SAND20XX-xxxxR

LDRD Project Number: 180817

LDRD Project Title: Coupling Chemical Energy with Protein Conformational Changes to Translocate Small Molecules Across Membranes

Project Team members: Josh Vermaas (UI-UC), Emad Tajkhorshid (UI-UC), Susan Rempe

Abstract

EmrE is a small, homodimeric membrane transporter that exploits the established pH gradient across the *E. coli* inner membrane to export polyaromatic cations that might otherwise inhibit cellular growth. While herculean efforts through experimental studies have established many fundamental facts about the specificity and rate of substrate transport in EmrE, the low resolution of the available structures have hampered efforts to tie those findings to the EmrE coupling mechanism between proton and small molecule substrates. Here we present a full three-dimensional structure of EmrE optimized against available cyro-EM data to delineate the critical interactions by which EmrE regulates its conformation. We use the generated structural model to conduct equilibrium and nonequilibrium molecular dynamics simulations to probe EmrE dynamics under different substrate loading states, representing different states in the transport cycle. The model is stable under extended simulation, and reveals that water dynamics within the EmrE lumen change substantially with the loading state. The water dynamics cause hydrogen bonding networks to shift radically when the protonation states change for a pair of solvent-exposed glutamate residues (E14) within the lumen of the transporter, which are proposed to act as proton binding sites during the transport cycle. One specific hydrogen bond from a tyrosine (Y60) of one monomer to a glutamate (E14) on the opposite monomer is especially critical, as it locks the protein conformation when the glutamate is deprotonated. Furthermore, the hydrogen bond provided by Y60 lowers the pK_a of the interacting glutamate relative to its partner on the opposite monomer such that it will protonate second, establishing the need for both glutamates to be protonated for the hydrogen bond to break and a substrate-free transition to take place.

Introduction

EmrE is a membrane transporter found in *Escherichia coli* that uses the cellular proton gradient across the bacterial inner membrane to export a variety of polyaromatic cations that may otherwise harm the cell.[1,2] EmrE acts as a homodimer of 110-residue monomers that each contain four transmembrane helices,[3,4] consistent with other members of the small multi-drug resistance (SMR) transporter family.[5] Structural studies using nuclear magnetic resonance (NMR) spectroscopy,[6,7] cryo-electron microscopy (cryo-EM),[3] and X-ray crystallography[8] have shown two distinguishable monomeric states that are simultaneously populated within each dimer. Experiments using Förster resonance energy transfer (FRET) conclusively demonstrate the antiparallel orientation of the monomers with respect to one another.[6] Based on this evidence, the proposed transport cycle of EmrE proceeds by an alternating access mechanism whereby the monomers swap conformations between these two states (Fig. 1).[6]

This transport mechanism presents an interesting structural dilemma shared by all secondary antiporters, which use the concentration gradient of one species to drive another against its gradient. Namely, the conformational transition when the transporter is empty must be forbidden (Fig. 1), otherwise transporter action would run down the gradient rather than fulfill its functional role. For antiporters with known structure, such as the glycerol 3-phosphate transporter GlpT, this requirement is reflected in a higher energetic barrier for the apo transition relative to a substrate-bound transition.[9] However, due to the rapid transition between conformational states in EmrE[2,10,11] and a dearth of complete EmrE structural models, no systematic study exists of how this conformational transition is controlled in EmrE at the atomic level. The glutamate (E14) residues in the transporter lumen that are thought to act as protonation sites as well as drug interaction partners during the transition[12,13] likely play a crucial role. However, their specific interaction partners that control the transition have not been elucidated previously, due to the incompleteness of prior structural models.

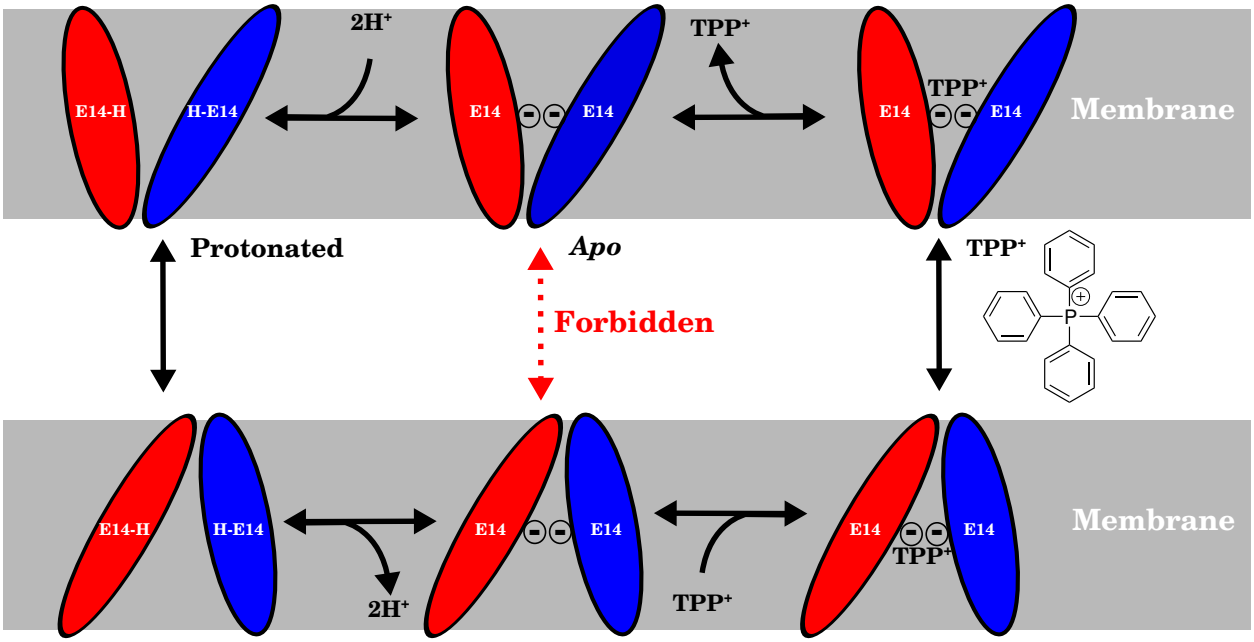

Figure 1: Schematic of the transport cycle of EmrE following the alternating access mechanism. EmrE monomers (red and blue ellipsoids) swap conformations and thus alternate the substrate/proton accessibility to the inside or outside of the cell when loaded with protons (left) or substrate (right). When the EmrE dimer is in its apo (middle) form, the transition should be forbidden in order to preserve the coupling of proton and substrate transport and prevent proton leakage across the membrane. This is in contrast to both the full protonated or substrate bound state (left and right columns, respectively), whose transition should be allowed (black arrows) for EmrE to fulfill its physiological function as a polyaromatic cation transporter.

To answer these questions explicitly, we constructed a complete and refined atomic model of EmrE using molecular dynamics flexible fitting (MDFF)[14] and interactive biased molecular dynamics techniques[15] to incorporate existing experimental data into the model. Through this procedure, the model expands upon the C_α positions provided in published crystallographic structures[8] to encompass the full sequence of EmrE, including side chain positions essential to a full structural model. These positions are consistent with experimental electron densities and structural proclivities of individual side chains. Using this starting model, we carried out a series of simulations where the loading state of EmrE was varied to mimic the different intermediate states of the transport cycle. To represent the drug-bound state, a tetraphenylphosphonium cation (TPP^+) was chosen from among the many substrates EmrE exports[2] due to crystallographic identification of the binding site location.[8] In addition to traditional equilibrium molecular dynamics simulations, replica exchange thermodynamic integration (RETI)[16] calculations were performed to determine the pK_a differences between the E14 residues in each monomer and assign specific causal factors to the observed experimental pK_a shift of 1.5–2.0 units[17] between E14 residues in each monomer. Non-equilibrium work calculations were also performed for a qualitative estimate of the conformational transition rates of different substrate loading states, using the inherent symmetry of EmrE to generate structural models for both inward- and outward-facing states of EmrE.

Over the course of the simulated trajectories, the secondary structure imposed by the modeling process remained intact over the course of 12.5 μs aggregate simulation time, with a more expansive helical architecture compared with prior models for EmrE.[8,18,19] The behavior of EmrE within the simulations is highly dynamic, and also highlight the pivotal role aromatic residues around the binding site play in controlling the observed conformational heterogeneity. Tyrosine Y60 from monomer B ($Y60^B$) in particular interacted frequently with E14 of the opposing monomer ($E14^A$). When combined with previous mutagenesis data[20] showing a deleterious effect for Y60F mutation on cell fitness, this finding suggests that the hydrogen bond formed between Y60 and E14 locks the conformational transition when E14 is deprotonated. In addition, this hydrogen bond lowers the pK_a of that specific E14 residue, thus forcing a second proton binding event before the interaction is broken and the conformational transition can take place. The identification of this hydrogen bond is a substantial new insight as to the mechanism of proton coupling to conformational change within EmrE.

Structural Model Construction

Construction of the structural model of the EmrE homodimer embedded into a lipid bilayer was challenging due to the low resolution of the crystallographic structure (~ 4 Å).[8] Only backbone alpha-carbon (C_α) positions were reported for a subset of residues, therefore requiring that the remaining backbone atoms as well as all side chains be modeled prior to simulation. In an initial naïve approach, the positions of the C_α atoms from the TPP^+-bound crystal structure (PDB: 3B5D)[8] were used to restrain a model of EmrE generated using a combination of `psfgen` and `Modeller`[22,23] software. In this manner, the available C_α positions act as a template to guide the positions of the side chains as they settle into their preferred rotameric states. The resulting structure from this naïve approach presented a number of unsatisfying features. The initial structure lacked α-helical secondary structure throughout the hydrophobic transmembrane regions of the protein (Fig. 2). Due to the strength of backbone hydrogen bonds in a low dielectric environment within the hydrophobic core of the membrane, transmembrane helices are expected to be largely α-helical.[24,25] Furthermore, the crystal structure placed a number of proline residues in the loop regions between helical domains (Fig. 2), rather than as helix terminators where they are far more commonly found.[26] Thus, considerable effort was invested to use available experimental information and trends from other membrane proteins to refine the model to recapitulate the *in vivo* state.

Beyond the crystal structure, a cryo-EM-based model of the EmrE dimer is available,[18] as well as the original electron density used to construct that model.[3] We combined this cryo-EM map with the Molecular Dynamics Flexible Fitting (MDFF)[14] approach to refine the initial structural model. MDFF uses a set of external forces determined by an electron density map determined by experiment to steer atoms to areas of high electron density.[14] The refinement proceeded through two broad steps: simulation of the naked protein in a high-dielectric ($\epsilon = 80$) implicit solvent model,[27,28] and embedding this model into an atomistic membrane representation for further relaxation with additional electron-density restraints. Splitting the

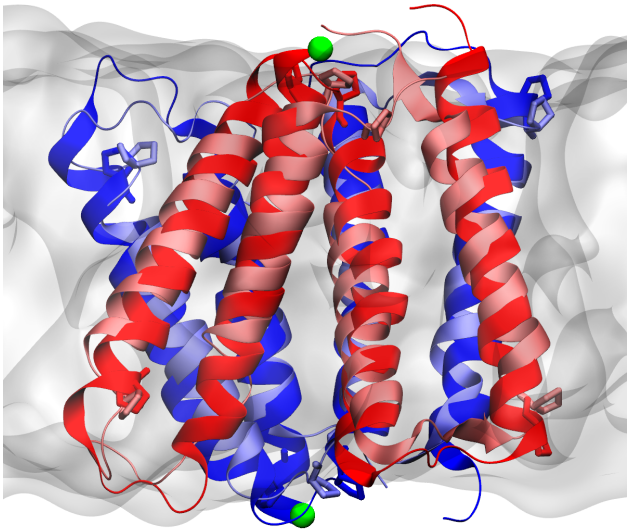

Figure 2: Direct structural comparison of the initial structure based on a naïve refinement of the original crystal structure (lighter colors) with the resulting newly refined structure obtained after interactive MDFF simulations (darker colors). The protein structure is represented as a cartoon, where α-helical secondary structure as determined by `Stride`[21] is clearly demarcated from loop regions. The extent of the hydrophobic acyl chains of the modeled lipid (DMPC) bilayer is shown as a transparent surface around the EmrE homodimer. The monomers are identified by their color (A is blue, B is red), using the chain identifier from the X-ray structure of the EmrE TPP^+-bound state (PDB: 3B5D),[8] which is also consistent with the cryo-EM based model (PDB: 2I68).[18] A green sphere has been drawn on the N-terminus of each monomer to help identify the loops. Proline residues 3, 32, 55, and 86 are also drawn to highlight their important role in terminating helices in the refined model. Supplementary animation 1 shows this representation rotated around the membrane normal. The backbone root mean square deviation (RMSD) between these two model structures is 3.8 Å.

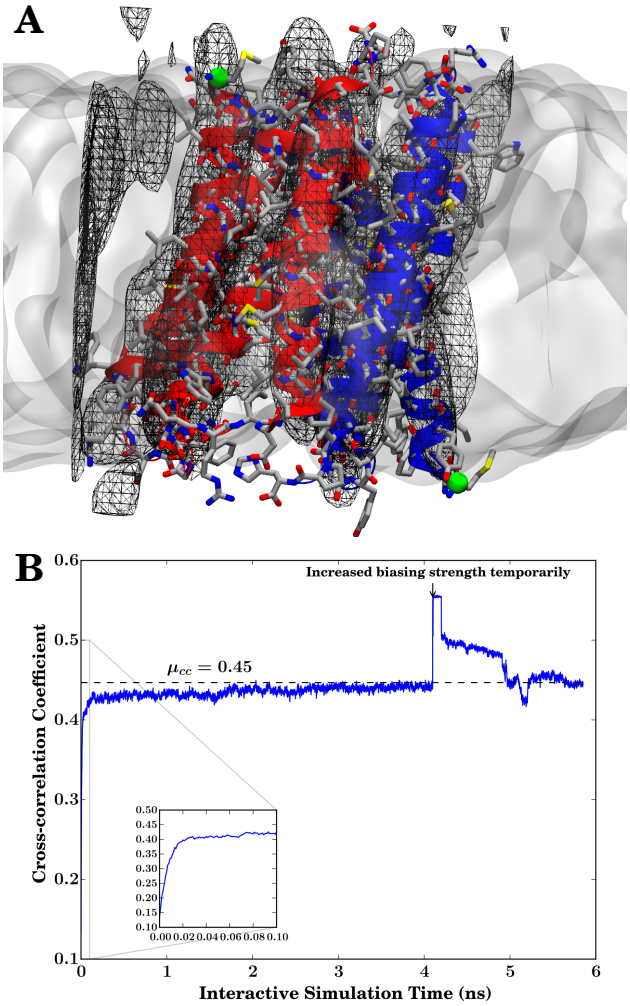

Figure 3: (A) The cryo-EM derived electron density (black wireframe) overlaid on the final structure as in Fig. 2, with the addition of a stick model for the side chain heavy atoms. Carbon atoms are gray, nitrogen atoms are blue, oxygen atoms are red, and sulfur atoms are yellow. The extent of the membrane hydrophobic core surrounding the protein is shown as a transparent surface. Note that the reported cryo-EM densities show artifacts outside the dimer. Supplementary animation 2 shows this representation rotated around the membrane normal. (B) Cross-correlation of simulated density maps from the simulation structure during interactive molecular dynamics[15] in implicit solvent against the experimental map over time. At t=0, the cross-correlation coefficient is 0.165 in the refined model based on cryo-EM,[18] compared with 0.138 for the naïve crystal structure model.[8] The mean cross-correlation at the end of the interactive MD in implicit solvent is reported ($\mu_{cc} = 0.45$), with the level shown by the dashed black line. The cross-correlation coefficient remains consistent during membrane equilibration.

refinement into two parts permitted large structural changes needed to locally refold EmrE to take advantage of the GPU implementation of implicit solvation in NAMD[29] to interactively correct secondary structure. After the large structural changes were achieved, further equilibration was carried out in an explicit membrane environment. During both steps, local structural restraints maintained the proper amino acid chirality by preventing the formation of *cis*-peptide bonds.[30]

The changes in protein structure were driven through fitting the heavy atoms of the EmrE dimer into the electron density (Fig. 3A) and applying interactive temporary forces via VMD.[15,31] The additional forces were designed to drive the system toward favorable secondary structure. A common scenario for adding temporary forces interactively involved flipping the orientation of carbonyl oxygen atoms to promote the formation of α-helical structures. Due to the geometrical constraints placed upon each residue by the protein environment, the flipping process has a high transition barrier, and therefore is unlikely to occur during conventional equilibration simulations. Other temporary interactive forces involved rotating peptide bonds to bring aromatic side chains into the hydrophobic core of the membrane, where those side chains have been shown to partition naturally into membrane bilayers.[32] By the exploratory nature of the refinement process, many parameters were changed during the simulation. Most notably, the coupling constant between the electron density and the heavy atoms (GSCALE in NAMD parlance) was not always held fixed, increasing briefly from 0.3 to 10 before being reduced back to 0.3 (Fig. 3B). In this case, high coupling between the atoms and the electron density drew protein termini to stray density elements on the periphery of the simulation box (Fig. 3A). Thus, while the cross-correlation coefficient in this case is lower (Fig. 3B), a reduced coupling constant yielded a compact EmrE dimer that was suitable to embed into a lipid environment. Nevertheless, the electron density correlation that results from this procedure (0.45) is a vast improvement on the starting value of 0.165, which is itself an improvement on the X-ray based structure (0.138).

The lipid embedding process was carried out in CHARMM-GUI,[33,34] where eighty 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) lipids were added to each of the top and bottom leaflets to form the full bilayer. This shorter lipid was chosen to mimic more closely the experimental conditions under which EmrE has been studied, where shorter lipid tails predominate,[6,7,10,11,35] rather than the longer lipids found in native bacterial *E. coli* membranes.[36] The membrane-protein system was solvated with 9,942 TIP3 water molecules[37] and enough sodium chloride (Na^+Cl^-) for a concentration of 150 mM. The total system size was approximately 50,700 atoms, with dimensions of 82 Å × 82 Å × 85 Å.

This membrane-embedded system was advanced forward through 18.8 ns of simulation time with biases applied non-interactively. NAMD 2.10[38] was used to propagate atomic coordinates with 2 fs time steps using the CHARMM36 force field for proteins[39] and lipids[40] with TIP3 water.[37] The equilibration simulation was maintained in a constant number, pressure, and temperature (NPT) ensemble using a Langevin thermostat[41,42] with a damping coefficient of 1/ps to maintain the temperature at 310 K and a Nosé-Hoover Langevin piston barostat[43,44] set to hold the membrane aspect ratio constant. The cryo-EM-derived electron density for EmrE,[3] applied as a grid potential,[14] maintained the overall topology of the helices as lipids equilibrated around them. Initially, a serine residue (S64) on Helix 3 of monomer "A" was oriented in such a way that it interacted directly with the membrane. Given that S64 of the "B" monomer was buried within the protein, such an orientation would require that helix 3 rotates substantially along the helical axis during the conformational transition. This unnatural conformation was alleviated through a rotation of 60 degrees along the helical axis using the Colvars module of NAMD[38,45] to allow S64 interaction with S64 of the opposite monomer (Fig. A1). The final membrane-embedded model is presented in Supporting Information in both binary (js) and human readable (pdb) formats.

Naturally, an accurate complete structural model would be preferred over a model that depended on manual refinement of the protein structure. In this instance, the available structural information is incomplete and contradictory. Alongside a TPP^+-bound structure, there are two apo crystal structures where two helices align perpendicularly to the others, and would lie along the membrane-water interface.[8] In this case, crystallographic contacts may have forced EmrE to adopt a non-native structure that perturbs its secondary structure, resulting in a compressed protein structure with partially unwound helices. The refinement process used here promotes helical secondary structure and helix extension across the full span of the bilayer hydrophobic region (Fig. 2). With additional experimental inputs, future models can correct for this helical bias. By leveraging available experimental inputs to guide the model, we arrive at a stable EmrE dimer corrects secondary structure deficiencies in the crystal structure. This represents a significant improvement upon previously reported EmrE models, where secondary structure disintegrates upon simulation.[19]

Simulation Protocols

The refined structural model described above forms the basis for studying the impact of protonation and substrate binding on the intramolecular interactions within EmrE, and how those interactions may regulate conformational changes consistent with alternating access transitions. The first step to prepare the structural model for simulation is to construct apo, substrate-bound, and singly and doubly protonated EmrE models. Three different EmrE protonation states were prepared using `psfgen`: 1) only protonating E14 of monomer "A" (A^+), 2) only protonating E14 of monomer "B" (B^+), and 3) protonating both E14 residues (A^+B^+). The tetraphenylphosphonium (TPP^+) cation was modeled in by aligning the TPP^+-bound structure to the newly refined structure and using the atomic coordinates found in the aligned 3B5D structure.[8] Waters within 1.5 Å of the newly placed TPP^+ molecule were removed.

In addition to the CHARMM36 force field for proteins,[39] lipids,[40] and TIP3 water,[37] parameters for the substrate TPP^+ were determined using CGenFF.[46] TPP^+ itself contains a phosphorus atom at the center of four benzene rings. This phosphorus atom has no analogous parameters in CGenFF,[47,48] and required parameterization. Using the ffTK force field development toolkit,[49] the charges and missing parameters were optimized to approximate the quantum interactions inherent to the TPP^+ cation in a classical force field, focusing on the central phosphorus atom. The parameters obtained are given as part of the Supporting Information.

Equilibrium Simulation

The five systems (apo, TPP^+-bound, and 3 different E14 protonation states) were equilibrated for 25 ns using `NAMD 2.10`.[38] During equilibration, the protein was restrained to the EM electron density with the coupling constant at 0.3. maintained the protein structure near that of its refined model structure. During equilibration under NPT conditions, the temperature was maintained at 310 K, and the pressure at 1 atm by Langevin dynamics and Langevin piston Nosé-Hoover methods, respectively.[50,51] The pressure operated on the membrane-normal and membrane-parallel dimensions separately. Short-range electrostatics were coupled to long-range particle mesh Ewald (PME)[52,53] electrostatics at 10 Å, with the PME grid set to a 1.2 Å spacing. Dynamics used a 2 fs time step, and the SETTLE algorithm[54] constrained the bond lengths to hydrogen atoms.

The equilibrated structures for each of the five simulation systems were converted into a `Gromacs`-compatible format using TopoGromacs[55] for further simulation. Using `Gromacs 5.0.4`,[56–58] five 500 ns simulations were performed for each of the five loading states of EmrE, resulting in an aggregate run time of 12.5 μs. The extensive sampling took advantage of new optimizations in Gromacs 5 that reduce computational walltimes for simple equilibrium simulations.[56] These simulations were carried out in a constant volume and temperature (NVT) ensemble, using a Nosé-Hoover thermostat[59–61] set to 310 K. Electrostatic interactions were computed as described above, with 1.2 Å grid spacing for PME after a 12 Å cutoff for short-range electrostatic and non-bonded interactions.

Driven EmrE Transitions

EmrE is unusual in that it is an antisymmetric homodimer,[6] which causes the outward- and inward-facing states of EmrE to be related by symmetry. Driven simulations were used to determine the relative ease of transition between the inward- and outward-facing states depending on protonation and substrate binding. These driven simulations used the end states from equilibrium simulation as their initial state, and used symmetry operations to construct the target state from these initial states. Biases were applied over 20 ns of simulation using the collective variables module of `NAMD`[45] such that the root mean square deviation (RMSD) to the target reached 2 Å. The TPP^+ molecule, if present, was translocated to the other side of the membrane by an additional bias. Combined, this protocol represents a simple way of driving the transition of EmrE from the outward- to inward-facing state. In this setup with no electrical or chemical gradient driving the transition, the free energy difference between the two states is implicitly zero. Thus, rather than using the non-equilibrium work for the transition to supply an upper bound on the free energy difference,[62] we use the non-equilibrium work instead to rank the height of the free energy barrier for each transition, connecting trends and interactions observed in equilibrium simulation and NMR experiments.[11]

Intermonomer E14 pK_a Calculations

Prior experiments identified a pK_a difference between the two glutamate residues present within the transmembrane portion of EmrE. These residues display pK_a values shifted by 1.2 to 1.7 units with respect to one another, depending on the temperature.[17] To determine which monomer has which pK_a, a replica exchange thermodynamic integration (RETI)[16] calculation was conducted in `NAMD 2.10`[38] to determine whether E14 from monomer A or B ($E14^A$ or $E14^B$) will be protonated first during the transport process. When the transition parameter $\lambda = 0$ in the perturbation calculation, the proton occupies $E14^B$ (equivalent to the B^+ state). That proton transits in an alchemical manner to $E14^A$ when $\lambda = 1$ (the A^+ state). `NAMD` uses the dual-topology paradigm for alchemical simulations. Since the partial charges on glutamate change after protonation, a 20 kcal/mol/Å^2 harmonic potential between equivalent heavy atoms of the two interconverting species was applied. This additional restraint eliminates unphysical conformational sampling that can result in slow convergence of the calculations.[63,64]

In order to optimize the transition rate while being mindful of the computational cost, the λ values for RETI simulation were chosen to be $\lambda \in \{0.0, 0.02, 0.08, 0.14, 0.2, 0.26, 0.32, 0.38, 0.44, 0.50, 0.56, 0.62, 0.68, 0.74, 0.80, 0.86, 0.92, 0.98, 1.0\}$, yielding exchange acceptance probabilities of at least 7% between adjacent replicas with frequent exchanges between replicas (Fig. A2). The initial coordinates were drawn from the A^+ and B^+ simulations (10 states from A^+, 9 from B^+) to minimize perturbation to the water network during the alchemical transition between these two states. Each of the 19 replicas was simulated for 10 ns to obtain the final result.

Analysis Methodology

Purpose-built `VMD`[31] scripts were written to analyze protein dynamics, water permeation, and connectivity during the equilibrium trajectories. These scripts utilized the python interface of `VMD`[31] to facilitate interoperability with the `NumPy`,[65] `SciPy`, `NetworkX`,[66] and `Matplotlib`[67] `python` packages for further analysis. The analyses performed include simple RMSD, root mean square fluctuation (RMSF), and distance measurements, as well as more elaborate evaluation of the geometries observed during simulation. To determine if water wires formed and if they might be conducive to proton transfer, directional networks of hydrogen bonds were constructed using `NetworkX`,[66] and evaluated for connectivity between distant water molecules on either side of the transporter.

Specifically, we say EmrE is leaky to water if a trail of water molecules whose oxygen atoms are within 3.2 Å of their neighbors exists such that they connect any pair of water molecules where $|z| > 11.8$ Å and $\Delta z > 23.6$ Å, which are on opposite sides of the lumen. Additionally, a single snapshot of EmrE is assumed to permit proton leakage if any two water molecules fulfilling the same criteria can be connected via a series of directional hydrogen bonds that might allow rapid proton traversal of the membrane span. The statistics are computed by evaluating the existence or absence of water pores or proton conduction pathways every 5 ps, the frequency at which snapshots from the trajectory were saved. Since proton-conduction pathways are uniquely short lived, their mean duration is likely overestimated.

A similar approach was employed to compute the water-mediated hydrogen bonds between residues within EmrE. For every snapshot from the trajectory, directional hydrogen bonding networks were generated based on the geometry of each protein residue or water molecule, using only the protein side chains, excluding hydrogen bonds to the backbone. The pathlength between two residues was determined from this directional network. The pathlength represents the number of edges in the path between two residues, or equivalently the number of intermediate waters in the interaction plus one. Note that since the hydrogen bonding pathway is directional from donor to acceptor, we only count the interaction if a donor-acceptor relationship might exist between the residues. This excludes water mediated interactions between two donors or two acceptors.

Contacts between TPP^+ and EmrE were evaluated using a distance-weighted contact criteria:

$$C_i = \sum_{j \in TPP^+} \frac{1}{1 + \exp\left(5\,\text{Å}^{-1}\left(d_{ij} - 4\,\text{Å}\right)\right)} \tag{1}$$

In this manner, every contact between protein heavy atoms (i) and heavy atoms in TPP^+ (j) can be aggregated together to visualize and quantify the strength of the contacts between residues and TPP^+. This

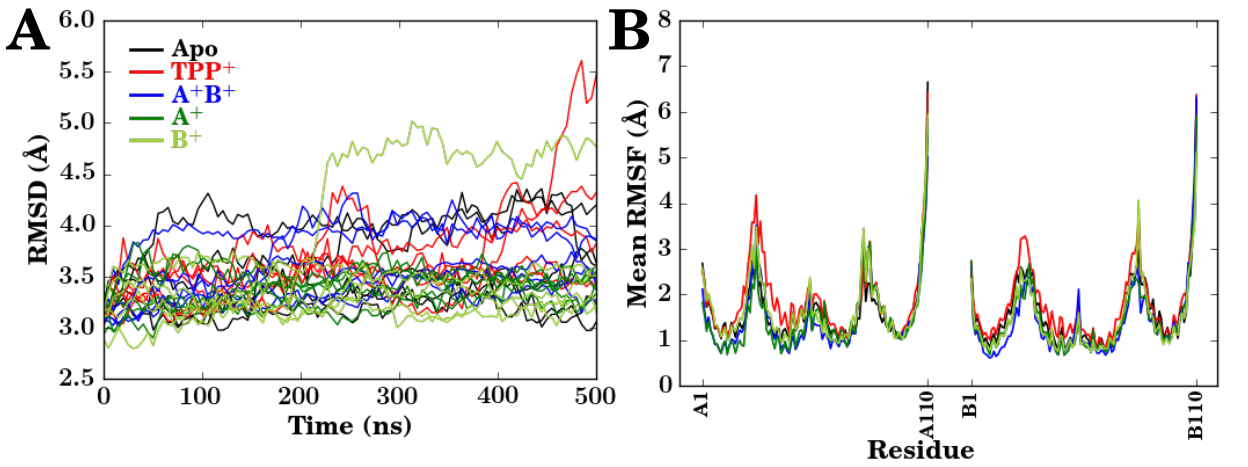

Figure 4: Structural deviation and fluctuation measured over the course of the trajectories. (A) RMSD of each trajectory compared against the 3B5D crystal structure. Due to limitations of the original crystal structure, the comparison strictly involves C_α positions reported in the crystal structure, and does not include side chain atoms. The color of the lines indicates the loading state of EmrE: black for the apo state, red for the TPP^+-bound state, blue for the doubly-protonated state (A^+B^+), green for the state where the proton is bound to monomer A (A^+), and yellow-green when the proton is bound to monomer B (B^+). (B) Mean RMSF per residue after trajectory alignment for each loading state of EmrE, using the same color scheme as in (A). For a detailed trajectory-by-trajectory analysis, see Fig. A3.

formulation has been used previously for quantifying lipid-protein[68] and protein-lignin interactions,[69] as it more strongly weights shorter interactions such as hydrogen bonds or π-stacking relative to longer-range, usually nonspecific contacts.

Results & Discussion

The model generated by the refinement approach is a marked improvement over the starting crystal structure, which features only C_α atoms. In a simple static test of the refined model, structure checks were conducted to assess the overall quality of the model. Crystallographers frequently use the molprobity score as a metric to assess their structures, with a score commensurate with the approximate resolution.[70] The molprobity score for the model was 0.89, a result that would be expected from a sub-Ångstrom resolution crystal structure,[70] compared to the actual crystal structure resolution of 3.8 Å.[8] In excess of 95% of side chains are in their favored rotameric and Ramachandran regions of conformational space, with only 2 outliers each in the 220 total amino acids of the structure. Separate analysis with PROCHECK[71] shows 93% of residues occupy their most favored region, compared with 73% in a recently generated EPR-based model.[72] Through dynamic simulation, additional features of the modeled structure become apparent, as described below.

Stability of the Modeled EmrE Structure

With the extensive structural remodeling that took place prior to simulation and the poor resolution of the starting structure, the model of EmrE generated here might be expected to suffer from instability and fall apart over extended simulation. While the RMSD with respect to the crystal structure can be high relative to other simulated membrane proteins (Fig. 4A), the majority of the simulations show RMSDs that are comparable to, or below, the crystal structure resolution (3.8 Å), as is generally the case for membrane proteins.[73] The two notable exceptions are one B^+ simulation and one TPP^+-bound simulation. In the TPP^+-bound case, lipids intercalate into the dimer interface formed between helix 2 of each monomer, and directly interact with TPP^+. This splits apart the dimer, leading to a large RMSD for this simulation

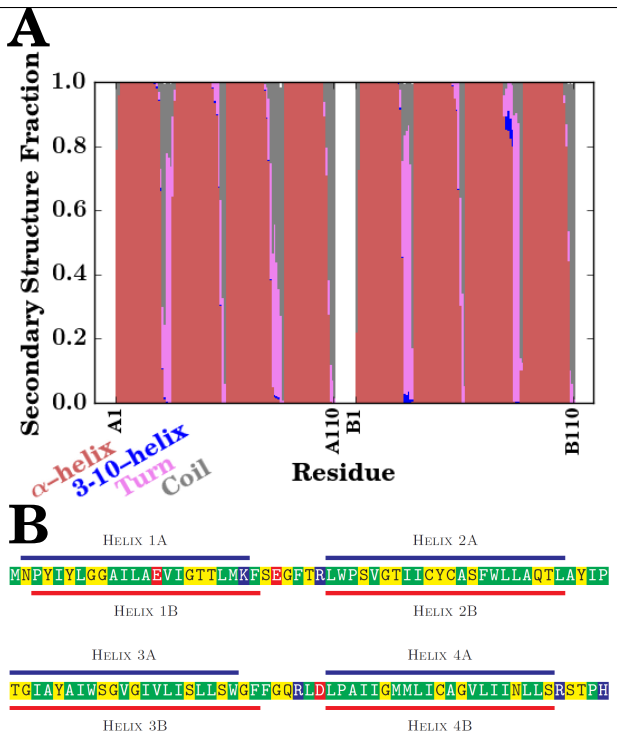

Figure 5: Representations of the predominant secondary structure over all trajectories. (A) A residue-by-residue breakdown of secondary structures present, and their relative predominance throughout the simulations as determined by `Stride`.[21] Each secondary structure type observed is shown as its own color in the cumulative barplot above; red for a conventional α-helix, blue for a 3-10-helix, pink for turns, and gray for random coils. (B) Residues that are α-helical 80% of the time were identified and grouped into individual α-helices. The helices formed in this model differ slightly between monomers, as indicated by the lines above (monomer A, blue) and below (monomer B, red) the EmrE sequence, which has been colored according to hydropathy[74] and typeset using TeXshade.[75] This mapping includes residues 2–22, 30–51, 56–76, and 85–105 of monomer A, and residues 3–23, 30–51, 56–78, and 85–105 of monomer B. As a comparison, the model of Fleishman[18] identifies residues 4–21, 34–52, 58–80, and 87–104 as helical. The conservative Fleishman helix definitions are used consistently in other analyses performed on the simulations.

(Supplementary animation 3). For the B^+ case, a helix rotation causes the formation of a continuous water channel, which increases the RMSD relative to the crystal (Supplementary animation 4).

Based on the per residue RMSF measurements conducted over these same trajectories (Figs. 4B and A3), no large-scale conformational changes are observed over the simulations. The RMSFs are comparable to what would be expected for a dynamic and diminutive α-helical membrane protein, with small fluctuations in the helical regions and larger fluctuations in the connecting loops and termini (Fig. 4B). Furthermore, the fluctuations that lead to spontaneous conformational changes move the protein away from the inverted model that would reflect an inward- to outward-facing state transition (Fig. A4). Since the duration of individual simulations (500 ns) is substantially shorter than measured turnover rates (ms time scale[10,11]), spontaneous transitions are unlikely to occur. Thus, biased simulations are needed to drive any transitions between outward- and inward-facing states.

An essential feature of membrane proteins is the secondary structure formed by the polypeptide spanning the low-dielectric environment of the membrane.[24,25] High RMSD can hide the dissolution of protein secondary structure, including the eight transmembrane α-helices of EmrE that collectively barely span the hydrophobic core of the DMPC bilayer into which they were placed (Fig. 2). By explicitly determining the secondary structure throughout the simulations using `Stride`,[21] we observed that the α-helical secondary

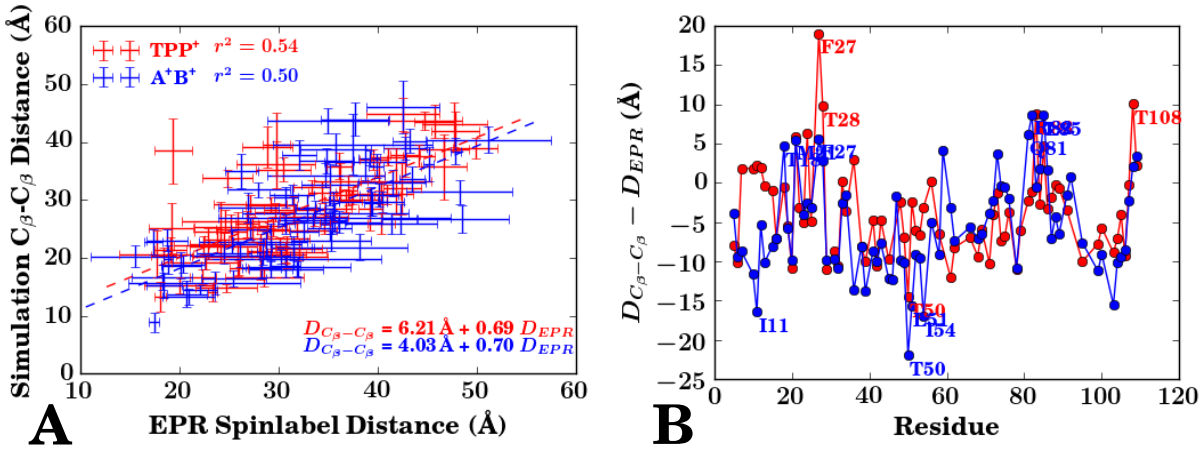

Figure 6: Comparison between measured β carbon distances from simulation and the equivalent spin label distances from EPR studies.[72] (A) Correlation plot between equivalent conditions, using the TPP^+-bound and pH 5 states from Dastvan et al.[72] as a comparison for the TPP^+-bound (red) and A^+B^+ (blue) simulated states. The equation for the dashed trendlines is given in the lower right corner, with the correlation coefficient given next to the in-figure legend. (B) Identification of residues with the largest deviation between our simulated distances and the EPR measurements. Individual residues that differ by more than 10 Å from the mean difference are annotated.

structure is maintained in all four transmembrane segments per monomer (Fig. 5A). The identity of the residues belonging to each helix differs slightly depending on the monomer (Fig. 5B), likely due to the different structure of the connecting loops between the individual helices pulling on the ends of the helices. Additional experimental restraints, e.g, those derived from future NMR experiments, would be instrumental in refining the exact secondary structure assignment of each residue, and improve the veracity of the simulated structure.

One example of an experimental observable that could be applied comes from recently published measurements of interresidue distances measured by electron paramagnetic resonance spectroscopy (EPR).[72] Due to their recent publication, these EPR distance restraints were not considered during model construction in this work. The distances measured through EPR suggest substantial conformational change between different loading states for EmrE.[72] Over our own simulations, which we can compare against these EPR results only because of the atomic detail of the model, large conformational differences were not observed between different loading states, as the distances between individual side chains are on average invariant with loading state (Fig. A5). The EPR experiments focused on two loading states for EmrE: a TPP^+-bound form for EmrE; and EmrE at pH 5, which corresponds to our doubly protonated A^+B^+ state. The distance between equivalent residues in both monomers was measured using site-directed spin labeling,[72] which we compared against the measured C_β-C_β distances from our own simulations (Fig. 6). The correlation between the EPR spin label distances and the distance between β carbons (C_β) is low (Fig. 6A), and the trendline has a slope different from unity. The difference in slope is due to spin labels being larger than individual residues, which result in EPR-measured distances that are larger than the actual C_β-C_β. The low correlation coefficient comes primarily from residues within loops of the protein with measured distances that are substantially different in each method (Fig. 6B). Since the cryo-EM densities were weakest between helices, it is unsurprising that the loop structures in our model have the weakest agreement with the EPR results. Incorporating spin label distributions into future refinements of EmrE would improve loop region structure.

Water Analysis

In investigating the different loading states of EmrE, an unexpected event occurred—the lumen of the doubly protonated A^+B^+ state spontaneously dehydrated (Fig. 7). Other protonation states show typical behavior for a membrane transporter, with lumen water that can directly exchange with bulk water on one side of

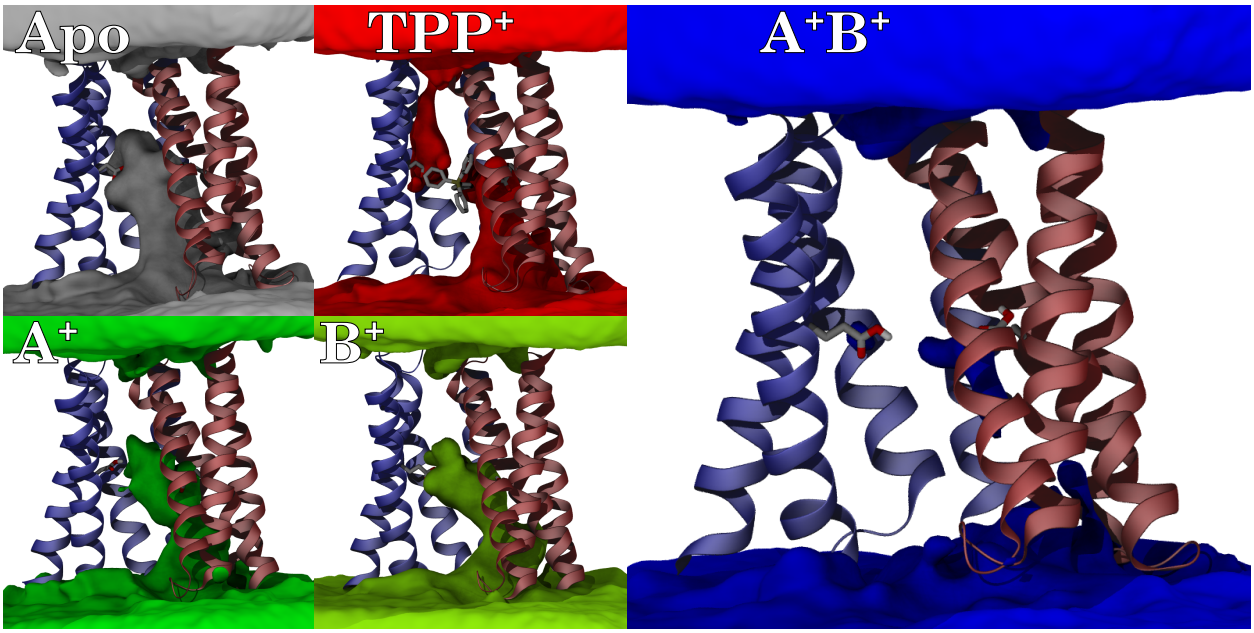

Figure 7: Visualization of the mean water occupancy across all five replicates within the lumen of EmrE under different loading conditions. Regions where water is frequently present are represented by isosurfaces of different colors (gray for the apo state, red for the TPP^+-bound state, green for A^+ and B^+ states, and blue for the A^+B^+ state), along with a cartoon representation of EmrE for context, including an explicit representation of E14 and TPP^+ where present. Animations of these views during rotation is presented in the Supporting Information (Supplementary Animations 5A-E).

the membrane. In contrast, the connection from the lumen to bulk water was severed in the A^+B^+ state, and the transporter spontaneously transitioned to a new state. Although this state looks occluded because the pocket of water in the lumen is disconnected from bulk solution, no large-scale conformational change has taken place (Fig. A4). Instead, water interactions with the protein appear too weak to maintain a water channel into the lumen in the A^+B^+ state, which lacks strong electrostatic interactions brought about by the charged glutamates. Thus, the water pathway spontaneously breaks. The rapid formation of an occluded intermediate is consistent with pH-dependent NMR studies, where low pH facilitates rapid conformational transition.[11]

The TPP^+-bound state (Fig. 7) appears weakly leaky, with a visible water pathway between both top and bottom bulk water regions to the TPP^+ binding site. Water leaks have been reported previously for other membrane transporters,[76] thus it is reasonable for a highly dynamic transporter such as EmrE to exhibit leaks. However, water leaks in EmrE are particularly interesting since EmrE transport is driven by a proton gradient. Thus, any significant water leaks could potentially serve as a conduit for protons across the membrane, short-circuiting the transport cycle (Fig. 1). The leaky states, where a water path exists between both sides of the membrane, occur only transiently (Table 1), accounting for less than 10% of the total simulation time. These leaky states are inhomogeneously distributed across the simulations conducted (Figs. A6 and A7). Critically, the leaky states are often not conducive to rapid proton translocation via a water wire. Frequently, one or more waters within the path orients sub-optimally for proton conductance due to interactions with the surrounding residues, and disrupts the chain of hydrogen bonds needed to form a conductive water wire. Thus the transient water wires observed in our simulations would not serve as efficient proton pathways,[77] thereby maintaining the proton gradient that drives substrate export.

Water dynamics are particularly sensitive to the loading state of EmrE. The TPP^+-bound or A^+B^+ states exchange lumen waters with bulk solution more slowly than do the other loading states tested (Table 1). For the A^+B^+ case, this delay stems from the irregularity with which the water molecules isolated within the lumen can exchange with bulk water during the sporadic formation of water pathways at the dimer interface, as evidenced by the low number of water molecules in the lumen at any given time (Fig. 8). TPP^+ blocks the

	Water Pore Existence		H^+ Conductance		Mean Lumen Water
	Probability	Mean Duration (ps)	Probability	Mean Duration (ps)	Exchange Time (ns)
Apo	11.6%	15.7 ± 0.1	0.3%	< 5.6	11.1 ± 0.4
TPP-bound	11.8%	12.9 ± 0.2	0.6%	< 6.1	18.8 ± 1.5
A^+B^+	3.9%	12.5 ± 0.1	0.7%	< 5.8	21.6 ± 3.6
A^+	0.7%	11.0 ± 0.2	0.1%	< 6.1	13.1 ± 0.8
B^+	4.1%	10.5 ± 0.1	0.1%	< 5.5	14.8 ± 1.0

Table 1: Formation propensity and lifetime of water pores in different loading states of EmrE. A water pore exists at a specific time-point if waters connect on either side of the membrane outside of the transporter lumen ($|z| > 11.8$ Å). Similarly, we say that EmrE is in a H^+ leaky state if a water wire exists across the lumen that might transfer a proton by a concerted Grotthuss-like mechanism. The final column measures the mean time taken for lumen water to exchange, as measured by determining the number of frames required for all lumen water molecules to have been replaced by water molecules from the bulk. Reported ranges indicate the standard error, not the standard deviation.

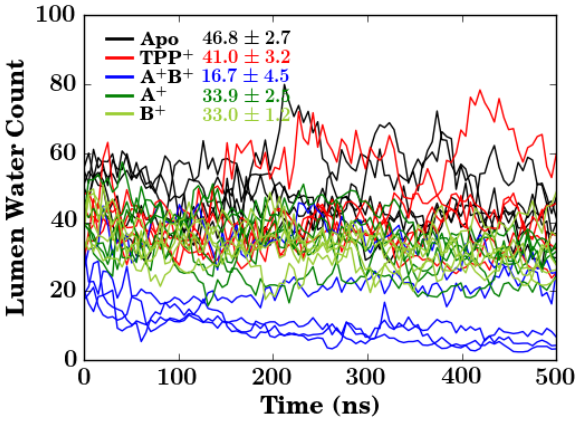

Figure 8: The number of water molecules within the lumen of EmrE as a function of time for the different loading environments. All 25 independent trajectories are shown simultaneously and are colored according to the EmrE loading state, with the mean number of waters reported beside the in-figure legend.

E14 Chain	E14 Partner		Hydrogen Bond Propensity				
			Apo	TPP-bound	$H_A^+H_B^+$	H_A^+	H_B^+
A	A	T18	63.2	65.8	75.2	60.4	63.5
		Y40	14.5	10.9	0.7	1.5	7.5
		S43	5.4	10.7	–	0.4	1.5
		W63	32.8	13.8	38.6	41.0	57.1
A	B	Y60	8.6	0.1	10.3	3.5	32.2
B	A	Y60	–	11.7	15.9	0.1	0.1
		R106	6.1	27.2	–	0.2	0.3
B	B	T18	57.8	58.7	71.4	60.1	54.2
		Y40	6.8	9.2	0.2	9.4	1.5
		S43	22.1	16.2	–	9.0	1.4
		W63	15.9	8.7	71.7	44.2	51.0

Table 3: Hydrogen bonds formed by protein side chains to E14, reported as a percentage of the total simulation time these hydrogen bonds existed. The results are grouped according to which pair of monomers are involved of the interaction. Note that unlike Table A1, donor and acceptor interactions to E14 are aggregated together into a single element on the table.

exchange of water through a different mechanism, trapping it within the lumen by limiting the accessibility of water molecules nearest to the substrate and thereby retarding their exchange. This observation is consistent with the model from Fig. 1, since slower water disconnected from the bulk would be the first step in forming an occluded state. Thus, viewed from the perspective of water dynamics, the A^+B^+ and TPP^+ bound states are closer to transitioning than the other loading states tested.

Residue-Specific Interactions in EmrE Structure

The current model provides new information about atomic interactions within EmrE. In particular, we can now elucidate how hydrogen bonding patterns change during the transport cycle. Table A1 details the interactions that exist throughout the anti-symmetric dimer structure modeled here. Many interactions, particularly those involving residues on the periphery of the protein, remain unchanged under the different loading states for EmrE tested here. Other robust interactions include those internal to the structure, such as polar side chains interacting with nearby backbone carboxy groups like the serine to alanine (S43–A10) interaction exemplified in Fig. 9.

The most significant changes in hydrogen bonding occur in the vicinity of E14 as it responds to different loading states (Table 3). Previous mutagenesis experiments have identified these residues as important to the function of EmrE.[20,78–81] Tryptophan W63, for instance, is highly conserved in the SMR transporter family (Fig. A9), and its mutation changes EmrE into a uniporter of cationic substrates.[78] Mutations to tyrosine Y40 or serine S43 reduce the effectiveness of EmrE in exporting toxic substrates,[20,79] and S43 has been implicated in the specificity of EmrE to its substrates.[80] The tyrosine residue Y60 is one of the few amino acids completely conserved in the SMR family (Fig. A9). The conservative mutation Y60F renders EmrE nonfunctional.[20,81] A reduced level of resistance to antibiotics is conferred by Y60T,[20] suggesting a role for the hydroxyl of Y60 in regulating the transport cycle. Taken together with the hydrogen bonding data presented in Table 3, the mutation studies identify critical roles for each of these residues in tuning the interactions of E14 and its surroundings, and establishes how these interactions are affected by the loading state of EmrE. For example, the proton from W63 interacts with the side chain of E14 from the same monomer if E14 is deprotonated, but stretches across to the backbone or other nearby residues instead if E14 is protonated (Fig. 9).

Measuring direct hydrogen bonds tells only part of the story. Water-mediated hydrogen bonding, in which water molecules bridge the gap between the donor and acceptor, also occurred between E14 and several

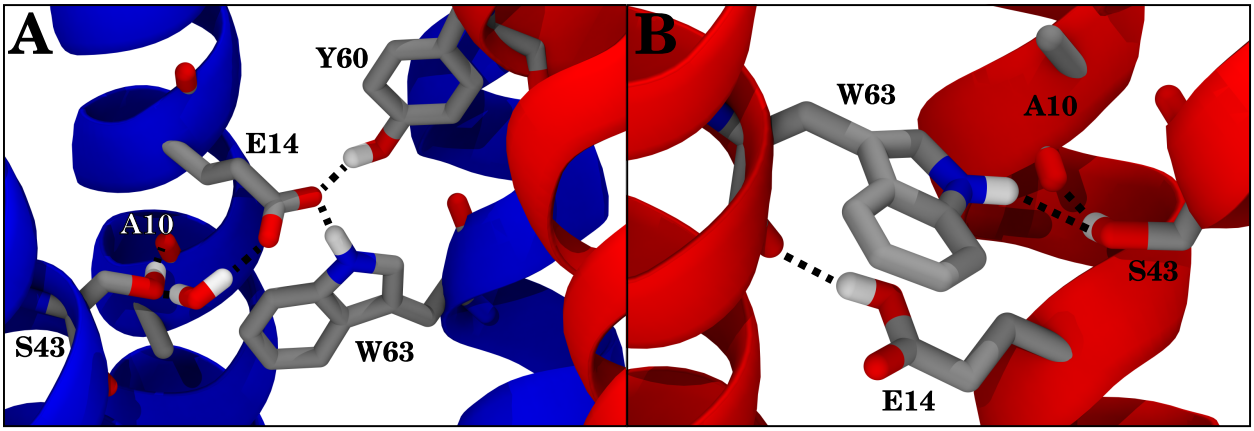

Figure 9: Examples of hydrogen bonding (black dashed lines) during simulation, and how they change with the protonation of E14. (A) Example interaction network surrounding deprotonated E14, highlighting multiple influential interactions to E14 of monomer A (blue cartoon). These include W63 from monomer A, Y60 from monomer B (red cartoon), and an example of water mediated interaction (S43). Helices 1 and 2 of monomer B have been omitted for clarity, as have hydrogens not directly involved in hydrogen bonds. (B) Example interaction network when E14 is protonated, detailing how E14 can act as a donor to the backbone of W63 rather than as an acceptor as it was in (A), which in turn changes W63 into a potential donor for S43.

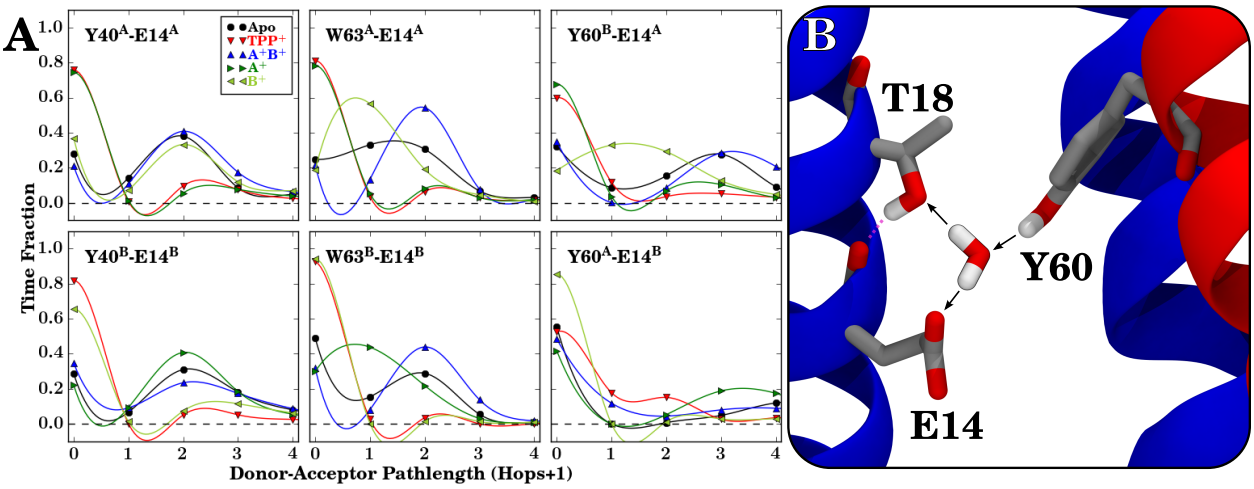

Figure 10: (A) Water mediated hydrogen bonding within EmrE for selected residue pairs, and (B) an example snapshot showing how these measurements are made. The pathlength between two residues is determined from a directional hydrogen bonding network (black arrows shown in the example), and the distribution of this pathlength over the aggregate trajectories for E14-Y40, E14-W63, and E14-Y60 pairs in each EmrE loading state is reported above in the labeled subpanels. The labels from these subpanels indicate the monomer of the residue in question (A or B) after the residue number. Note that since the hydrogen bonding pathway is directional, we only count the interaction if a donor-acceptor relationship might exist between the residues. Thus in the example pictured, Y60 has a pathlength of 2 to both T18 and E14, as they are connected via the shown water, while T18 has a pathlength of 0 to E14, signifying that the two side chains are disconnected since the bridging water acts as the donor to both residues. To assist in determining the location of the datapoints for each loading type and the visualize the overall pattern, the datapoints have been connected with an interpolating polynomial in the plots in (A).

other residues: Y40, W63, and Y60 (Fig. 10A). The water-mediated hydrogen bonding pattern observed within a single monomer (Y40 and W63) remained unchanged across monomers, with the similar patterns in both the A and B monomers suggesting that the interactions are retained throughout the transport cycle (Fig. 10A). Critically, the interaction between Y60 from monomer B ($Y60^B$) and E14 from monomer A ($E14^A$) (Figs. 9A and 10B) is clearly inequivalent to the interaction between the opposite set of monomers. Additionally, this interaction is strongest when $E14^A$ is deprotonated (the B^+, TPP^+, and apo states), suggesting that the protonation of monomer A acts as an "electrostatic lock" that couples the monomers together and prevents conformational change when $E14^A$ is deprotonated. This mechanistic hypothesis was tested by driven conformational changes, as discussed below.

Contacts between the C-terminus of monomer A and $E14^B$ unexpectedly appear in some states (Table 3). These interactions do not always exist. Instead, they are the result of heterogeneous populations of conformations in the C-terminal regions of EmrE. Due to the closure of the lumen to one side, only the C-terminus of monomer A can make these interactions (Fig. A8, Supplementary animation 6). As a result, the C-terminus of both monomers effectively experience independent environments (Fig. A8), and these environments may change rapidly as the terminus is exposed to bulk solution or the protein lumen. The environmental diversity of the C-terminus may affect the terminal histidine, whose exposure to different environments can vary greatly. The heterogeneity of histidine exposure may result in alternative protonation states rather than the assumed N_ϵ protonation in this model. Further experimental studies[82] or quantum mechanical simulation[83] would be needed to explore this phenomenon.

TPP^+ Dynamics and Membrane Binding

With the newly refined atomic resolution structure including an explicit TPP^+ molecule borrowed from the low-resolution crystal structure, we are uniquely positioned to elucidate the nature of the interactions between TPP^+ and its binding site within EmrE. The observed binding site for TPP^+ was consistent across the trajectories, focusing primarily on the adjacent aromatic residues (Fig. 11). TPP^+ came into contact with a number of other residues as well, including some on the C-terminal helix of monomer A (Fig. 11). The nonspecific nature of the observed interaction aligns well with the polyspecific nature of EmrE transport.[2] EmrE requires only two features of its substrates: 1) the substrate must be a cation so it is attracted to the electrostatic potential created by both E14 residues, and 2) the substrate should have aromatic groups to satisfy the π-stacking requirements of neighboring aromatic residues.

As mentioned previously, lipids play an active role in the structural dynamics of EmrE, and consequently may influence the TPP^+ binding pathway. Lipids wriggle their tails between helical interfaces and can influence protein structure strongly, as observed in some high RMSD cases (Supplementary animation 3). Indeed, the portal formed by cracking apart the interface between Helix 2 of both monomers opens in a subset of simulations (Fig. 12). We propose two explanations for this phenomenon: 1) the residue packing along that interface is non-optimal in the proposed model, resulting in lipids wedging them apart; or 2) these gaps exist *in vivo* as well, and might serve as the conduit for TPP^+ binding and unbinding as it arrives via a membrane-derived route.

The lipid binding regions at the interface between the two Helix 2 are large enough to accommodate TPP^+ entry in some cases (Fig. 12, Supplementary animation 7). As a hydrophobic cation, TPP^+ is enriched in the membrane by a factor of 100 relative to solution.[84] Thus, as in other membrane proteins whose substrates or ligands arrive via a membrane-embedded route,[85–87] EmrE may increase its efficiency by allowing direct access of the substrate through the gap between neighboring Helix 2 from each monomer. However, since the pathway for TPP^+ binding is currently unresolved, further study will be required to distinguish between the two alternative hypotheses.

Driving the Transition of EmrE

The lock formed by the interaction between $Y60^B$ and $E14^A$ would be expected to slow the transition for the apo state relative to the protonated states. This hypothesis can be tested through driven simulations, where the average nonequilibrium work needed to induce the transition can be compared between different loading states to deduce which have the lowest barrier to conformational transition. Any driven simulation requires an initial and a final state. The initial states for the transition are seeded from the final state of

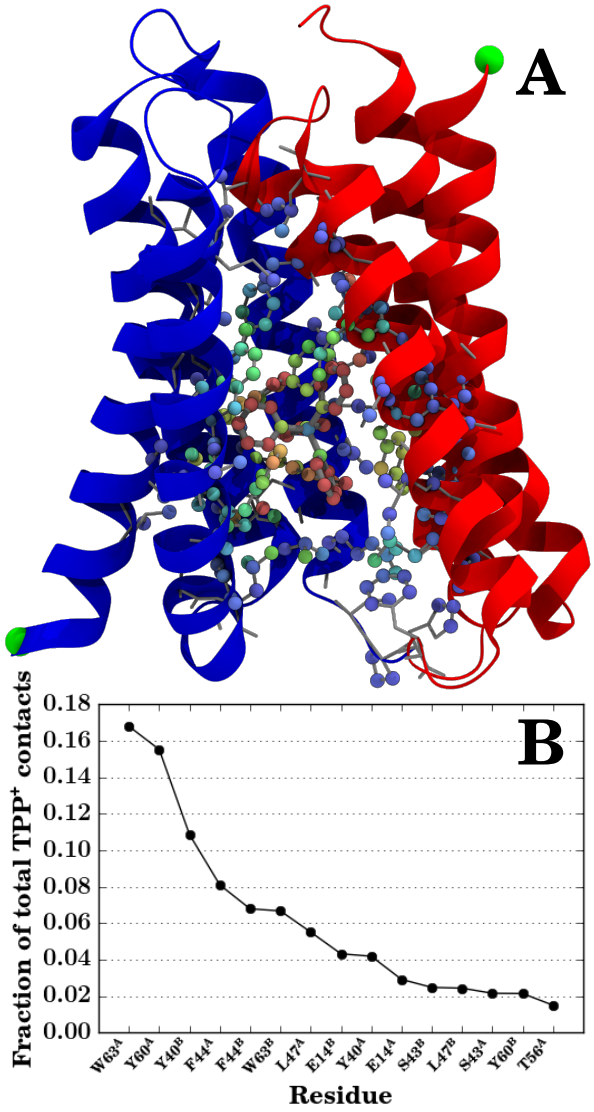

Figure 11: TPP^+ contact analysis to elements within EmrE. (A) Pictorial representation of the contact data, where each heavy atom that contacted TPP^+ (and TPP^+ itself) is colored according to the number of contacts. Bluer atoms had fewer contacts, and redder atoms had more contacts. A cartoon representation of the protein, and gray licorice representations of every side chain are also included in the figure for context. (B) Ordered listing of the top 15 amino acids with the most contacts with TPP^+, expressed as a fraction of the total number of TPP^+ contacts over all trajectories as computed by Eq. 1.

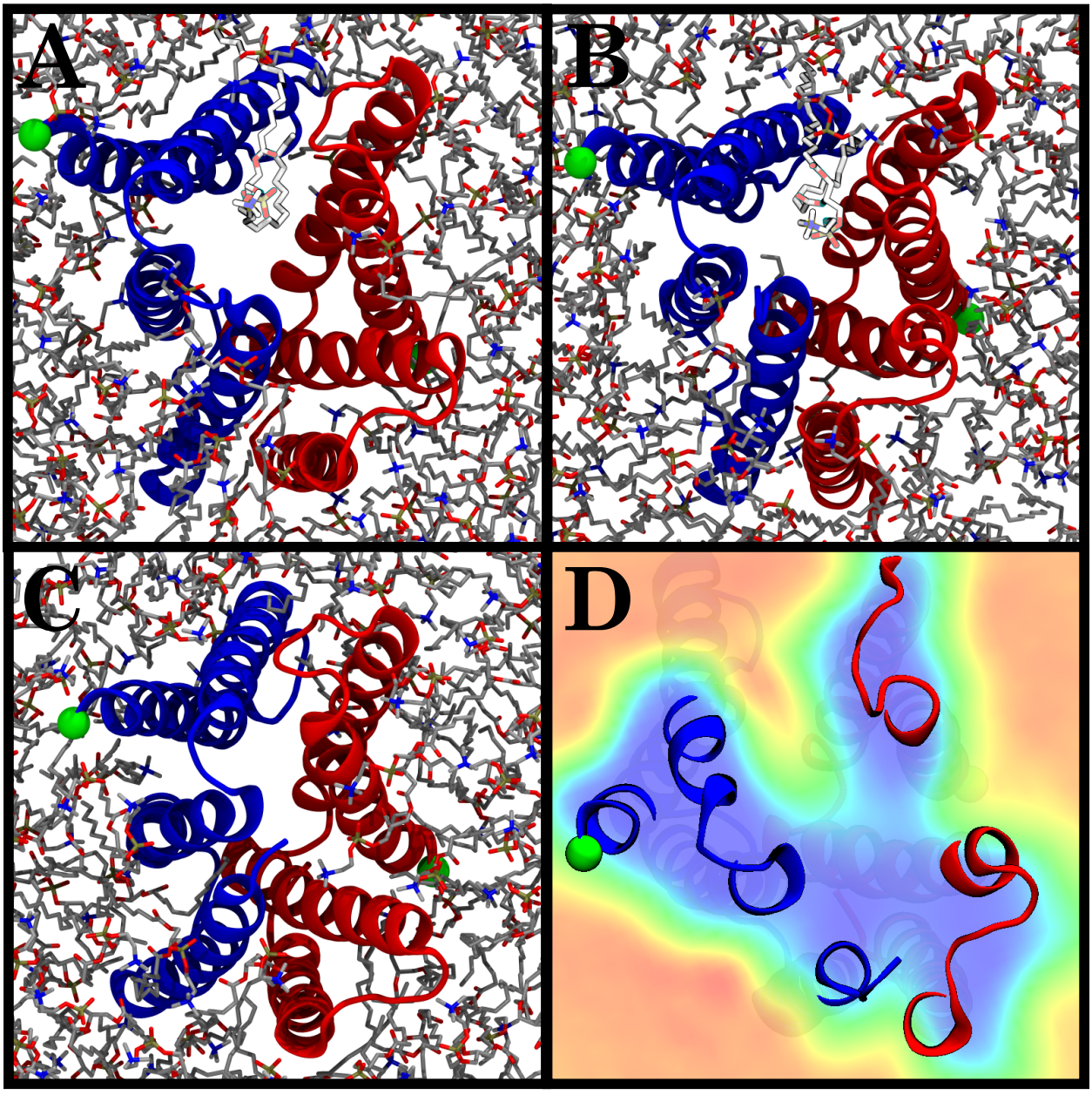

Figure 12: Lipid intercalation into the interface between Helix 2 monomers. (A & B) Snapshots from the trajectory where an intercalated lipid (highlighted lighter-colored lipid) splits the monomers apart (blue for monomer A, red for monomer B, with the N-terminus tagged in green). This is a stochastic process, and the intercalation is not always present (C). (D) The lipid occupancy across all simulation conditions for a specific slice along the membrane normal, with higher lipid occupancies colored in red, and lower occupancies in blue, with the protein provided for context. Sequential slices along the membrane normal are displayed in Supplementary animation 7, highlighting the specificity of lipid intercalation to the "open" side of the transporter. The snapshots shown all have the "open" side of the transporter (the side accessible to solution) oriented toward the viewer.

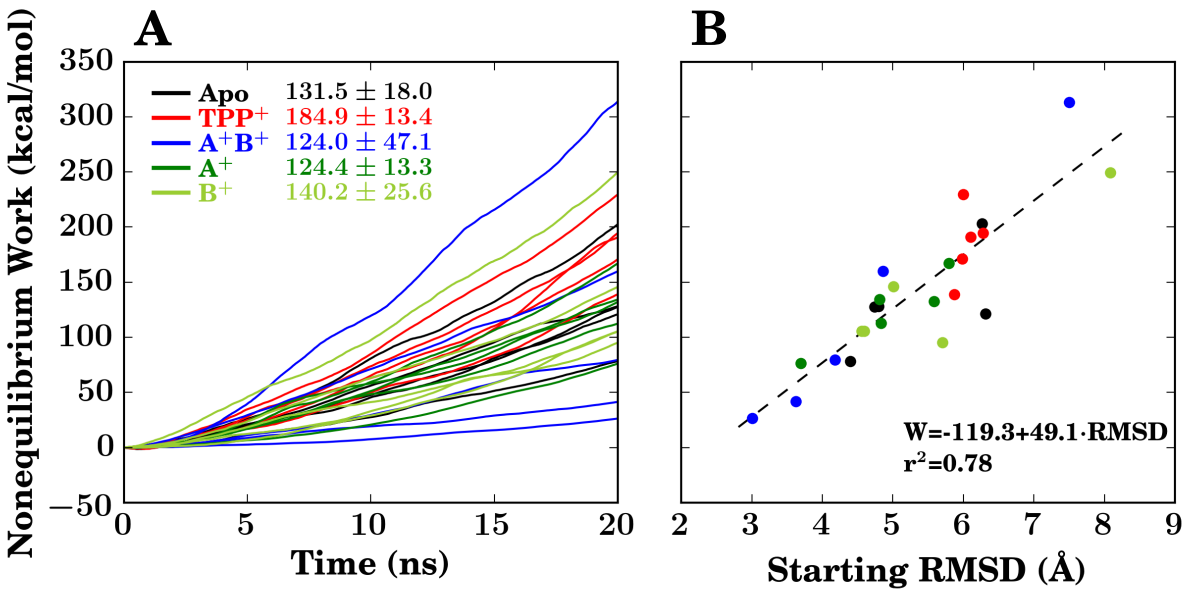

Figure 13: Nonequilibrium work profiles (A) associated with the conformational transition of EmrE for each loading state, and (B) the relationship between the total nonequilibrium work and the starting RMSD difference between the initial states (taken from the equilibrium trajectories), and the final states (which are symmetry related to the initial state). The mean and standard deviation of the nonequilibrium work required for the transition in each of the 5 independent loading states are reported adjacent to the figure legend, using the standard color scheme (black is apo, red for TPP^+, blue for A^+B^+, and dark and lighter green for A^+ and B^+). Fit parameters describing the linear relationship between the nonequilibrium work and the starting RMSD are presented within panel B.

the equilibrium trajectories. Exploiting the unique symmetries presented by an antisymmetric homodimer, swapping the conformations of each monomer generates the target structure. A biased simulation then drives the initial structure to the target structure by minimizing the RMSD to the target, with an example transition shown in Supplementary animation 8. Using RMSD as a collective variable is known to generate transitions that may be nonphysical.[88] For the particular case of EmrE in the absence of a gradient driving conformational change, we know that the true free energy change must be zero, as the initial and target structures are related by symmetry to one another. Thus what we measure here in the transition is purely the dissipative component of the non-equilibrium work, with less work implying lower transition barriers along the reaction coordinate, which can be qualitatively compared with measured transition rates from NMR[6,11] and the understood transport cycle (Fig. 1).

The nonequilibrium work results (Fig. 13) contrast with the transport cycle of EmrE (Fig. 1). A forbidden apo transition should have the highest nonequilibrium work of the states tested, signifying a high barrier to the transition. Instead, the TPP^+-bound transition requires the most work (Fig. 13A). The higher work for the TPP^+-bound case suggests that the TPP^+-bound transition should be "forbidden," in contrast with its physiological role. Although the high barrier for TPP^+-bound transition is inconsistent with the simple transport model from Fig. 1, it is consistent with measured transition rates from previous NMR studies.[6,11] Protonation of EmrE experimentally increases the conformational transition rate in the drug-free state from $40\,s^{-1}$ to $220\,s^{-1}$,[11] both of which are much faster than the conformational exchange rate of $5\,s^{-1}$ when TPP^+ is bound to EmrE.[6] Likewise in our simulations, the barrier for the A^+B^+ state is lowest, followed by the singly protonated A^+ and the deprotonated apo or singly protonated B^+ states, which are distantly followed by the TPP^+ bound state. Thus, our new EmrE model accurately recapitulates available experimental transition rates, and for the first time allows us to connect these observed rates to specific interactions induced by the protonation state of E14, as well as the dehydration of the lumen upon E14 protonation.

The work values can also guide us towards additional insight as to the distribution of states observed

$\Delta G^{B^+ \rightarrow A^+}$ (kcal·mol^{-1})	ΔpK_a^{B-A}
1.53 ± 0.12	1.08 ± 0.08

Table 4: Computed ΔG and ΔpK_a for the movement of a proton from E14^B to E14^A in replica exchange thermodynamic intergration (RETI) calculations. $\Delta pK_a = \Delta G \cdot \log_{10}(e)/RT$

in equilibrium. For example, the A^+B^+ state yields both the highest and the lowest nonequilibrium work value (Fig. 13B). The underlying reason for this result is the RMSD difference from the initial to the final state. Since the starting conformation comes from the end of the equilibrium trajectory, and the target state is symmetry-related to the initial state, equilibrium conformations that are more symmetric than others undergo smaller conformational changes to exchange between these defined states. It must be emphasized that these conformational exchanges are enough to flip the accessibility of the lumen from one side of the membrane to the other, no matter the difference in RMSD between the initial and final states. Three of the four conformations that have the smallest difference in RMSD between the starting and ending states are A^+B^+ states, which indicate that these states are the closest to being symmetric. Although the time scales of our simulations are too short to prove or disprove this hypothesis, this observation qualitatively agrees with a recently published EPR study that suggests that the doubly protonated A^+B^+ state would be more symmetric than the apo state.[72]

Computational Determination of ΔpK_a

As discussed above, the E14 residues form specific, possibly water-mediated interactions with surrounding residues. These interactions change with protonation state, and critically, differ between individual monomers. The asymmetry in interactions should cause a noticeable pK_a shift between the two E14 residues. Recent NMR studies reported the pK_a shift as 1.7 ± 0.2 units in the temperature range we are simulating, with one pK_a at 6.8 and the other at 8.5.[17] Mechanistically, these pK_a values suggest that one E14 residue is often protonated, but leaves open the question of which residue is always protonated and which is only selectively protonated. This is a very difficult question to answer experimentally, as the symmetry of EmrE makes assignment of pK_a to a specific E14 very difficult. Instead, we leverage our atomic model to determine the order of protonation (and thereby assign pK_as) in two ways: 1) the interactions observed, and 2) a RETI calculation of proton transfer between B^+ and A^+.

The structural argument for shifted E14 pK_a values is that the hydrogen bonding patterns of E14 to other residues are largely equivalent between monomers A and B, except with respect to the interaction with Y60 from the opposite monomer (Fig. 10). Since Y60^B could in principle donate a proton to E14^A, but Y60^A cannot donate a proton to E14^B, E14^A should be more willing to lose a proton as it already partially shares the proton from Y60^B. Therefore, the free energy change of the proton transfer process from B^+ to A^+ should be positive, and any singly protonated state should place the proton on the B monomer, creating the B^+ state.

We measure this effect explicitly through the alchemical transition of a proton on E14^B transferring to E14^A (Table 4). For the RETI calculation, the ΔG for moving the proton is positive as expected based on the asymmetrical interaction between Y60 and E14, highlighting that monomer B will be protonated to a greater degree than monomer A. In fact, since the measured 1.7 pK_a unit shift[17] is equivalent to a 2.4 kcal·mol^{-1} ΔG, the overall accuracy of the RETI calculations performed is to within 1 kcal·mol^{-1}, approximately the limit of our current classical force fields. Connecting this evidence to the nonequilibrium work measurements, favoring the B^+ state over the A^+ state would raise the barrier to transition, and would delay conformational transition until the second proton binds, and would be a mechanistic way of slowing the transition of the singly protonated state.

Conclusions

After substantial refinement based on the electron density maps provided by previous cryo-EM experiments,[3] we now have a stable dimeric structure of EmrE that is sequence complete and includes side chains of all residues. Based on traditional crystallographic metrics, the quality of the resulting model is in line with a sub-Ångstrom resolution crystal structure. This high quality of the structure is in large part due to the interactive refinement that extended the transmembrane helices to span the full membrane bilayer, and positioned side chains such as proline consistent with protein population averages.

With this new model in hand, we can provide new insights into the regulation of the EmrE transport cycle that had previously been unattainable. Our simulations indicate that transferring the proton from the B^+ state to create A^+ requires approximately enough energy ($1.5\,\mathrm{kcal{\cdot}mol^{-1}}$) to break a single hydrogen bond,[89] definitively indicating that $E14^B$ protonates first, and that $E14^A$ will accept the second proton. This is due to the asymmetry of cross-monomer Y60–E14 interactions, with $Y60^B$ in a position to hydrogen bond with $E14^A$, but not vice-versa. Nonequilibrium work measurements indicate a lower barrier to transition whenever $E14^A$ is protonated, which breaks the $Y60^B$–$E14^A$ interaction. This evidence strongly suggests that Y60 is the "electrostatic lock" that couples protonation to conformational change, indicating why Y60 is so strongly conserved across the larger SMR family. In addition to the electostatic lock, the hydration of the EmrE lumen also responds strongly to changes in the protonation state, effectively spontaneously transitioning into an occluded state without large-scale conformational change when both $E14^A$ and $E14^B$ are protonated.

Future experimental data sets, particularly from NMR spectra and recent EPR data,[72] could be used in conjunction with the current model for further refinement. Future successor models that incorporate these new datasets may shed light on questions that have not been addressed. These include lipid intercalation into the dimer interface, the ingress and egress pathway of the substrate, and under what protonation conditions substrate binding can take place. Another open question is how the behavior of substrate and protonation state change under *in vivo* conditions, which include a substantial electrical potential[90] opposing the egress of charged substrate. Answering these questions will be essential to further exploring how drug export is coupled to protonation in EmrE.

Acknowledgement

J.V.V. and S.B.R. gratefully acknowledge support from the Sandia National Laboratories Campus Executive Program, which is funded by the Laboratory Directed Research and Development (LDRD) Program. This project made extensive use of the high performance computing resources provided by Sandia National Laboratories. The project also received support from the Defense Threat Reduction Agency-Joint Science and Technology Office for Chemical and Biological Defense (IAA number DTRA10027IA-3167, S.B.R.). The work was performed, in part, at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Los Alamos National Laboratory (Contract DE-AC52-06NA25396) and Sandia National Laboratory. Sandia is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. DOE National Nuclear Security Administration under Contract No. DE-AC04-94AL85000. This research is also supported by the National Institutes of Health, through grants P41-GM104601 and U54-GM087519 to E.T. We would also like to acknowledge Katherine Henzler-Wildman, Chao Wu, Reza Dastvan, and Nate Traaseth for helpful conversations that guided the research direction. Additionally, we thank Reza Dastvan and Hassane Mchaourab for providing the spin-label distances in a text format for comparison with our own results.

References

[1] Yerushalmi, H.; Lebendiker, M.; Schuldiner, S. EmrE, an Escherichia coli 12-kDa multidrug transporter, exchanges toxic cations and H+ and is soluble in organic solvents. *J. Biol. Chem.* **1995**, *270*, 6856–63.

[2] Morrison, E. A.; Henzler-Wildman, K. A. Transported Substrate Determines Exchange Rate in the Multidrug Resistance Transporter EmrE. *J. Biol. Chem.* **2014**, *289*, 6825–6836.

[3] Ubarretxena-Belandia, I.; Baldwin, J. M.; Schuldiner, S.; Tate, C. G. Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J.* **2003**, *22*, 6175–6181.

[4] Butler, P.; Ubarretxena-Belandia, I.; Warne, T.; Tate, C. The Escherichia coli Multidrug Transporter EmrE is a Dimer in the Detergent-solubilised State. *J. Mol. Biol.* **2004**, *340*, 797–808.

[5] Bay, D. C.; Rommens, K. L.; Turner, R. J. Small multidrug resistance proteins: A multidrug transporter family that continues to grow. *Biochim. Biophys. Acta - Biomembr.* **2008**, *1778*, 1814–1838.

[6] Morrison, E. A.; DeKoster, G. T.; Dutta, S.; Vafabakhsh, R.; Clarkson, M. W.; Bahl, A.; Kern, D.; Ha, T.; Henzler-Wildman, K. A. Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature* **2012**, *481*, 45–50.

[7] Gayen, A.; Banigan, J. R.; Traaseth, N. J. Ligand-Induced Conformational Changes of the Multidrug Resistance Transporter EmrE Probed by Oriented Solid-State NMR Spectroscopy. *Angew. Chemie Int. Ed.* **2013**, *52*, 10321–10324.

[8] Chen, Y.-J.; Pornillos, O.; Lieu, S.; Ma, C.; Chen, A. P.; Chang, G. X-ray structure of EmrE supports dual topology model. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18999–9004.

[9] Moradi, M.; Enkavi, G.; Tajkhorshid, E. Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate transporter. *Nat. Commun.* **2015**, *6*, 8393.

[10] Cho, M.-K.; Gayen, A.; Banigan, J. R.; Leninger, M.; Traaseth, N. J. Intrinsic Conformational Plasticity of Native EmrE Provides a Pathway for Multidrug Resistance. *J. Am. Chem. Soc.* **2014**, *136*, 8072–8080.

[11] Gayen, A.; Leninger, M.; Traaseth, N. J. Protonation of a glutamate residue modulates the dynamics of the drug transporter EmrE. *Nat. Chem. Biol.* **2016**, *12*, 141–145.

[12] Muth, G. W.; Ortoleva-Donnelly, L.; Strobel, S. A. A Single Adenosine with a Neutral pKa in the Ribosomal Peptidyl Transferase Center. *Science* **2000**, *289*, 947–950.

[13] Yerushalmi, H.; Schuldiner, S. An Essential Glutamyl Residue in EmrE, a Multidrug Antiporter from Escherichia coli. *J. Biol. Chem.* **2000**, *275*, 5264–5269.

[14] Trabuco, L. G.; Villa, E.; Mitra, K.; Frank, J.; Schulten, K. Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure* **2008**, *16*, 673–683.

[15] Stone, J. E.; Gullingsrud, J.; Grayson, P.; Schulten, K. A System for Interactive Molecular Dynamics Simulation. 2001 ACM Symposium on Interactive 3D Graphics. New York, 2001; pp 191–194.

[16] Woods, C. J.; Essex, J. W.; King, M. A. The Development of Replica-Exchange-Based Free-Energy Methods. *J. Phys. Chem. B* **2003**, *107*, 13703–13710.

[17] Morrison, E. A.; Robinson, A. E.; Liu, Y.; Henzler-Wildman, K. A. Asymmetric protonation of EmrE. *J. Gen. Physiol.* **2015**, *146*, 445–461.

[18] Fleishman, S. J.; Harrington, S. E.; Enosh, A.; Halperin, D.; Tate, C. G.; Ben-Tal, N. Quasi-symmetry in the Cryo-EM Structure of EmrE Provides the Key to Modeling its Transmembrane Domain. *J. Mol. Biol.* **2006**, *364*, 54–67.

[19] Padariya, M.; Kalathiya, U.; Baginski, M. Structural and dynamic changes adopted by EmrE, multidrug transporter protein–Studies by molecular dynamics simulation. *Biochim. Biophys. Acta - Biomembr.* **2015**, *1848*, 2065–2074.

[20] Rotem, D.; Steiner-Mordoch, S.; Schuldiner, S. Identification of Tyrosine Residues Critical for the Function of an Ion-coupled Multidrug Transporter. *J. Biol. Chem.* **2006**, *281*, 18715–18722.

[21] Frishman, D.; Argos, P. Knowledge-based secondary structure assignment. *Proteins* **1995**, *23*, 566–579.

[22] Martí-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sánchez, R.; Melo, F.; Šali, A. Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.

[23] Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M.; Eramian, D.; Shen, M.-y.; Pieper, U.; Sali, A. *Curr. Protoc. Protein Sci.*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007.

[24] Barrera, F. N.; Fendos, J.; Engelman, D. M. Membrane physical properties influence transmembrane helix formation. *Proc. Natl. Acad. Sci. USA* **2015**, *109*, 14422–14427.

[25] Popot, J.-L.; Engelman, D. M. Helical Membrane Protein Folding, Stability, and Evolution. *Annu. Rev. Biochem.* **2000**, *69*, 881–922.

[26] Shelar, A.; Bansal, M. Sequence and conformational preferences at termini of α-helices in membrane proteins: Role of the helix environment. *Proteins Struct. Funct. Bioinforma.* **2014**, *82*, 3420–3436.

[27] Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

[28] Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes With a Modified Generalized Born Model. *Proteins: Struct., Func., Bioinf.* **2004**, *55*, 383–394.

[29] Tanner, D. E.; Phillips, J. C.; Schulten, K. GPU/CPU Algorithm for Generalized Born / Solvent-Accessible Surface Area Implicit Solvent Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2521–2530.

[30] Trabuco, L. G.; Villa, E.; Schreiner, E.; Harrison, C. B.; Schulten, K. Molecular Dynamics Flexible Fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **2009**, *49*, 174–180.

[31] Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

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

[33] Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2007**, *2*, e880.

[34] Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comp. Chem.* **2008**, *29*, 1859–1865.

[35] Dutta, S.; Morrison, E. A.; Henzler-Wildman, K. A. Blocking Dynamics of the SMR Transporter EmrE Impairs Efflux Activity. *Biophys. J.* **2014**, *107*, 613–620.

[36] Oursel, D.; Loutelier-Bourhis, C.; Orange, N.; Chevalier, S.; Norris, V.; Lange, C. M. Lipid composition of membranes ofEscherichia coli by liquid chromatography/tandem mass spectrometry using negative electrospray ionization. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1721–1728.

[37] Jorgensen, W.; Chandrasekhar, J.; Maudura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

[38] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comp. Chem.* **2005**, *26*, 1781–1802.

[39] Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

[40] Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell Jr., A. D.; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.

[41] Kubo, R.; Toda, M.; Hashitsume, N. *Statistical Physics II: Nonequilibrium statistical mechanics*; Springer: New York, 1991.

[42] Brünger, A. T.; Brooks III, C. L.; Karplus, M. Stochastic Boundary Conditions for Molecular Dynamics Simulations of ST2 Water. *Chem. Phys. Lett.* **1984**, *105*, 495–498.

[43] Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177.

[44] Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

[45] Fiorin, G.; Klein, M. L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013**, *111*, 3345–3362.

[46] Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, Jr., A. D. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comp. Chem.* **2010**, *31*, 671–690.

[47] Vanommeslaeghe, K.; MacKerell, Jr., A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.

[48] Vanommeslaeghe, K.; Raman, E. P.; MacKerell, Jr., A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.

[49] Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid parameterization of small molecules using the Force Field Toolkit. *J. Comp. Chem.* **2013**, *34*, 2757–2770.

[50] Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

[51] Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

[52] Darden, T.; York, D.; Pedersen, L. G. Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

[53] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

[54] Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Molecules. *J. Comp. Chem.* **1992**, *13*, 952–962.

[55] Vermaas, J. V.; Hardy, D. J.; Stone, J. E.; Tajkhorshid, E.; Kohlmeyer, A. TopoGromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD. *J. Chem. Inf. Model.* **2016**, *56*, 1112–1116.

[56] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

[57] Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

[58] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comp. Chem.* **2005**, *26*, 1701–1718.

[59] Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

[60] Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

[61] Nose, S. Constant-temperature molecular dynamics. *J. Phys. Condens. Matter* **1990**, *2*, SA115–SA119.

[62] Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.

[63] Axelsen, P. H.; Li, D. Improved convergence in dual-topology free energy calculations through use of harmonic restraints. *J. Comput. Chem.* **1998**, *19*, 1278–1283.

[64] Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.

[65] Van Der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30.

[66] Hagberg, A.; Swart, P.; Chult, D. Exploring network structure, dynamics, and function using NetworkX. Proc. 7th Python Sci. Conf. Pasedena, CA USA, 2008; pp 11–15.

[67] Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. in Sci. and Eng.* **2007**, *9*, 90–95.

[68] Vermaas, J. V.; Tajkhorshid, E. Conformational heterogeneity of α-synuclein in membrane. *Biochim. Biophys. Acta Biomembr.* **2014**, *1838*, 3107–3117.

[69] Vermaas, J. V.; Petridis, L.; Qi, X.; Schulz, R.; Lindner, B.; Smith, J. C. Mechanism of lignin inhibition of enzymatic biomass deconstruction. *Biotechnol. Biofuels* **2015**, *8*, 217.

[70] Chen, V. B.; Arendall III, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst. D* **2010**, *66*, 12–21.

[71] Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **1993**, *26*, 283–291.

[72] Dastvan, R.; Fischer, A. W.; Mishra, S.; Meiler, J.; Mchaourab, H. S. Protonation-dependent conformational dynamics of the multidrug transporter EmrE. *Proc. Natl. Acad. Sci.* **2016**, *113*, 1220–1225.

[73] Law, R. J.; Capener, C.; Baaden, M.; Bond, P. J.; Campbell, J.; Patargias, G.; Arinaminpathy, Y.; Sansom, M. S. P. Membrane protein structure quality in molecular dynamics simulation. *J. Mol. Graph. Model.* **2005**, *24*, 157–165.

[74] Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.

[75] Beitz, E. TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e. *Bioinformatics* **2000**, *16*, 135–139.

[76] Li, J.; Shaikh, S. A.; Enkavi, G.; Wen, P.-C.; Huang, Z.; Tajkhorshid, E. Transient formation of water-conducting states in membrane transporters. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7696–7701.

[77] Hassanali, A.; Giberti, F.; Cuny, J.; Kühne, T. D.; Parrinello, M. Proton transfer through the water gossamer. *Proc. Natl. Acad. Sci.* **2013**, *110*, 13723–13728.

[78] Brill, S.; Falk, O. S.; Schuldiner, S. Transforming a drug/H+ antiporter into a polyamine importer by a single mutation. *Proc. Natl. Acad. Sci.* **2012**, *109*, 16894–16899.

[79] Wang, J.; Rath, A.; Deber, C. M. Functional response of the small multidrug resistance protein EmrE to mutations in transmembrane helix 2. *FEBS Lett.* **2014**, *588*, 3720–3725.

[80] Brill, S.; Sade-Falk, O.; Elbaz-Alon, Y.; Schuldiner, S. Specificity Determinants in Small Multidrug Transporters. *J. Mol. Biol.* **2015**, *427*, 468–477.

[81] Yerushalmi, H.; Lebendiker, M.; Schuldiner, S. Negative Dominance Studies Demonstrate the Oligomeric Structure of EmrE, a Multidrug Antiporter from Escherichia coli. *J. Biol. Chem.* **1996**, *271*, 31044–31048.

[82] Li, S.; Hong, M. Protonation, Tautomerization, and Rotameric Structure of Histidine: A Comprehensive Study by Magic-Angle-Spinning Solid-State NMR. *J. Am. Chem. Soc.* **2011**, *133*, 1534–1544.

[83] Jiao, D.; Rempe, S. B. Combined Density Functional Theory (DFT) and Continuum Calculations of pK_a in Carbonic Anhydrase. *Biochemistry* **2012**, *51*, 5979–5989.

[84] Flewelling, R.; Hubbell, W. Hydrophobic ion interactions with membranes. Thermodynamic analysis of tetraphenylphosphonium binding to vesicles. *Biophys. J.* **1986**, *49*, 531–540.

[85] Boiteux, C.; Vorobyov, I.; French, R. J.; French, C.; Yarov-Yarovoy, V.; Allen, T. W. Local anesthetic and antiepileptic drug access and binding to a bacterial voltage-gated sodium channel. *Proc. Natl. Acad. Sci.* **2014**, *111*, 13057–13062.

[86] Hurst, D. P.; Grossfield, A.; Lynch, D. L.; Feller, S.; Romo, T. D.; Gawrisch, K.; Pitman, M. C.; Reggio, P. H. A lipid pathway for ligand binding is necessary for a cannabinoid G protein-coupled receptor. *J. Biol. Chem.* **2010**, *285*, 17954–17964.

[87] Mayne, C. G.; Arcario, M. J.; Mahinthichaichan, P.; Baylon, J. L.; Vermaas, J. V.; Navidpour, L.; Wen, P.-C.; Thangapandian, S.; Tajkhorshid, E. The cellular membrane as a mediator for small molecule interaction with membrane proteins. *Biochim. Biophys. Acta Biomembr.* **2016**, *1858*, 2290–2304, Biosimulations of lipid membranes coupled to experiments.

[88] Moradi, M.; Tajkhorshid, E. Computational recipe for efficient description of large-scale conformational changes in biomolecular systems. *J. Chem. Theory Comput.* **2014**, *10*, 2866–2880.

[89] Myers, J.; Pace, C. Hydrogen bonding stabilizes globular proteins. *Biophys. J.* **1996**, *71*, 2033–2039.

[90] Bot, C. T.; Prodan, C. Quantifying the membrane potential during E. coli growth stages. *Biophys. Chem.* **2010**, *146*, 133–137.

Appendix

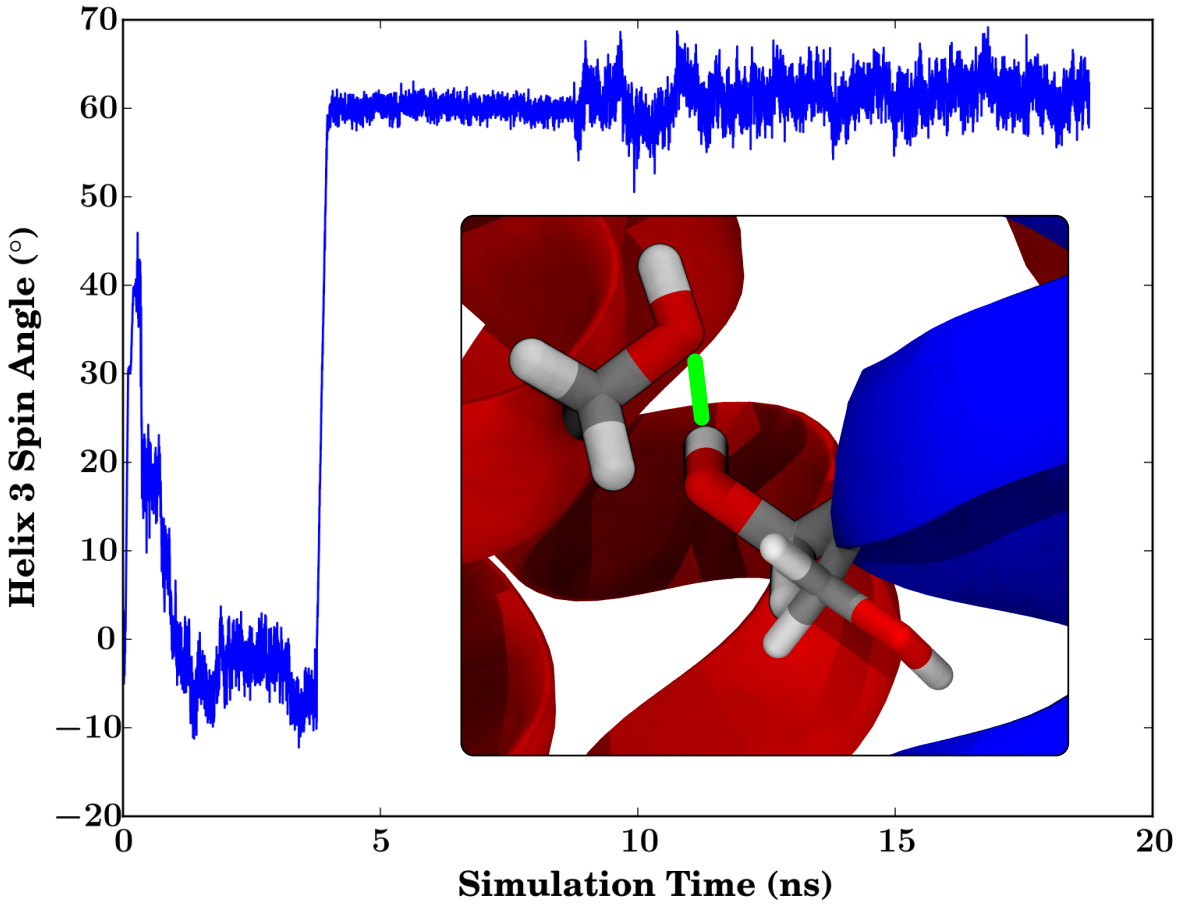

Figure A1: Time series showing the rotation of the A monomer of helix 3 until S64 no longer faces the membrane. This transition was attempted twice using the Colvars module of NAMD. The initial 30 degree rotation was unstable. The 60 degree rotation was restrained for a time, and was stable over longer simulation (at approximately 9 ns). The comparison between the initial and final state is shown in the inset. Initially, S64 of the A monomer (thinner stick representation) pointed into the membrane. After rotation, S64 can interact with the S64 of monomer B (green line).

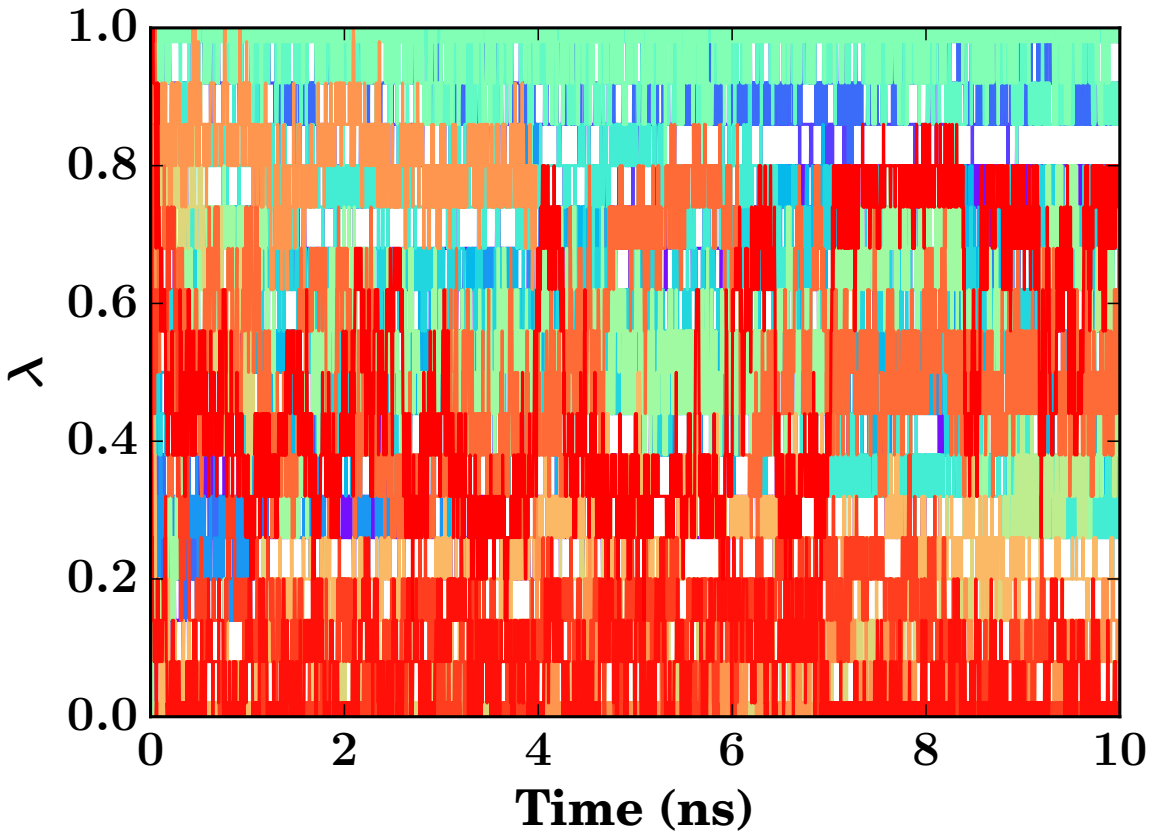

Figure A2: Time series showing how each replica exchanges through values of the alchemical transition parameter (λ) space over time. Each replica is drawn as its own color, based on the initial value for λ of the replica (redder for larger λ, bluer for smaller λ). Exchanges are frequent on the time scale of the simulation, and no isolated λ values are observed.

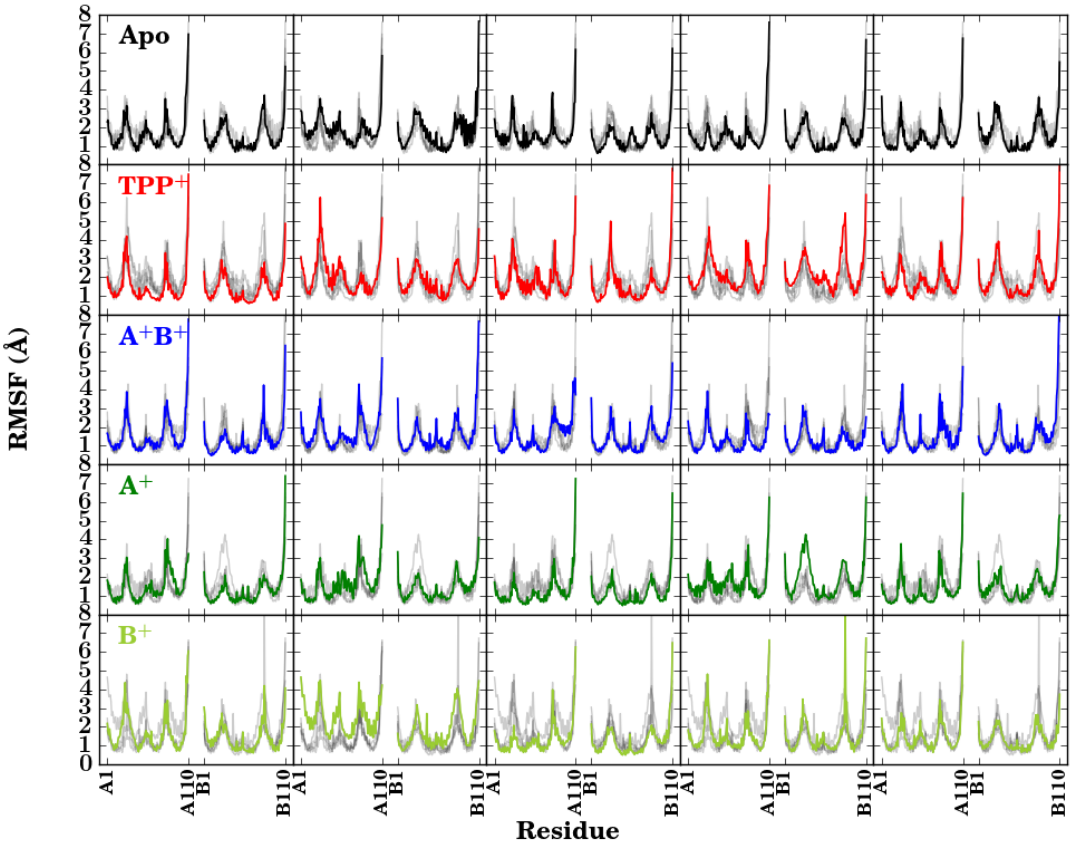

Figure A3: RMSF per residue for every equilibrium trajectory run. Each different EmrE loading is one row of the matrix, with the 5 separate trajectories occupying the columns of the matrix. The colored lines indicate the RMSF for that specific trajectory, while the gray background lines provide context for how the other copies of the same loading state behaved.

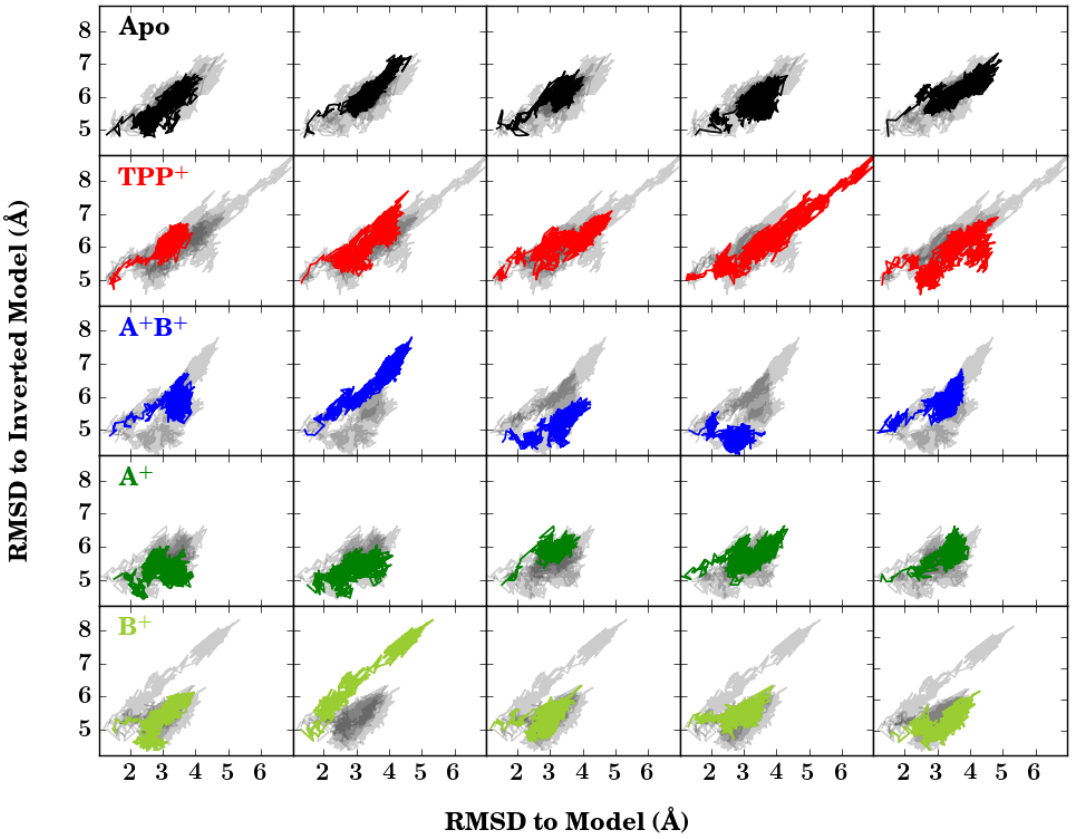

Figure A4: RMSD matrix highlighting the states sampled by the equilibrium trajectories with respect to both the original model (x-axis) and the inverted model (y-axis). Motion towards the inverted model (low RMSD states) would be suggestive of the transition EmrE undergoes as part of its functional cycle.

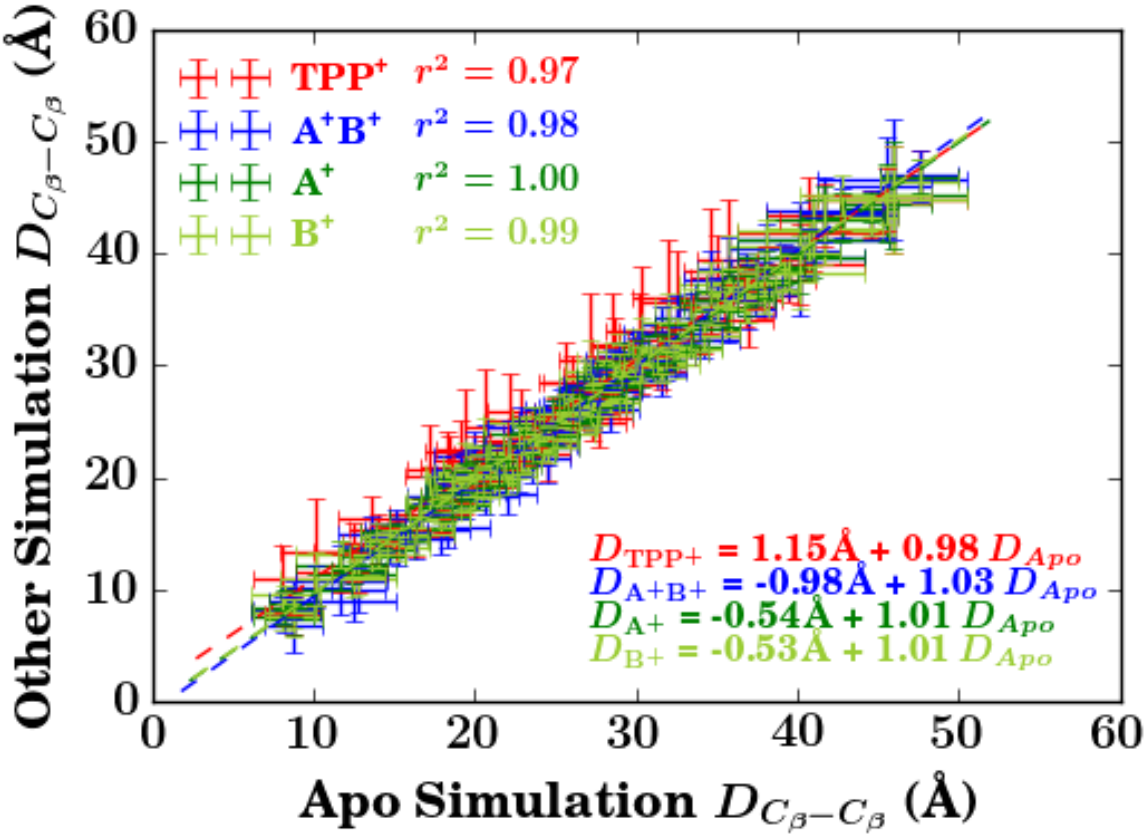

Figure A5: Comparison of the distances observed between C_β-C_β from residue pairs on opposite monomers within simulation. This figure is equivalent to Fig. 6A, but compares the distances from different loading conditions against one another.

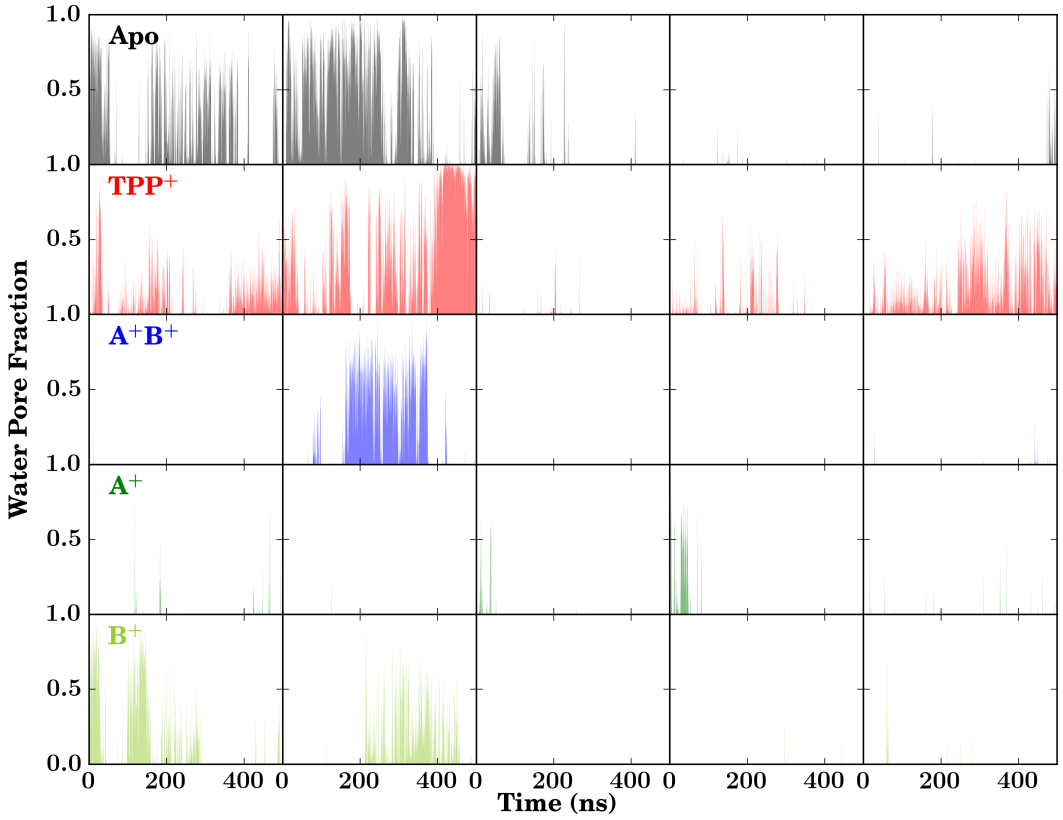

Figure A6: Time resolved water pore formation for each simulation trajectory. Since the time resolution of the data (5 ps steps) far exceeds the limits of what is easily resolvable graphically, the individual datapoints are aggregated together in lots of 50, such that each feature is equivalent to 250 ps of simulation time. Each remaining point thus represents the fraction of time within that simulation snapshot that the water pore, as defined in Table 1 of the main text, was formed during the 250 ps of simulation.

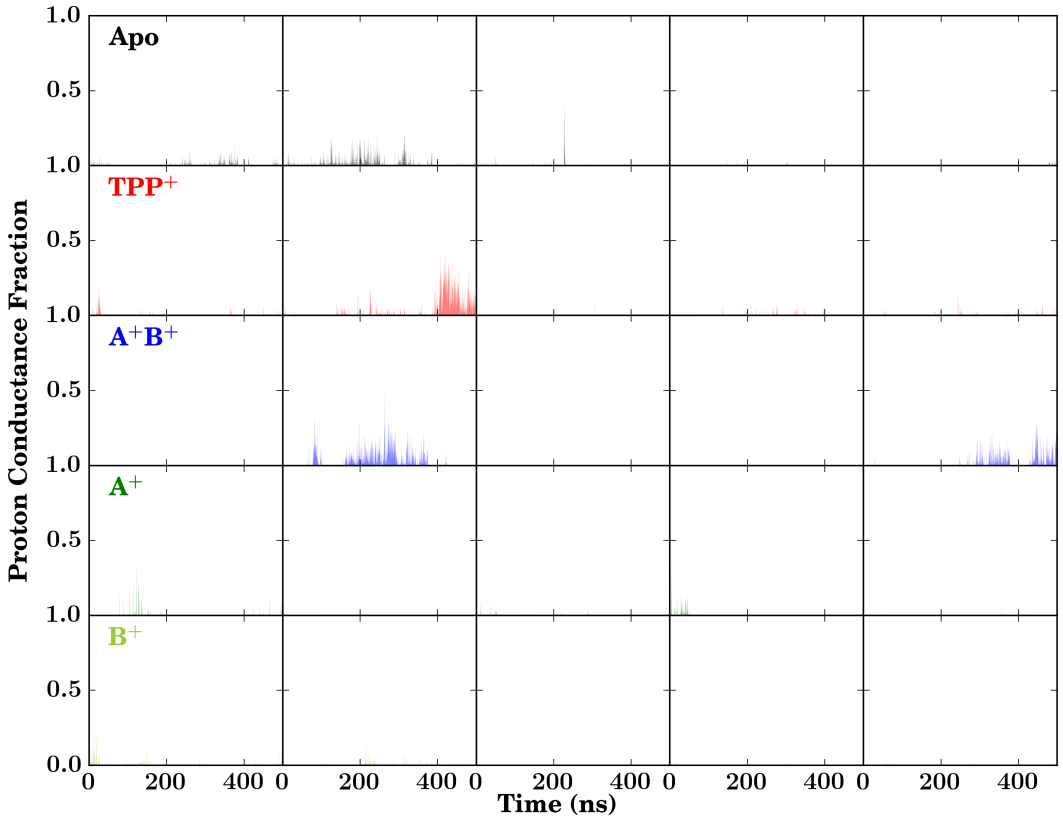

Figure A7: Time resolved existence of water pores that can conduct protons for each simulation trajectory. Since the time resolution of the data (5 ps steps) far exceeds the limits of what is easily resolvable graphically, the individual datapoints are aggregated together in lots of 50, such that each feature is equivalent to 250 ps of simulation time. Each remaining point thus represents the fraction of time within that simulation snapshot that the water pore, as defined in Table 1 of the main text, was formed and could conduct protons during the 250 ps of simulation.

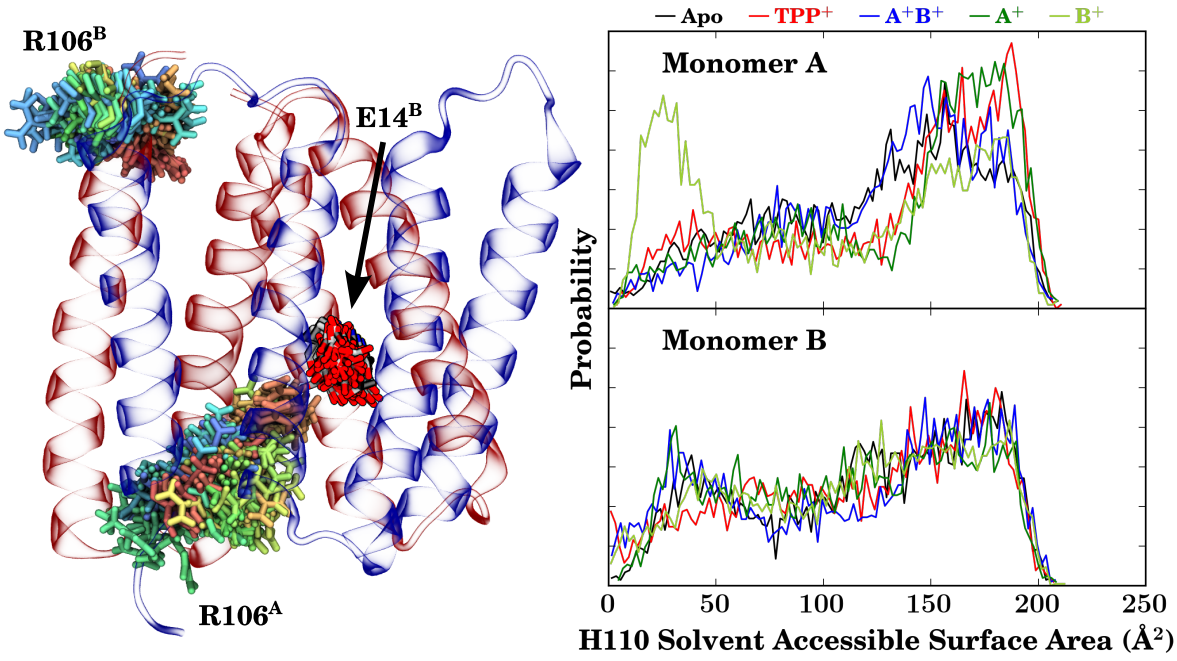

Figure A8: Demonstration of the different motions of the C-terminus. (Left) A comparison of the motion of R106 from each monomer relative to the lumen, represented here by $E14^{B}$. The position of R106 across an apo trajectory is shown through a rainbow of residues superimposed on the same dimeric structure (transparent cartoon). A rotation of this view around the membrane normal is shown in Supplmentary animation 6. (Right) Probability distribution of the solvent-accessible surface area of the terminal H110 residue.

Table A1: Hydrogen bond propensity for observed donor-acceptor pairs, normalized such that 100% equates to a single hydrogen bond being formed between two residues for the entire simulation. The hydrogen bonds reported here are only for cases where an amino acid sidechain would be the donor, thereby excluding the helix-stabilizing backbone hydrogen bonds.

				Sidechain Hydrogen Bond Propensity				
Acceptor		Donor		Apo	TPP-bound	$H_A^+H_B^+$	H_A^+	H_B^+
A	Y6	A	N2	7.5	7.4	9.2	5.6	6.6
	A10		S43	81.8	65.6	84.4	73.9	81.5
	E14		T18	63.2	65.8	75.2	60.4	63.5
	E14		Y40	14.5	10.9	0.7	1.5	7.5
	E14		S43	5.4	10.7	–	0.4	1.5
	E14		W63	32.8	13.8	3.1	5.1	57.1
	V15		T19	90.4	88.9	85.9	89.8	87.3
	G17		T36	1.5	9.0	7.0	3.1	3.5
	T18		Y40	5.2	13.1	33.7	19.5	16.9
	L20		S24	14.4	4.0	43.6	28.2	38.8
	S24		R29	0.5	1.6	11.2	7.0	13.1
	E25		K22	3.7	–	29.9	20.3	19.8
	E25		R29	7.6	0.2	9.0	5.1	9.8
	G26		R29	2.9	4.1	24.9	20.6	14.0
	T28		S24	2.0	2.7	0.1	5.5	1.3
	R29		S33	74.5	56.4	59.8	74.2	67.9
	P32		T36	87.4	78.4	82.4	88.5	86.7
	G35		C39	63.9	62.5	60.8	61.8	65.1
	I37		C41	56.9	59.6	55.5	54.9	53.0
	C39		S43	1.3	1.0	5.6	7.7	1.8
	Y40		T18	3.7	2.0	3.4	5.5	3.4
	S43		W63	–	2.1	31.2	28.6	–
	W45		Q49	33.9	30.5	37.2	36.5	31.2
	L46		T50	86.3	88.0	71.0	78.7	86.1
	L47		T50	3.8	0.5	16.4	11.6	1.1
	G57		N102	–	0.3	10.4	–	–
	Y60		S64	68.8	51.5	74.3	60.9	60.9
	Y60		N102	9.3	5.2	0.2	12.0	15.2
	W63		E14	–	–	35.5	35.9	–
	S64		N102	1.9	4.2	8.7	14.4	7.4
	I68		S72	83.8	86.2	87.8	90.0	90.5
	I71		S75	63.5	64.9	63.1	67.1	67.6
	S75		Q81	13.6	2.6	1.8	3.6	2.4
	W76		Q81	6.9	9.1	1.7	0.3	0.8
	W76		R82	–	5.4	2.8	5.1	–
	F78		Q81	10.9	10.2	1.5	2.5	1.6
	F79		Q81	3.2	8.9	4.2	5.4	4.1
	G80		S75	5.7	3.6	0.4	0.1	0.1
	Q81		R82	5.6	5.0	2.0	0.4	0.7
	D84		R82	11.4	21.6	21.5	28.7	18.4
	M91		C95	33.5	50.6	42.6	40.8	28.9
	V98		N102	5.0	2.7	13.3	3.0	2.8
	I101		S105	58.4	58.0	53.7	63.4	61.1
	N102		R106	28.5	26.6	10.6	34.4	17.5
	N102		T108	–	–	9.8	–	–
	L103		S107	38.9	34.2	27.0	27.4	38.3
	L104		S107	2.1	4.0	4.8	7.8	2.1
	L104		T108	15.1	19.0	11.1	6.9	15.2
	R106		T56	1.4	2.2	13.5	6.7	3.7
	T108		H110	8.5	9.0	9.2	9.6	11.1

	H110		T56	–	–	11.3	–	–
	H110		R106	–	–	29.2	–	–
A	E14	B	Y60	8.6	0.1	10.3	3.5	32.2
	T18		Y60	1.3	–	7.8	9.2	2.9
	Y60		Y40	–	–	4.5	3.3	7.0
	S64		S64	0.1	–	8.8	0.3	0.6
	S75		T56	6.5	11.8	3.2	4.3	1.9
	Q81		T56	–	–	0.1	9.1	0.3
	D84		T108	4.5	10.2	13.3	12.1	13.8
	N102		S72	7.8	–	0.1	–	3.1
	H110		K22	–	–	15.4	10.7	–
	H110		R82	32.8	32.5	43.1	65.4	32.3
B	E14	A	Y60	–	11.7	15.9	0.1	0.1
	E14		R106	6.1	27.2	–	0.2	0.3
	A48		K22	–	–	0.3	8.0	0.1
	Q49		K22	–	3.4	–	3.1	5.4
	L51		K22	7.4	2.3	4.3	6.3	10.9
	T56		K22	3.4	1.1	6.2	3.8	2.8
	Y60		Y40	–	–	5.2	0.4	0.6
	Y60		S64	0.6	7.4	0.8	1.7	0.4
	S64		N102	10.3	2.8	2.5	0.1	0.5
	S72		N102	4.8	1.1	0.2	3.8	11.9
	S72		S105	0.3	0.4	1.6	10.6	–
	D84		R106	–	–	19.9	5.3	–
	D84		T108	7.1	–	1.8	0.5	0.8
	H110		R82	93.3	38.7	65.7	54.5	68.3
B	Y4	B	N102	6.9	10.7	9.3	8.3	8.6
	Y6		N2	6.1	5.8	4.7	3.5	6.2
	A10		S43	59.6	65.5	87.0	74.0	65.0
	E14		T18	57.8	58.7	71.4	60.1	54.2
	E14		Y40	6.8	9.2	0.2	9.4	1.5
	E14		S43	22.1	16.2	–	9.0	1.4
	E14		W63	15.9	8.7	2.0	44.2	0.4
	V15		T19	87.2	83.5	87.8	89.3	81.9
	G17		T36	5.1	13.4	1.8	3.4	6.4
	T18		Y40	6.1	13.1	32.5	6.7	4.2
	L20		S24	17.2	17.8	35.9	21.4	8.1
	M21		S24	15.1	22.0	13.0	10.6	20.6
	S24		T28	7.6	1.6	1.4	4.2	3.1
	E25		T28	26.7	44.6	25.9	17.1	32.2
	E25		R29	91.0	74.7	38.0	34.0	93.3
	G26		S24	4.1	–	2.2	19.2	15.5
	R29		S33	82.1	77.3	68.8	75.6	83.0
	P32		T36	82.0	72.3	84.9	84.2	84.9
	S33		S24	2.2	10.2	1.5	1.3	7.6
	G35		C39	62.5	58.9	61.6	62.3	61.2
	I37		C41	61.4	54.4	50.1	62.4	51.9
	C39		S43	1.1	2.0	1.8	3.0	8.7
	Y40		T18	2.3	2.0	3.8	4.2	5.6
	S43		W63	5.6	0.9	61.7	3.9	41.8
	W45		Q49	24.0	22.7	33.3	31.5	31.4
	L46		T50	93.3	92.5	66.3	91.4	83.3
	L47		T50	0.1	0.3	23.7	1.7	9.2
	Y60		S64	58.4	68.5	57.9	57.6	49.2
	W63		E14	–	–	69.7	–	50.6
	I68		S72	78.3	87.8	85.3	87.1	84.3
	I71		S75	74.7	71.6	70.2	77.4	73.2

W76	Q81	2.6	5.1	5.2	0.3	1.0
F79	R82	12.4	8.4	6.9	5.6	10.2
G80	R82	3.2	2.4	1.2	0.8	11.8
Q81	W76	11.6	5.9	8.0	12.2	20.3
Q81	R82	7.2	6.7	8.5	3.0	3.2
D84	R82	58.9	68.5	100.1	119.9	25.5
M91	C95	57.2	56.6	60.4	53.0	58.5
V98	N102	8.1	8.1	7.6	4.7	6.7
I101	S105	46.8	29.1	44.4	45.4	55.5
N102	Y4	9.9	3.8	6.3	6.4	4.5
N102	R106	17.1	13.7	17.6	7.2	18.8
L103	S107	40.9	36.9	40.9	43.0	39.3
L104	S107	9.6	2.1	2.0	2.5	2.7
L104	T108	26.6	32.7	23.0	24.7	23.8
S105	Y4	1.3	5.4	2.5	1.7	2.1
T108	H110	7.3	8.6	6.2	7.8	7.1
P109	H110	3.6	2.7	4.6	5.2	3.3
H110	R106	–	16.7	–	–	–

Figure A9: Sequence alignment for diverse members of the small multi-drug resistance (SMR) transporter family, highlighting regions of sequence conservation above 75%. Seqences taken from reviewed sequences within the Uniprot sequence database.