Supporting information for:

Free Energy Landscape of the Complete Transport Cycle in a Key Bacterial Transporter

Balaji Selvam,[†] Shriyaa Mittal,[‡] and Diwakar Shukla[*,‡,†,¶]

†*Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL*

‡*Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL*

¶*Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL*

E-mail: diwakar@illinois.edu

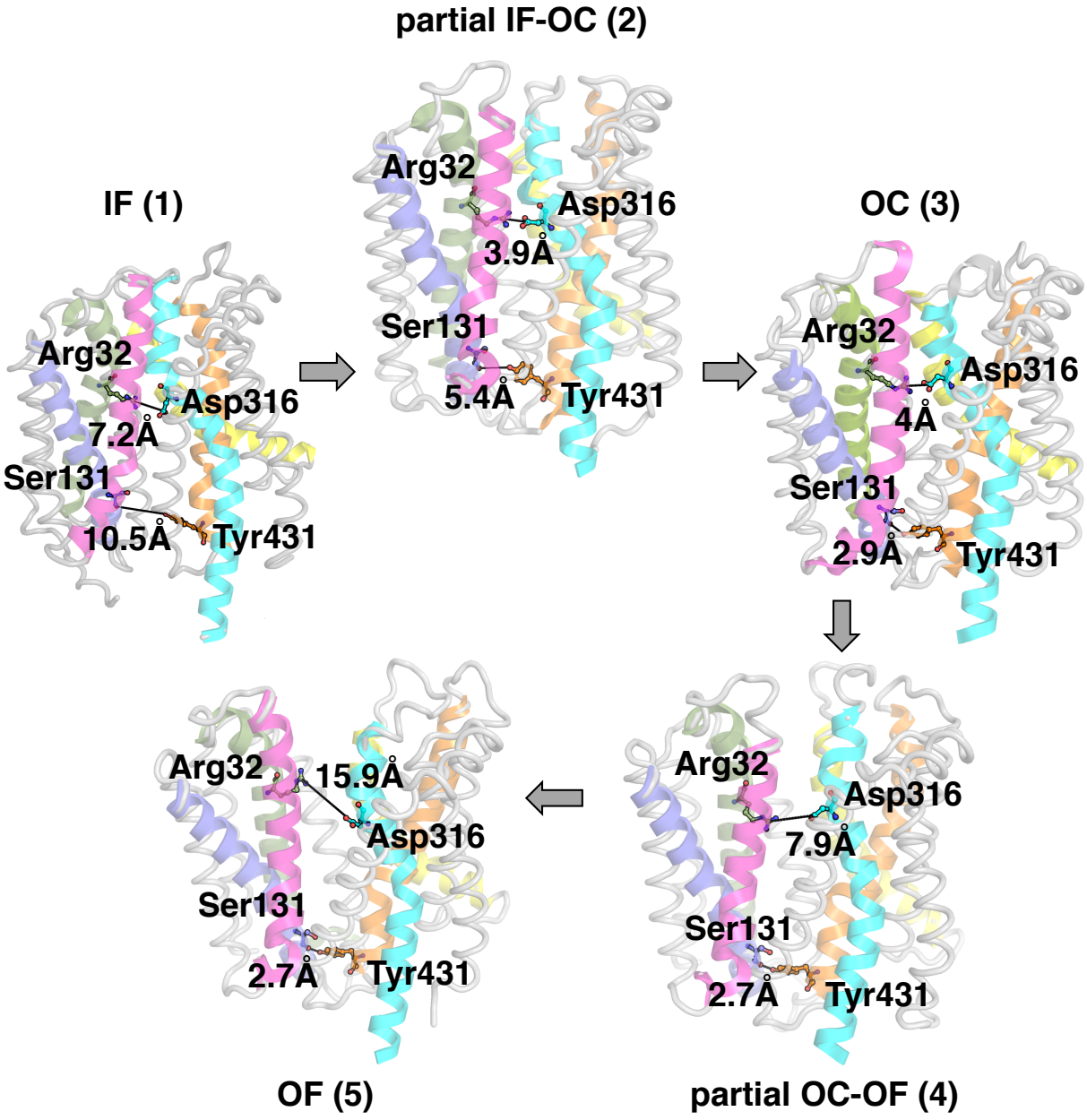

Figure S1: Most dominant conformational transition of $PepT_{So}$ from IF to OF, via partial IF-OC, OC and partial OC-OF states. The extracellular and intracellular gating residues and their distances in each state have been indicated. The numbers in brackets refer to the energy minima in the free-energy landscape.

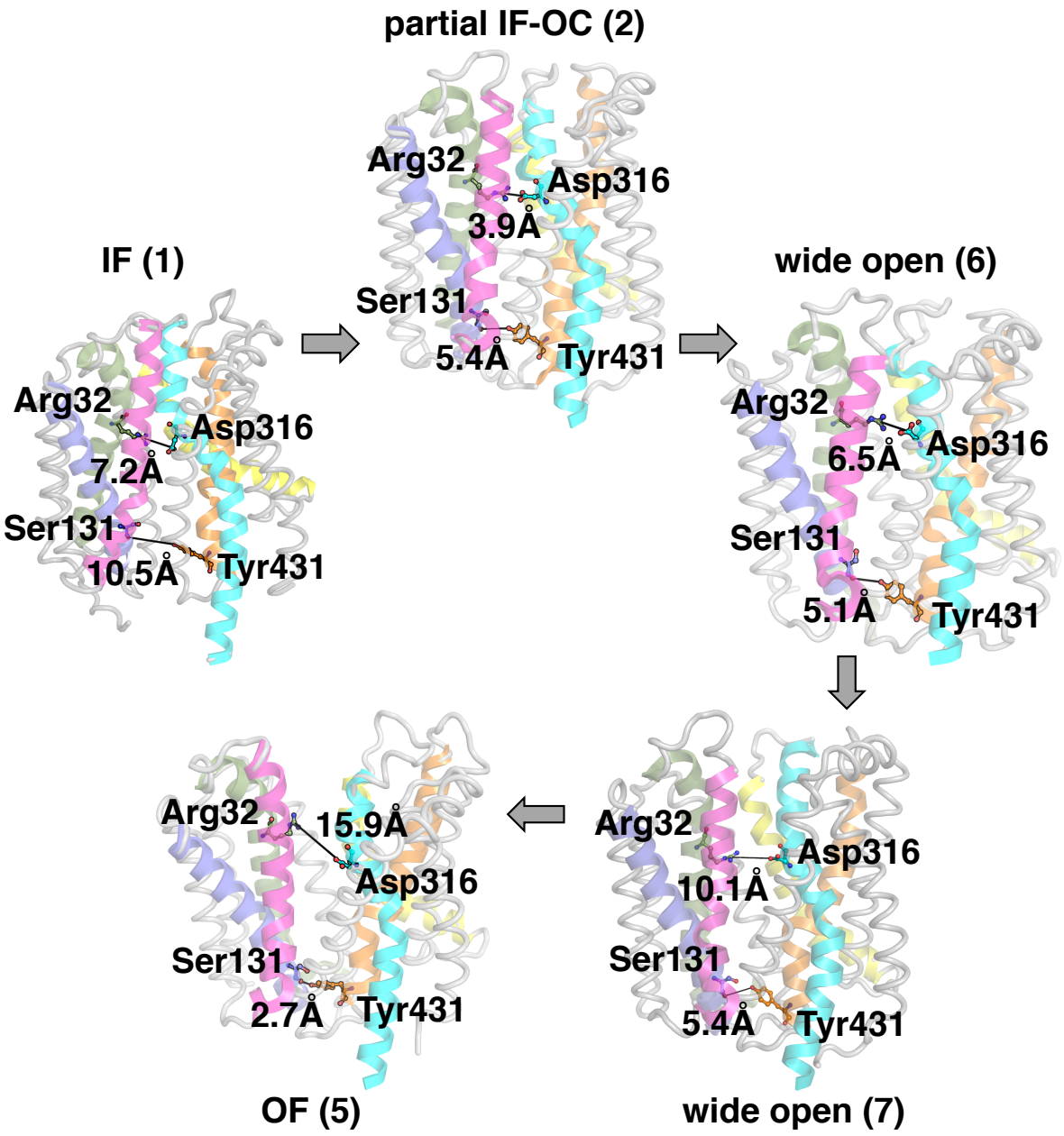

Figure S2: Another possible conformational transition of $\mathrm{PepT_{So}}$ from IF to OF, via partial IF-OC and wide open states. The extracellular and intracellular gating residues and their distances in each state have been indicated. The numbers in brackets refer to the energy minima in the free-energy landscape.

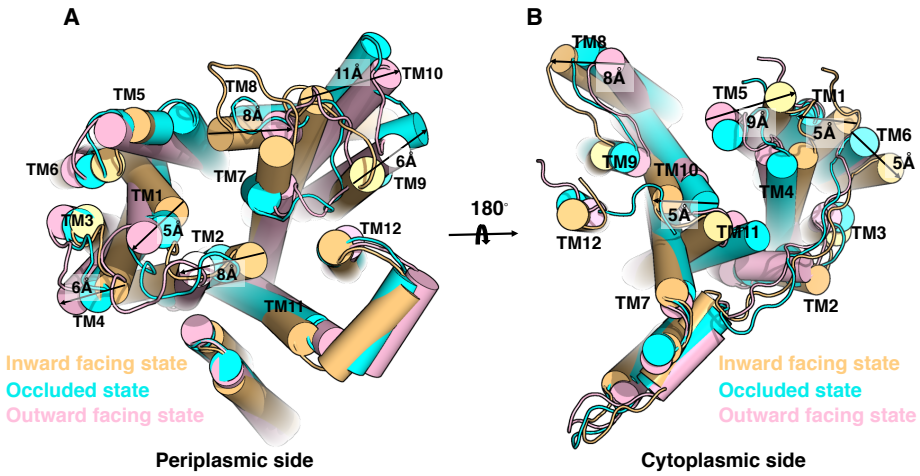

Figure S3: Conformational transition of helical tips of PepT$_{\mathbf{So}}$ IF (yellow), OC (cyan) and OF (magenta) states. Helical tips of three states show drastic changes in TM1, 2, 4, 7, 8, 9 and 10 at both (A) extracellular and (B) intracellular sides.

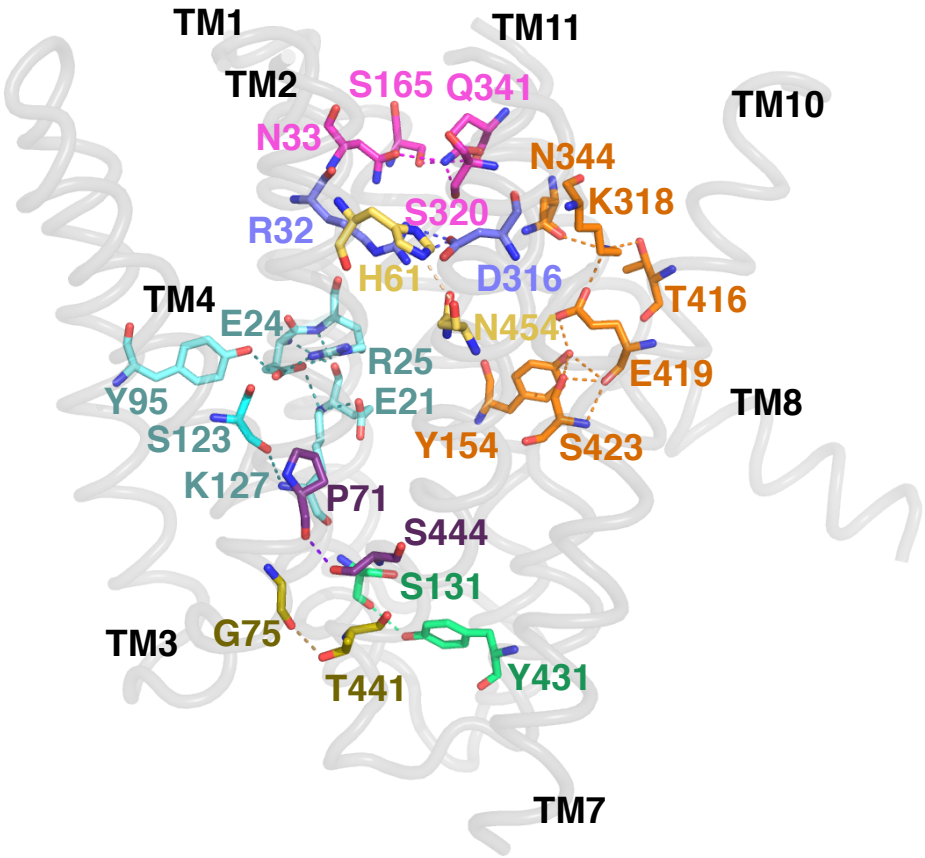

Figure S4: MD predicted PepT$_{\mathbf{So}}$ OC state. Detailed description of residues involved in the interactions between the helices that stabilize the OC state of PepT*So* are shown. Asn33 (TM1), Ser320 (TM7), Gln341 (TM8), Arg32 (TM1), Asp310 (TM7), His61 (TM2) and Asn454 (TM11) form a hydrogen bond network and lock the extracellular side. The conserved residue Glu419 (TM10) form an extensive network of polar interaction and stabilize the conformation of C domain. The ExxERxxxY motif on TM1 forms ionic interaction with Lys127 and neighboring residues. The intracellular side of the transporter is locked by hydrogen bond interaction between Ser131 (TM4)-Tyr431 (TM10). Pro71 (TM2)-Ser444 (TM11) and Gly75 (TM2)-Thr441 (TM11) interactions also favor the conformation of the OC state.

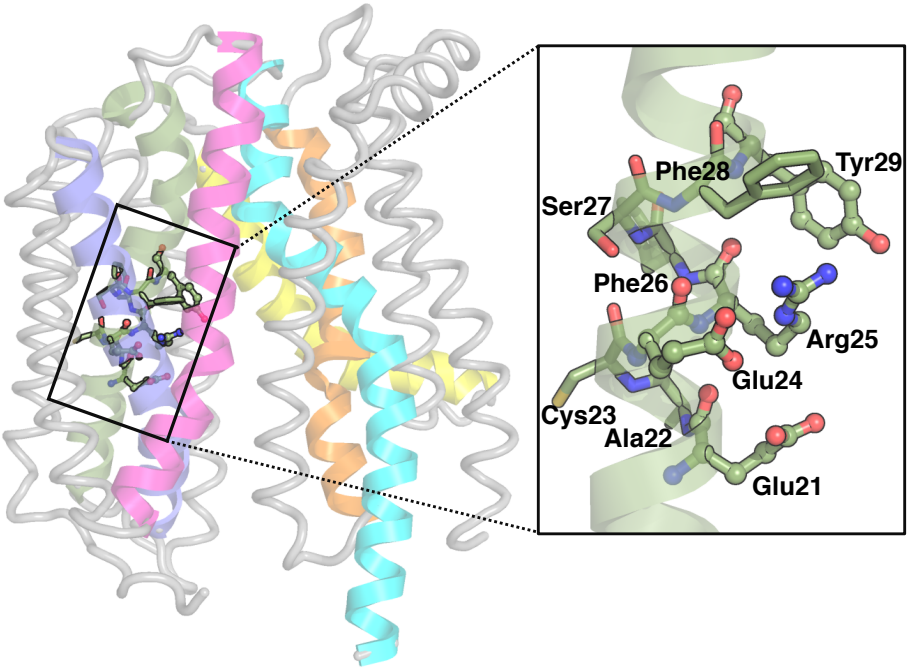

Figure S5: The motif ExxERxxxY is conserved in the POT family of transporter on TM1.

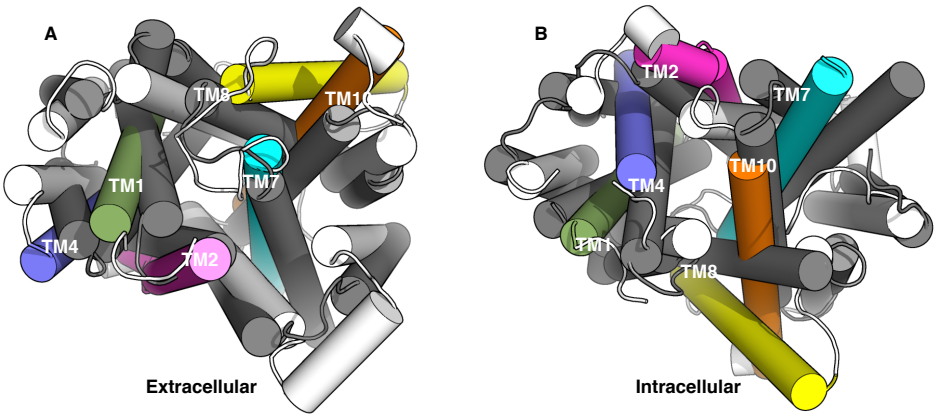

Figure S6: Comparison of predicted $\text{PepT}_{\mathbf{So}}$ OC state (colored) with the EmrD OC crystal structure (PDB: 2GFP,[S1] black) viewed on the (A) extracellular and (B) intracellular side.

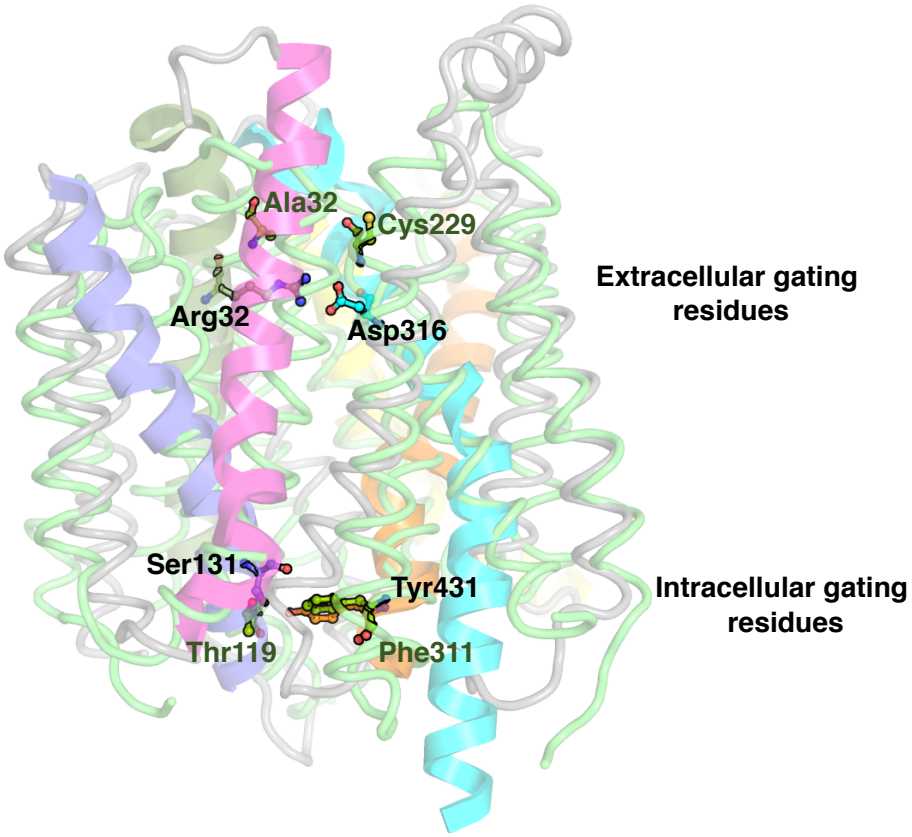

Figure S7: Gating residues for the predicted $PepT_{So}$ OC state (colored) and EmrD OC crystal structure (PDB: 2GFP,[S1] green) are indicated.

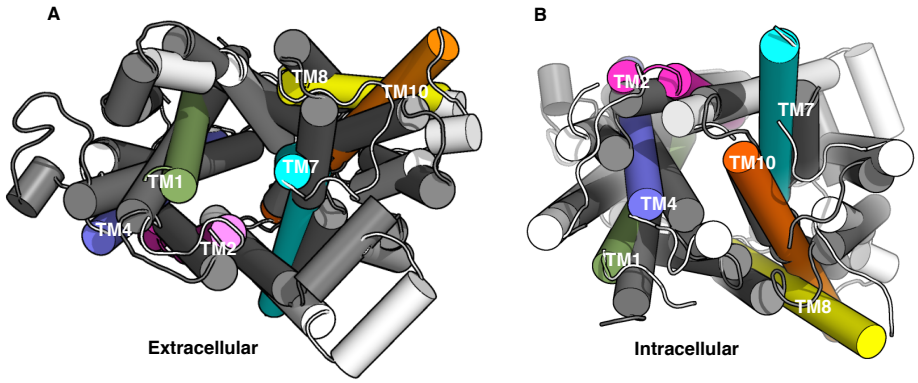

Figure S8: Comparison of predicted $PepT_{So}$ OC state (colored) with the XylE OC crystal structure (PDB: 4GBY,[S2] black) viewed on the (A) extracellular and (B) intracellular side.

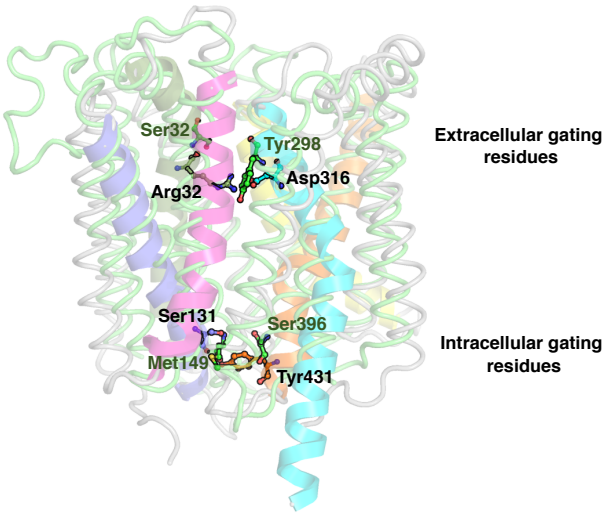

Figure S9: Gating residues for the predicted PepT$_{\mathbf{So}}$ OC state (colored) and XylE OC crystal structure (PDB: 4GBY,[S2] green) are indicated.

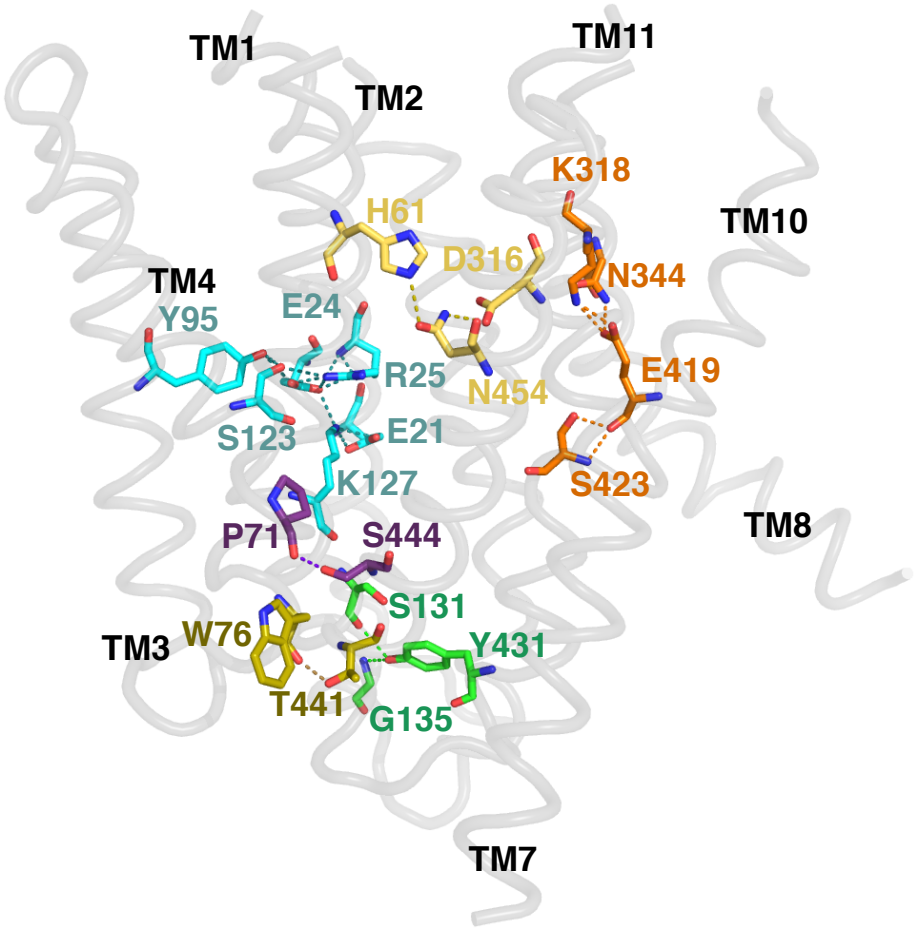

Figure S10: MD predicted $PepT_{So}$ OF state. Detailed description of residues involved in the interaction between the helices that stabilizes the OF state of $PepT_{So}$ are shown. The residues His61 (TM2), Asn454 (TM11) and Asp316 form a hydrogen bond network and stabilize the extracellular part of OF state. The conserved residues Gln419 (TM10) and ExxERxxxY (TM1) contacts are similar as in the predicted OC state. The residues Ser131 (TM4), Gly134 (TM4), Tyr431 (TM10), Trp76 (TM2) and Thr441 (TM11) form interactions that lock the OF state and close the pore channel on the intracellular side.

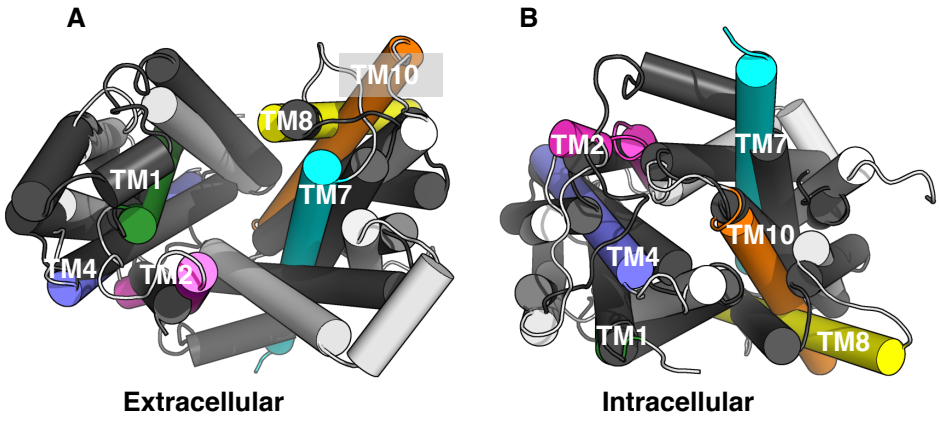

Figure S11: Comparison of FucP OF (PDB: 3O7P,[S3] black) with predicted OF structure viewed on the (A) extracellular and (B) intracellular side.

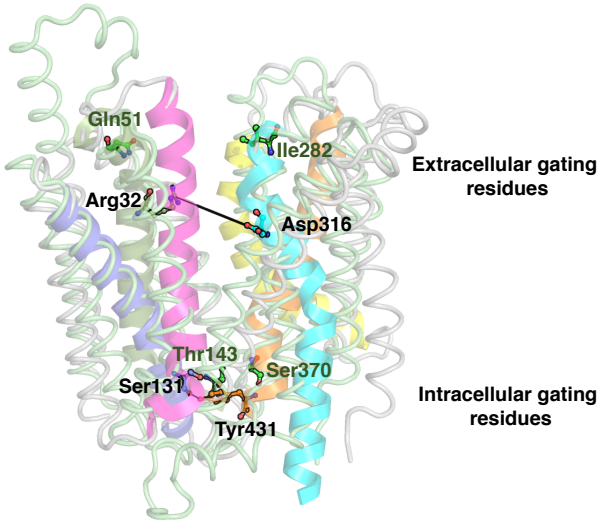

Figure S12: Gating residues for the predicted $PepT_{So}$ OF state (colored) and FucP OF crystal structure (PDB: 3O7Q,[S3] green) are indicated.

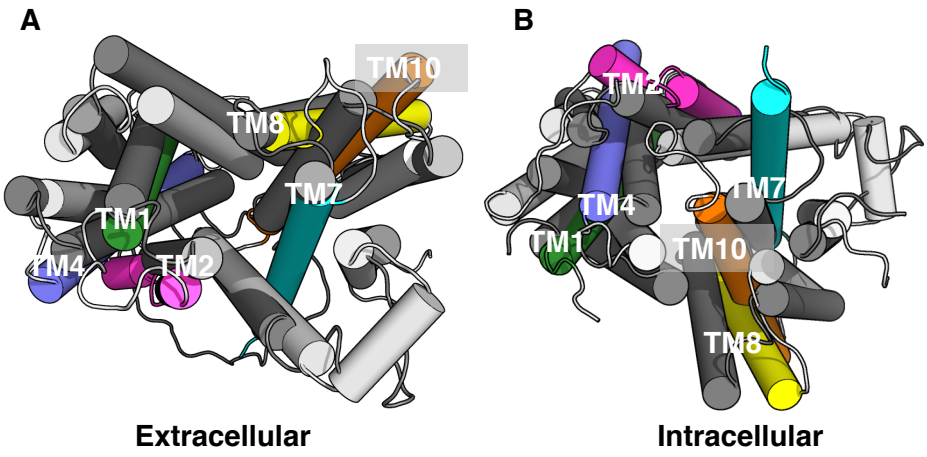

Figure S13: Comparison of $^{\mathbf{RSM}}\mathrm{PepT}_{\mathbf{So}}$ with OF predicted structure. The RSM modeled $\mathrm{PepT}_{\mathbf{So}}$[S4] (black) OF structure was compared with $\mathrm{PepT}_{\mathbf{So}}$ OF MD predicted structure. Transmembrane helices 1, 2, 4, 7, 8 and 10 align well with RSM modeled $\mathrm{PepT}_{\mathbf{So}}$ structure at both (A) extracellular and (B) intracellular ends.

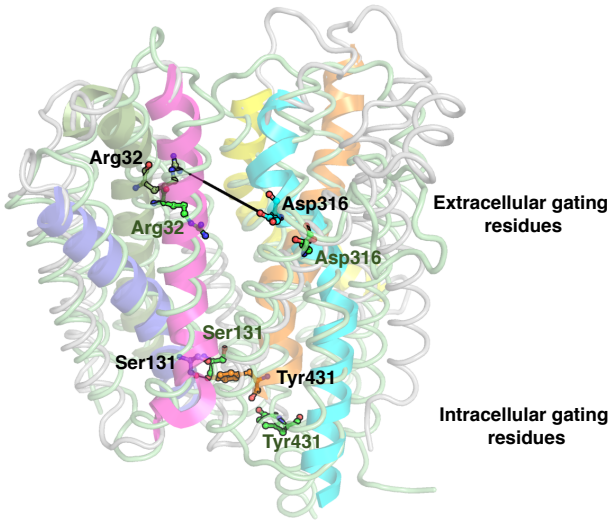

Figure S14: Gating residues for the predicted PepT$_{\mathbf{So}}$ OF state (colored) and RSM modeled PepT$_{\mathbf{So}}{}^{S4}$ (green) are indicated.

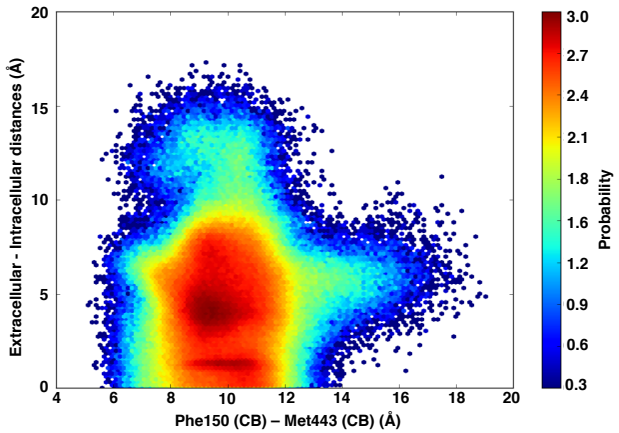

Figure S15: Raw MD simulation data was projected on the difference between the extracellular and intracellular residue pairs and the distance between the residue pairs Phe150-CB (TM5)-Met443-CB (TM11).

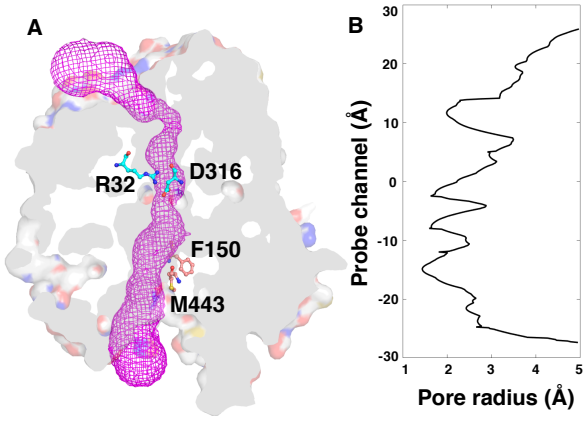

Figure S16: (A) The residue pair Arg32-CZ (TM1)-Asp310-CG (TM7) and Phe150-CB (TM5)-Met443-CB (TM11) interactions are indicated, which characterize the partial IF-OC state. (B) The channel pore radius for the partial IF-OC state.

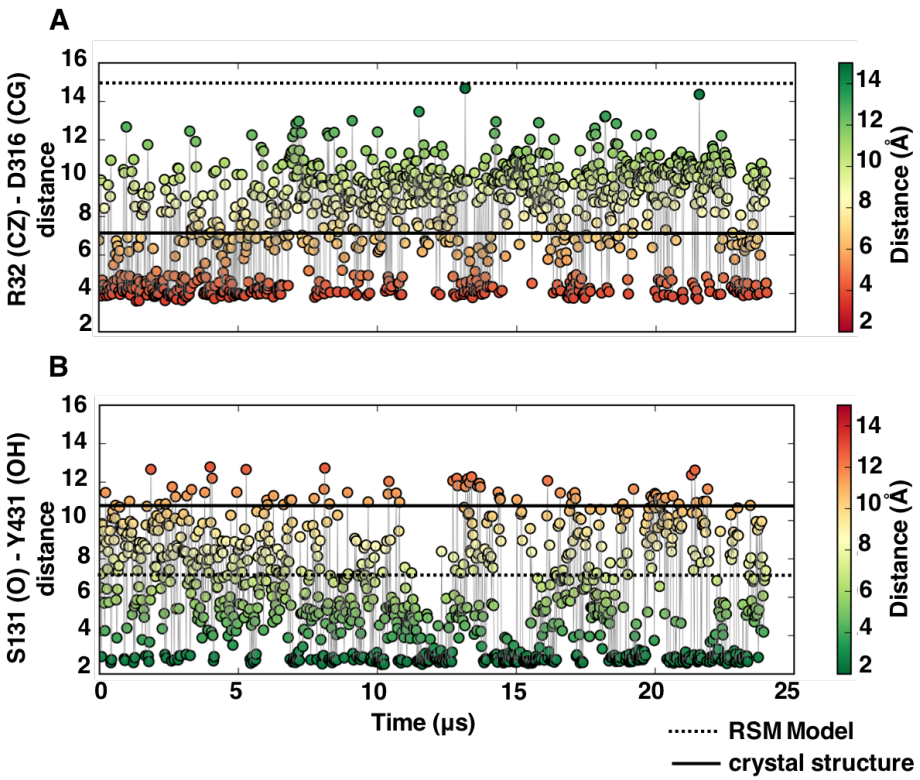

Figure S17: Kinetics of the conformational changes and their timescales. (*A*) The extracellular and (*B*) intracellular distances are shown as a function of time. The color bar indicates the extent of opening and closing of the gating residues of $\text{PepT}_{\mathbf{So}}$. The distances in the crystal structure and the RSM model are indicated in black and dotted lines, respectively.

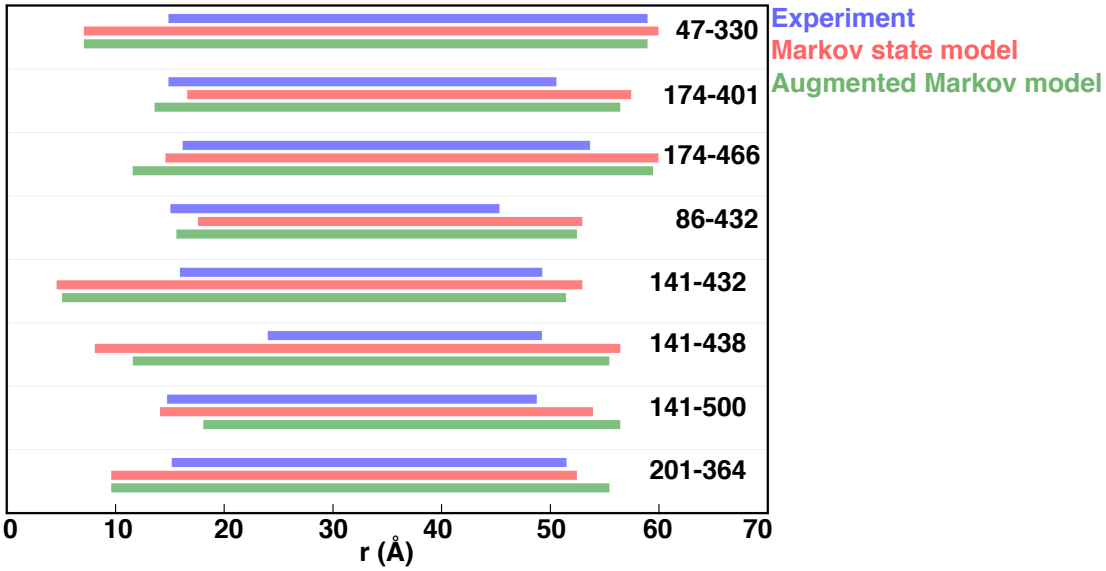

Figure S18: MD simulation predicted DEER distance distribution ranges (green and red) are compared to the experimental DEER distance distribution range (blue). Red and green simulation predictions are based on Markov state model and augmented Markov model, respectively.

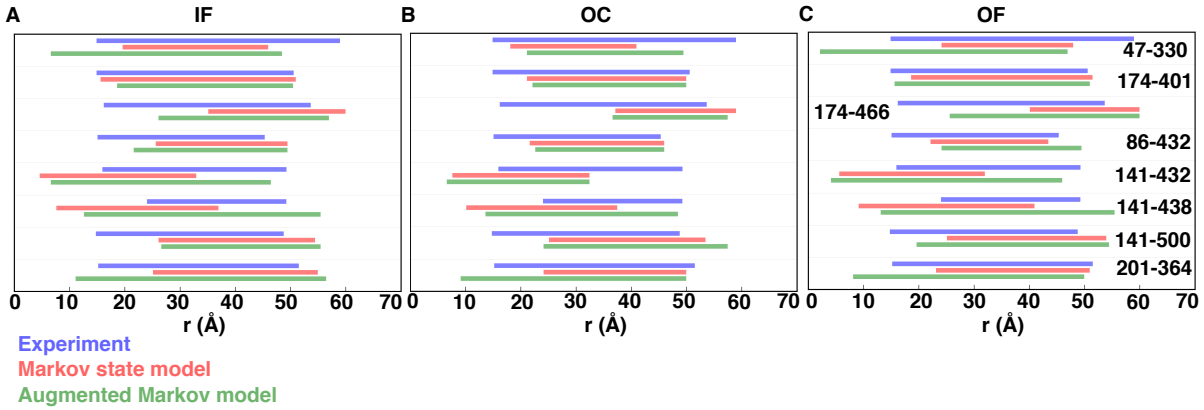

Figure S19: MD simulation predicted DEER distance distribution ranges (green and red) are compared to the experimental DEER distance distribution range (blue) for the (A) IF, (B) OC, and (C) OF states from the $\text{PepT}_{\mathbf{So}}$ conformational landscape. Red and green simulation predictions are based on Markov state model and augmented Markov model, respectively.

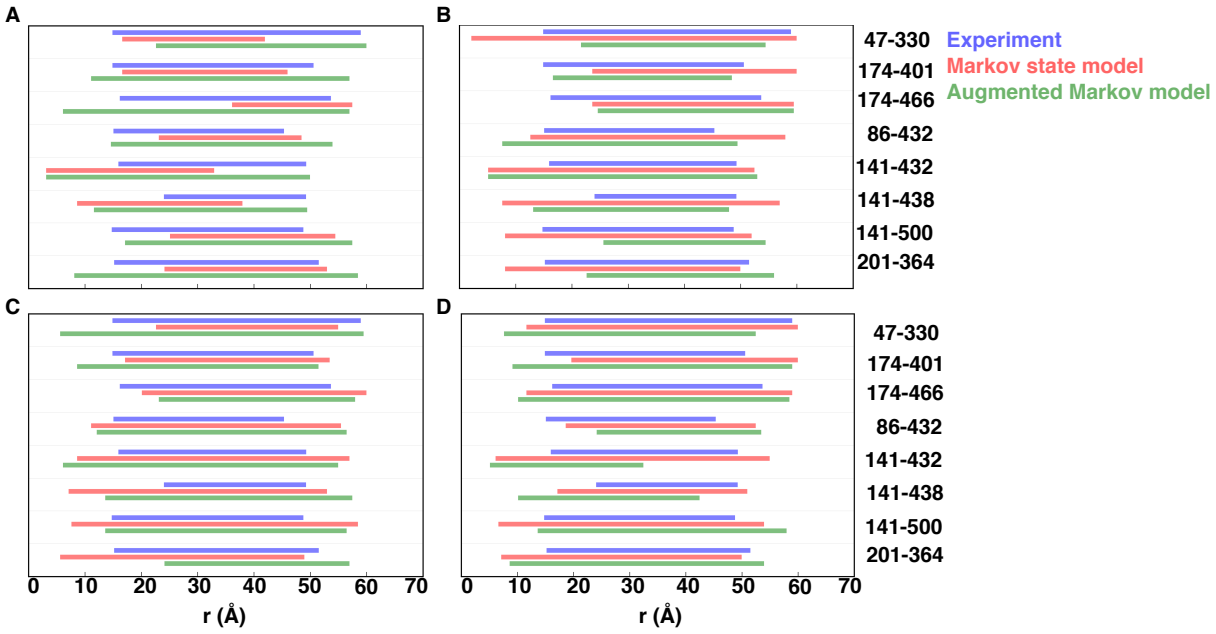

Figure S20: MD simulation predicted DEER distance distribution ranges (green and red) are compared to the experimental DEER distance distribution range (blue) for the (*A*) partial IF-OC, (*B*) partial OC-OF, and (*C*,*D*) wide-open states from the $\text{PepT}_{\mathbf{So}}$ conformational landscape. Red and green simulation predictions are based on Markov state model and augmented Markov model, respectively.

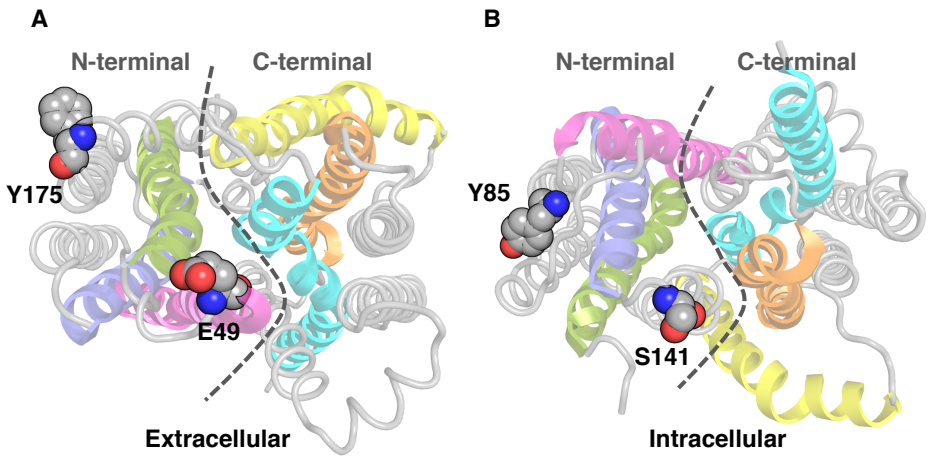

Figure S21: DEER residue-pairs that are assigned a low 'optimal probe score'. The low ranked choices on the (A) extracellular and (B) intracellular side are shown. It is intuitive understanding that the the indicated residue-pair choices would lead to low ranked MSMs as they are placed on only the N domain of the $\mathrm{PepT}_{\mathbf{So}}$ and interactions between the domains are not captured.

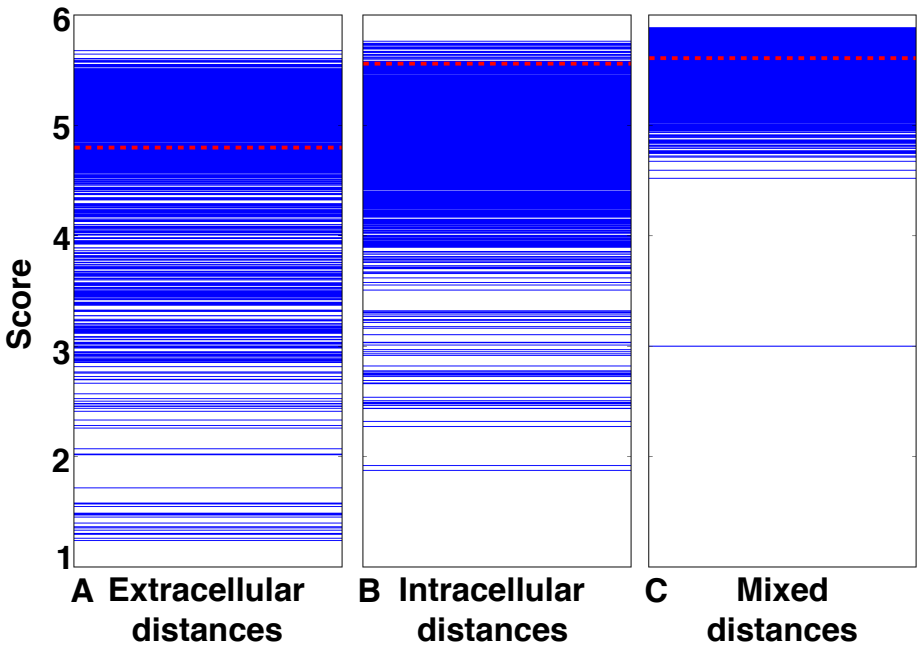

Figure S22: GMRQ scores for MSM constructed based on residue-pair distance choices on (A) extracellular only, (B) intracellular only, and (C) both sides of $\text{PepT}_{\mathbf{So}}$. The red line indicates the MSM score for the experimental residue distance pairs.

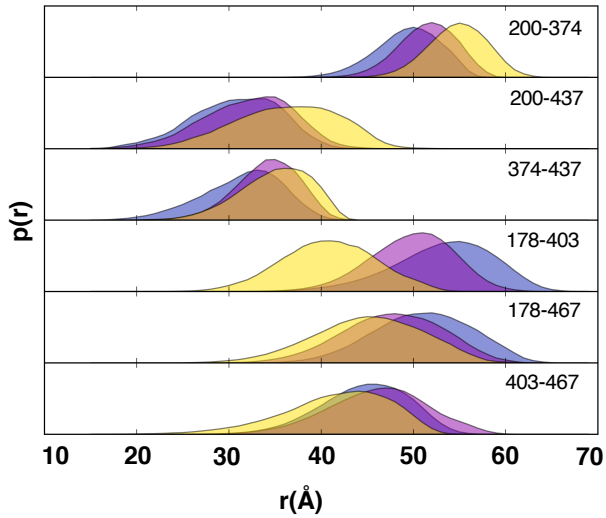

Figure S23: DEER distance distributions for the highest ranking choices. IF, OC and OF plots are represented in yellow, violet and blue, respectively. The distances between the residue pairs 200, 374 and 437 are chosen on the intracellular side and the residues 178, 403 and 467 are chosen on the extracellular side of $\mathrm{PepT}_{\mathbf{So}}$.

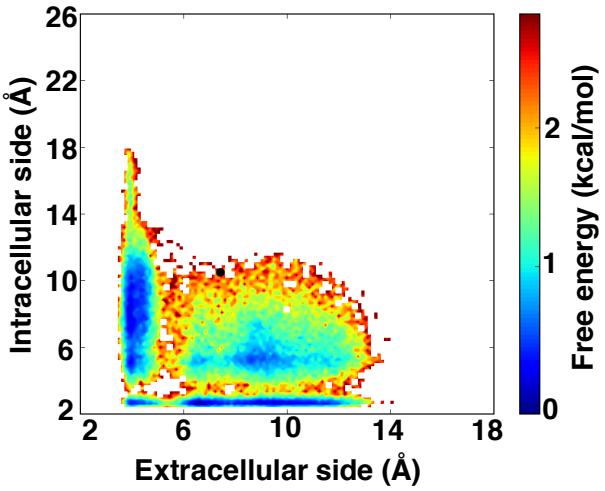

Figure S24: PepT$_{\mathbf{So}}$ two dimensional free energy plot for the raw data obtained using 5 μs accelerated molecular dynamics simulations. The black dot indicates the crystal structure of PepT$_{\mathbf{So}}$ (PDB: 4UVM[S4]) in the IF state.

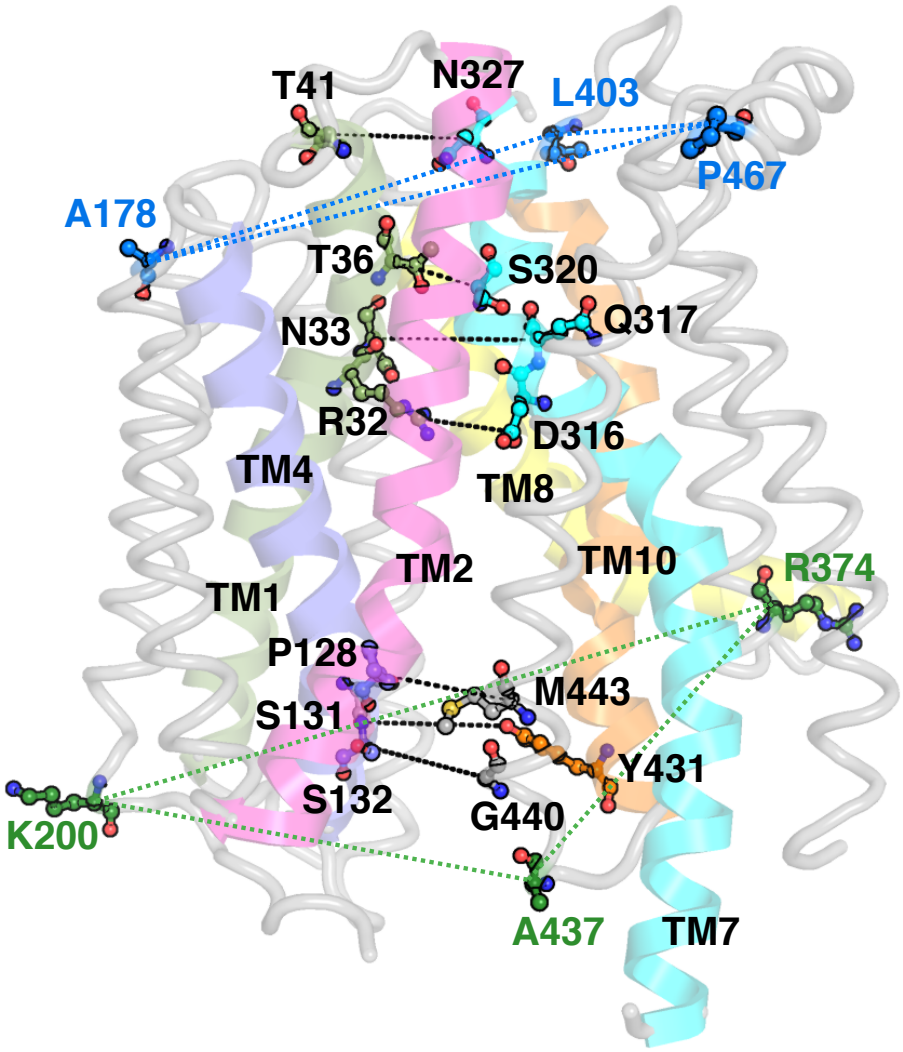

Figure S25: Extracellular, transmembrane and intracellular residue pair distances used for MSM construction.

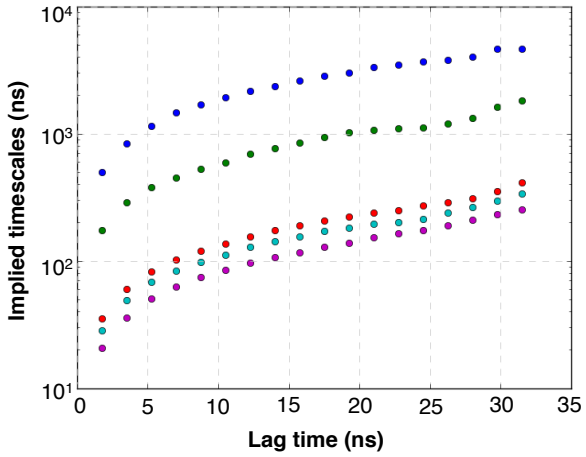

Figure S26: Implied timescales from transition probability matrix of the MSM. Eigenvalues of the transition probability matrix correspond to the dominant rates of transition in the 200 state model. The top 5 eigenvalues for the MSM are shown here which converged at a lag time of 24 ns.

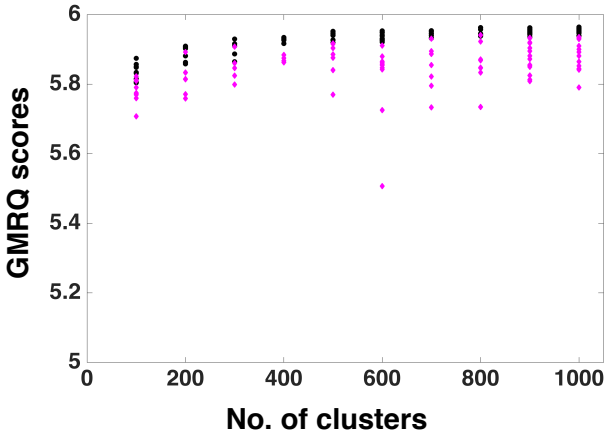

Figure S27: Comparison of the maximum GMRQ scores of MSM built using variable cluster numbers. 200 clusters yields the highest GMRQ score and hence was used for all MSM construction and analysis. The black and pink dots correspond to the scores for training and testing datasets to calculate GMRQ, respectively.

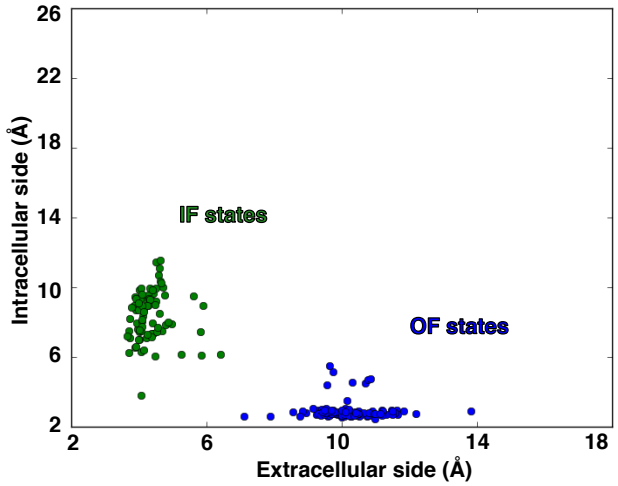

Figure S28: Sampled conformations from the IF and OF microstates from the MSM.

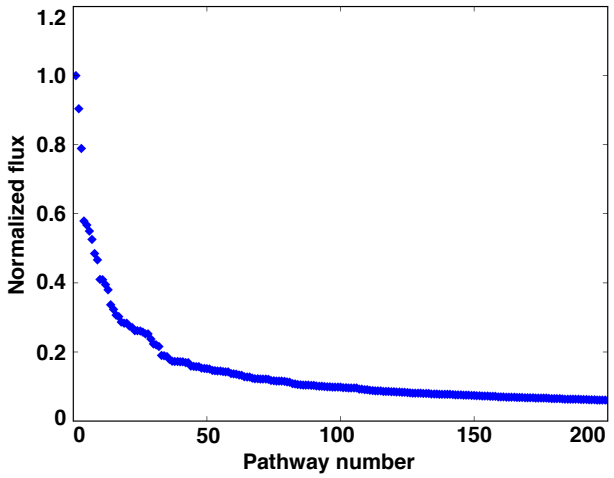

Figure S29: Normalized flux values for the top 200 reactive paths between IF and OF microstates in the MSM. There are several paths with high flux and large number of pathways with lower flux values.

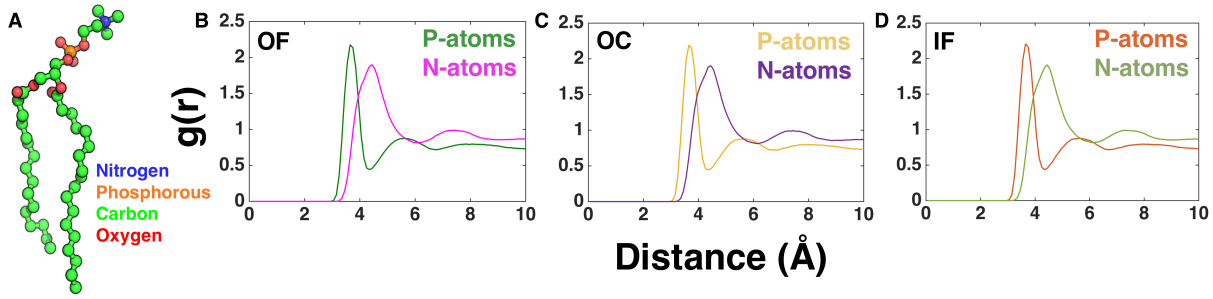

Figure S30: (A) A single POPC lipid molecule indicating positions of nitrogen (blue), phosphorous (orange), carbon (green), and oxygen (red) atoms. Radial distribution of water around lipid bilayer head group atoms in (B) OF, (C) OC, and (D) IF state. The orientation of water molecules around the phosphate and nitrate groups are calculated using VMD 1.9.2.

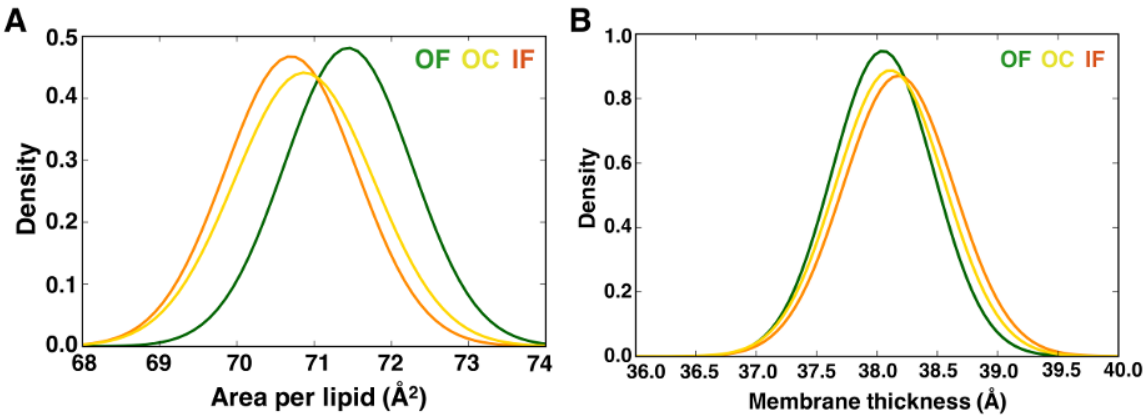

Figure S31: (A) Probability distribution of area per lipid. To obtain the probability distribution, a normal distribution is fitted to the histogram obtained for 500 structures for each state, OF (μ=71.4;σ=0.83), OC (70.9;0.9) and IF (70.7;0.85). (B) Probability distribution of membrane thickness of the POPC lipid bilayer for OF (green), OC (yellow), and IF (orange) states. A normal distribution is fitted to the histogram obtained for 500 structures for each state, OF (μ=38.05;σ=0.42), OC (38.11;0.45) and IF (38.18;0.46).

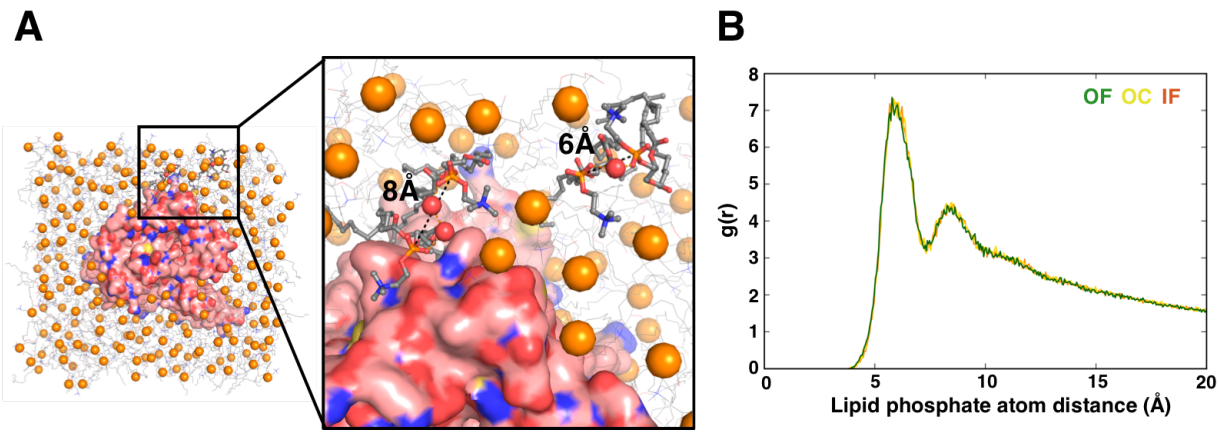

Figure S32: (A) Lipid bilayer molecules in an MD snapshot shows water mediated lipid molecule stabilization. B) The radial distribution plots of phosphate atoms distances in the head group of lipid molecules.

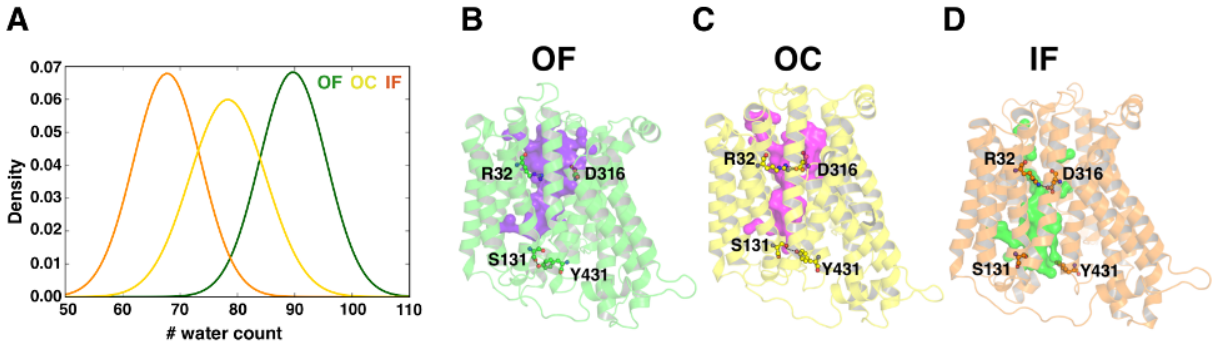

Figure S33: (A) Probability distribution of water molecules in the protein tunnel. To obtain the probability distribution, a normal distribution is fitted to the histogram obtained for 500 structures for each state, OF (μ=89.7;σ=5.84), OC (78.3;6.65) and IF (67.8;5.87). The water conducting channels are visualized for (B) OF, (C) OC, and (D) IF states.

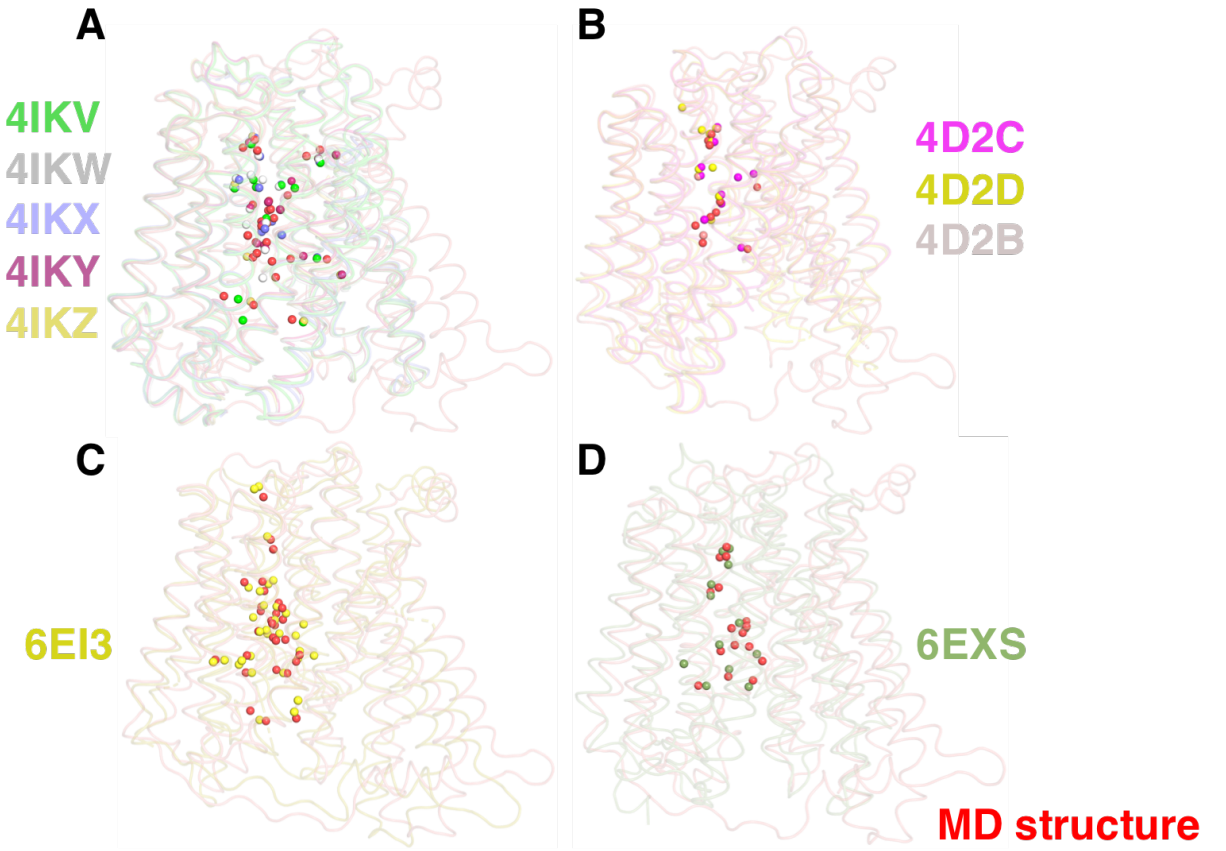

Figure S34: Comparison of water molecules inside the MD structure (red) protein pore channel with water molecules in the crystal structures of (A) GkPOT,[S5] (B) PepT$_{\mathbf{St}}$,[S6] (C) PepT$_{\mathbf{Xc}}$[S7] and (D) PepT$_{\mathbf{Sh}}$.[S8]

Table S1: Adaptive sampling rounds for accelerated MD simulations

Number of rounds	Time (ns)
Round 1	135
Round 2	135
Round 3	108
Round 4	107
Round 5	97
Round 6	87
Round 7	66
Round 8	54
Round 9	135
Round 10	108
Round 11	81
Round 12	68
Round 13	95
Round 14	169
Round 15	164
Round 16	139
Round 17	90
Round 18	157
Round 19	136
Round 20	145
Round 21	118
Round 22	110
Round 23	2171

Table S2: Adaptive sampling rounds for classical MD simulations

Number of rounds	Time (μs)
Round 1	6.4
Round 2	6.8
Round 3	40.5

Table S3: Constraints used for augmented Markov models (σ values are fixed at 0.1)

Distance (k)	Residue numbers	m_k (Å)
1	47, 330	39.875078
2	174, 401	41.78273
3	174, 466	37.278798
4	86, 432	41.883655
5	141, 432	36.666667
6	141, 438	28.412256
7	141, 500	43.81818
8	201, 364	44.585633

References

(S1) Yin, Y.; He, X.; Szewczyk, P.; Chang, G. Structure of the Multidrug Transporter EmrD from Escherichia coli. *Science* **2006**, *312*, 741–744.

(S2) Sun, L.; Zeng, X.; Yan, C.; Sun, X.; Gong, X.; Rao, Y.; Yan, N. Crystal Structure of a Bacterial Homologue of Glucose Transporters GLUT1-4. *Nature* **2012**, *490*, 361–366.

(S3) Dang, S.; Sun, L.; Huang, Y.; Lu, F.; Liu, Y.; Gong, H.; Wang, J.; Yan, N. Structure of a Fucose Transporter in an Outward-Open Conformation. *Nature* **2010**, *467*, 734–738.

(S4) Fowler, P. W.; Orwick-Rydmark, M.; Radestock, S.; Solcan, N.; Dijkman, P. M.; Lyons, J. A.; Kwok, J.; Caffrey, M.; Watts, A.; Forrest, L. R.; Newstead, S. Gating Topology of the Proton-Coupled Oligopeptide Symporters. *Structure* **2015**, *23*, 290–301.

(S5) Doki, S.; Kato, H. E.; Solcan, N.; Iwaki, M.; Koyama, M.; Hattori, M.; Iwase, N.; Tsukazaki, T.; Sugita, Y.; Kandori, H.; Newstead, S.; Ishitani, R.; Nureki, O. Structural Basis for Dynamic Mechanism of Proton-Coupled Symport by the Peptide Transporter POT. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 11343–11348.

(S6) Lyons, J. A.; Parker, J. L.; Solcan, N.; Brinth, A.; Li, D.; Shah, S. T.; Caffrey, M.; Newstead, S. Structural Basis for Polyspecificity in the POT Family of Proton-Coupled Oligopeptide Transporters. *EMBO Rep.* **2014**, *15*, 886–893.

(S7) Parker, J. L.; Li, C.; Brinth, A.; Wang, Z.; Vogeley, L.; Solcan, N.; Ledderboge-Vucinic, G.; Swanson, J. M. J.; Caffrey, M.; Voth, G. A.; Newstead, S. Proton Movement and Coupling in the POT Family of Peptide Transporters. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 13182–13187.

(S8) Minhas, G. S.; Bawdon, D.; Herman, R.; Rudden, M.; Stone, A. P.; James, A. G.;

Thomas, G. H.; Newstead, S. Structural Basis of Malodour Precursor Transport in the Human Axilla. *eLife* **2018**, *7*.