBS

Adaptive Poisson-Boltzmann Solver (APBS) is open source software developed by the Baker group (Baker et al., 2001) for electrostatic calculations in biomolecular systems. APBS binaries, source code and documentation are available at www.poissonboltzmann.org. The electrostatics model used by APBS is detailed in section 2 of Baker, Bashford, and Case (2006). This supplement gives the explicit formulae used by APBS to calculate electrostatic energies for various APBS input parameters according to the APBS version 1.3 source code.

3.1 APBS solution method

APBS uses a finite difference method to numerically solve the Poisson-Boltzmann equation (Eq. 1 of main text) for the electric potential ϕ at points on a regular three dimensional grid. This requires discretizing the physical properties of the system over the same grid, and discretizing the boundary conditions on the grid boundary. Specifically, the free charge density $\rho_f(\mathbf{x})$, the dielectric constant $\epsilon(\mathbf{x})$, and the ion accessibility $kappa(\mathbf{x})$ are discretized over the grid and are stored in three dimensional arrays referred to as "maps". The *kappa* map should not be confused with the Debye-Hückel parameter κ (more on this below). The other physical inputs to the solver are the temperature T, and details of the ions in solution (charge z_j, concentration $\bar{n}_j$, and radius for each ion species).

For a typical APBS calculation, the charge density, dielectric constant and ion accessibility maps are computed from a molecule in PQR format and from keywords describing the solution dielectric (***sdie***), the molecular dielectric (***pdie***), discretization of the charge distribution (***chgm***), charge, concentration, and radius for each ion species (***ion***), and

construction of the solvent accessible surface area (***srad***, ***sdens***, ***swin*** and ***srfm***). For example, the flags to the keyword ***chgm*** specify how the point charge of an atom in the PQR file is partitioned to nearby grid points in the charge density map. See the APBS documentation for more details on how the keywords above effect map generation.

Alternatively, APBS can read the maps from input files, skipping the map generation step and proceeding straight to solving for the potential ϕ. APBSmem works by calling APBS to generate and write maps, inserting a membrane by altering the dielectric, charge, and ion accessibility maps, and calling APBS using the altered maps, thereby solving for ϕ for a molecule in a membrane environment.

3.2 APBS energy equations and terms

For the non-linear Poisson-Boltzmann equation, APBS calculates the total electrostatic energy from three components: the charge-potential interaction energy (or fixed charge energy, E_f), the dielectric polarization energy (E_d), and the counter-ion distribution energy (or mobile charge energy, E_m):

$$\begin{aligned} E_{total} &= 2E_f - E_d - E_m \\ &= \int d^3x \left(\rho_f \phi - \frac{\epsilon}{8\pi} (\nabla\phi)^2 - kT \sum_{\text{ion}j} \bar{n}_j e^{-\frac{V_j}{kT}} \left(e^{-\frac{ez_j\phi}{kT}} - 1 \right) \right). \end{aligned} \tag{S1}$$

The sum is over each counter-ion species j, where the number density of counter-ion species j is $\bar{n}_j$. The potential V_j represents steric interactions between ion species j and atoms of the molecule, and is part of the modified Debye-Hückel term $\bar{\kappa}(\mathbf{x})$:

$$\bar{\kappa}^2(\mathbf{x}) = \frac{4\pi e^2}{kT} \sum_j \bar{n}_j z_j^2 e^{-\frac{V_j}{kT}}, \tag{S2}$$

where z_j is the charge of species j in atomic units. APBS assumes that V_j is the same for all ions in solution, and that the solvent accessible surface presents an infinite potential: $V_j = 0$ outside the molecular surface, $V_j = \infty$ inside the molecular surface. Thus, the steric factor $e^{-V_j/kT}$ is one outside the molecule, and zero inside the molecule, and is represented by APBS as the ion accessibility map *kappa*. See the APBS keyword ***srfm*** for details on how the solvent accessible surface is defined.

In the linear limit, the mobile charge energy reduces to

$$E_m = \int d^3x \; \frac{1}{2} \bar{\kappa}^2 \phi^2 \tag{S3}$$

For linear media, the total energy is equal to the work required to assemble the charges: $E_{total} = E_f = E_d + E_m$. This is not generally true for non-linear media. APBSmem sets the APBS keyword ***calcenergy*** to the flag ***comps*** causing APBS to print E_{total}, E_f, E_d, E_m, and the per-atom energy contribution E_{f_a} for each atom a.

3.2.1 Fixed charge energy

APBS reports the per-atom fixed-charge energy E_{f_a} for each atom a in the PQR file:

$$E_{f_a} = \frac{1}{2} q_a \Phi_a(r_a), \tag{S4}$$

where q_a is the charge of atom a, and $\Phi_a(r_a)$ is the trilinear interpolation of the three dimensional solution array for the potential, $\phi[i,j,k]$, to the coordinates r_a of atom a. If

the charge map is derived from a molecule input from a PQR file, the total fixed charge energy is calculated as the sum over each atom in the molecule:

$$E_f = \sum_{\text{Atom } a} E_{f_a}. \tag{S5}$$

If instead a charge density, $\rho[i,j,k]$, is supplied by specifying a charge map in the APBS input file, such as in APBSmem calculations, APBS calculates the total fixed charge energy as the volume integral of the charge density times the potential by the approximation:

$$E_f = \frac{1}{2} h_x h_y h_z \sum_{i,j,k} \rho[i,j,k]\, \phi[i,j,k], \tag{S6}$$

where h_x, h_y and h_z are the x, y and z grid lengths, respectively.

3.2.2 Dielectric energy

The formula APBS uses for the dielectric energy term is

$$\begin{aligned} E_d = \frac{1}{2} h_x h_y h_z \sum_{i,j,k} \Big\{ & \epsilon_x[i,j,k]\,(\phi[i+1,j,k] - \phi[i,j,k])^2 + \\ & \epsilon_y[i,j,k]\,(\phi[i,j+1,k] - \phi[i,j,k])^2 + \\ & \epsilon_z[i,j,k]\,(\phi[i,j,k+1] - \phi[i,j,k])^2 \Big\}, \end{aligned} \tag{S7}$$

where ϵ_x, ϵ_y and ϵ_z are dielectric maps which are centered half a grid point in the x, y and z directions, respectively, relative to the charge density and potential maps.

3.2.3 Mobile charge energy

For the non-linear Poisson-Boltzmann equation, the mobile ion energy term is evaluated by summing over each ion species m:

$$E_m = \frac{1}{2} kT\, h_x h_y h_z \sum_{\text{ion } m} \sum_{i,j,k} \bar{n}_m \big(e^{-z_m \phi[i,j,k]/kT} - 1\big). \tag{S8}$$

For the linear case, the Debye-Hückel parameter $\bar{\kappa}$ is taken to be constant, and ion accessibility map, *kappa*, restricts the mobile charge energy calculation to outside of the solvent accessible molecular surface:

$$E_m = \frac{1}{2} h_x h_y h_z \sum_{i,j,k} \bar{\kappa}^2 \phi^2[i,j,k]\, kappa[i,j,k]. \tag{S9}$$

3.3 Case I: Ion transfer free energy

We downloaded the PDB structure (PDB: 3j5q) from RCSB, selected Ion solvation as the calculation type in the GUI, and loaded the structure into PQR File 1 with the Browse button. We then centered the channel with the following commands: Orient/Translate by -20 Å in z, Orient/Rotate about x by 180°, and then Orient/Rotate about z by 45°. This last rotation was done to better fit the simulation volume. We then set all parameters in the GUI according to the information in Table 1, and we selected Preview to ensure that the membrane interface was properly drawn around TRPV1 (light grey boundaries in panel

A). Rather than choosing a second PQR File 2, we created an ion from Ion/Create ion in the menu with an initial position on the z-axis at -60 Å and the charge (+2) and radius (1.03 Å) of a Ca^{2+} cation. Next, we selected Ion/Step ion and entered a final location in the extracellular space +60 Å making a path with 100 steps.

3.4 Case II: Per-residue interaction energies with ions & small molecules

APBSmem calculates the fixed charge energy for each residue R in a PQR file by summing the per-atom fixed-charge energies (equation S4) over each of the n atoms in residue R:

$$E_f^R = \sum_{i=1}^{n} E_{f_i} = \frac{1}{2} \sum_{i=1}^{n} q_i \Phi_{P,I}(r_i), \tag{S10}$$

where q_i is the charge of the i^{th} atom, r_i is its position, $\Phi_{P,I}$ is the electrostatic potential of the protein-ion/small molecule system embedded in the membrane. This energy value is formally divergent because it evaluates the potential at the position of the atoms, so APBSmem isolates the residue-ion interaction energy by subtracting off the protein-protein energies:

$$\Delta E_f^R = \frac{1}{2} \sum_{i=1}^{n} q_i \left(\Phi_{P,I}(r_i) - \Phi_P(r_i) \right) \tag{S11}$$

where Φ_P is the electrostatic potential of the protein embedded in the membrane without the ion. These per-residue interaction energies are saved to a log file in the output directory.

3.5 Case III: Per-residue solvation energies

Per-residue solvation energies, ΔG_R^{env}, are calculated as the difference of the per-residue fixed-charge energies of residue R in the membrane $E_{f,R}^M$ and in solution $E_{f,R}^S$:

$$\begin{aligned} \Delta G_R^{env} &= E_{f,R}^M - E_{f,R}^S. \\ &= \frac{1}{2} \sum_{i=1}^{n} q_i \left(\Phi^M(r_i) - \Phi^S(r_i) \right), \end{aligned} \tag{S12}$$

where the sum runs over the n atoms in residue R, and Φ^M and Φ^S are the electrostatic potentials of the protein in the membrane and solution, respectively. For these calculations interaction energies are reported in units of kcal/mol and written to a space-delimited file. When multiple calculations are called at one time, which happens for an ion stepping calculation, each new calculation is recorded as a new column. This text file can be directly loaded into a spreadsheet or plotting program for further analysis, as below.

3.6 Numeric focusing

To accurately model the electrostatics in a volume, the boundary conditions need to be well approximated. Typically this requires placing the boundary far away from any charges in the system. Since APBS solves the Poisson-Boltzman equation on a regular three dimensional grid, placing the boundary far away from the region of interest means, due to memory and time constraints, that the resolution may be worse than desired, *i.e.*: grid spacings larger than structures of interest.

To increase resolution in the region of interest, APBS employs a method called focusing: one or more focusing calculations may be performed after an initial calculation, with each focusing volume being a subset of the volume of the prior calculation. The solution to the Poisson-Boltzman equation for the potential at one level of focusing is used to set the potential on the boundary of the volume at the subsequent level, after which the potential in the interior is calculated.

Then to evaluate an energy term (such as the dielectric energy term) over the whole volume, the term is evaluated over the finest grid available in each region of the volume. Suppose for example that a calculation is performed using two focusing levels. Let the initial solution be ϕ_0 over total volume V_0, the first focus level solution be ϕ_1 over subset V_1, and the final focus solution ϕ_2 over subset V_2: $V_2 \subset V_1 \subset V_0$. Then the energy term is evaluated using ϕ_2 over V_2, ϕ_1 over $V_1 \setminus V_2$, and ϕ_0 over $V_0 \setminus V_1$. ($A \setminus B$ is the set difference: $A \setminus B = \{x \in A \mid x \notin B\}$, *i.e.* the volume A excluding the volume B.)

Per-atom energies are reported for each atom at each focus level. APBS assigns atoms outside the current focus volume a per-atom energy of zero. For these atoms, APBSmem reports the energy at the smallest focusing volume containing that atom. In practice, this will be that last focus level for which the per-atom energy is non-zero.

3.7 Technical notes on PB calculations

We wish to note a few technical details concerning the calculations. First, the choice in atomic parameter set does not heavily influence the ion solvation energies when the ion and channel are well hydrated, which is the case for large pores such as VDAC (Choudhary et al., 2010). However, for TRPV1 the parameter set is very important when the ion is less than a few Ångstroms from the protein. In these cases, the spline-based methods, such as SWANSON, produce much smoother profiles along the pathway, especial in regions where the ion approaches the protein surface. Regardless, it is important to remember that the parameter set used must be properly matched with the surface representation for the protein, i.e. PARSE with the molecular surface representation (***mol***) and SWANSON with spline-based representations (***spl2*** or ***spl4***). APBSmem does not let one choose parameters that are incompatible with each other. At present, the membrane dielectric boundaries are currently not smoothed. We will incorporate this feature into future releases. Second, it is always best to use the non-linear Poisson Boltzmann method (***npbe***) over the linear solver (***lpbe***), but this method requires longer calculation times. Moreover, when the electrostatic potentials are small and energies are on the order of a kcal/mol the linear solution is generally similar to the non-linear solution, and this is true of the OmpF profile but not the TRPV1 profile. Finally, the number of grid points determines the fidelity of the final solutions. Formally, one should carry out convergence testing, in which profiles are plotted with increasing numbers of grid points in order to determine if the solutions approach a constant value. A rule of thumb is that the grid spacing at the finest level should be less than 0.5 Å in all directions.

The ion profile calculations in the main text are relatively fast, requiring approximately 7 to 8 minutes per ion location on a 2011 MacBook Pro using the parameters listed in Table 1. At lower resolutions with 653 grid points, calculations take 30 seconds per step. Thus, initial diagnostic estimates of the profiles in Figure 3 can be generated in less than half an hour followed by production quality calculations at higher resolution (grid spacing < 0.5 Å), which may then require about 12 hours.

We caution users that continuum methods can break down for narrow pores when ions approach the protein surface; however, if care is taken and the results are not over interpreted, valuable information can still be gleaned from the calculations.

4 Membrane Induced Changes in $\mathrm{p}K_\mathrm{a}$ Values

As shown in Figure S1, it is useful to calculate these shifts in terms of thermodynamic cycles. We consider two cycles: first, the energy required to protonate/deprotonate the residue of interest in solution (cycle 1), and second, the energy required to protonate/deprotonate the residue in the presence of the membrane (cycle 2). We use the program PROPKA to compute the values along cycle 1, and below we describe how to use APBSmem to estimate the energies along cycle 2.

PROPKA uses an empirical method to quickly estimate $\mathrm{p}K_\mathrm{a}$ shifts based on the residue's solvent accessibility, hydrogen bonding, and short-range charge-charge interactions (Li et al., 2005; Olsson et al., 2011). We call this shift along cycle 1 $\Delta\mathrm{p}K_\mathrm{a}^1$. While PROPKA is fast and reliable, it is not equipped to handle changes in dielectric environment such as moving into a membrane, and it does not account for long-range electrostatic interactions. Yang and coworkers developed another method based on solutions to the PB equation to calculate expected $\Delta\mathrm{p}K_\mathrm{a}^2$ values. In this framework, the $\mathrm{p}K_\mathrm{a}$ of a charge group within a protein is given by:

$$\Delta\mathrm{p}K_\mathrm{a} = -\gamma_i \Delta\Delta G_i^{env} 2.3 k_B T, \tag{S13}$$

where $\Delta\Delta G_i^{env}$ is the change in electrostatic energy associated with ionizing the group in the protein versus alone in solution, and γ_i is -1 or 1 for an acidic or basic group, respectively. While this method was intended for calculating protein-induced shifts, we adapt it here to compute membrane induced $\mathrm{p}K_\mathrm{a}$ shifts along cycle 2, which we call $\Delta\mathrm{p}K_\mathrm{a}^2$. To do this, APBSmem solves the PB equation to determine the change in total electrostatic energy for inserting a protein into the membrane with residue i deprotonated, $\Delta G_i^{env}(A)$, and the change in energy with residue i protonated, $\Delta G_i^{env}(AH)$. The membrane induced energy shift, $\Delta\Delta G_i^{env}$, is then:

$$\Delta\Delta G_i^{env} = -\gamma_i \left\{ \Delta G_i^{env}(A) - \Delta G_i^{env}(AH) \right\}, \tag{S14}$$

which determines $\mathrm{p}K_\mathrm{a}^2$ according to equation S13. The total modified $\mathrm{p}K_\mathrm{a}$ of each residue is then:

$$\mathrm{p}K_\mathrm{a} = \mathrm{p}K_\mathrm{a}^0 + \Delta\mathrm{p}K_\mathrm{a}^1 + \Delta\mathrm{p}K_\mathrm{a}^2, \tag{S15}$$

where $\mathrm{p}K_\mathrm{a}^0$ is the experimentally determined $\mathrm{p}K_\mathrm{a}$ of the isolated residue.

5 Homology Models of Ste6p*

Models of the transmembrane domains (TM1 & TM2) of the ATP-binding cassette (ABC) transporter Ste6p* were constructed with Modeller9v13 (Sali and Blundell., 1993) using the P-glycoprotein transporter (PDB: 3g5u) as a template structure and the alignment provided in Figure S3. There are several ABC transporters whose structures have been solved, which could have been used as templates; however, the sequence identities with Ste6p* are modest with values ranging from 20-27%. Moreover, the transmembrane segments are even less conserved. Among the best match for TM1 and TM2 is the P-glycoprotein transporter from mouse (PDB: 3g5u), which has 7.7% and 14.8% identity, respectively. We initially used ClustalW (Thompson et al., 1994) to align the sequences, but we also identified the TM stretches independently with the programs TMHMM (Krogh et al., 2001) and TMPRED (Hofman and Stoffel, 1993). Additional hand adjustments were then carried out to close up gaps and maximally align the second transmembrane regions. The final alignment used for construction of the first two TM segments of Ste6p* and the wild-type TM2 segment

is shown in Figure S3. The mutant TM2 segment was then constructed using the wild-type TM2 model as a template. Please note that for the TM insertion energy calculations provided in the main text, these alignments are not crucial since we simply assume that both TM2 segments take on roughly straight helical configurations.

6 Insertion Energy Heat Map Construction with Command Line Scripting

We constructed insertion energy maps by repeatedly calling APBSmem in a shell script, altering the orientation of the TM segments between each call. We initially aligned the helices with their long axes in the z-direction, and chose a pivot point (P_x, P_y, P_z) at the N-terminal end of the helices as the origin of rotations. Looping over values of ϕ, the angle about the long axis, and θ, the angle away from the membrane normal, the shell script first called APBSmem to create a PQR file with the protein in the new (ϕ, θ) orientation, and then the script called APBSmem again to calculate the insertion energy.

Creating a PQR file in a new orientation was performed by a single call to APBSmem, chaining together several orientation operations:

1. translating the protein by (-P_x, -P_y, -P_z) to put the center of rotation at the origin,
2. rotating the protein about the z axis by ϕ,
3. rotating the protein about the x axis by θ, and
4. translating the protein back by (P_x, P_y, P_z).

The above operation is performed using command line options: "-t -P_x -P_y -P_z -z ϕ -x θ -t P_x P_y P_z". Note that the italicized arguments are actual values, while the non-italicized arguments are keywords. For example, for a pivot point of (0,0,10Å), and initial PQR file named `TM1.pqr`, the full command line to rotate by $\phi = 60$ degrees and $\theta = 45$ degrees may appear as:

```
java -cp apbsmem.jar apbsmem.orient.Orient -t 0 0 -10 -z 60 -x 45 -t 0 0 10 TM1.pqr
```

Usage information for orientation operations is printed if no command line arguments to apbsmem.orient.Orient are specified.

Supplemental References

N. A. Baker, D. Bashford, and D. A. Case. Implicit solvent electrostatics in biomolecular simulation. In *New algorithms for macromolecular simulation*, pages 263–295. Springer, 2006.

O. P. Choudhary, R. Ujwal, W. Kowallis, R. Coalson, J. Abramson, and M. (2010). Grabe. The electrostatics of vdac: implications for selectivity and gating. *Journal of Molecular Biology*, 396(3):580–592, 2010.

K. Hofman and W. Stoffel. Tmbase - a database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler*, 374:166, 1993.

A. Krogh, B. Larsson, G. Von Heijne, and E. L. L. Sonnhammer. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3):567–580, 2001.

M. Nina, D. Beglov, and B. Roux. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *The Journal of Physical Chemistry B*, 101(26):5239–5248, 1997.

A. W. Schüttelkopf and D. M. F. van Aalten. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.*, D60:1355–1363, 2004.

J. D. Thompson, D. G. Higgins, and T. J. Gibson. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22):4673–4680, 1994.