Supporting Information

Parker et al. 10.1073/pnas.1710727114

SI Materials and Methods

Reconstitution of $PepT_{So}$. For reconstitution, $PepT_{So}$ purified in the detergent decyl maltoside (19) was mixed in a 50:1 ratio (lipid:protein) with preformed lipid vesicles composed of POPE and POPG (in a 3:1 ratio). After incubation at room temperature for 30 min the protein–lipid mix was diluted into a large volume of reconstitution buffer (50 mM potassium phosphate, pH 7.0) and proteoliposomes were harvested by ultracentrifugation (>200,000 × *g*) for 2 h. Pelleted proteoliposomes were resuspended at 0.5 μg/μL (protein) and dialyzed extensively against reconstitution buffer (24 h with two changes of buffer). Proteoliposomes were recovered and subjected to three rounds of freezing and thawing before storage at −80 °C.

Details for comparison between WT and variant proteins. The experiments were set up as per the general outline using 3 μg of protein per time point and a final concentration of 0.2 mM dialanine. For each sample, at least four time points were taken and this was repeated in triplicate. For simplicity, the amount of peptide transported after 2 min is shown in Fig. 1*A*.

Details of counterflow experiments. Proteoliposomes were prepared as above but the buffer used consisted of 50 mM Tris, pH 7.5. The internal buffer contained 2 mM of the substrate of interest which was loaded into the liposomes by freezing and thawing five times. The external buffer was 50 mM Tris, pH 7.5, with 100 μM dialanine containing [^{3}H]dialanine as a reporter. For simplicity, the amount of peptide transported after 5 min is shown in Fig. 1*A*.

Details for pH optimum experiments. The experiments were set up as per the general outline, to cover the pH range required bis-Tris and Trizma were used and a final concentration of 0.2 mM dialanine. At least eight time points (3 μg of protein) were taken for each pH and the rate calculated from the linear portion of transport; this was repeated in triplicate to obtain a mean rate.

Details of steady-state accumulation of di- and trialanine. The experiments were set up as per the general outline, and a final concentration of 0.5 mM di- or trialanine was used with fixed amount of [^{3}H]dialanine or [^{14}C]trialanine in the outside buffer. For $PepT_{So}$ experiments, 2 μg of protein was used per time point. To drive uptake with only a pH gradient, an acetate-induced pH gradient system was used (36, 38) (dashed lines in Fig. 2*B*), rather than those made from different buffers as implemented in previous assays for $PepT_{St}$ (13). Therefore, the outside buffer used was 10 mM Tris, pH 7.5, 2 mM $MgSO_4$, and 120 mM KCl with valinomycin to dissipate any charge buildup. To drive uptake with a pH gradient and membrane potential ($\Delta\mu H^+$) (solid lines in Fig. 2*B*) the following outside buffer was used: 10 mM Tris, pH 7.5, 2 mM $MgSO_4$, and 120 mM NaCl with valinomycin to generate a potassium diffusion gradient and hence a membrane potential in addition to the acetate-induced pH gradient. For comparison with $PepT_{St}$, the optimal buffers for this protein were used, which consisted of internal buffer 10 mM potassium phosphate, pH 6.5, 2 mM $MgSO_4$, and 110 mM potassium acetate and outside buffer 10 mM sodium phosphate, pH 6.5, 2 mM $MgSO_4$, and 110 mM sodium chloride with valinomycin to drive transport with $\Delta\mu H^+$. One microgram of protein was used per time point as $PepT_{St}$ shows higher activity than $PepT_{So}$.

Details of competition experiments. The experiments were set up as per the general outline; a final concentration of 40 μM di-alanine was used with 0.5 mM cold competitor. For each competing substrate at least four time points were taken, and this was repeated in triplicate. For simplicity, the amount of peptide transported after 2 min as a percent of no competitor is shown.

Generation of $PepT_{Xc}$ Truncations. A truncated form of $PepT_{Xc}$ was generated by cloning residues 1–514 ($PepT_{Xc}$M514) into the NdeI/BamHI sites of pWaldoD; expression, purification, and reconstitution were carried out as detailed for the WT protein. We noticed no difference in protein expression or stability.

Classical MD. All of the MD simulations were initiated from the high-resolution crystal structure of $PepT_{Xc}$ by embedding the protein into an 82-Å × 82-Å lipid bilayer composed of a 3:1 ratio of POPE and POPG lipids and then solvating the system with a 25-Å layer of TIP3P water on either side of the membrane. To compensate for the excess charge and mimic the experimental solution, 28 sodium cations were added to the water. All residues other than Glu425 and Glu26 were assigned standard protonation states based on pK_a calculations performed by PROPKA. The CHARMM-CMAP and CHARMM 36 force field was employed to describe the protein and lipid interactions. Particle mesh Ewald with a cutoff of 12.0 Å and a precision of 10^{-5} was used to treat the electrostatic interactions. A cutoff of 12.0 Å was chosen for the Lennard-Jones potentials with a switching function starting at 10.0 Å. At first, the alpha carbons and other protein heavy atoms were constrained by 40 kJ/mol/Å^2 and 20 kJ/mol/Å^2 harmonic restraints, respectively, to allow the lipids, water molecules, and ions to relax. The force constants were gradually reduced to zero during 37.5 ns of equilibration in the constant NPT ensemble. In the simulation with Glu425 protonated, the system was equilibrated for another 250 ns in the constant NPT ensemble at 303.15 K and under 1 atm for better density equilibration. The positions and velocities of the particles were updated using the leap-frog algorithm with a 2-fs long time step. All of the bonds involving hydrogen atoms were restrained using the LINC algorithm. The temperature and the pressure were controlled by a modified Berendsen thermostat and the semiisotropic Parrinello–Rahman barostat respectively. Then equilibration in the constant NVT ensemble for an extra 100 ns was followed by a production run for 100 ns. The velocity Verlet integrator and the Nosé–Hoover chain temperature coupling were used for the constant NVT simulations. A similar simulation protocol was used for two duplicates of the His67 protonated system. In one simulation, 500 ns of constant NPT followed equilibration and then 1.1 μs of constant NVT simulation was collected and analyzed. In a separate simulation, 550 ns of constant NPT MD followed equilibration. The reported structural analysis is based on the second simulation, but both simulations resulted in almost identical conformational changes, verifying that this transition is reproducible. All of the classical MD simulations were conducted in the GROMACS package.

QM/MM Umbrella Sampling. Hybrid QM/MM multiscale MD simulations were performed in the CP2K package. To capture the delocalized nature of an excess proton charge defect, the following definition for the CEC was used:

$$\vec{r}_{CEC} = \sum_{i=1}^{N_H} \vec{r}^{\,H_i} - \sum_{j=1}^{N_X} w^{X_j} \vec{r}^{\,X_j} - \sum_{i=1}^{N_H} \sum_{j=1}^{N_X} f_{sw}\left(d_{X_jH_j}\right)\left(\vec{r}^{\,H_i} - \vec{r}^{\,X_j}\right), \quad \text{[S1]}$$

where the second sum goes over all of the heavy atoms in the QM region (i.e., the water and the carboxylic oxygens); the first sum goes over all of the hydrogen atoms connected to the heavy atoms, while the last term contains all of the pairs of the hydrogen atoms and the heavy atoms. The weighting factor w^{X_j}, which is the coordination

number of hydrogens to oxygen in the deprotonated state of each species, was set to 2 for water molecules and 0 for carboxyl oxygens. The switching function f_{sw} is a function of the distance between hydrogen atoms and oxygen atoms and has the functional form of $f_{sw}(x) = 1/(1+\exp[(x-r_{sw})/d_{sw}])$, where the parameters r_{sw} and d_{sw} are 1.25 Å and 0.04 Å.

The proton transport reaction coordinate between Asp322 and Glu425 is characterized by a projection collective variable defined by a ratio of distances between the CEC and the carboxyl oxygen atoms in Asp322 and Glu425. In this definition, the carboxyl oxygens on Asp322 and Glu425 are denoted as O_D and O_E, respectively. We define the following weights:

$$w_1 = \frac{1}{e^{(r_{D1}-r_{D2})/\sigma}+1}, w_2 = \frac{1}{e^{(r_{D2}-r_{D1})/\sigma}+1},$$

$$w_3 = \frac{1}{e^{(r_{E1}-r_{E2})/\sigma}+1}, w_4 = \frac{1}{e^{(r_{E2}-r_{E1})/\sigma}+1},$$

where σ is set to 0.025 Å; r_{D1}, r_{D2}, r_{E1}, and r_{E2} are the distances between the CEC and the closer (denoted 1) and further (denoted 2) carboxyl oxygens in Asp322 and Glu425, respectively. The distances between the CEC and O_D, O_E were thus approximated as $r_D = w_1 r_{D1} + w_2 r_{D2}$ and $r_E = w_3 r_{E1} + w_4 r_{E2}$ and the distance between O_D and O_E was approximated as $r_{DE} = w_1 w_3 r_{D1E1} + w_1 w_4 r_{D1E2} + w_2 w_3 r_{D2E1} + w_2 w_4 r_{D2E2}$, where $r_{D1E1}, r_{D1E2}, r_{D2E1}$ and r_{D2E2} are the distances between the corresponding oxygen atoms. The collective variable, denoted as ξ_R, is then defined as the ratio of the length of the projection of the vector connecting O_D with the CEC onto the vector connecting O_D and O_E over the distance between O_D and O_E. Based on the approximations made above, ξ_R was calculated by the law of cosines:

$$\xi_R = \frac{r_D^2 + r_{DE}^2 - r_E^2}{2\, r_{DE}^2}. \quad \textbf{[S2]}$$

With this definition, our CV is zero when Asp322 is protonated and one when Glu425 is protonated.

The QM/MM MD simulations were started with structures extracted from either MS-RMD simulations or from an adjacent umbrella window. The equilibrated structure from the MD simulations with Glu425 protonated was chosen to initialize the MS-RMD simulations and umbrella sampling was used to equilibrate the system with the excess proton moving along the reaction coordinate for 250 ps to 1 ns depending on the window. The side chains of Asp322, Glu425, Arg37, Lys324, and all of the water molecules in three solvation shells of the carboxyl groups were treated as quantum (QM) atoms. A QM box size of at least 6 Å larger than the actual ranges of QM atoms was assigned to make sure that wall potentials restraining atoms in the QM box would have minor effects on the QM atoms. The electronic structure of the QM region was described using DFT with the BLYP functional and the D3 dispersion correction (39). To save computational cost, the Goedecker–Teter–Hutter pseudopotentials were used to describe the core electrons. The Gaussian TZV2P basis set and plain wave basis set with a cutoff of 360 Ry were used to expand the Kohn–Sham orbitals and the electron density, respectively. The classical part of the system was described by the CHARMM36 force field and the QM and MM electrostatic coupling was treated by the Gaussian expansion of the electrostatic potential algorithm. Constant temperature ensembles were simulated using velocity Verlet to integrate the equations of motion and a Nosé–Hoover chain to control the temperature at 303.15 K. Forty-six windows with a spacing of ~0.025 in ξ_R were simulated for ~15 ps after ~5-ps-long equilibration and the collective variable was collected at every time step (0.5 fs).

The PMF was constructed by dTRAM, which does not require the assumption of global equilibrium and the independency between frames, as WHAM does. A bootstrap of 200 times of resampling was performed to estimate the 68% confidence interval.

The proton transport rate constant was estimated using transition state theory (40):

$$k_{rxn} = \frac{\omega_0}{2\pi}\exp\left(-\frac{\Delta F^{\ddagger}}{k_B T}\right), \quad \textbf{[S3]}$$

where $\Delta F^{\ddagger}$ is the free energy barrier height for the proton transfer reaction from Asp322 to Glu425, and ω_0 is the frequency of the principle oscillation of the reactant, given by

$$\omega_0 = \sqrt{\frac{1}{m_{eff}}\frac{\partial^2 PMF(\xi_R)}{\partial \xi_R^2}}, \quad \textbf{[S4]}$$

where the second-order derivative was evaluated at the local minimum representing the reactant (specially $\xi_R = 0$). In the above equation, the effective mass corresponding to the collective variable was calculated according to the equipartition theorem:

$$m_{eff} = \frac{k_B T}{v^2}, \quad \textbf{[S5]}$$

where the velocity of the collective variable, v, was calculated using the time series sampled at the minimum.

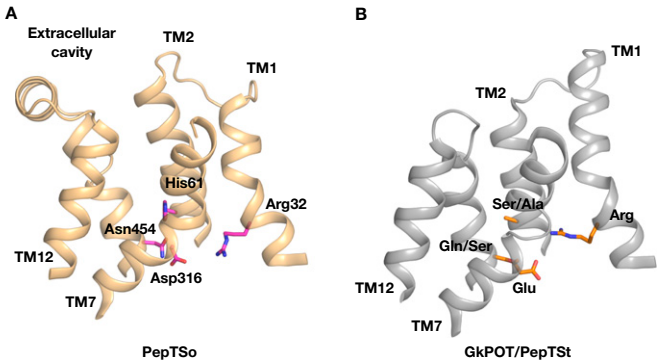


Fig. S1. Conservation of the TM2 histidine in mammalian and mammalian-like POT family transporters. The extracellular gate architecture for $PepT_{So}$ (*A*) (PDB ID code 2XUT) and $PepT_{St}$ (*B*) (PDB ID code 4D2D) are shown. In mammalian and mammalian-like POT family transporters there is always a His, Asp, Arg triad at the extracellular gate, whereas in non-mammalian-like members, such as GkPOT, $PepT_{St}$, and $PepT_{So2}$ the extracellular gate region is different.

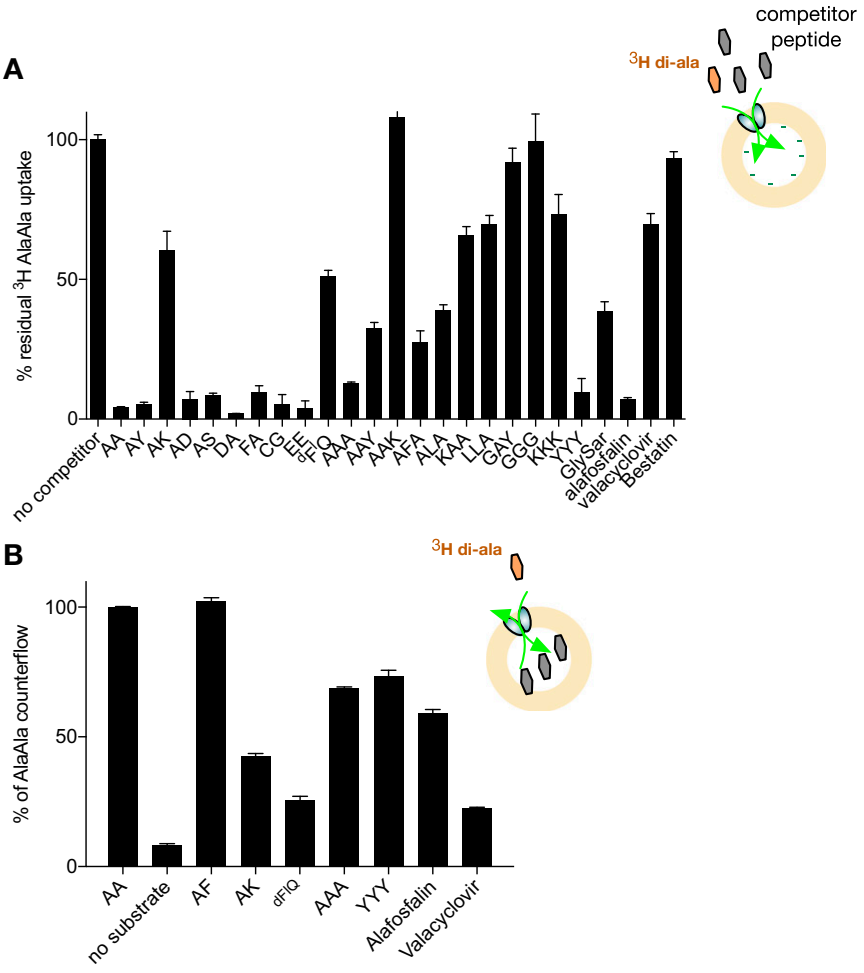


Fig. S2. $PepT_{So}$ can transport a wide range of di- and tripeptides as well as prodrugs and β -lactam antibiotics. (*A*) A wide range of di- and tripeptides as well as valacyclovir and bestatin can compete with dialanine for uptake by $PepT_{So}$ into proteoliposomes. (*B*) Counterflow activity of $PepT_{So}$ highlighting that valacyclovir is a substrate of this transporter. Using [^{3}H]dialanine as a reporter we can see that $PepT_{So}$ does indeed recognize numerous different peptide substrates containing different charges and sizes. There is indeed some specificity observed by this transporter with the hydrophobic-containing dipeptide being recognized very efficiently and the lysine-containing one very poorly, a similar substrate profile as observed for other POT family members including the human PepT1 and PepT2. We also see that numerous different tripeptides can be recognized and, indeed, transported, for example one of the largest natural substrates, the tripeptide Tyr-Tyr-Tyr. Indeed, also reporter substrates commonly used for studying mammalian peptide transporters are also recognized and transported, dFlQ (D-Phe-L-Gln) and alafosfalin, a phosphonodipeptide with antibacterial properties.

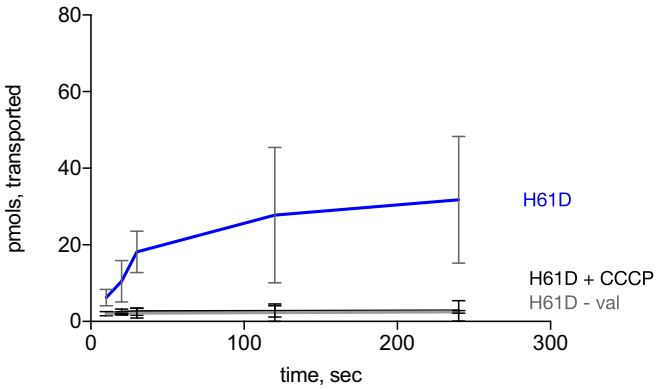
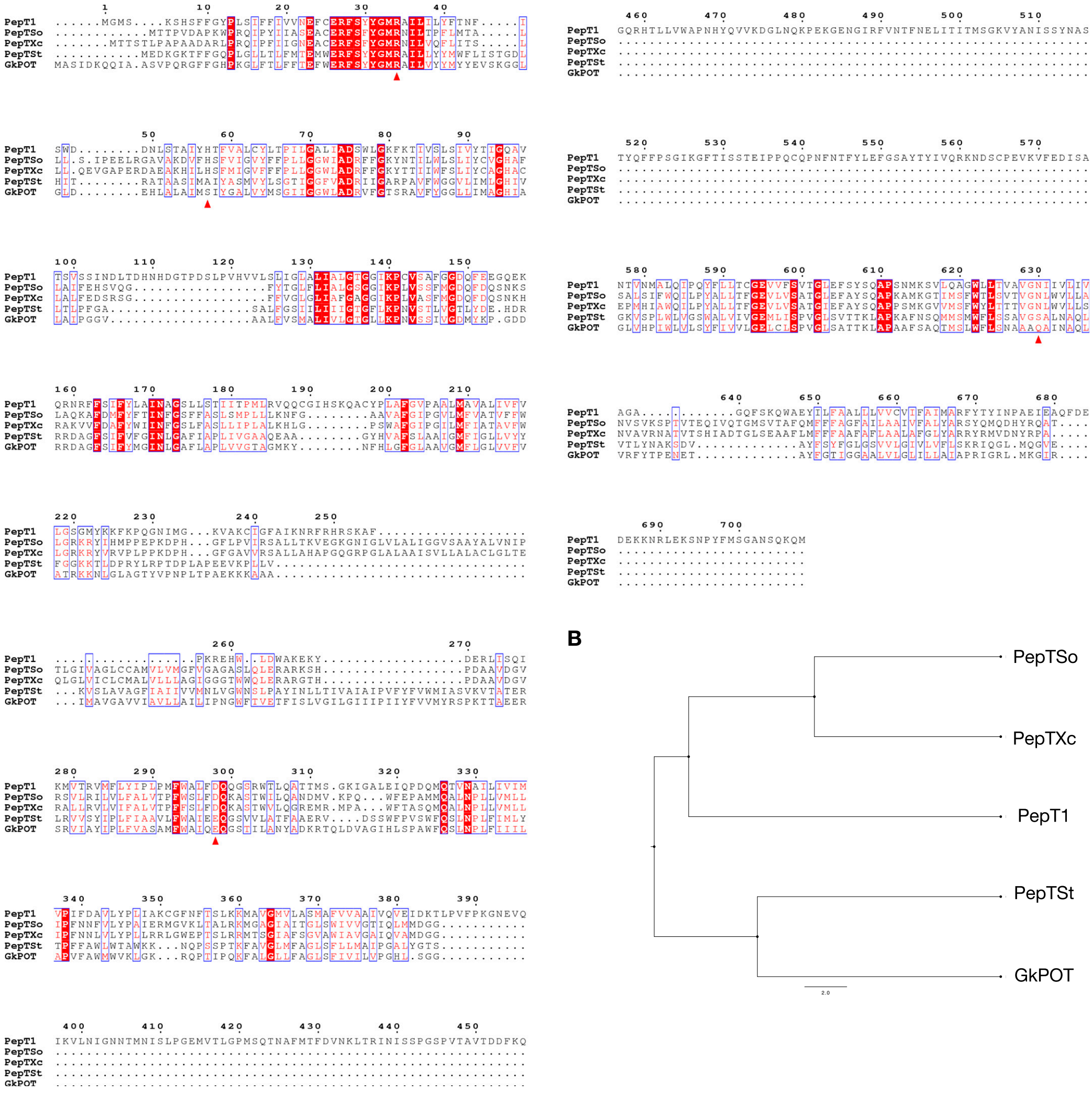


Fig. S3. Proton-driven uptake of dialanine over time for His61Asp, mutant. The His61Asp variant, while showing low activity compared with WT (Fig. 3*B*), does still transport dialanine.

A



B

Fig. S4. Conservation of the extracellular gate residues in mammalian and mammalian-like POT family transporters. (*A*) Sequence alignment using Clustal Omega of the $PepT_{So}$ (UniProt ID Q8EKT7) $PepT_{Xc}$ (UniProt ID Q3BVH5), $PepT_{St}$ (UniProt ID Q5M4H8), and GkPOT (UniProt ID Q5KYD1) with human PepT1 (UniProt ID P46059). The conserved extracellular gate residues are highlighted with red triangles. (*B*) Phylogenetic tree showing the average distance between the sequences using percent identity. The mammalian-like members cluster with PepT1.

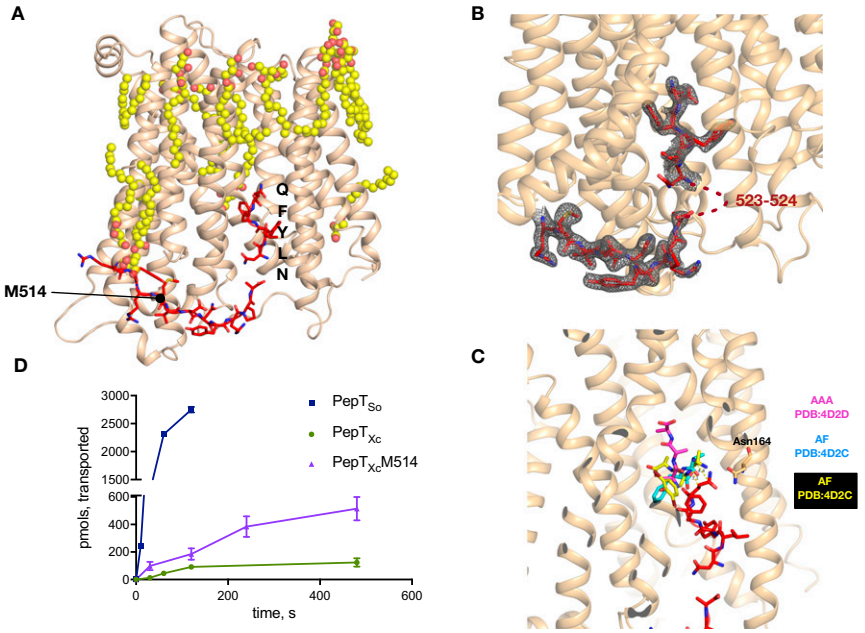


Fig. S5. Crystal structure of PepT$_{Xc}$. (*A*) Cartoon representation of PepT$_{Xc}$ showing the MAG9.9 lipid molecules. The C terminus of the protein is colored red and shown as sticks. The TEV protease recognition site can be seen clearly in the electron density maps in *B*. The truncation point (M514) used in *D* is marked. (*B*) The 2Fo − Fc electron density maps calculated from the refinement in Buster (v2.10.2), contoured at 1 sigma. It appears as though two residues are disordered in the structure (523–524) as these are not visible in the electron density maps. We think these may have "looped out," given the steric constraints imposed by the position of residues 522 and 525. (*C*) Overlay of the currently determined peptide complexes for POT family members PepT$_{St}$ and PepT$_{So2}$ onto PepT$_{Xc}$. The interaction of the terminal glutamine in the TEV recognition sequence with Asn164 in PepT$_{Xc}$ is clearly visible. The equivalent position to Asn164 in both PepT$_{St}$ and PepT$_{So2}$ was identified as forming part of the peptide-binding site, although different roles were ascribed to this side chain. (*D*) Transport activity of PepT$_{Xc}$ and a C-terminus truncation in a reconstituted liposome system are shown. Transport was driven by an inwardly directed proton electrochemical gradient ($\Delta\mu H^+$), inside negative and alkaline. In comparison with PepT$_{So}$ the transport activity of the full-length PepT$_{Xc}$ protein is lower. Truncation of the C terminus at Met-514 increases the activity of the protein, indicating that the flexible C terminus is likely to inhibit transport due to the interactions observed in the crystal structure.

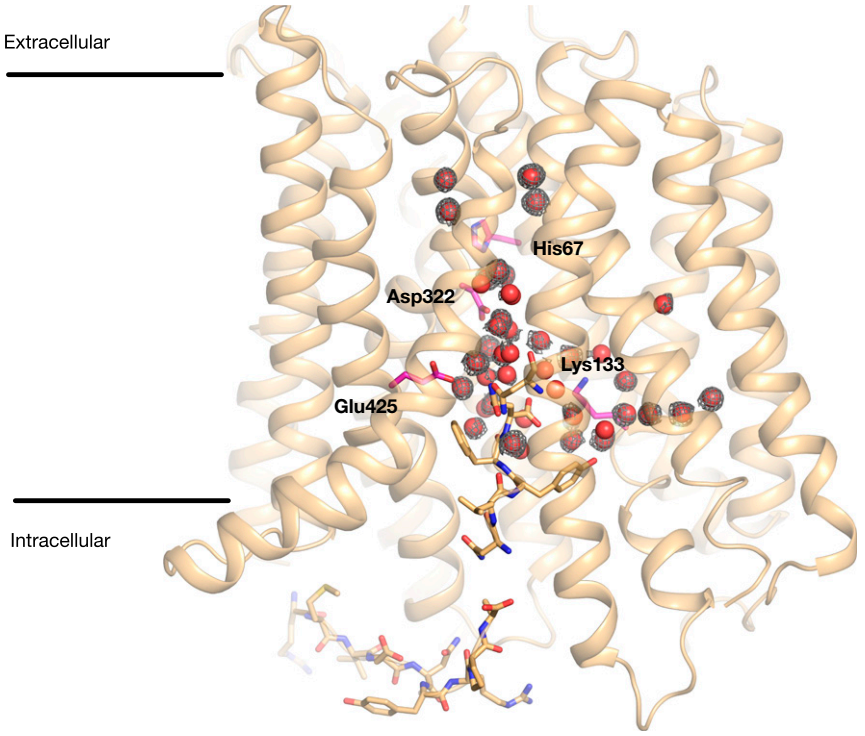


Fig. S6. Water molecules form networks of interactions within the binding site of $PepT_{Xc}$. Cartoon representation of $PepT_{Xc}$ showing the location of the water molecules in the central binding site. The refined 2Fo − Fc electron density map is shown around the water molecules contoured at 1 sigma.

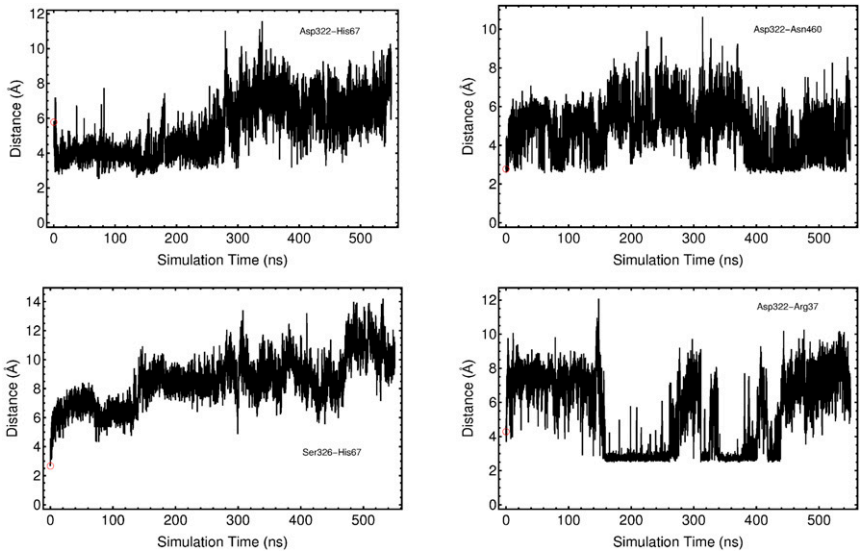


Fig. S7. Distances between several key residues in the His67 protonated simulation. All of the distances were calculated based on hydrogen binding moieties. The corresponding distances in the crystal structure are shown in red cycles. Changes in distances during system minimization and 37.5 ns of equilibration are not shown.

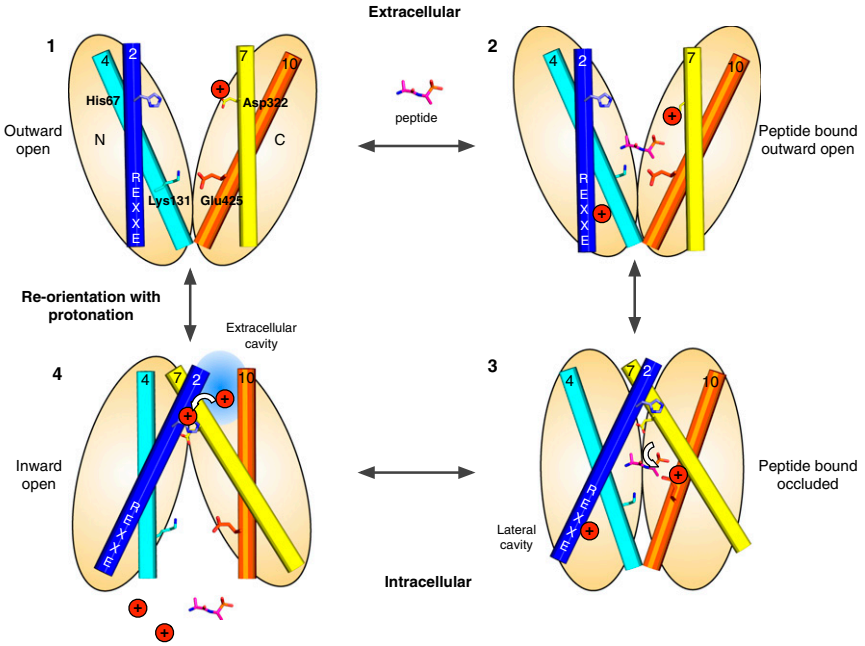


Fig. S8. A role for protonation in facilitating the transition from inward to outward facing states in the transport cycle. Transport is predominantly controlled through the opening and closing of two gates that sit on either side of a central binding site. The gates are made of helices 1,2,4,5 from the N-terminal bundle and 7,8,10,11 from the C-terminal bundle. Salt-bridge interactions are proposed to coordinate the open/closed state of the gates; shown here are His67-Asp322 and Lys131-Glu425. Sequential proton binding and release is proposed to regulated their open and closed state, thus driving the transport or peptides into the cell. The ExxER motif on TM1 is indicated. Our data suggest that in mammalian-like members the extracellular cavity created in the inward-facing state allows water to protonate the TM2 histidine from the extracellular side (step 4). This event destabilizes the extracellular gate, resulting in the reorientation of the transporter to the outward facing state (step 1). At this point the transporter is already protonated and able to receive further peptides for transport (steps 2–3). The white arrows indicate the proposed movement of protons, which we show is facilitated through ordered water networks.

Table S1. Data collection and statistics

Data collection	
Space group	C2
Cell dimensions	
a, b, c, Å	138.06, 65.23, 70.34
α, β, γ, °	90, 96.84, 90
Wavelength, Å	0.96859
Resolution,* Å	25.27–2.10
CC1/2, %	99.9 (78.2)
R_{merge}	9.1 (60.1)
R_{pim}	3.5 (35.4)
I/σI	15.2 (2.5)
Completeness, %	99.9 (99.9)
Redundancy	8.2 (4.6)
Refinement	
Resolution, Å	24.65–2.10
No. of reflections	36,377
R_{work}/R_{free}	18.09/21.35 (21.01/27.01)
ESD from Luzzati plot, Å	0.235
Mean B value, $Å^2$	45.29
Rmsd	
Bond lengths, Å	0.010
Bond angles, °	0.94
Ramachandran statistics, favored/outliers, %	98.04/0
MolProbity score	100th percentile

*Highest-resolution shell shown in parentheses.