Proton Coupling and the Multiscale Kinetic Mechanism of a Peptide Transporter

Chenghan Li,[1] Zhi Yue,[1] Simon Newstead,[2*] Gregory A. Voth[1*]

[1]Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, University of Chicago, Chicago, United States of America

[2]Department of Biochemistry, University of Oxford, Oxford, United Kingdom

* Corresponding authors: simon.newstead@bioch.ox.ac.uk, gavoth@uchicago.edu

ABSTRACT

The proton electrochemical gradient drives substrate transport across the cell membrane via a diverse set of secondary active transporters. Proton coupled peptide transporters (POTs) are important for peptide transport in prokaryotes and eukaryotic cells, where they mediate the uptake of di- and tri-peptides in addition to drug and pro-drug molecules. Previously, we captured a POT transporter from *Staphylococcus hominis*, $PepT_{Sh}$, in a cytoplasm-facing, inward open state (Minhas et al., 2018). Biochemical experiments have further revealed several critical residues for proton coupled transport; however, the precise role played by these residues in coupling proton binding to conformational changes as well as the timescales for proton transfers have remained obscure. Here, we employed multiscale modeling, including classical molecular dynamics, reactive molecular dynamics, and enhanced free energy sampling to characterize proton coupling within this transporter. We show directly that proton binding to a glutamate on TM7 opens the extracellular gate. The inward proton flow is found to induce movement of the peptide towards the cytosol by varying the protonation state of a second conserved glutamate on TM10. We also show that proton movement between TM7 and TM10 is thermodynamically driven and kinetically permissible, revealing a mechanism for proton movement inside the transporter.

INTRODUCTION

Proton coupled oligopeptide transporters (POTs) utilize a membrane pH gradient to drive cellular uptake of di-/tri-peptides and their analogs with homologs found in all bacterial and eukaryotic genomes (Daniel & Kottra, 2004; Smith et al., 2013). Mammalian cells contain two POT family transporters, PepT1 (SLC15A1) and PepT2 (SLC15A2), which are responsible for the bulk uptake and retention of dietary peptides in the small intestine and kidneys respectively (Daniel & Kottra, 2004). PepT1 and PepT2 can also transport prodrug molecules and are increasingly recognized as important targets for rational drug design to improve drug pharmacokinetics (Giacomini et al., 2010; Lin et al., 2015; Rubio-Aliaga & Daniel, 2002). The POT proteins use an alternating access mechanism, where peptide transport is realized through conformational switching between two major conformations, termed inward-facing (IF) and outward-facing (OF) states (Solcan et al., 2012). An occluded (OC) state has also been observed (Newstead et al., 2011; Quistgaard et al., 2016), although this is likely to be transitory. In OF and IF conformations the ligand-binding site is accessible either from the extracellular or the intracellular environment, while access is prohibited from both sides in the OC state (Drew & Boudker, 2016). The alternating access cycle in POTs has been determined following structural and biochemical studies on bacterial homologs (Newstead, 2017) and is rationalized on the basis of alternating formation and breaking of conserved salt bridge interactions, which drive the structural changes following peptide and proton binding (Newstead, 2017; Parker et al., 2017).

Several conserved side chains have been identified that play key roles in the transport mechanism, in particular it was found that the salt bridge between a highly-conserved glutamate/aspartate on transmembrane helix (TM) 7 and the arginine/lysine on TM1, stabilizes the closed state of the extracellular gate (Huang et al., 2003; Newstead, 2015; Solcan et al., 2012). Protonation of the glutamate/aspartate was predicted to break the salt bridge and trigger the IF-to-OF conformational change, allowing ligand access from the extracellular side (Parker et al., 2017; Solcan et al., 2012), but direct evidence for this crucial part of the transport mechanism is still lacking. Molecular dynamics (MD) simulations have been used to investigate the protein and ligand dynamics and coupling in POTs (Aduri et al., 2019; Batista et al., 2019; Doki et al., 2013; Immadisetty et al., 2017; Parker et al., 2017; Selvam et al., 2018). However, many of these MD based studies contain contradictory findings, ascribing different functions for these side chains in different POT family

homologs, and highlighting the challenges of sufficient sampling of the extended phase space spanned by protein, ligand, protons and their associated hydration. It was reported in an MD study on GkPOT from *Geobacillus kaustophilus* for example, that the charge state of the TM7 Glu has no effect on the conformational state of the transporter (Immadisetty et al., 2017). A later study that employed enhanced conformational sampling found the OF conformation of $PepT_{So}$ from *Shewanella oneidensis* to be more stable than the IF one when the TM7 Glu is deprotonated (Selvam et al., 2018), suggesting the TM1-TM7 salt bridge cannot determine the extracellular gate closure. A more recent computational study systematically evaluated the impact of the TM7 Glu and ligand on the conformational free energy landscape of $PepT_{St}$ from *Streptococcus thermophilus*, and found the IF and OC states were the most stable, independent of the TM7 Glu and ligand (Batista et al., 2019), inconsistent with the expectation for a stable apo OF state to allow efficient access to ligands in the periplasmic space. All of these studies were conducted in a fixed charge and non-reactive (fixed chemical bonding topology) manner that limited their ability to fully sample the complete transport cycle and especially the proton coupling mechanism.

Previously, we discovered a homolog of PepT1 from the bacterium *Staphylococcus hominis*, termed as $PepT_{Sh}$, transports a peptide-like thioalchohol precursor, cysteinylglycine-3-methyl-3-sulfanylhexan-1-ol (S-Cys-Gly-3M3SH) (Minhas et al., 2018). The compound is secreted by the apocrine gland, mostly in the armpit and groin regions of the body (Collins, 1989). Uptake of the S-Cys-Gly-3M3SH ligand into *S.hominis* results in its biotransformation into an odorous volatile responsible for human body odor (Minhas et al., 2018). $PepT_{Sh}$ shares high sequence similarity with PepT1 and PepT2 and can also transport prodrugs such as valacyclovir and 5-aminolevulinic acid (Minhas et al., 2018; Minhas & Newstead, 2019, 2020). Hence, $PepT_{Sh}$ represents an ideal model for understanding the proton coupling mechanism across the bacterial and human homologues with the POT family. Our previous work captured the IF state of $PepT_{Sh}$ bound with S-Cys-Gly-3M3SH as well as several drugs (Minhas et al., 2018; Minhas & Newstead, 2019), although the OF and OC states were not captured. It therefore remains challenging to complete the functional cycle of proton-coupled ligand transport starting from static IF structures.

In this work, we performed extensive all-atom MD and multiscale reactive MD simulations (MS-RMD; previously termed multistate empirical valence bond or MS-EVB) (Lee et al., 2016; Wu et al., 2008) combined with enhanced free energy sampling to elucidate how proton transport (PT)

through key residues drives conformational changes and ligand translocation. We have demonstrated here that it is indeed the titrations on the TM7 and TM10 glutamates that trigger the global conformational change and ligand release in $PepT_{Sh}$. Extending previous studies, we have also evaluated the energetics associated with proton movements and found the PT free energies are subtly modulated by the transporter conformational state and the bound ligand. Our findings, which are consistent with a previous study looking at extracellular gate dynamics in $PepT_{Xc}$ (Parker et al., 2017), reveal a coupled and cooperative multiscale kinetic mechanism between the transporter conformation, proton, and ligand motions, which represents a major step toward a complete and quantitative description of the full transport cycle.

RESULTS

TM10 Glu is crucial for ligand binding and its titration controls ligand release

Our earlier $PepT_{Sh}$ crystal structure shows the transporter coupled to S-Cys-Gly-3M3SH in the IF state (**Figure 1**A). The hydrophobic tail of the ligand was discovered to be inserted into a pocket created by TM7 and TM10 (**Figure 1**B). Although the E418 residue is not directly coordinated to the ligand in the crystal structure, mutagenesis of the TM10 Glu in $PepT_{Sh}$, as well as in $PepT_{St}$ (Solcan et al., 2012) and human PepT1 (Xu et al., 2009) has suggested the Glu is important for proton coupling and ligand recognition. This prompted us to run two simulations with protonated and deprotonated E418 (TM10 Glu in $PepT_{Sh}$) from the crystal structure. We defined a collective variable (CV) as the minimal distance between TM9 backbone atoms and the ligand heavy atoms to assess the binding of the ligand tail in the pocket (see **Figure 1**B). We found that if E418 is protonated, the simulations could reproduce the crystal ligand binding pose, as demonstrated by the time evolution and distribution of the distance CV (**Figure 1**C). When E418 is deprotonated, on the other hand, the ligand exits the TM7-TM10 binding pocket, yielding a lower binding posture that differs somewhat from the crystal (**Figure 2**A). These results are consistent with the optimal conditions for the ligand bound crystal structure of pH 5.0.

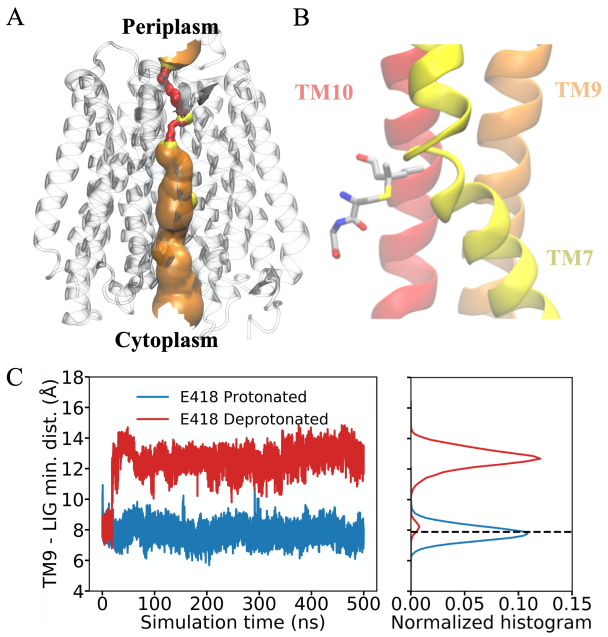


Figure 1. Role of E418 in ligand binding. (A) The inward-facing holo crystal structure captured in previous work (Minhas et al., 2018). The pore radius profile was computed by the HOLE program (Smart et al., 1996). The region that forbids water (pore radius < 1.15 Å) is colored red, the region that allows single water permeation (1.15 Å < pore radius < 2.30 Å) is colored yellow, and orange indicates pore radius > 2.30 Å. This color scheme will be used for the following molecular figures. (B) The position of the ligand and the TM7-TM10 pocket as well as TM9 in the crystal structure. (C) The minimum distance between the ligand and TM9 backbone atoms in simulations with protonated E418 and deprotonated E418. The value in the crystal structure was indicated by the dashed horizontal line.

To further understand the role of E418 in ligand binding, we computed the minimal distance between the carboxyl of E418 and the N-terminus of the ligand. The protonated E418 simulation, which serves as a control, faithfully reproduces the weak interaction pattern between E418 and ligand seen in the crystal structure, as evidenced by the crystal distance value falling within the MD distribution measured. A deprotonated E418, however, interacts with the ligand quickly (~50 ns) by forming a salt bridge with it (**Figure 2**B), causing its hydrophobic tail to leave the TM7-TM10 pocket as previously stated. Interestingly, despite these discrepancies in its binding pose, the bound ligand remains stable in the binding site over the course of the 3.3-μs MD simulation, possibly due to the stability gained from forming the salt bridge with E418 compensating for the loss from leaving the TM7-TM10 pocket. The residue E418 is well-solvated in the IF conformational state, and the water network fully connects the Glu to the cytosolic bulk, so a pH gradient should be able to drive the proton of E418 down to the cytosol. As such, the crystal

structure could represent a prelude to the deprotonated E418 state, in which the ligand transfers into the lower binding location after the proton dissociates from E418.

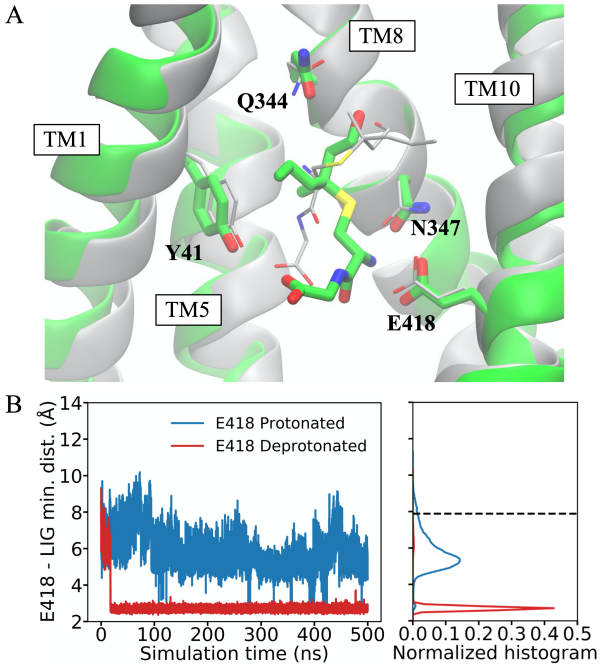


Figure 2. The binding mode of the ligand when E418 is deprotonated. (A) A super-position of the crystal structure (grey) and the equilibrated structure with deprotonated E418. (B) The minimum distance between E418 carboxyl and the ligand N-terminus in simulations with protonated and deprotonated E418. The value in the crystal structure is indicated by the dashed horizontal line.

We next protonated the TM10 Glu to model PT driven by the pH gradient from the periplasmic side to the Glu, and initiated MD trajectories from equilibrated configurations sampled when the Glu is deprotonated. We will quantify and illustrate the feasibility of this hypothetical PT in the following sections, but we now report the consequences of this protonation state change. As shown in **Figure 3**A, protonation of E418 makes the ligand unstable and eventually triggers the ligand release into the cytosol. Notably, the salt bridge between E418 and the ligand breaks almost instantly upon the protonation, but the ligand remains near the binding site for several hundreds of nanoseconds. During this metastable period, we looked at the interacting residues with the ligand (**Figure 3**C) and compared them to the interaction pattern in the stable ligand-bound state when E418 is deprotonated (**Figure 3**B). We found that the protonation indeed weakens the interaction between the ligand and E418, and as a result, the Q344 becomes the dominant residue interacting with the ligand via a hydrogen bond to its sidechain hydroxyl group. The missing hydroxyl in di-/tri-peptides could be one of the reasons why S-Cys-Gly-3M3SH is transported at a slower rate

than these other ligands (Minhas et al., 2018). The lack of contacts with E418 and N347 (**Figure 3**D) resulted in faster ligand release in a replicate than the other two (**Figure 3**A red vs. green and blue), emphasizing the importance of the residues for ligand binding, which is consistent with the mutagenesis results that E418A and N347A decreased or eliminated transport efficiency (Minhas et al., 2018).

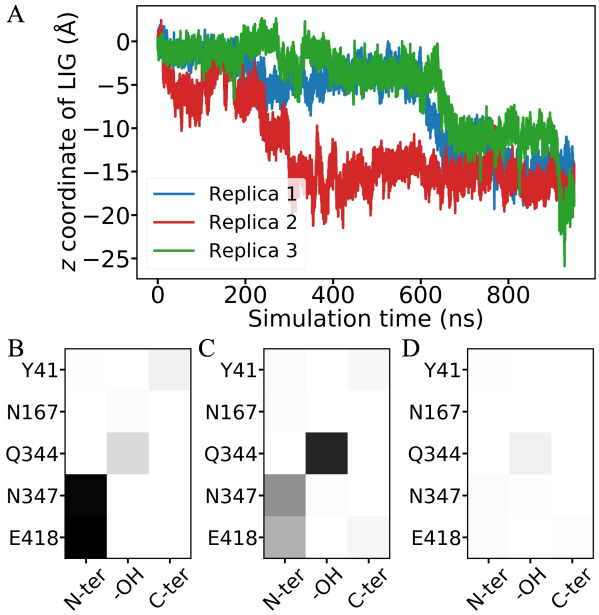


Figure 3. Proton-induced ligand release. (A) The *z* coordinate of ligand geometric center with respect to the middle of the membrane. (B) The contact map between ligand functional groups and binding-site residues (B) in the bound state with deprotonated E418, (C) in the first 200-ns metastable state in replicas 1 & 3, and (D) in the first 20-ns of replica 2 with E418 protonated.

Titration of TM7 Glu triggers conformational change to outward-facing (OF) state

The highly conserved TM7 glutamate/aspartate–TM1 arginine/lysine pair and the TM7 serine–TM2 histidine pair are two of the critical salt bridge and hydrogen bond interactions that maintain the closure of the POT extracellular gate. We have previously shown (Parker et al., 2017) that the extracellular gate in $PepT_{Xc}$, a POT family homologue from *Xanthomonas campestris*, is regulated by the interaction between histidine and serine, and protonation of histidine causes the gate to open by disrupting the histidine-serine hydrogen bond. The histidine and serine residues are conserved in $PepT_{Xc}$ and $PepT_{So}$, a POT from *Shewanella oneidensis*, as well as in mammalian POTs; however, the TM2 histidine is absent from certain bacterial POTs, such as $PepT_{St}$, and $PepT_{Sh}$

studied in this work. Instead, the salt bridge between TM7 Glu/Asp and TM1 Arg/Lys, which can be seen in virtually every IF structure of POTs, might serve as an alternate mechanism for controlling the extracellular gate conformation. We performed MD simulations using deprotonated E311 (TM7 Glu in $PepT_{Sh}$) to stabilize the salt bridge and, as predicted, the salt bridge remained stable in the simulations and the transporter populated the IF state (**Figure 4**A blue histogram). Notably, the system also samples an inward-facing occluded (IF-OC) state featuring a partially closed intracellular gate in our 3.3-μs simulation (**Figure 4**B, 4C, and 4D), and we discovered that this state is metastable since the transporter returns to the IF state after ~ 400 ns in the IF-OC state (**Figure 4**C). The biological relevance of the IF-OC state will be discussed in the following paragraphs.

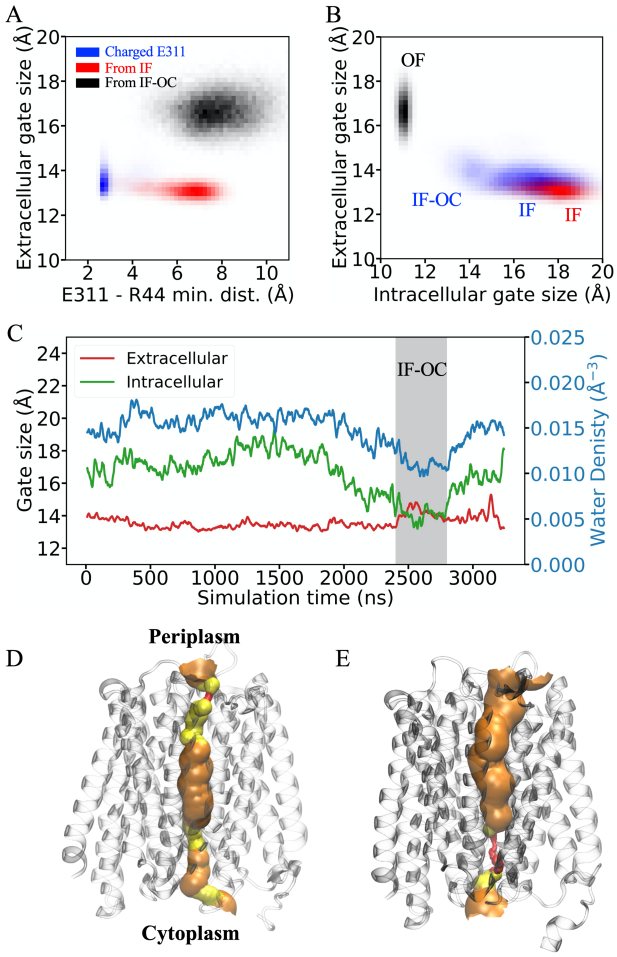


Figure 4. Proton-induced conformational change. (A) Two-dimensional histogram of the minimum distance between E311 and R44 heavy atoms, and the extracellular gate size of the simulation with deprotonated E311 (blue), with protonated E311 initiated from an inward-facing conformation (red), and with protonated E311 initiated from an inward-facing occluded conformation (black). (B) Two-

dimensional histogram of the intracellular and extracellular gate sizes. (C) The gate sizes and the water density around the intracellular gate in the deprotonated E311 simulation. The region corresponding to the inward-facing occluded state is highlighted by grey. A running average with a 20-ns window was performed on the time series. (D) The pore radius profile of the MD-sampled inward-facing occluded state. (E) The pore radius profile of the MD-sampled outward-facing state.

To model PT from the periplasm to E311, we changed the Glu to be protonated starting from the equilibrated IF structure and, as a result, the salt bridge between E311 and R44 (on TM1) became unstable and the two residues separated. Interestingly, in our 500-ns simulations of two independent runs, the transporter remained in the IF state (**Figure 4**A and 4B red histogram; Supplementary Figure 1B), which is indeed consistent with prior MD investigations. This contradicts the idea that protonation of the TM7 Glu opens the extracellular gate, since we found only a weak relationship between the TM7 Glu–TM1 Arg distance and the extracellular gate size, shown by a wide range of Glu-Arg distances correlating to a narrow gate size distribution.

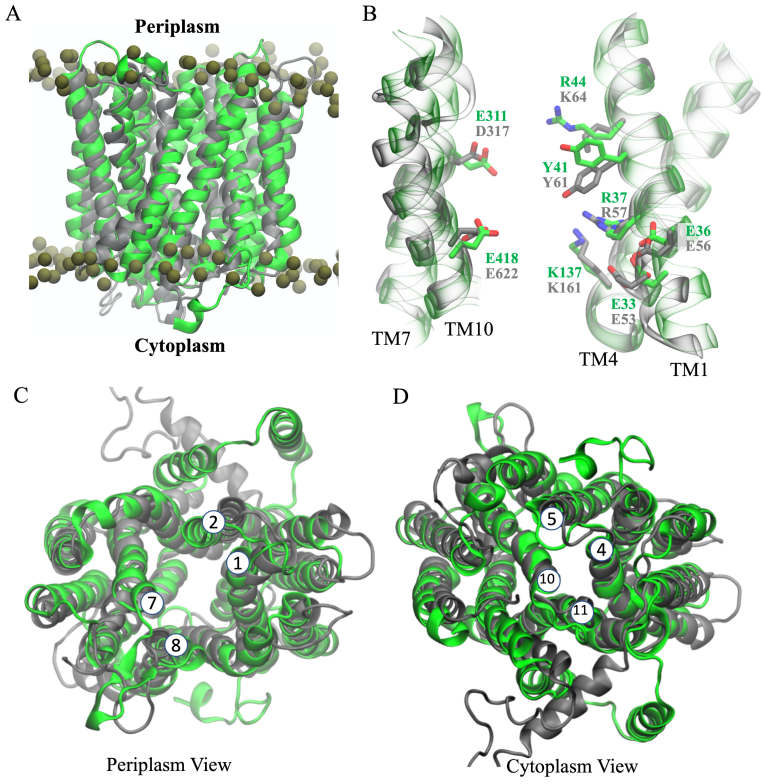


Figure 5. Structural comparison between $PepT_{Sh}$ and PepT2. (A) Superposition of MD-sampled outward-facing $PepT_{Sh}$ (green) with the cryo-EM outward-facing PepT2 (grey). The membrane phosphorus atoms are represented by dark yellow spheres. Comparison for (B) key residue, (C) extracellular gate and (D) intracellular gate between $PepT_{Sh}$ and PepT2. The helices that form the gates are labeled with numbers in (C) and (D).

However, in simulations initiated from the IF-OC state, the transporter undergoes a fast conformational change into the OF state (~ 100 ns) following the protonation of E311, as seen in two separate runs (Supplementary Figure 1A). The slow shutting of the intracellular gate is thought to be the cause for the restriction of a straight transition from IF to OF state. Notably, there is a clear correlation between the intracellular gate size and the water density between helix pair tips TM4,5 and TM10,11, the helices that form the intracellular gate (**Figure 4**C), indicating that fluctuations in hydration may be important for intracellular gate closure. The water collective motion can represent a slow degree of freedom, and an enhanced conformational sampling that omits it may suffer from strong hysteresis and produce inaccurate free energetics (see Supplementary Figure 2 as an example of slow convergence of conformational free energies). Interestingly, when starting from the equilibrated OF structures, deprotonation of E311 does not result in the closing of the extracellular gate within our 550 ns × 2 replicas (Supplementary Figure 3), implying that the closing of the gate is consistently slow as the intracellular gate, which may be the rate-limiting step of the entire functional cycle. This behavior is in contrast with a kinetic model (Mayes et al., 2018; Yue et al., 2021) in which multiple stochastic paths are possible; in this case the closing of the gate may provide a kinetic funnel bottleneck through which all paths are focused.

The equilibrated OF conformation was superimposed with a recently resolved OF structure of rat PepT2 (PDB: 7NQK)(Parker et al., 2021) to determine if this OF structure was an artifact of the MD simulations. The two structures were aligned using a maximum likelihood approach (Theobald & Steindel, 2012; Theobald & Wuttke, 2008) and they overlap extremely well (C^{α} maximum likelihood RMSD = 0.95 Å), as illustrated in **Figure 5**, and a deeper inspection of the critical residues TM 7 and 10 glutamates/aspartates and TM1 arginine/lysine, as well as the well-conserved ExxER/K motif, reveals a significant overlap between those two structures. These observations support that the OF state sampled from our MD simulations are a physical state of the transporter. As derived from the findings presented above, we directly observe that the titration of TM7 E311 initiates the opening of the extracellular gate via the breaking of the E311-R44 salt bridge, and that the underlying mechanism is more complex than previously anticipated. Although a protonated E311 loses the salt bridge with R44, potentially allowing the extracellular gate structural flexibility, this is insufficient for the transporter to directly switch from an IF to an OF state. The rate-limiting step for the transition is to find an IF-OC state and potentially the coupled

hydration fluctuation that may take microseconds or more. Once the IF-OC state has been reached, the transition to the OF state occurs quickly, in roughly 100 nanoseconds.

Facile PT between TM7 and TM10 via water

Given that TM7 E311 requires losing a proton to seal the extracellular gate while TM10 E418 requires a proton to deliver the ligand into the cytosol, it is a plausible assumption that the proton needs to be transported from E311 to E418. According to the results we showed previously, the proton can drive the conformational change and ligand transport, and now we want to explore how the conformation and ligand, in turn, effect this PT. Toward that goal, we implemented the MS-RMD methodology to simulate explicit proton transport for an all-atom reactive potential energy surface (proton transporting via the Grotthuss shuttle mechanism) in both the IF and OF states, as well as the apo and holo forms, to calculate the free energetics of this PT step with these coupled motions. Interestingly, as seen in **Figure 6**A, the two potentials of mean force (PMF; free energy profiles) of PT in apo OF and apo IF differ only slightly. The difference in the free energy well corresponding to a protonated E311 reveals that the proton on E311 in the IF state is about 1 kcal/mol less stable than that in OF, which is a result of a closer positively charged R44 in the IF state.

Another notable variation in PT behavior is the location of the free energy barrier maximum, which is closer to E311 in apo IF than in apo OF. In the IF state, the closure of the extracellular gate restricts hydration around E311, resulting in less hydration (solvation) for the excess proton (H^+) in that region and as a result, an earlier barrier is seen in the PMF. In spite of these differences, both PMFs reveal that the proton is more thermodynamically stable when bonded to E418 than E311, showing an E311 → E418 proton movement, and this preference will be even more significant when an inward proton gradient across the membrane is taken into account.

In addition to the thermodynamic preference, the calculated rate constants (**Table 1**) confirm that the PT is facile in both apo conformational states. When comparing to the turnover rate of the protein *in vitro* ~ 2 seconds (Minhas et al., 2018), this PT will not be rate-limiting in the transporter functional cycle.

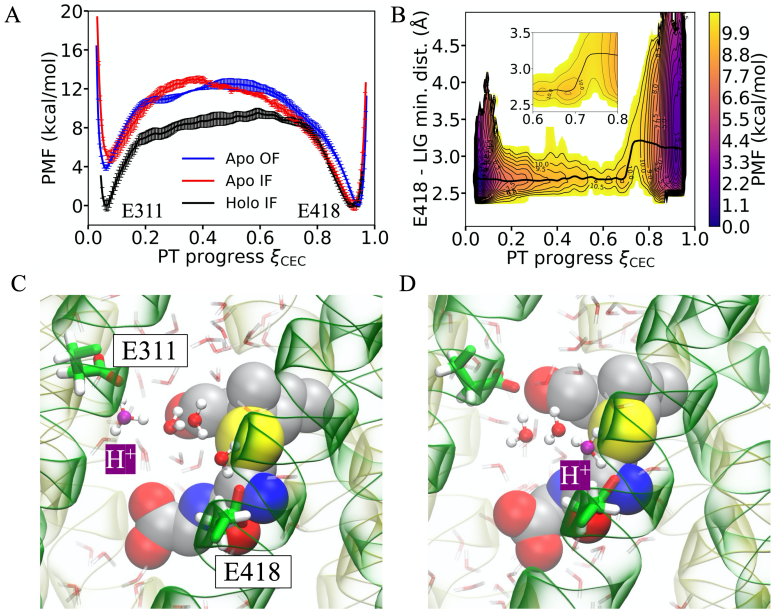


Figure 6. Characterization of proton transport between TM7 and TM10. (A) Potential of mean force (free energy profile) for proton transport between E311 and E418. (B) Two-dimensional potential of mean force of proton transport between E311 and E418 with the minimum distance between E418 carboxyl and ligand N-terminal nitrogen. The inset is a zoom-in showing the strongly coupled region between proton and ligand. (C) & (D) Molecular figures showing Grotthuss proton shuttling mechanism when the ligand is present. The most probable hydronium oxygen is highlighted in purple. The ligand is shown in the Van-der-Waals representation. The N-terminal bundle of the protein is shown in transparent yellow and the C-terminal bundle is shown in transparent green.

As mentioned in the preceding section, the TM10 E418 is correlated to ligand transport, and its protonation results in cytosolic ligand release. Here, we again used the MS-RMD method in conjunction with enhanced sampling to investigate how the proton could be transferred from E311 to E418 in the presence of the ligand. The presence of ligand does not significantly alter the overall shape of the PT free energy, as seen in **Figure 6**A, but it does affect the relative stability between protonated E311 and E418. Because of the energy gained through forming a salt bridge between deprotonated E418 and the ligand N-terminus, the excess proton is now equally stable on both glutamates, and the PT direction will be fully determined by the direction of the proton gradient. Interestingly, the bulky sidechain does not fully remove water between the glutamates, allowing an excess proton to travel through connected water wires via Grotthuss shuttling (Grotthuss, 1806; Knight & Voth, 2012) (**Figure 6**C & 6D). As a result, the PT rate is slower than in the apo case (**Table 1**), but it is still a feasible PT and would not constitute the rate-limiting step of the entire cycle. This facile PT mechanism entails a "decoupling" between proton and ligand transport in the sense that the ligand is not needed to participate and has only a modest impact on proton movement,

thus adding to our understanding of why the transporter may use the proton gradient to transport a wide range of ligands. The joint free energy profile of PT progress with the E418–ligand distance (**Figure 6**B), on the other hand, depicts the coupling between proton and ligand. When a proton is bound to E418, the PMF shows more flexibility in the direction of E418–ligand distance, consistent with the observations in classical MD that protonation of E418 eventually triggers ligand release. At equilibrium, the system motion tends to follow the minimum free energy path (MFEP), and the ramped slope of the MFEP, especially in the range of $0.6 < \xi_{\mathrm{CEC}} < 0.8$ (**Figure 6**B inset), reveals that the motion of the proton can drive the motion of the ligand and that the ligand can also be the driving force of the proton.

Table 1. Calculated proton transport rates between TM7 E311 and TM10 E418.

System	Apo IF	Apo OF	Holo IF
$k_{\mathrm{E311\rightarrow E418}}$ ($\mu\mathrm{s}^{-1}$)	19.2±0.8	1.6±0.4	0.54±0.26
$k_{\mathrm{E418\rightarrow E311}}$ ($\mu\mathrm{s}^{-1}$)	$1.5 \pm 1.4\times10^{-2}$	$3.8 \pm 1.6\times10^{-3}$	0.41±0.39

DISCUSSION

POT family proteins represent secondary active transporters that make use of the proton electrochemical gradient to transport various peptides and their analogs into cells. Unraveling the mechanism of how transport is coupled to the proton gradient is important for understanding and improving drug transport, pharmacokinetics, and oral bioavailability. However, it remains challenging to experimentally track the molecular motions in real-time at an atomistic resolution, thus limiting a direct experimental examination of the detailed transport process and discovery of the important interactions. MD simulations, on the other hand, are a valuable tool for studying complicated molecular systems and processes to enable direct insight into atomic motions as well as the calculations of associated free energetics and kinetics.

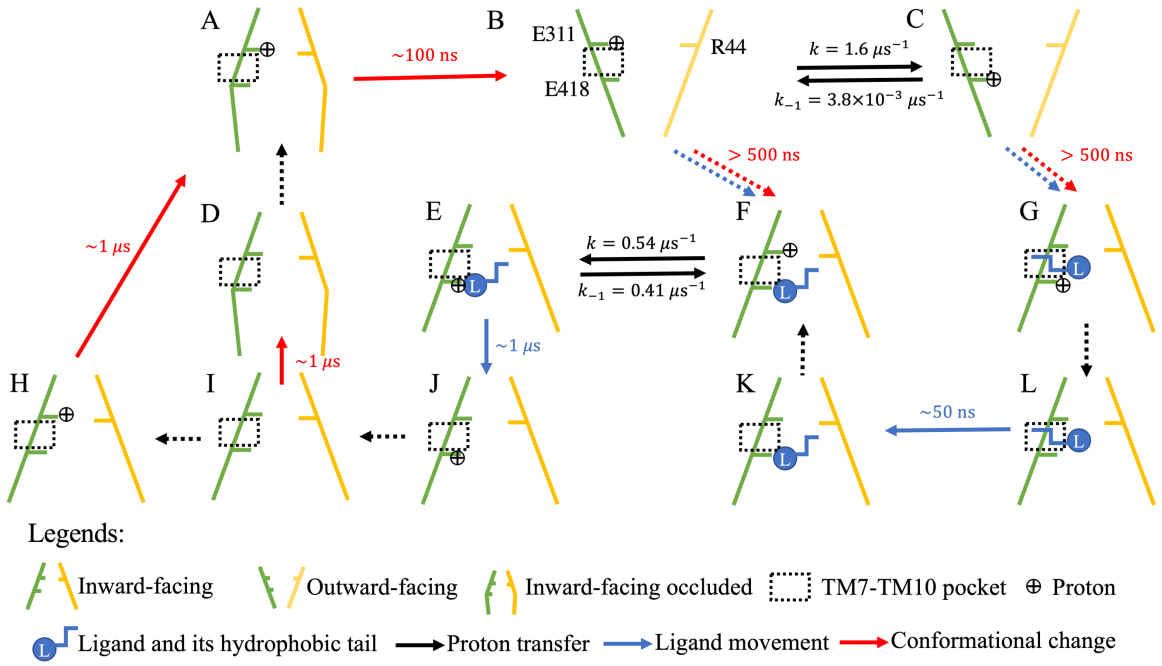


Figure 7. Schematic diagram of the transporter functional conformational and kinetic cycle. The dashed arrows represent the transitions whose reaction rates are not known yet from the performed simulations. The N-terminal bundle is represented by yellow sticks and the C-terminal bundle is colored in green. Note the PT rates were computed without a pH gradient.

In this study and building upon our prior structural and biochemical characterizations, we used comprehensive classical and reactive MD along with enhanced free energy sampling to elucidate the proton coupling mechanism of $PepT_{Sh}$, a member of the POT family. As summarized in **Figure 7**, we suggest based on our results a schematic functional cycle where the proton flow mediated by the transporter drives the conformational switching and ligand movement via altering the TM7 and TM10 glutamate charge states. However, we should emphasize that proteins are stochastic molecular machines and the hopping between the microstates shown in **Figure 7** must be considered to be probabilistic. Moreover, there may exist multiple possible transition pathways connecting the states as hypothesized for other transporters, such as ClC-ec1 (Mayes et al., 2018; Yue et al., 2021) and PiPT (Liu et al., 2021). As such, it should also be the case in $PepT_{Sh}$, and for example, in one pathway (**Figure 7**B→F→E→J), that one single proton reaches TM10 E418 after ligand binding and activates its release. Alternatively, the proton may possibly reach E418 before the ligand binds, and the ligand tail inserts into the TM7-TM10 pocket, resulting in the state captured in the crystal structure (PDB:6EXS) (**Figure 7**B→C→G). The proton then dissociates from E418 and enters the cytosol (**Figure 7**G→L), forming an E418–ligand salt bridge (**Figure**

7L→K), which is followed by another proton binding to E418 and releasing the ligand (**Figure 7**K→F→E→J). We note that in the case of ClC-ec1, the rate-limiting step is not completely restricted to a specific elementary process but is more diverse among multiple pathways, while here in this peptide transporter it may be always the slow global conformational change that dominates over all the various pathway rates. In **Figure 7**, we mainly outlined the primary functioning pathways that best represent the data at hand, but a more complete quantitative description of the whole functional cycle may be required to fully understand the coupling between protein, ligand, and proton motions, as well as their pH dependence and stoichiometry. To achieve that, even more extensive and converged conformational enhanced sampling of the protein is necessary to accurately estimate the transition energy and rates.

Because hydration may also play a critical and entangled role in conformational changes, a reaction coordinate describing it should be properly defined and employed in the free energy sampling. However, the inclusion of additional degrees of freedom in such enhanced sampling may result in an excessive computing overhead when dealing with a high-dimensional CV space. In this scenario, machine learning and statistical approaches (M. Sultan & Pande, 2017; Mardt et al., 2018; McCarty & Parrinello, 2017; Zhang, Wang, et al., 2018) could be effective in reducing the sampling dimensionality and intelligently selecting the most significant phase space region upon which to focus.

We note that the information supplied here provides a limited understanding of the critical importance of the ExxER/K motif. The preliminary data in Supplementary Figure 4 shows that the intracellular gate becomes more flexible with a deprotonated E33 since the IF-OC state was observed earlier than the protonated E33 simulation (**Figure 4**C), but a more rigorous conformational free energy landscape is needed to draw a firm conclusion. In the current simulations, the charged E33 serves as an attraction to stabilize the unusual interaction between R37 and K137 observed in $PepT_{Sh}$ as well as in the cryo-EM rat PepT2 structure. We further note that the current force field-based description may be inaccurate for a possible proton sharing between R37 and K137, but a quantum mechanical treatment of that motif would be too computationally expensive to achieve adequate conformational sampling. In future work it may be possible to train an effective potential from highly accurate electronic structure calculations that corrects the conventional force fields for better modeling this novel Arg-Lys interaction, in light

of the recent developments deep learning potentials using *ab initio* calculations (Zhang, Han, Wang, Saidi, et al., 2018; Zhang, Han, Wang, Car, et al., 2018).

METHODS

Classical molecular dynamics simulations

The MD computational model was constructed using the CHARMM-GUI (Jo et al., 2007; Jo et al., 2008) from the holo IF crystal structure of $PepT_{Sh}$ (PDB: 6EXS) by embedding the protein into a 90-Å × 90-Å bilayer composed of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipids and then solvating the system with a 20-Å layer of TIP3P water (Durell et al., 1994) on both sides of the membrane with 0.15 M sodium chloride. The protonation states of ionizable residues were assigned according to pK_a predictions from constant-pH MD (cpHMD) simulations (detailed in below) and are listed in Supplementary Table 1. The convergence of cpHMD pK_a calculations for crucial residues are shown in Supplementary Figure 5. The CHARMM 22/CMAP (MacKerell Jr. et al., 1998; Mackerell Jr. et al., 2004a, 2004b) force field was employed to describe the protein and the CHARMM36 (Klauda et al., 2010) was used for lipid interactions. The ligand was modeled as a dipeptide consisting of S-Cys-3M3SH (CSM) and a glycine. CSM was added as a non-standard amino acid to the force field, and was parameterized by CHARMM General Force Field (CGenFF, version 4.0) (Vanommeslaeghe et al., 2010) using ParamChem (version 2.2.0). The CGenFF parameters for the backbone and sidechain atoms up to S^{γ} were replaced by those for cysteine from CHARMM22 force field. The atomic charge of C^{β} was adjusted for charge neutrality (Appendix I). Particle mesh Ewald (PME) (Darden et al., 1993; Essmann et al., 1995) with a cutoff of 12.0 Å and a precision of 10^{-5} was used to compute the electrostatic interactions. The Lennard-Jones (LJ) non-bonded interactions was force-switched from 10 Å to 12 Å. The system was equilibrated following the CHARMM-GUI protocol (Wu et al., 2014), followed by a 200-ns additional equilibration with 1000 kJ/nm^2 harmonic restraints on protein heavy atoms. In the additional equilibration and production runs, the system was integrated by the leap-frog algorithm with a 2-fs long time step, and all the bonds involving hydrogen atoms were restrained using the LINCS algorithm (Hess et al., 1997). The temperature was controlled at 303.15 K by the Nosé-Hoover thermostat (Hoover, 1985; Nosé, 1984) and the pressure was controlled at 1 atm by the Parrinello-Rahman semiisotropic barostat (Parrinello & Rahman, 1981). The simulation time

and the initial configurations of production runs were summarized in Supplementary Table 1. All of the classical MD simulations were conducted in the GROMACS package (Abraham et al., 2015).

Multiscale reactive molecular dynamics (MS-RMD) simulations

The MS-RMD approach was well documented in refs (Lee et al., 2016; Wu et al., 2008). In brief, the MS-RMD approach provides an efficient way to simulate molecular systems with explicit modeling of chemical reactions (varying bonding topology in time). This was achieved by considering the system as a linear combination of diabatic states $\{|i\rangle\}$, each of which corresponds to a different bonding topology. The Hamiltonian of the system is then expressed in the following diabatic state representation:

$$\mathbf{H}^{\mathrm{RMD}} = \sum_{ij} |i\rangle\, h_{ij} \langle j| \qquad 1$$

The diagonal term h_{ii} is taken to be the energy function described by the classical force field, namely the CHARMM22/CMAP for proteins and CHARMM36 for lipids. The off-diagonal element, h_{ij}, is modeled by a physically inspired ansatz in a MM form. The detailed definitions of these terms are provided in the ref (Li & Voth, 2021). The ground state of the reactive system can be obtained through solving the following secular equation "on the fly" as a function of nuclear configuration, such that

$$\mathbf{H}^{\mathrm{RMD}}\boldsymbol{c} = E\boldsymbol{c} \qquad 2$$

The eigenvector $\boldsymbol{c} = \{c_i\}$ with the lowest eigen energy is the adiabatic wave-function of the ground state. The atomic forces, as the energy gradient, are c_i weighted diabatic forces according to the Hellmann-Feynman theorem:

$$\boldsymbol{F} = \sum_{ij} c_i c_j \boldsymbol{F}_{ij} \qquad 3$$

The diabatic matching approach (Li & Voth, 2021) was used to parametrize the ionizable MS-RMD model for glutamates described by the CHARMM22 force field and the parameters are summarized in Supplementary Table 2.

The MS-RMD simulations were initiated from classical MD equilibrations. The electrostatics was computed by the particle-particle particle-mesh method (Hockney & Eastwood, 1988) with a

cut-off of 10 Å and an accuracy criterion of 10^{-4}. The non-bonded LJ potential was energy-switched from 8 Å to 10 Å. The system was integrated by the Nosé-Hoover chain thermostat to maintain a 303.15 K temperature in the NVT ensemble using a timestep of 1 fs. The MS-RMD simulations were performed by the LAMMPS MD package (Plimpton, 1995) coupled to RAPTOR (Yamashita et al., 2012) to enable chemical reactions.

Enhanced free energy sampling and rate calculations

The PT between E311 and E418 in the apo form was enhanced by the well-tempered metadynamics (WT-MTD) approach (Barducci et al., 2008). The CV used in the WT-MTD was defined as a distance ratio:

$$\xi_{\mathrm{CEC}} = \frac{r_{\mathrm{E311}}}{r_{\mathrm{E311}} + r_{\mathrm{E418}}} \qquad 4$$

where r_{E311} and r_{E418} are the minimum distance between the center of excess charge (CEC) and carboxyl oxygen atoms of E311 or E418. Due to the charge delocalized nature of a hydrated excess proton, its effective position is tracked by the center of excess charge (CEC) introduced by the excess proton defect, or equivalently, the "electron hole" created by the extra proton nucleus. In MS-RMD, the CEC is defined as the c_i weighted center of charge (COC) of the proton-carrier species in each diabatic state (Cuma et al., 2001),

$$\boldsymbol{r}_{\mathrm{CEC}} = \sum_i c_i^2 \boldsymbol{r}_i^{\mathrm{COC}} \qquad 5$$

The minimum distance between CEC and carboxyl oxygens was implemented by a softmin function, defined as

$$\mathrm{softmin}(r_1, r_2) = -\frac{1}{\kappa}\log\left[\exp\left(-\kappa(r_1 - \bar{r})\right) + \exp(-\kappa(r_2 - \bar{r}))\right] + \bar{r} \qquad 6$$

where $\kappa = 40\ \text{Å}^{-1}$, r_1 and r_2 denote the CEC separation from the two carboxyl oxygen atoms, and $\bar{r} = (r_1 + r_2)/2$. The initial Gaussian height in WT-MTD was 0.8 kcal/mol and scaled according to a bias factor of 12. The Gaussians were deposited on the ξ_{CEC} dimension every 1 ps with a fixed width of 0.01. In both apo IF and apo OF states, two replicates of metadynamics were run for at least 10 ns.

The PT between E311 and E418 in holo IF state was sampled by umbrella sampling (Torrie & Valleau, 1977) owing to its convenience in distributing sampling tasks on multiple computer nodes.

The umbrella window centers were placed from $\xi_{\mathrm{CEC}} = 0$ to $\xi_{\mathrm{CEC}} = 1$ every 0.025 with harmonic force constants ranging from 2000 kcal/mol to 3500 kcal/mol. The simulation time for each window ranged from 370 ps to 7 ns depending on its convergence, resulting in a 57-ns simulation time in total. The enhanced sampling functionality was provided by PLUMED 2 (Tribello et al., 2014) coupled to LAMMPS MD engine and the RAPTOR MS-RMD software.

The free energy surfaces (potentials of mean force; PMFs) were computed from metadynamics or umbrella sampling data using the dynamic histogram analysis method (DHAM) (Rosta & Hummer, 2015), and so was the Markov transition matrix in the CV space. The CV-position-dependent diffusion constants $D(\xi)$ were computed following the same procedure in ref (Sicard et al., 2021) using the transition matrix, and the reaction rate constant was computed as the inverse of the mean first passage time (MFPT) (Szabo et al., 1980),

$$k_{\mathrm{E311\rightarrow E418}} = \tau_{\mathrm{MFPT}}^{-1} = 1/\int_{\xi_{\mathrm{E311}}}^{\xi_{\mathrm{E418}}} \mathrm{d}\xi' \int_{\xi_l}^{\xi'} \mathrm{d}\xi''\, D(\xi')^{-1} e^{-\beta\left(F(\xi')-F(\xi'')\right)} \qquad 7$$

assuming Smoluchowski dynamics of for the variable ξ_{CEC}. In eq 7, the ξ_{E311} and ξ_{E418} are the CV values that correspond to free energy minima of protonated E311 and E418 respectively, ξ_l is the CV value that corresponds to the lower boundary of the E311 free energy well, $\beta = 1/k_B T$ is the inverse temperature, and $F(\xi)$ is the PMF. The reverse PT rate was computed via the detailed balance relation

$$k_{\mathrm{E418\rightarrow E311}} x_{\mathrm{E418}} = k_{\mathrm{E311\rightarrow E418}} x_{\mathrm{E311}} \qquad 8$$

where x is equilibrium mole fraction computed as the integral of Boltzmann factor $e^{-\beta F(\xi)}$ in the E311 basin or the E418 basin. The errors reported for WT-MTD PMFs and PT rates were computed as the standard deviation between two replicates, while the errors in umbrella sampling counterparts were computed from the standard deviation of the last 5 blocks of equally partitioning the trajectories into 6 blocks.

Characterization of gate sizes and gate hydration

The extracellular and intracellular gate sizes were calculated as the tip distance between TM1,2 and TM7,8 and between TM4,5 and TM10,11. We first defined four virtual atoms as backbone geometric centers of (1) residues 46-52 and 64-70, (2) residues 316-324 and 342-347, (3) residues

140-146 and 154-160, and (4) residues 426-432 and 438-444, to represent the tip positions of TM1,2, TM7,8, TM4,5 and TM10,11 respectively. The extracellular gate size was then defined as the distance between virtual atoms (1) and (2), and the intracellular gate size was similarly defined as the distance between virtual atoms (3) and (4).

The intracellular gate water density was defined as the number of water molecules in a quadrangular prism divided by its volume. The prism base was defined as the *xy*-plane (membrane plane) projection of the tip residues of TM4,5,10 and 11, represented by four virtual atoms defined as the backbone geometric center of residues 140-146, 154-160, 426-432 and 438-444. The height of the prism was defined as the *z* range of the backbone atoms used to define the virtual centers.

CpHMD simulation

To enable titration, doubly protonated HIS residues were used, dummy hydrogen atoms were added to ASP, GLU, and C-terminus (CT) carboxylates in *syn* positions. The membrane-enabled (Chen et al., 2016) hybrid-solvent Constant pH MD (CpHMD) (Wallace & Shen, 2011) simulations were conducted using CHARMM (Brooks et al., 2009) (version c42b2) with the pH-based replica exchange (pH-REX) enhanced-sampling protocol (Wallace & Shen, 2011). The simulations represented protein, lipids, and waters using the CHARMM22/CMA force field, the CHARMM36 model, the CHARMM-modified TIP3P model, respectively. The CpHMD parameters for the N-terminus (NT) (Appendix II and Supplementary Figure 6) and CT (Appendix III and Supplementary Figure 7) were derived using the protocols described by Lee *et al*. (Lee et al., 2004) and Khandogin *et al.* (Khandogin & Brooks III, 2005), respectively. The LJ interactions were force-switched (Steinbach & Brooks, 1994) from 8 to 12 Å. Electrostatic interactions were computed using the PME method with a real-space cutoff of 12 Å and a sixth-order interpolation with 1-Å grid spacing. The non-bonded neighbor list was updated heuristically. To allow for a 2-fs step length, SHAKE (Ryckaert et al., 1977) was used to restrain the bonds connecting hydrogen atoms. All simulations were conducted with periodic boundary conditions at 303.15 K by the Nosé-Hoover thermostat, and 1 atm by the Langevin piston pressure-coupling algorithm (Feller et al., 1995).

In CpHMD, a fictitious titration coordination λ that updates simultaneously with the spatial coordinates is coupled to each ionizable site to describe its charge state. The λ particles with a

mass of 10 atomic mass unit are propagated using the extended Langevin algorithm (Loncharich et al., 1992) with a collision frequency of 5 ps^{-1} and updated every 10 MD steps to allow for water relaxation. The hybrid-solvent CpHMD uses the leapfrog Verlet integrator (Hockney, 1970) to propagate the conformational dynamics in explicit solvent and lipid molecules. The implicit solvent and membrane modelled by the generalized-Born (GB) model GBSW (Im, Feig, et al., 2003; Im, Lee, et al., 2003) with optimized GB input radii (Chen et al., 2006) is used to calculate the electrostatic hydration forces on λ particles. An infinite low-dielectric-constant (ε = 2) slab with a high-dielectric-constant (ε = 80) exclusion cylinder aligned with the membrane normal is used to represent protein-embedded membrane in GBSW. The implicit-membrane had a thickness of 35 Å and the radius of the exclusion cylinder was 25 Å. The dielectric constant (ε) was switched from 2 to 80 within 2.5 Å from both membrane surfaces. An ionic strength of 0.150 M was used for the Debye–Hückel term (Srinivasan et al., 1999) in GB. All Asp, Glu, His, Lys, Arg, Cys, Tyr residues, as well as the NT and CT of the ligand were allowed to ionize. A cylindrical restraint with a force constant of 1 kcal/(mol•$Å^2$) was applied to the center of mass of the protein heavy atoms via the MMFP utility in CHARMM (Brooks et al., 2009) to prevent the protein from lateral drift. Ions were excluded from the hydrophobic membrane region (–16.5 Å < Z < 16.5 Å) by a planar restraint of 5 kcal/(mol•$Å^2$) via MMFP. In pH-REX, exchanges between neighboring pH replicas were attempted every 500 MD steps. Decision of accepting or refusing an exchange attempt was determined using the Metropolis criterion (Metropolis & Ulam, 1949). Two independent runs were conducted for apo-$PepT_{Sh}$: one with and the other without titrations on Lys/Arg/Cys/Tyr residues (Supplementary Table 3). The 1st apo run placed 40 replicas from pH 2.0 to 11.75 with an interval of 0.25 and lasted 20 ns per replica. The 2nd apo run placed 40 replicas from pH 2.0 to 9.0 with an interval of 0.125 or 0.25 and lasted 20 ns per replica. The apo run of dipeptide CSM-GLY (i.e., ligand) placed 20 replicas from pH 7.0 to 16.5 with an interval of 0.5 and lasted for 5 ns. The holo run (i.e., $PepT_{Sh}$ with CSM-GLY) placed 40 replicas from pH 2.0 to 11.75 with an interval of 0.25 and lasted 10 ns per replica.

Calculation of pK_a

The pK_a was computed by fitting the unprotonated fraction S^{unprot} *vs.* pH to the Hill equation (Hill, 1910),

$$S^{\mathrm{unprot}} = \frac{1}{1 + 10^{n(\mathrm{p}K_{\mathrm{a}} - \mathrm{pH})}} \tag{9}$$

where the Hill coefficient n describes the steepness of the transition region in a titration curve. S^{unprot} was determined by counting the population of protonated (defined as those with $\lambda \leq 0.1$) and deprotonated ($\lambda \geq 0.9$) states for every pH.

ACKNOWLEDGEMENTS

This research was supported by the National Institute of General Medical Sciences (NIGMS) of the U.S. National Institutes of Health (NIH) through grant R01 GM053148. Computational resources were provided by the Research Computing Center (RCC) at the University of Chicago. We thank Prof. Jessica Swanson for helpful comments during the early stages of this research.

REFERENCES

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, & Lindahl E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Softwarex, 1−2*, 19−25. doi:10.1016/j.softx.2015.06.001

Aduri NG, Montefiori M, Khalil R, Gajhede M, Jørgensen FS, & Mirza O. (2019). Molecular Dynamics Simulations Reveal the Proton: Peptide Coupling Mechanism in the Bacterial Proton-Coupled Oligopeptide Transporter YbgH. *ACS Omega, 4*(1), 2040-2046.

Barducci A, Bussi G, & Parrinello M. (2008). Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical Review Letters, 100*(2), 020603. doi:ARTN 020603

10.1103/PhysRevLett.100.020603

Batista MR, Watts A, & José Costa-Filho A. (2019). Exploring Conformational Transitions and Free-Energy Profiles of Proton-Coupled Oligopeptide Transporters. *Journal of Chemical Theory and Computation, 15*(11), 6433-6443.

Brooks BR, Brooks III CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, & Karplus M. (2009). CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry, 30*(10), 1545−1614. doi:10.1002/jcc.21287

Chen J, Im W, & Brooks III CL. (2006). Balancing Solvation and Intramolecular Interactions: Toward a Consistent Generalized Born Force Field. *J. Am. Chem. Soc., 128*(11), 3728−3736. doi:10.1021/ja057216r

Chen W, Huang Y, & Shen J. (2016). Conformational Activation of a Transmembrane Proton Channel from Constant pH Molecular Dynamics. *Journal of Physical Chemistry Letters, 7*(19), 3961−3966. doi:10.1021/acs.jpclett.6b01853

Collins K. (1989). Sweat glands: eccrine and apocrine. In *Pharmacology of the Skin I* (pp. 193-212): Springer.

Cuma M, Schmitt UW, & Voth GA. (2001). A multi-state empirical valence bond model for weak acid dissociation in aqueous solution. *Journal of Physical Chemistry A, 105*(12), 2814-2823. doi:10.1021/jp0038207

Daniel H, & Kottra G. (2004). The proton oligopeptide cotransporter family SLC15 in physiology and pharmacology. *Pflügers Archiv, 447*(5), 610-618.

Darden T, York D, & Pedersen L. (1993). Particle mesh Ewald: An N•log(N) method for Ewald sums in large systems. *Journal of Chemical Physics, 98*(12), 10089−10092. doi:10.1063/1.464397

Doki S, Kato HE, Solcan N, Iwaki M, Koyama M, Hattori M, Iwase N, Tsukazaki T, Sugita Y, Kandori H, Newstead S, Ishitani R, & Nureki O. (2013). Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proceedings of the National Academy of Sciences of the United States of America, 110*(28), 11343-11348. doi:10.1073/pnas.1301079110

Drew D, & Boudker O. (2016). Shared molecular mechanisms of membrane transporters. *Annual Review of Biochemistry, 85*, 543-572.

Durell SR, Brooks BR, & Ben-Naim A. (1994). Solvent-Induced Forces between Two Hydrophilic Groups. *Journal of Physical Chemistry, 98*(8), 2198−2202. doi:10.1021/j100059a038

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, & Pedersen LG. (1995). A smooth particle mesh Ewald method. *Journal of Chemical Physics, 103*(19), 8577−8593. doi:10.1063/1.470117

Feller SE, Zhang Y, Pastor RW, & Brooks BR. (1995). Constant pressure molecular dynamics simulation: The Langevin piston method. *Journal of Chemical Physics, 103*(11), 4613−4621. doi:10.1063/1.470648

Giacomini KM, Huang SM, Tweedie DJ, Benet LZ, Brouwer KLR, Chu XY, Dahlin A, Evers R, Fischer V, Hillgren KM, Hoffmaster KA, Ishikawa T, Keppler D, Kim RB, Lee CA, Niemi M, Polli JW, Sugiyama Y, Swaan PW, Ware JA, Wright SH, Yee SW, Zamek-Gliszczynski MJ, Zhang L, & Transporter I. (2010). Membrane transporters in drug development. *Nature Reviews Drug Discovery, 9*(3), 215-236. doi:10.1038/nrd3028

Grotthuss CJDv. (1806). Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann. Chim.*(58), 54-73.

Hess B, Bekker H, Berendsen HJC, & Fraaije JGEM. (1997). LINCS: A Linear Constraint Solver for Molecular Simulations. *Journal of Computational Chemistry, 18*(12), 1463−1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hill AV. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol., 40*(Suppl), iv−vii. doi:10.1113/jphysiol.1910.sp001386

Hockney RW. (1970). The Potential Calculation and Some Applications. *Methods Comput. Phys., 9*, 136−211.

Hockney RW, & Eastwood JW. (1988). *Computer simulation using particles* (Special student ed.). Bristol England ; Philadelphia: A. Hilger.

Hoover WG. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A, 31*(3), 1695−1697. doi:10.1103/PhysRevA.31.1695

Huang YF, Lemieux MJ, Song JM, Auer M, & Wang DN. (2003). Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science, 301*(5633), 616-620. doi:Doi 10.1126/Science.1087619

Im W, Feig M, & Brooks III CL. (2003). An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins. *Biophysical Journal, 85*(5), 2900−2918. doi:10.1016/S0006-3495(03)74712-2

Im W, Lee MS, & Brooks III CL. (2003). Generalized Born Model with a Simple Smoothing Function. *Journal of Computational Chemistry, 24*(14), 1691−1702. doi:10.1002/jcc.10321

Immadisetty K, Hettige J, & Moradi M. (2017). What can and cannot be learned from molecular dynamics simulations of bacterial proton-coupled oligopeptide transporter GkPOT? *The Journal of Physical Chemistry B, 121*(15), 3644-3656.

Jo S, Kim T, & Im W. (2007). Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *Plos One, 2*(9), e880. doi:10.1371/journal.pone.0000880

Jo S, Kim T, Iyer VG, & Im W. (2008). CHARNIM-GUI: A Web-Based Grraphical User Interface for CHARMM. *Journal of Computational Chemistry, 29*(11), 1859−1865. doi:10.1002/jcc.20945

Khandogin J, & Brooks III CL. (2005). Constant pH Molecular Dynamics with Proton Tautomerism. *Biophysical Journal, 89*(1), 141−157. doi:10.1529/biophysj.105.061341

Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, Vorobyov IV, MacKerell Jr. AD, & Pastor RW. (2010). Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *Journal of Physical Chemistry B, 114*(23), 7830−7843. doi:10.1021/jp101759q

Knight C, & Voth GA. (2012). The curious case of the hydrated proton. *Accounts of Chemical Research, 45*(1), 101-109.

Lee MS, Salsbury Jr. FR, & Brooks III CL. (2004). Constant-pH Molecular Dynamics Using Continuous Titration Coordinates. *Proteins-Structure Function and Genetics, 56*(4), 738−752. doi:10.1002/prot.20128

Lee S, Liang R, Voth GA, & Swanson JM. (2016). Computationally efficient multiscale reactive molecular dynamics to describe amino acid deprotonation in proteins. *Journal of Chemical Theory and Computation, 12*(2), 879-891.

Li C, & Voth GA. (2021). Accurate and Transferable Reactive Molecular Dynamics Models from Constrained Density Functional Theory. *Journal of Physical Chemistry B*. doi:10.1021/acs.jpcb.1c05992

Lin L, Yee SW, Kim RB, & Giacomini KM. (2015). SLC transporters as therapeutic targets: emerging opportunities. *Nature Reviews Drug Discovery, 14*(8), 543-560.

Liu Y, Li C, Gupta M, Verma N, Johri AK, Stroud RM, & Voth GA. (2021). Key computational findings reveal proton transfer as driving the functional cycle in the phosphate transporter PiPT. *Proceedings of the National Academy of Sciences, 118*(25).

Loncharich RJ, Brooks BR, & Pastor RW. (1992). Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of *N*-Acetylalanyl-*N*'-Methylamide. *Biopolymers, 32*(5), 523−535. doi:10.1002/bip.360320508

M. Sultan M, & Pande VS. (2017). tICA-metadynamics: accelerating metadynamics by using kinetically selected collective variables. *Journal of Chemical Theory and Computation, 13*(6), 2440-2447.

MacKerell Jr. AD, Bashford D, Bellott M, Dunbrack Jr. RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, & Karplus M. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *Journal of Physical Chemistry B, 102*(18), 3586−3616. doi:10.1021/jp973084f

Mackerell Jr. AD, Feig M, & Brooks III CL. (2004a). Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *Journal of Computational Chemistry, 25*(11), 1400−1415. doi:10.1002/jcc.20065

MacKerell Jr. AD, Feig M, & Brooks III CL. (2004b). Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am .Chem. Soc., 126*(3), 698−699. doi:10.1021/ja036959e

Mardt A, Pasquali L, Wu H, & Noé F. (2018). VAMPnets for deep learning of molecular kinetics. *Nature Communications, 9*(1), 1-11.

Mayes HB, Lee S, White AD, Voth GA, & Swanson JMJ. (2018). Multiscale Kinetic Modeling Reveals an Ensemble of Cl-/H+ Exchange Pathways in ClC-ec1 Antiporter. *Journal of the American Chemical Society, 140*(5), 1793-1804. doi:10.1021/jacs.7b11463

McCarty J, & Parrinello M. (2017). A variational conformational dynamics approach to the selection of collective variables in metadynamics. *The Journal of Chemical Physics, 147*(20), 204109.

Metropolis N, & Ulam S. (1949). The Monte Carlo Method. *J. Am. Stat. Assoc., 44*(247), 335−341. doi:10.1080/01621459.1949.10483310

Minhas GS, Bawdon D, Herman R, Rudden M, Stone AP, James AG, Thomas GH, & Newstead S. (2018). Structural basis of malodour precursor transport in the human axilla. *Elife, 7*, e34995.

Minhas GS, & Newstead S. (2019). Structural basis for prodrug recognition by the SLC15 family of proton-coupled peptide transporters. *Proceedings of the National Academy of Sciences, 116*(3), 804-809.

Minhas GS, & Newstead S. (2020). Recent advances in understanding prodrug transport through the SLC15 family of proton-coupled transporters. *Biochemical Society Transactions, 48*(2), 337-346.

Newstead S. (2015). Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters. *Biochimica Et Biophysica Acta-General Subjects, 1850*(3), 488-499. doi:10.1016/j.bbagen.2014.05.011

Newstead S. (2017). Recent advances in understanding proton coupled peptide transport via the POT family. *Current opinion in structural biology, 45*, 17-24.

Newstead S, Drew D, Cameron AD, Postis VL, Xia X, Fowler PW, Ingram JC, Carpenter EP, Sansom MS, & McPherson MJ. (2011). Crystal structure of a prokaryotic homologue of the mammalian oligopeptide–proton symporters, PepT1 and PepT2. *The EMBO journal, 30*(2), 417-426.

Nosé S. (1984). A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics, 52*(2), 255−268. doi:10.1080/00268978400101201

Parker JL, Deme JC, Wu Z, Kuteyi G, Huo J, Owens RJ, Biggin PC, Lea SM, & Newstead S. (2021). Cryo-EM structure of PepT2 reveals structural basis for proton-coupled peptide and prodrug transport in mammals. *Science Advances, 7*(35), eabh3355.

Parker JL, Li C, Brinth A, Wang Z, Vogeley L, Solcan N, Ledderboge-Vucinic G, Swanson JMJ, Caffrey M, Voth GA, & Newstead S. (2017). Proton movement and coupling in the POT family of peptide transporters. *Proceedings of the National Academy of Sciences of the United States of America, 114*(50), 13182-13187. doi:10.1073/pnas.1710727114

Parrinello M, & Rahman A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics, 52*(12), 7182−7190. doi:10.1063/1.328693

Plimpton S. (1995). Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *Journal of Computational Physics, 117*(1), 1-19. doi:DOI 10.1006/jcph.1995.1039

Quistgaard EM, Löw C, Guettou F, & Nordlund P. (2016). Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nature Reviews Molecular Cell Biology, 17*(2), 123-132.

Rosta E, & Hummer G. (2015). Free energies from dynamic weighted histogram analysis using unbiased Markov state model. *Journal of Chemical Theory and Computation, 11*(1), 276-285.

Rubio-Aliaga I, & Daniel H. (2002). Mammalian peptide transporters as targets for drug delivery. *Trends in Pharmacological Sciences, 23*(9), 434-440. doi:Doi 10.1016/S0165-6147(02)02072-2

Ryckaert J-P, Ciccotti G, & Berendsen HJC. (1977). Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *Journal of Computational Physics, 23*(3), 327−341. doi:10.1016/0021-9991(77)90098-5

Selvam B, Mittal S, & Shukla D. (2018). Free energy landscape of the complete transport cycle in a key bacterial transporter. *ACS central science, 4*(9), 1146-1154.

Sicard F, Koskin V, Annibale A, & Rosta E. (2021). Position-Dependent Diffusion from Biased Simulations and Markov State Model Analysis. *Journal of Chemical Theory and Computation, 17*(4), 2022-2033.

Smart OS, Neduvelil JG, Wang X, Wallace BA, & Sansom MSP. (1996). HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph., 14*(6), 354−360. doi:10.1016/s0263-7855(97)00009-x

Smith DE, Clémençon B, & Hediger MA. (2013). Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Molecular Aspects of Medicine, 34*(2-3), 323-336.

Solcan N, Kwok J, Fowler PW, Cameron AD, Drew D, Iwata S, & Newstead S. (2012). Alternating access mechanism in the POT family of oligopeptide transporters. *Embo Journal, 31*(16), 3411-3421. doi:10.1038/emboj.2012.157

Srinivasan J, Trevathan MW, Beroza P, & Case DA. (1999). Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theoretical Chemistry Accounts, 101*(6), 426−434. doi:10.1007/s002140050460

Steinbach PJ, & Brooks BR. (1994). New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *Journal of Computational Chemistry, 15*(7), 667−683. doi:10.1002/jcc.540150702

Szabo A, Schulten K, & Schulten Z. (1980). First passage time approach to diffusion controlled reactions. *The Journal of Chemical Physics, 72*(8), 4350-4357.

Theobald DL, & Steindel PA. (2012). Optimal simultaneous superpositioning of multiple structures with missing data. *Bioinformatics, 28*(15), 1972-1979.

Theobald DL, & Wuttke DS. (2008). Accurate structural correlations from maximum likelihood superpositions. *PLoS computational biology, 4*(2), e43.

Torrie GM, & Valleau JP. (1977). Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics, 23*(2), 187-199. doi:Doi 10.1016/0021-9991(77)90121-8

Tribello GA, Bonomi M, Branduardi D, Camilloni C, & Bussi G. (2014). PLUMED 2: New feathers for an old bird. *Computer Physics Communications, 185*(2), 604-613. doi:10.1016/j.cpc.2013.09.018

Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, & Vorobyov I. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry, 31*(4), 671-690.

Wallace JA, & Shen JK. (2011). Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange. *Journal of Chemical Theory and Computation, 7*(8), 2617−2629. doi:10.1021/ct200146j

Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, Qi Y, Lee J, Monje-Galvan V, Venable RM, Klauda JB, & Im W. (2014). CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. *Journal of Computational Chemistry, 35*(27), 1997−2004. doi:10.1002/jcc.23702

Wu Y, Chen H, Wang F, Paesani F, & Voth GA. (2008). An improved multistate empirical valence bond model for aqueous proton solvation and transport. *Journal of Physical Chemistry B, 112*(2), 467-482. doi:10.1021/jp076658h

Xu L, Haworth IS, Kulkarni AA, Bolger MB, & Davies DL. (2009). Mutagenesis and cysteine scanning of transmembrane domain 10 of the human dipeptide transporter. *Pharmaceutical Research, 26*(10), 2358-2366.

Yamashita T, Peng Y, Knight C, & Voth GA. (2012). Computationally Efficient Multiconfigurational Reactive Molecular Dynamics. *Journal of Chemical Theory and Computation, 8*(12), 4863-4875. doi:10.1021/ct3006437

Yue Z, Bernardi A, Li C, Mironenko AV, & Swanson JM. (2021). Toward a Multipathway Perspective: pH-Dependent Kinetic Selection of Competing Pathways and the Role of the Internal Glutamate in Cl–/H+ Antiporters. *The Journal of Physical Chemistry B*.

Zhang L, Han J, Wang H, Saidi WA, & Car R. (2018). End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. *arXiv preprint arXiv:1805.09003*.

Zhang L, Wang H, & E W. (2018). Reinforced dynamics for enhanced sampling in large atomic and molecular systems. *The Journal of Chemical Physics, 148*(12), 124113.

Zhang LF, Han JQ, Wang H, Car R, & Weinan E. (2018). Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Physical Review Letters, 120*(14). doi:ARTN 143001

10.1103/PhysRevLett.120.143001

Supplementary Materials for Unraveling the Proton Coupling Mechanism of a Peptide Transporter

Supplementary Figures and Tables

Supplementary Table 1. Simulation details of classical MD

Simulation ID	Protonation States (E33/E311/E418)[a]	Initial Configuration[b]	Total Length (ns)	Data usage (time/Figure)[c]
1	+/-/+	Crystal (6EXS)	500	0-500 ns/Figure 1C
				0-500 ns/Figure 2B
2	+/-/-	Crystal (6EXS)	3300	0-500 ns/Figure 1C
				500 ns/Figure 2A
				0-500 ns/Figure 2B
				0-500 ns/Figure 3B
				0-3300 ns/Figure 4A
				0-3300 ns/Figure 4B
				0-3300 ns/Figure 4C
				2690.2 ns/Figure 4D
3-1	+/-/+	550 ns (simulation 1)	1000	0-1000 ns/Figure 3A
				0-200 ns/Figure 3C
3-2	+/-/+	800 ns (simulation 1)	1000	0-1000 ns/Figure 3A
				0-20 ns/Figure 3D
3-3	+/-/+	1050 ns (simulation 1)	1000	0-1000 ns/Figure 3A
				0-20 ns/Figure 3C
4-1	+/+/-	2500 ns (simulation 1)	500	100-500 ns/Figure 4A
				100-500 ns/Figure 4B
				500 ns/Figure 4E
				500 ns/Figure 5
				0-500 ns/ Figure S1A
4-2	+/+/-	2450 ns (simulation 1)	500	0-500 ns/ Figure S1A
5-1	+/+/-	3250 ns (simulation 1)	500	0-500 ns/Figure 4A
				0-500 ns/Figure 4B
				0-500 ns/ Figure S1B
5-2	+/+/-	2100 ns (simulation 1)	500	0-500 ns/ Figure S1B
6-1	+/-/-	150 ns (simulation 4-1)	550	0-500 ns/ Figure S3

6-2	+/-/-	250 ns (simulation 4-1)	550	0-500 ns/ Figure S3
7	-/-/-	Crystal (6EXS)	1000	0-1000 ns/ Figure S4

[a] Protonated is indicated by "+", and deprotonated is indicated by "-". All the other ionizable residues were assigned their default protonation states except H22, H179, H187, H253, H399, and ligand terminuses were protonated according to CpHMD pK_a calculations. [b] The number indicates the simulation was initiated from that time of the simulation specified in the parenthesis. "Crystal" means the simulation was started from the 200-ns equilibration with protein restraints as described in the Method section. [c] The time segments or time points to be used for making the specified figure.

Supplementary Table 2. MS-RMD parameters for Glu

B	1.94530	V_{ii}	-151.2996
b	1.40003	$\epsilon^{\mathrm{LJ}}_{\mathrm{OE-HH}}$	0.173646
b'	1.08892	$\sigma^{\mathrm{LJ}}_{\mathrm{OE-HH}}$	1.35219
C	1.90167	$\epsilon^{\mathrm{LJ}}_{\mathrm{Ow-HEP}}$	0.544690
c	1.29037	$\sigma^{\mathrm{LJ}}_{\mathrm{Ow-HEP}}$	1.35577
c_1	-25.0477	$\epsilon^{\mathrm{LJ}}_{\mathrm{OE-OH}}$	0.160421
c_2	2.95380	$\sigma^{\mathrm{LJ}}_{\mathrm{OE-OH}}$	3.09726
c_3	1.36184	$\epsilon^{\mathrm{LJ}}_{\mathrm{OEP-Ow}}$	0.0775357
D	143.003	$\sigma^{\mathrm{LJ}}_{\mathrm{OEP-Ow}}$	3.06560
α	1.8		
r_0	0.975		

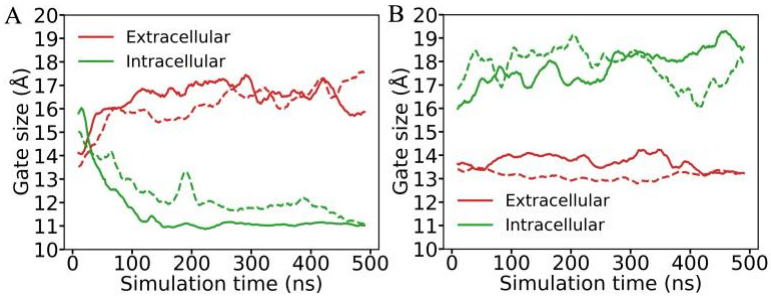
The units of the listed parameters use kcal/mol as the energy unit and Å as the length unit.

Supplementary Table 3. pK_a's from pH-REX hybrid-solvent CpHMD

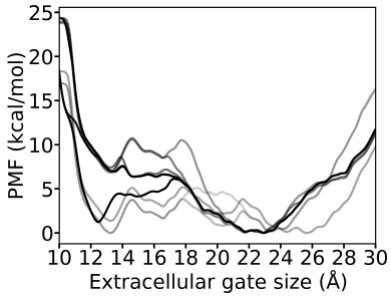
Residue	pK_a values[a]		
	Apo, 1st[b]	Apo, 2nd[c]	Holo[b]
D63	2.8 (0.9)	2.5 (0.7)	2.8 (0.9)
D90	3.9 (0.8)	4.0 (0.9)	4.0 (0.9)
D152	2.1 (0.9)	2.1 (0.8)	2.3 (1.0)
D156	3.6 (0.9)	3.4 (1.0)	3.5 (0.9)
D182	2.5 (0.8)	2.2 (0.8)	2.5 (0.9)
D258	2.5 (0.7)	2.5 (0.9)	2.4 (0.8)
D284	3.6 (0.9)	3.7 (0.8)	3.8 (0.9)
D287	3.5 (0.8)	3.3 (0.8)	2.7 (0.9)
D326	4.3 (0.8)	4.4 (0.9)	4.3 (0.8)
E33	**6.3 (0.6)**	**6.3 (0.7)**	**6.7 (0.6)**
E36	**4.2 (0.7)**	**4.0 (0.8)**	**3.9 (0.8)**
E150	3.7 (1.0)	3.7 (1.0)	3.8 (0.9)
E226	3.4 (0.9)	3.3 (0.9)	3.1 (0.9)
E227	2.3 (0.8)	2.1 (0.9)	2.4 (1.3)
E289	4.1 (0.8)	4.2 (0.8)	4.7 (0.9)
E311	**4.2 (1.0)**	**4.3 (1.0)**	**4.4 (0.8)**
E323	3.1 (0.8)	3.1 (0.9)	3.2 (0.9)
E340	4.7 (0.7)	4.6 (0.8)	4.7 (0.7)
E418	**5.8 (0.8)**	**5.9 (0.8)**	**3.8 (0.6)**
H22	7.7 (0.8)	7.7 (0.8)	7.4 (0.7)
H56	5.8 (0.8)	5.7 (0.8)	5.4 (0.8)
H110	5.2 (0.7)	5.4 (0.7)	5.3 (0.6)
H179	7.6 (0.9)	7.6 (0.9)	7.6 (0.9)
H187	7.8 (0.8)	7.9 (0.8)	7.8 (0.8)
H253	7.6 (0.8)	7.7 (0.9)	7.5 (0.8)
H399	7.1 (0.9)	7.1 (0.9)	7.1 (0.9)
K17	10.2 (0.9)	-	10.3 (0.9)
K64	Stay protonated	-	Stay protonated
K137	Stay protonated	-	Stay protonated
K210	10.4 (0.8)	-	10.3 (0.8)
K225	10.3 (0.8)	-	10.4 (0.8)
K228	9.9 (0.8)	-	9.8 (0.7)
K230	11.4 (0.6)	-	11.5 (0.5)
K283	10.7 (0.9)	-	10.8 (0.9)
K294	9.4 (0.7)	-	9.5 (0.8)
K364	10.7 (0.6)	-	9.6 (0.5)
K367	10.2 (0.8)	-	10.4 (0.8)
K368	10.3 (0.8)	-	10.4 (0.8)
K375	Stay protonated	-	Stay protonated
K431	10.1 (0.7)	-	10.4 (0.8)
K435	10.6 (0.9)	-	10.7 (0.9)
K461	11.3 (0.7)	-	11.4 (0.6)
K464	10.5 (0.8)	-	10.6 (0.8)

K492	10.0 (0.7)	-	9.9 (0.7)
K495	10.4 (0.8)	-	10.3 (0.7)
R24	Stay protonated	-	Stay protonated
R37	Stay protonated	-	Stay protonated
R44	Stay protonated	-	Stay protonated
R91	Stay protonated	-	Stay protonated
R96	Stay protonated	-	Stay protonated
R146	Stay protonated	-	Stay protonated
R154	Stay protonated	-	Stay protonated
R184	Stay protonated	-	Stay protonated
R209	Stay protonated	-	Stay protonated
R229	Stay protonated	-	Stay protonated
R281	Stay protonated	-	Stay protonated
R290	Stay protonated	-	Stay protonated
R292	Stay protonated	-	Stay protonated
R324	Stay protonated	-	Stay protonated
R337	Stay protonated	-	Stay protonated
C112	11.1 (0.6)	-	11.1 (0.6)
C414	10.9 (0.8)	-	11.6 (1.3)
C420	9.5 (0.5)	-	Stay protonated
Y40	Stay protonated	-	Stay protonated
Y41	11.3 (0.6)	-	Stay protonated
Y50	Stay protonated	-	Stay protonated
Y52	11.0 (0.7)	-	11.1 (0.8)
Y74	Stay protonated	-	Stay protonated
Y79	10.5 (0.7)	-	11.3 (0.9)
Y148	Stay protonated	-	Stay protonated
Y163	10.7 (0.8)	-	10.4 (1.0)
Y204	Stay protonated	-	Stay protonated
Y231	11.4 (0.4)	-	11.1 (0.3)
Y250	Stay protonated	-	Stay protonated
Y251	Stay protonated	-	Stay protonated
Y275	11.5 (0.6)	-	11.5 (0.9)
Y320	Stay protonated	-	Stay protonated
Y387	Stay protonated	-	Stay protonated
Y397	11.2 (0.6)	-	10.9 (0.7)
Y411	Stay protonated	-	Stay protonated
Y471	Stay protonated	-	Stay protonated
NT-CSM[d]	16.0 (0.9)		Stay protonated
CT-GLY[d]	8.6 (0.9)		9.8 (0.6)

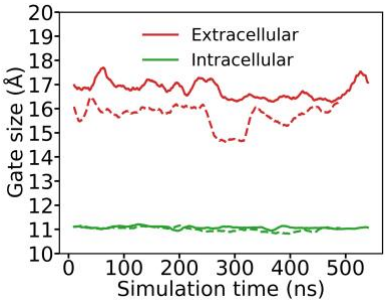
[a] Parenthesized are Hill coefficients. "Stay protonated" indicates a residue, though ionizable in CpHMD, did not titrate within the pH range studied. [b] pH range 2.0–11.75. [c] pH range 2.0–9.0. Lys/Arg/Cys/Tyr permanently protonated. [d] pH range 7.0–16.5.



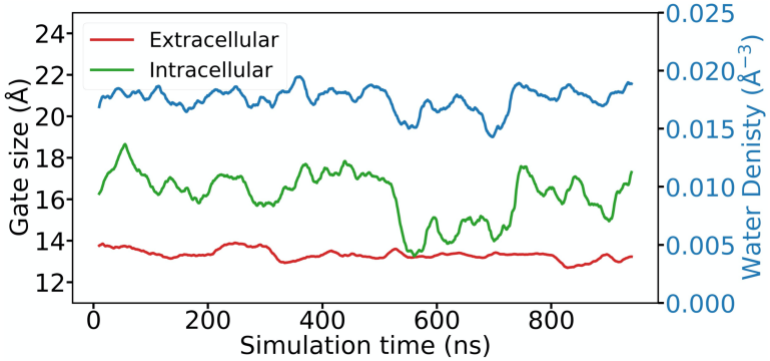
Supplementary Figure 1. Time evolution of gate sizes in protonated E311 simulations. (A) Simulations initiated from the inward-facing occluded state. The solid and dashed curves represent two independent replicas respectively. (B) Simulations initiated from the inward-facing state with two independent replicas (solid and dashed). A running average with a 20-ns window was performed on the time series.



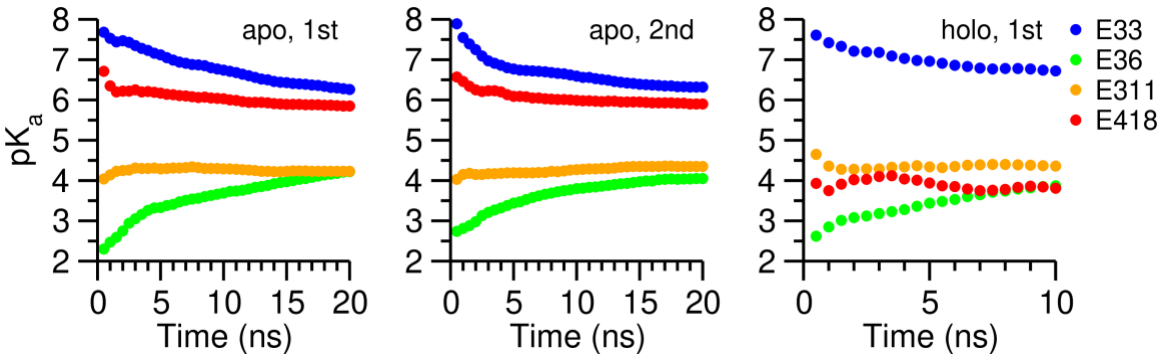
Supplementary Figure 2. Time evolution of potential of mean force of extracellular gate size in a well-tempered metadynamics simulation. The transparency of the curve color indicates the accumulative simulation time from 1 μs to 3 μs with a 200-ns spacing. The metadynamics biased both the extracellular and intracellular gate sizes, and used an initial Gaussian height = 0.6 kcal/mol. The bias factor was 25, Gaussian widths = (0.25 Å, 0.25 Å), and Gaussian were deposited every 100 ps.



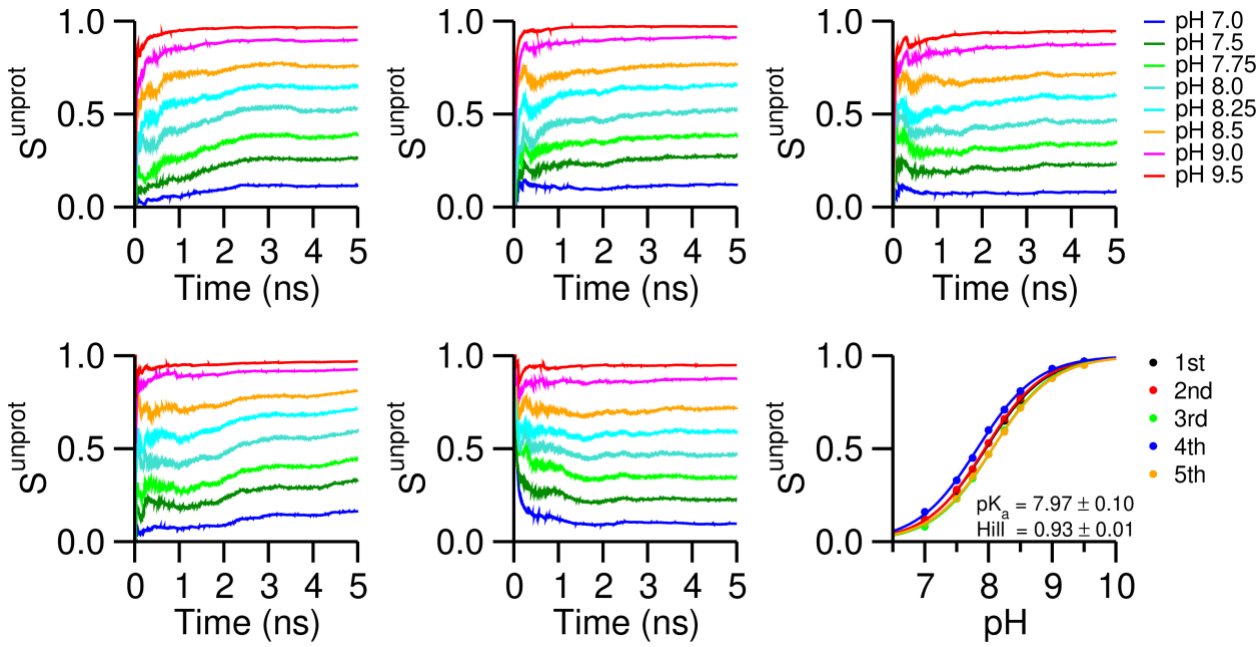
Supplementary Figure 3. Time evolution of gate sizes in deprotonated E311 simulations initiated from the outward-facing state. The solid and dashed curves represent two independent runs. A running average with a 20-ns window was performed on the time series.



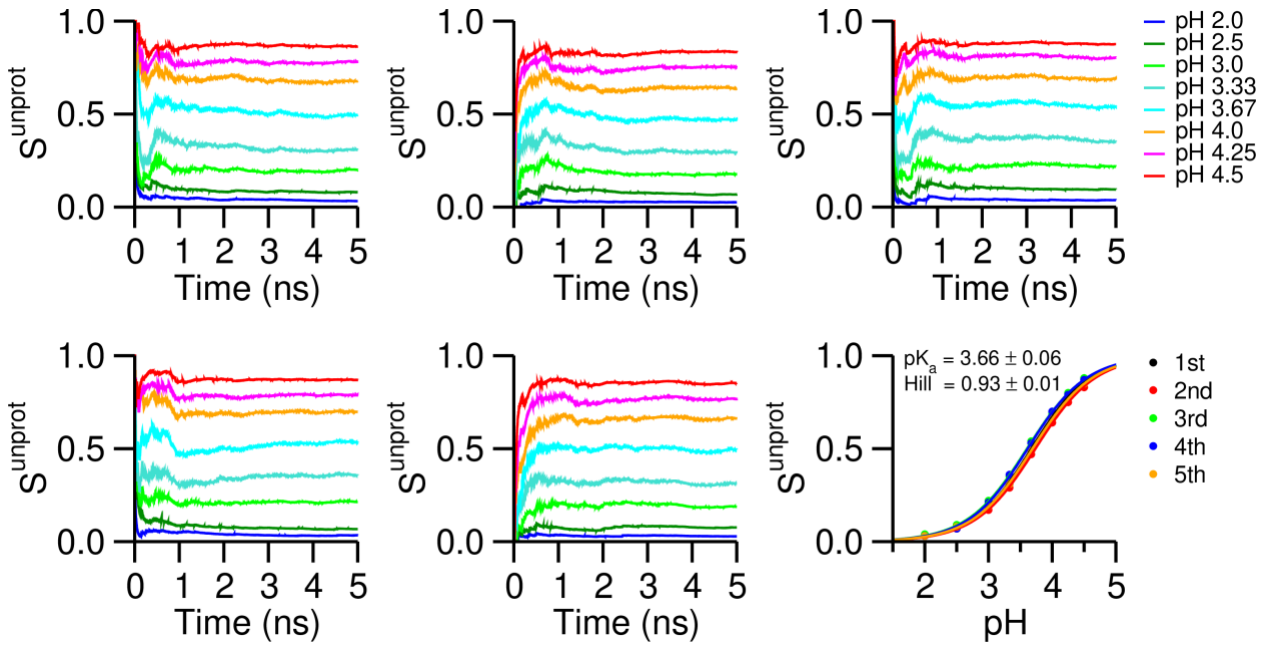
Supplementary Figure 4. Time evolution of gate sizes and the water density around the intracellular gate in a simulation with deprotonated E33. A running average using a 20-ns window was performed on the time series.



Supplementary Figure 5. Convergence of the pK_a values for E33, E36, E311, and E418 in $PepT_{Sh}$ calculated by pH-REX CpHMD. pK_a's were calculated cumulatively versus simulation time.



Supplementary Figure 6. Validation of the hybrid-solvent CpHMD parameters for N-terminus (NT). To calculated the pK_a of NT in aqueous, 5 independent pH-REX CpHMD simulations were performed. The titration curves are plotted in the bottom right panel with pK_a and Hill coefficient *n* reported (mean ± standard deviation). Other panels plot the convergence of the unprotonated fraction S^{unprot} cumulatively calculated *vs.* time.



Supplementary Figure 7. Validation of the hybrid-solvent CpHMD parameters for C-terminus (CT). To calculated the pK_a of CT in aqueous, 5 independent pH-REX CpHMD simulations were performed. The titration curves are plotted in the bottom right panel with pK_a and Hill coefficient *n* reported (mean ± standard deviation). Other panels plot the convergence of the unprotonated fraction S^{unprot} cumulatively calculated *vs.* time.

Appendix I: CHARMM topology and parameter for S-Cys-3M3SH

```
* Toppar stream file generated by
* CHARMM General Force Field (CGenFF) program version 2.2.0
* For use with CGenFF version 4.0
* To model dipeptide S-Cys-Gly-3M3SH, define S-Cys-3M3SH a new non-standard amino acid “CSM”.
* Check Minhas et al., eLife, 2018, 7: e34995
* Generated by Zhi (Shane) Yue, Gregory A. Voth lab, UChicago, 02/18/2019
*

read rtf card append
* Topologies generated by
* CHARMM General Force Field (CGenFF) program version 2.2.0
*
36 1

! "penalty" is the highest penalty score of the associated parameters.
! Penalties lower than 10 indicate the analogy is fair; penalties between 10
! and 50 mean some basic validation is recommended; penalties higher than
! 50 indicate poor analogy and mandate extensive validation/optimization.

DECL -CA
DECL -C
DECL -O
DECL +N
DECL +HN
DECL +CA

RESI CSM            0.000
GROUP
ATOM N      NH1    -0.470 ! from C22 CYS
ATOM HN     H       0.310 ! from C22 CYS
ATOM CA     CT1     0.070 ! from C22 CYS
ATOM HA     HB      0.090 ! from C22 CYS
GROUP
ATOM CB     CT2    -0.132 ! from C22 CYS, ATOM CHARGE adjusted for charge neutrality
ATOM HB1    HA      0.090 ! from C22 CYS
ATOM HB2    HA      0.090 ! from C22 CYS
ATOM SG     S      -0.092 ! penalty = 21.575 from MSH CGENFF, ATOM TYPE replaced by C22 CYS
ATOM CD     CG301   0.238 ! penalty = 27.831 from MSH CGENFF
ATOM CE1    CG331  -0.331 ! penalty = 10.272 from MSH CGENFF
ATOM HE11   HGA3    0.090 ! penalty =  0.810 from MSH CGENFF
ATOM HE12   HGA3    0.090 ! penalty =  0.810 from MSH CGENFF                HE12
ATOM HE13   HGA3    0.090 ! penalty =  0.810 from MSH CGENFF                |
ATOM CE2    CG321  -0.246 ! penalty = 14.778 from MSH CGENFF     |      HE11-CE1-HE13
ATOM HE21   HGA2    0.090 ! penalty =  0.900 from MSH CGENFF  HN-N          |
ATOM HE22   HGA2    0.090 ! penalty =  0.900 from MSH CGENFF     |   HB1     |   HE21  HZ21  HH21
ATOM CZ2    CG321  -0.180 ! penalty =  0.901 from MSH CGENFF     |   |       |   |     |     |
ATOM HZ21   HGA2    0.090 ! penalty =  0.000 from MSH CGENFF  HA-CA--CB--SG--CD--CE2---CZ2---CH2-HH22
ATOM HZ22   HGA2    0.090 ! penalty =  0.000 from MSH CGENFF     |   |       |   |     |     |
ATOM CH2    CG331  -0.270 ! penalty =  0.045 from MSH CGENFF     |   HB2     |   HE22  HZ22  HH23
ATOM HH21   HGA3    0.090 ! penalty =  0.000 from MSH CGENFF   O=C           |
ATOM HH22   HGA3    0.090 ! penalty =  0.000 from MSH CGENFF     |      HE31-CE3-HE32
ATOM HH23   HGA3    0.090 ! penalty =  0.000 from MSH CGENFF                |
ATOM CE3    CG321  -0.249 ! penalty = 14.786 from MSH CGENFF           HZ31-CZ3-HZ32
ATOM HE31   HGA2    0.090 ! penalty =  0.900 from MSH CGENFF                |
ATOM HE32   HGA2    0.090 ! penalty =  0.900 from MSH CGENFF                OH3
ATOM CZ3    CG321   0.051 ! penalty =  1.065 from MSH CGENFF                 \
ATOM HZ31   HGA2    0.090 ! penalty =  0.000 from MSH CGENFF                  HH3
ATOM HZ32   HGA2    0.090 ! penalty =  0.000 from MSH CGENFF
ATOM OH3    OG311  -0.649 ! penalty =  0.090 from MSH CGENFF
ATOM HH3    HGP1    0.420 ! penalty =  0.000 from MSH CGENFF
GROUP
ATOM C      C       0.510 ! from C22 CYS
ATOM O      O      -0.510 ! from C22 CYS

BOND N   HN
BOND N   CA
BOND CA  HA
BOND C   CA
BOND C   +N
BOND CA  CB
BOND CB  HB1
BOND CB  HB2
BOND CB  SG
BOND SG  CD
BOND CD  CE1
BOND CE1 HE11
BOND CE1 HE12
BOND CE1 HE13
BOND CD  CE2
BOND CE2 HE21
BOND CE2 HE22
BOND CE2 CZ2
BOND CZ2 HZ21
BOND CZ2 HZ22
BOND CZ2 CH2
BOND CH2 HH21
BOND CH2 HH22
BOND CH2 HH23
BOND CD  CE3
BOND CE3 HE31
BOND CE3 HE32
BOND CE3 CZ3
BOND CZ3 HZ31
BOND CZ3 HZ32
BOND CZ3 OH3
BOND OH3 HH3
DOUBLE C O
IMPR N -C CA HN  C CA +N O
CMAP -C  N  CA  C   N  CA  C  +N
DONOR HN N
DONOR HH3 OH3
ACCEPTOR O C
IC -C   CA   *N   HN    1.3479 123.9300  180.0000 114.7700  0.9982 ! from C22 CYS
IC -C   N    CA   C     1.3479 123.9300  180.0000 105.8900  1.5202 ! from C22 CYS
IC N    CA   C    +N    1.4533 105.8900  180.0000 118.3000  1.3498 ! from C22 CYS
IC +N   CA   *C   O     1.3498 118.3000  180.0000 120.5900  1.2306 ! from C22 CYS
IC CA   C    +N   +CA   1.5202 118.3000  180.0000 124.5000  1.4548 ! from C22 CYS
IC N    C    *CA  CB    1.4533 105.8900  121.7900 111.9800  1.5584 ! from C22 CYS
IC N    C    *CA  HA    1.4533 105.8900 -116.3400 107.7100  1.0837 ! from C22 CYS
IC N    CA   CB   SG    1.4533 111.5600  180.0000 113.8700  1.8359 ! from C22 CYS
IC SG   CA   *CB  HB1   1.8359 113.8700  119.9100 107.2400  1.1134 ! from C22 CYS
IC SG   CA   *CB  HB2   1.8359 113.8700 -125.3200 109.8200  1.1124 ! from C22 CYS
```

```
IC CA   CB   SG   CD    1.5460 110.2800  180.0000  98.9400  1.8180 ! from C22 MET CB-CG-SD-CE, S-CG301 updated
IC CB   SG   CD   CE1   1.8180  95.0000  180.0000 114.5000  1.5380 ! CT2-S,       CT2-S-CG301,       CT2-S-CG301-CG331,       S-CG301-CG331,     CG301-CG331
IC CE1  SG   *CD  CE2   1.5380 114.5000  120.0000 114.5000  1.5380 ! CG331-CG301, CG331-CG301-S,     CG331-S-CG301-CG321,     S-CG301-CG321,     CG301-CG321
IC CE1  SG   *CD  CE3   1.5380 114.5000 -120.0000 114.5000  1.5380 ! CG331-CG301, CG331-CG301-S,     CG331-S-CG301-CG321,     S-CG301-CG321,     CG301-CG321
IC SG   CD   CE1  HE11  1.8180 114.5000    0.0000 110.1000  1.1110 ! S-CG301,     S-CG301-CG331,     S-CG301-CG331-HGA3,      CG301-CG331-HGA3,  CG331-HGA3
IC HE11 CD   *CE1 HE12  1.1110 110.1000  120.0000 110.1000  1.1110 ! HGA3-CG331,  HGA3-CG331-CG301,  HGA3-CG301-CG331-HGA3,   CG301-CG331-HGA3,  CG331-HGA3
IC HE11 CD   *CE1 HE13  1.1110 110.1000 -120.0000 110.1000  1.1110 ! HGA3-CG331,  HGA3-CG331-CG301,  HGA3-CG301-CG331-HGA3,   CG301-CG331-HGA3,  CG331-HGA3
IC SG   CD   CE2  CZ2   1.8180 114.5000    0.0000 113.5000  1.5300 ! S-CG301,     S-CG301-CG321,     S-CG301-CG321-CG321,     CG301-CG321-CG321, CG321-CG321
IC CZ2  CD   *CE2 HE21  1.5300 113.5000  120.0000 110.1000  1.1110 ! CG321-CG321, CG321-CG321-CG301, CG321-CG301-CG321-HGA2,  CG301-CG321-HGA2,  CG321-HGA2
IC CZ2  CD   *CE2 HE22  1.5300 113.5000 -120.0000 110.1000  1.1110 ! CG321-CG321, CG321-CG321-CG301, CG321-CG301-CG321-HGA2,  CG301-CG321-HGA2,  CG321-HGA2
IC CD   CE2  CZ2  CH2   1.5380 113.5000    0.0000 115.0000  1.5280 ! CG301-CG321, CG301-CG321-CG321, CG301-CG321-CG321-CG331, CG321-CG321-CG331, CG321-CG331
IC CH2  CE2  *CZ2 HZ21  1.5280 115.0000  120.0000 110.1000  1.1110 ! CG331-CG321, CG331-CG321-CG321, CG331-CG321-CG321-HGA2,  CG321-CG321-HGA2,  CG321-HGA2
IC CH2  CE2  *CZ2 HZ22  1.5280 115.0000 -120.0000 110.1000  1.1110 ! CG331-CG321, CG331-CG321-CG321, CG331-CG321-CG321-HGA2,  CG321-CG321-HGA2,  CG321-HGA2
IC CE2  CZ2  CH2  HH21  1.5300 115.0000    0.0000 110.1000  1.1110 ! CG321-CG321, CG321-CG321-CG331, CG321-CG321-CG331-HGA3,  CG321-CG331-HGA3,  CG331-HGA3
IC HH21 CZ2  *CH2 HH22  1.1110 110.1000  120.0000 110.1000  1.1110 ! HGA3-CG331,  HGA3-CG331-CG321,  HGA3-CG321-CG331-HGA3,   CG321-CG331-HGA3,  CG331-HGA3
IC HH21 CZ2  *CH2 HH23  1.1110 110.1000 -120.0000 110.1000  1.1110 ! HGA3-CG331,  HGA3-CG331-CG321,  HGA3-CG321-CG331-HGA3,   CG321-CG331-HGA3,  CG331-HGA3
IC SG   CD   CE3  CZ3   1.8180 114.5000    0.0000 113.5000  1.5300 ! S-CG301,     S-CG301-CG321,     S-CG301-CG321-CG321,     CG301-CG321-CG321, CG321-CG321
IC CZ3  CD   *CE3 HE31  1.5300 113.5000  120.0000 110.1000  1.1110 ! CG321-CG321, CG321-CG321-CG301, CG321-CG301-CG321-HGA2,  CG301-CG321-HGA2,  CG321-HGA2
IC CZ3  CD   *CE3 HE32  1.5300 113.5000 -120.0000 110.1000  1.1110 ! CG321-CG321, CG321-CG321-CG301, CG321-CG301-CG321-HGA2,  CG301-CG321-HGA2,  CG321-HGA2
IC CD   CE3  CZ3  OH3   1.5380 113.5000    0.0000 110.1000  1.4200 ! CG301-CG321, CG301-CG321-CG321, CG301-CG321-CG321-OG311, CG321-CG321-OG311, CG321-OG311
IC OH3  CE3  *CZ3 HZ31  1.4200 110.1000  120.0000 110.1000  1.1110 ! OG311-CG321, OG311-CG321-CG321, OG311-CG321-CG321-HGA2,  CG321-CG321-HGA2,  CG321-HGA2
IC OH3  CE3  *CZ3 HZ32  1.4200 110.1000 -120.0000 110.1000  1.1110 ! OG311-CG321, OG311-CG321-CG321, OG311-CG321-CG321-HGA2,  CG321-CG321-HGA2,  CG321-HGA2
IC CE3  CZ3  OH3  HH3   1.5300 110.1000    0.0000 106.0000  0.9600 ! CG321-CG321, CG321-CG321-OG311, CG321-CG321-OG311-HGP1,  CG321-OG311-HGP1,  OG311-HGP1

END

read param card flex append
* Parameters generated by analogy by
* CHARMM General Force Field (CGenFF) program version 2.2.0
*

! Penalties lower than 10 indicate the analogy is fair; penalties between 10
! and 50 mean some basic validation is recommended; penalties higher than
! 50 indicate poor analogy and mandate extensive validation/optimization.

BONDS
CG301  S       198.00     1.8180 ! C3H , from CG321 SG311, penalty = 12, SG311 replaced by S

ANGLES
CG321  CG301  S        58.00    114.50 ! C3H , from CG321 CG321 SG311, penalty = 12,  SG311 replaced by S
CG331  CG301  S        58.00    114.50 ! C3H , from CG331 CG321 SG311, penalty = 12,  SG311 replaced by S
CG301  S      CT2      34.00     95.00 ! C3H , from CG321 SG311 CG321, penalty = 1.8, SG311 replaced by S, CG321 replaced by CT2

DIHEDRALS
S      CG301  CG321  CG321      0.1950  3     0.00 ! C3H , from CG321 CG321 CG321 SG311,  penalty = 12,   SG311 replaced by S
S      CG301  CG321  HGA2       0.0100  3     0.00 ! C3H , from SG311 CG321 CG321 HGA2,   penalty = 12,   SG311 replaced by S
S      CG301  CG331  HGA3       0.1600  3     0.00 ! C3H , from SG311 CG321 CG331 HGA3,   penalty = 12,   SG311 replaced by S
CG321  CG301  S      CT2        0.2400  1   180.00 ! C3H , from CG321 CG321 SG311 CG321,  penalty = 12,   SG311 replaced by S, CG321 replaced by CT2
CG321  CG301  S      CT2        0.3700  3     0.00 ! C3H , from CG321 CG321 SG311 CG321,  penalty = 12,   SG311 replaced by S, CG321 replaced by CT2
CG331  CG301  S      CT2        0.2400  1   180.00 ! C3H , from CG321 CG321 SG311 CG321,  penalty = 12.9, SG311 replaced by S, CG321 replaced by CT2
CG331  CG301  S      CT2        0.3700  3     0.00 ! C3H , from CG321 CG321 SG311 CG321,  penalty = 12.9, SG311 replaced by S, CG321 replaced by CT2
CG301  CG321  CG321  CG331      0.1950  3     0.00 ! C3H , from CG301 CG321 CG321 CG321,  penalty = 0.9
CG301  CG321  CG321  OG311      0.1950  3     0.00 ! C3H , from CG321 CG321 CG321 OG311,  penalty = 1.8
CT1    CT2    S      CG301      0.1950  3     0.00 ! C3H , from CG324 CG321 SG311 CG321,  penalty = 2.4,  SG311 replaced by S, CG321 replaced by CT2, CG314 replaced
by CT1
HA     CT2    S      CG301      0.2800  3     0.00 ! C3H , from HGA2  CG321 SG311 CG321,  penalty = 1.8,  SG311 replaced by S, CG321 replaced by CT2, HGA2  replaced
by HA

IMPROPERS

END
RETURN
```

Appendix II: CHARMM hybrid-solvent CpHMD parameters for N-terminus

```
* Parameter file for HPHMD
* Charges generated based on C22 PRES NTER & NNEU
* Parameterized by Zhi (Shane) Yue, Gregory A. Voth lab, UChicago, February 2019
* Syntax:
* RESNAME EXPT_PKA PARA PARB BARR
*  ATOM_NAME PROT_CHARGE UNPROT_CHARGE PROT_RAD UNPROT_RAD
*

! -------------------------
! Experimental pKa: 8.00 +/- 0.03
! Measured at 298.15 K, 1 atm, 0.1 M NaCl using pentapeptide NH3+-AAAAA-CONH2
! Check Thurlkill et al., Protein Sci.,2006, 15(5): 1214-1218
! -------------------------
! 5 independent pH-REX hybrid-solvent CpHMD runs
!
! CHARMM version: c42b2
! C22+CMAP CHARMM Force Field
! Truncated octahedron box, 39 Angstrom
! 1 NTER-AAAAA-CT2, 1463 CHARMM-modified TIP3P waters, no ions
! 300.00 K (Nosé-Hoover), 1 atm (Langevin piston)
!
! GB settings:
! gbsw hybrid sgamma 0.000 nang 50 conc 0.000 temp 300.00 -
!      sele SOLUTE end
!
! Nonbonded settings:
! nbond elec atom cdie vdw vatom vswitch -
!       ctonnb 10.0 ctofnb 12.0 cutnb 14.0 cutim 16.0 -
!       ewald pmew fftx 40 ffty 40 fftz 40 kappa 0.34 spline order 6 -
!       inbfrq -1 imgfrq -1
! -------------------------
! pKa(calc) = 7.97 +/- 0.10
! n  (calc) = 0.93 +/- 0.01
!
! Fraction of mixed state: <= 24%
! -------------------------
! To use with other terminal residues such as NH3+-XXX-,
! make a copy of the entry below and replace “Ala” with “XXX”
! -------------------------

NTALA 8.00 -78.5636 1.04627435067512 1.5
N    -0.30  -0.97         !     HT1                     HT1
HT1   0.33   0.33         !    /                       /
HT2   0.33   0.33         ! --N--HT3* (+)  <=====>  --N
HT3   0.33   0.00 1.0 0.0 !    \                       \
                           !     HT2                     HT2
END
```

Appendix III: CHARMM hybrid-solvent CpHMD parameters for C-terminus

```
* Parameter file for HPHMD
* Charges generated based on C22 PRES CTER & CNEU
* Parameterized by Zhi (Shane) Yue, Gregory A. Voth lab, UChicago, February 2019
* Syntax:
* RESNAME EXPT_PKA PARA PARB BARR
*  ATOM_NAME PROT_CHARGE UNPROT_CHARGE PROT_RAD UNPROT_RAD
* RESNAME EXPT_PKA PARA PARB BARR PARA10 PARB10 BARTAU
*                  R1   R2   R3   R4     R5     R6
*  ATOM_NAME PROT_CHARGE UNPROT_CHARGE PROT_RAD UNPROT_RAD
*

! -------------------------
! Experimental pKa: 3.67 +/- 0.03
! Measured at 298.15 K, 1 atm, 0.1 M NaCl using pentapeptide CH3CONH-AAAAA-COO-
! Check Thurlkill et al., Protein Sci.,2006, 15(5): 1214-1218
! -------------------------
! 5 independent pH-REX hybrid-solvent CpHMD runs
!
! CHARMM version: c42b2
! C22+CMAP CHARMM Force Field
! Truncated octahedron box, 41 Angstrom
! 1 ACE-AAAAA-CTRP2, 1772 CHARMM-modified TIP3P waters, no ions
! 300.00 K (Nosé-Hoover), 1 atm (Langevin piston)
!
! GB settings:
! gbsw hybrid sgamma 0.000 nang 50 conc 0.000 temp 300.00 -
!      sele SOLUTE end
!
! Nonbonded settings:
! nbond elec atom cdie vdw vatom vswitch -
!       ctonnb 10.0 ctofnb 12.0 cutnb 14.0 cutim 16.0 -
!       ewald pmew fftx 40 ffty 40 fftz 40 kappa 0.34 spline order 6 -
!       inbfrq -1 imgfrq -1
! -------------------------
! pKa(calc) = 3.66 +/- 0.06
! n  (calc) = 0.93 +/- 0.01
!
! Fraction of mixed state: <= 30%
! -------------------------
! To make C-terminus titratable in CpHMD, apply patch "PRES CTRP2"
! Check documentation phmd.doc for CpHMD topology & parameter
! -------------------------
! To use with other terminal residues such as -XXX-COO-,
! make a copy of the entry below and replace "Ala" with "XXX"
! -------------------------

CTALA1 3.67 -90.202 0.183009 2.0
C      0.72   0.34         !    OT1-HC1*              OT1 (-)
OT1   -0.61  -0.67         !   /                     /
OT2   -0.55  -0.67         ! --C          <=====>  --C
HC1    0.44   0.00 1.0 0.0 !   \\                    \\
HC2    0.00   0.00         !    OT2                   OT2

CTALA2 3.67 -90.2589 0.182468 2.0 -29.8537 0.498661
            -26.7031 56.9114 -29.9575 0.490501 -90.1841 0.180540580420495
C      0.72   0.34         !    OT1                   OT1
OT1   -0.55  -0.67         !   //                    //
OT2   -0.61  -0.67         ! --C          <=====>  --C
HC1    0.00   0.00         !   \                     \
HC2    0.44   0.00 1.0 0.0 !    OT2-HC2*              OT2 (-)

END
```