Supporting Information

Khafizov et al. 10.1073/pnas.1209039109

SI Materials and Methods

Molecular Dynamics Simulations of BetP. Molecular dynamics (MD) simulations were carried out on two different protein structures.

(1) An X-ray crystal structure of the BetP trimer in inward-occluded conformations with resolution of 3.35 Å [Protein Data Bank (PDB) ID 2WIT] (1), including residues 56–272 and 275–595 of chain A. Residues 559–589 and 569–589 of protein chains B and C were modeled using the conformation of the helix in chain A. Details of the trimer simulations, including the lipid bilayer setup, have been provided previously (2). Briefly, protonation states and side-chain rotamers were assigned according to calculations performed on the X-ray structure using the MCCE2.0 program (3), and the protein was inserted into a 50:50 diastereoisomeric mixture of 632 palmitoyl oleoyl phosphatidylglycerol (POPG) lipid molecules (4), hydrated with 52,834 water molecules, and containing 673 sodium and 54 chloride ions to neutralize the overall charge of the systems and to maintain an ~0.1-M salt concentration. The final system was ~152 × 152 × 105 Å in size.

A substrate betaine molecule and two sodium ions were placed in putative binding pockets of each protomer of the trimer according to previous proposals (Fig. S2*B*) (1). Equilibration involved 5,000 steps of steepest descents minimization, followed by 1.5 ns of MD simulation in which all nonhydrogen atoms of the protein, ligand, and bound sodium ions were constrained to their initial positions using springs with a force constant of 1 kcal·mol^{-1}·Å^2. Production simulations were carried out in triplicate for 30 ns each.

(2) From an X-ray crystal structure of a BetP trimer at 3.1-Å resolution (PDB ID 4AIN) (Fig. S6) (5), the coordinates of a protomer (chain B) in a fully occluded conformation were used; this structure did not include the crystallographically unresolved residues 1–55 and 558–595. The missing residues in loop 2—D131, E132, A133, and P134—were modeled in using COOT. Residues H173, H188, and H482 were protonated at their Nε atom, and E161 was assumed to be neutral. The sodium at Na2 and the betaine molecule were placed according to positive peaks observed in the Fo-Fc difference density map, whereas the sodium at Na1′ initially was placed equidistant between the hydroxyl oxygen atoms of T246, T250, and S376. A water molecule was placed adjacent to the Na1′.

The complete monomer-bilayer system was set up using GRIFFIN (6). The resultant model contained 219 POPG lipids, 15,016 water molecules, 288 sodium ions, and 41 chloride ions and had dimensions of 91 × 91 × 96 Å. After 1,000 steps of conjugate gradients energy minimization, 20 ns of MD simulation was carried out in which all nonhydrogen atoms of the protein, ligand, and bound sodium ions were constrained to their initial positions using springs with progressively smaller force constants starting at 15 kcal·mol^{-1}·Å^2. Production simulations were carried out in triplicate for >70 ns each.

All MD calculations were performed using the program NAMD (7). Periodic boundary conditions were used. A real-space cutoff of 10 Å was used for both van der Waals and long-range electrostatics; the distance at which the switching function began to take effect was 8 Å. The time step was 2 fs. The SHAKE algorithm (8) was used to fix all bond lengths. Constant temperature (300 K) was set with a Langevin thermostat (9), with a coupling coefficient of 0.2 ps^{-1}. A Nosé–Hoover Langevin barostat (10) was used to apply constant pressure normal to the membrane plane, with an oscillation period of 200 fs and the damping time scale set to 50 fs. The surface area in the membrane plane was kept constant.

The all-atom Charmm27 force field was used for protein (11, 12) and ions (13), and TIP3P was used for water molecules (14). Force field parameters for POPG were kindly provided by H. Jang (National Cancer Institute, Frederick, MD) (15). Parameters for betaine combined the Charmm27 parameters for the POPC lipid headgroup and the glutamic acid side chain.

The MD trajectories were analyzed with Charmm, Gromacs (16, 17), and Visual Molecular Dynamics (18). The first 10 ns of the trajectories were excluded from the analysis of the structural deviations. Calculation of rmsd was carried out after first superposing the same atoms on the reference structure.

Sequence Search and Multiple-Sequence Alignment. Using sequences of known structures with the five transmembrane-helix inverted-topology repeat, LeuT-like (FIRL) fold as the query, we extracted sequences from the National Center for Biotechnology Information nonredundant database of protein sequences (as of August 2010) using a PSI-BLAST search. Five PSI-BLAST iterations were run, with a next-round cutoff of 0.005, a final E-value cutoff of 10^{-4}, and retaining the 250 best hits of each iteration. Sequences containing nonstandard amino acid types were removed, as were those containing large insertions or unusual deletions of transmembrane (TM) domains. A multiple-sequence alignment of the homologs was constructed from the filtered set of PSI-BLAST results using MUSCLE with default settings (19). CD-hit (20) then was used to reduce the set to sequences that share no more than 90% sequence identity with one other, after which the sequences were realigned.

Solid Supported Membrane-Based Electrophysiology. Solid supported membrane-based electrophysiology was performed as described (21) using 40 μL of proteoliposomes at a protein concentration of 1 mg/mL Briefly, experiments were carried out at room temperature (22 °C). The solution-exchange protocol consisted of four phases with a total duration of 4.8 s: (*i*) nonactivating solution (2.5 s), (*ii*) activating solution (0.8 s), (*iii*) nonactivating solution (0.5 s), and (*iv*) resting solution (1.0 s). The resting solution was composed of 250 mM KPi buffer (pH 7.5). Nonactivating and activating solutions also contained 250 mM KPi buffer (pH 7.5). The osmolality of nonactivating and activating solutions was kept constant. Activation by betaine was performed using a nonactivating solution with 50 μM glycine plus 500 mM NaCl and an activating solution with 50 μM betaine plus 500 mM NaCl. Activation by sodium was performed by 1,000 mM KCl plus 50 μM betaine in the nonactivating buffer and 900 mM KCl plus 100 mM NaCl and 50 μM betaine in the activating buffer. Thereby an osmolar gradient was established when switching from the resting to the nonactivating solution at the start of the experiment. Note that during the final phase and between experiments (several minutes) the cuvettete contained resting solution to ensure low osmolar conditions in the proteoliposomes. Transient currents were recorded at the concentration jumps taking place from nonactivating to activating buffers. The current amplifier was set to a gain of 10^9–10^{10} V/A, and low-pass filtering was at 300–1,000 Hz.

1. Ressl S, Terwisscha van Scheltinga AC, Vonrhein C, Ott V, Ziegler C (2009) Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP. *Nature* 458 (7234):47–52.

2. Perez C, Khafizov K, Forrest LR, Krämer R, Ziegler C (2011) The role of trimerization in the osmoregulated betaine transporter BetP. *EMBO Rep* 12(8): 804–810.

3. Alexov EG, Gunner MR (1997) Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys J* 72(5):2075–2093.
4. Zhao W, Róg T, Gurtovenko AA, Vattulainen I, Karttunen M (2007) Atomic-scale structure and electrostatics of anionic palmitoyloleoylphosphatidylglycerol lipid bilayers with Na^+ counterions. *Biophys J* 92(4):1114–1124.
5. Perez C, Koshy C, Yildiz Ö, Ziegler C (2012) Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter BetP. *Nature*, 10.1038/nature11403.
6. Staritzbichler R, Anselmi C, Forrest LR, Faraldo-Gómez JD (2011) GRIFFIN: A versatile methodology for optimization of protein-lipid interfaces for membrane protein simulations. *J Chem Theory Comput* 7(4):1167–1176.
7. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26 (16):1781–1802.
8. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comput Phys* 23(3):327–341.
9. Adelman SA, Doll JD (1976) Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *J Chem Phys* 64(6):2375–2388.
10. Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) constant-pressure molecular-dynamics simulation - the Langevin piston method. *J Chem Phys* 103(11):4613–4621.
11. Feller SE, Gawrisch K, MacKerell AD, Jr. (2002) Polyunsaturated fatty acids in lipid bilayers: Intrinsic and environmental contributions to their unique physical properties. *J Am Chem Soc* 124(2):318–326.
12. Mackerell AD, Jr., Feig M, Brooks CL, 3rd (2004) 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. *J Comput Chem* 25(11):1400–1415.
13. Nina M, Beglov D, Roux B (1997) Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J Phys Chem* 101(26): 5239–5248.
14. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* 79(2): 926–935.
15. Jang H, Ma B, Woolf TB, Nussinov R (2006) Interaction of protegrin-1 with lipid bilayers: Membrane thinning effect. *Biophys J* 91(8):2848–2859.
16. Berendsen HJC, Vanderspoel D, Vandrunen R (1995) Gromacs - a message-passing parallel molecular-dynamics implementation. *Comput Phys Commun* 91(1–3):43–56.
17. Lindahl E, Hess B, Van Der Spoel D (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J Mol Model* 7(8):306–317.
18. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–28.
19. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
20. Li W, Jaroszewski L, Godzik A (2001) Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* 17(3):282–283.
21. Schulz P, Garcia-Celma JJ, Fendler K (2008) SSM-based electrophysiology. *Methods* 46 (2):97–103.

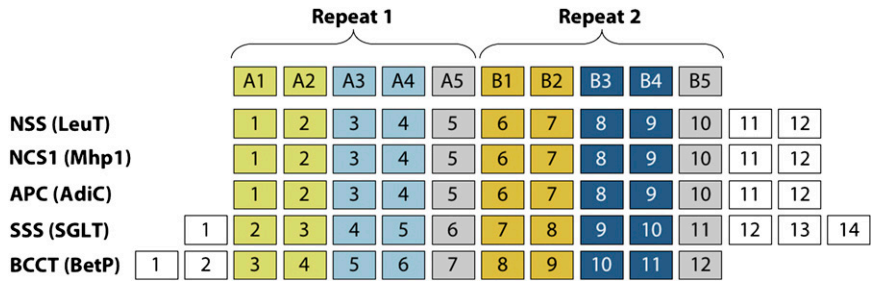

Fig. S1. Schematic showing the numbering of the transmembrane helices in LeuT-fold transporters in five sequence families and the corresponding names used in the main text (A1–5 and B1–5). APC, amino acid/polyamine/organocation; BCC, betaine/carnitine/choline; NCS1, nucleobase/cation symporter-1; NSS, neurotransmitter/sodium symporter; SSS: solute/sodium symporter.

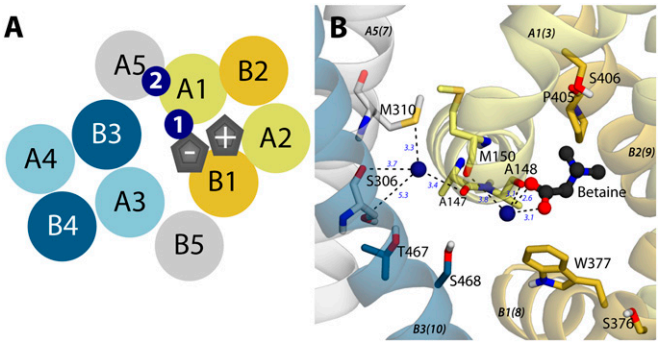

Fig. S2. Proposed sodium ion-binding sites in BetP from Ressl et al. (1). (*A*) Sites shown schematically. (*B*) Close-up of the binding pocket as used in MD simulations, with TM helices colored according to the schematic and the substrate betaine colored black, as viewed from the periplasm. Sodium ions are represented by dark blue spheres, and selected amino acids are shown by sticks. The proposed Na1 site (pNa1) included coordination by the betaine carboxylate group and by carbonyl oxygens of A148 and M150 in helix A1, whereas the proposed Na2 site (pNa2) included direct coordination by side chains of S306 and M310 in helix A5 and the carbonyl oxygen of A147 in helix A3 but not by residues in helix B3. The positions of T467 and S468 in helix B3 are shown for reference; the distances from the ion at pNa2 are 6.1 and 8.1 Å, respectively.

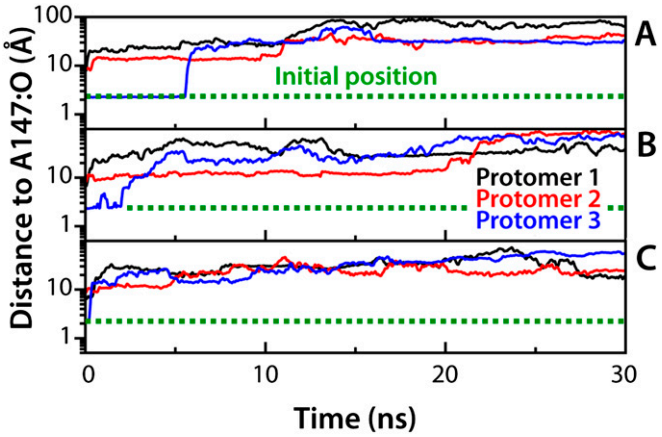

Fig. S3. Stability coordination of a sodium ion at position pNa2 proposed by Ressl et al. (1) as a function of time during MD simulations of the BetP crystal structure, 2WIT. Stability coordination is measured as the average distance of the ion from the backbone carbonyl oxygen atom of A147 for three independent simulations (*A*–*C*), and for each of the three protomers in the trimer (1–3, colored black, red, and blue, respectively). The distance in the initial structure is shown for reference (green dashed line). Note the log scale on the *y* axis.

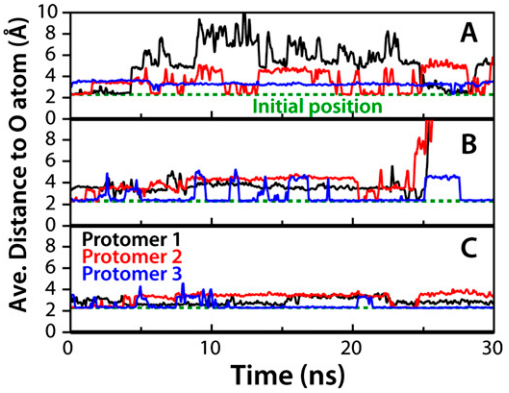

Fig. S4. Stability coordination of a sodium ion at position pNa1 proposed by Ressl et al. (1) as a function of time during MD simulations of the BetP crystal structure, 2WIT. Stability coordination is measured as the average distance of the ion from the coordinating surrounding oxygen atoms for three independent simulations (*A*–*C*) and for each of the three protomers in the trimer (1–3, colored black, red, and blue, respectively). The distance in the initial structure (green dashed line) is shown for reference.

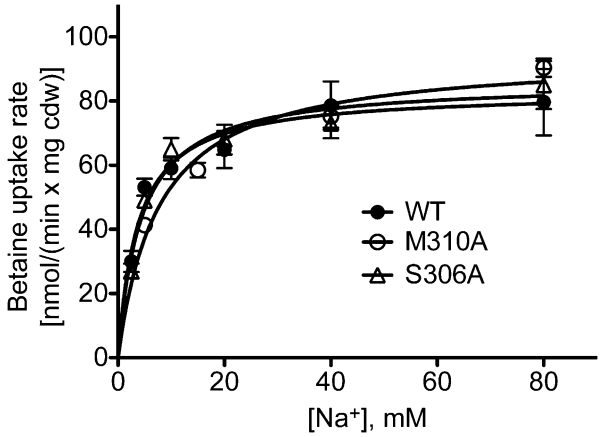

Fig. S5. Sodium dependence of betaine uptake in mutations at sodium-binding site pNa2. Uptake rates in nmol·min^{-1}· mg^{-1} cell dry weight (cdw) were measured as a function of the external sodium concentration in *Escherichia coli* MKH13 cells expressing BetP mutants. Each point shows the average of at least three independent experiments. Error bars represent SD.

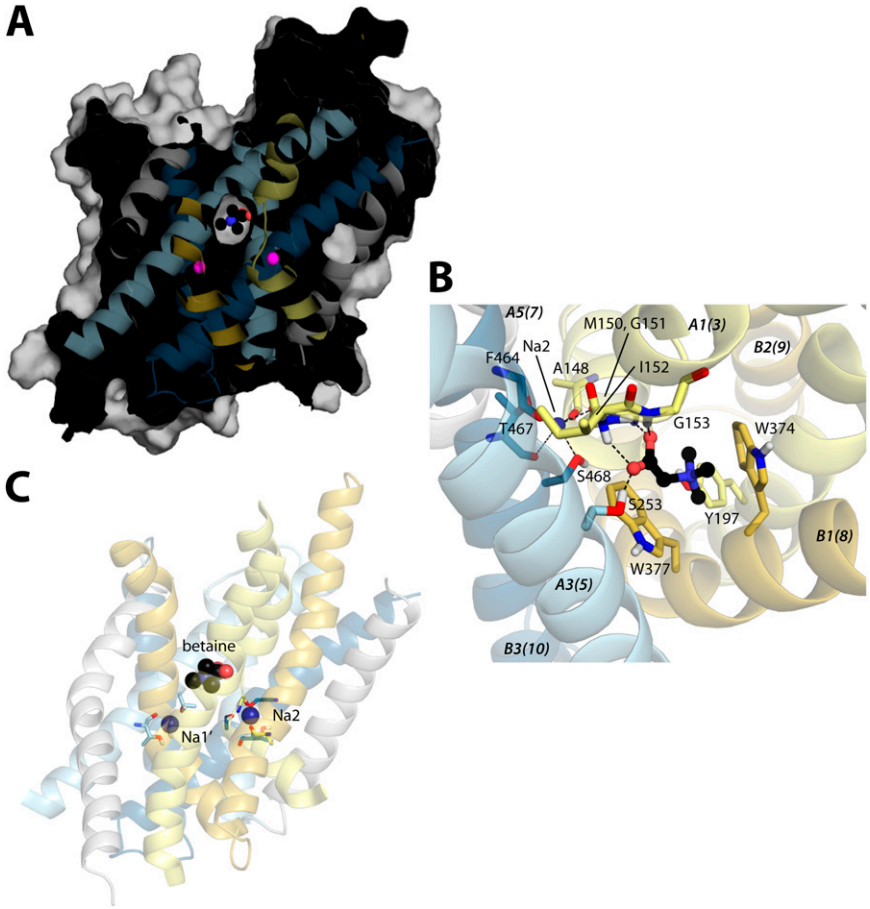

Fig. S6. The structure of BetP in PDB ID 4AIN protomer B. (*A*) Solvent-accessible surface of BetP and its ligands, colored according to Fig. 1, with sodium ions in magenta, and viewed from the plane of the membrane. The periplasm is at the top of the figure. Slicing the surface through the center of the protein reveals that the center of the protein is occluded to substrate (and probably even to water) from both sides of the membrane. (*B*) Coordination of betaine and its position relative to the Na2 site; the view is similar to that in Fig. S2. (*C*) View from the plane of the membrane showing that the Na1′ and Na2 sites are at a similar depth in the structure and are closer to the cytoplasmic side than the betaine-binding site.

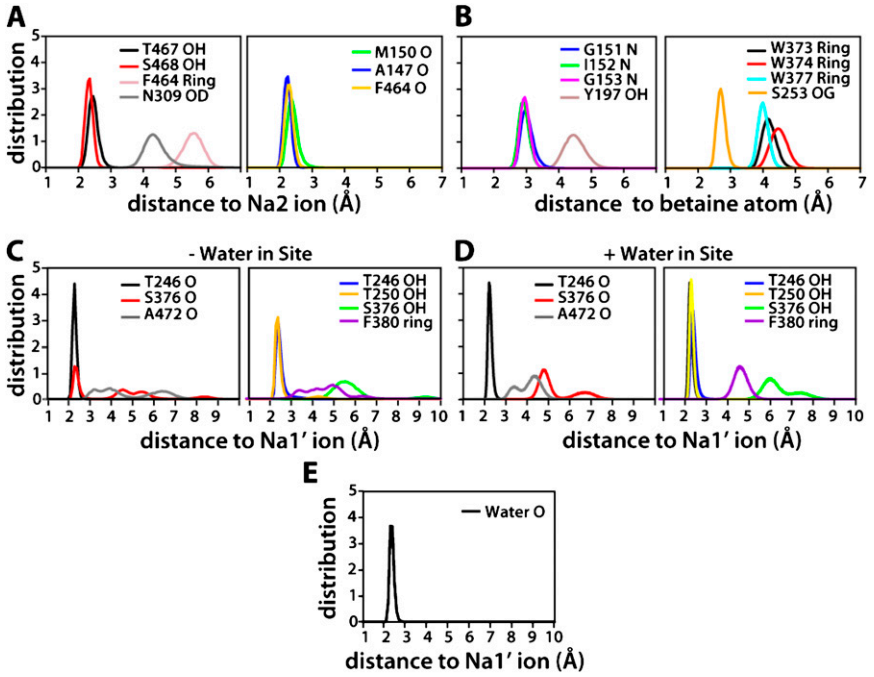

Fig. S7. Protein–ligand interactions in MD simulations of an occluded BetP monomer (PDB ID 4AIN). (*A*) Coordination of a sodium ion at the Na2-binding site. Distances are calculated between the ion and a side-chain hydroxyl (OH), amide (OD), or backbone oxygen (O) or the average position of the carbon atoms in the phenyl ring of F464 (Ring). (*B*) Distances between betaine carboxylate oxygen atoms and backbone amide nitrogen (N) or the hydroxyl oxygen (OG) of S254 are consistent with hydrogen-bonding. Coordination of the betaine trimethylamine is calculated from the betaine N atom to the geometric center of tryptophan (W) side chains (Ring) or the phenoxyl oxygen of Y197. Data in *A* and *B* are taken as distributions from six replica trajectories covering a total simulation time of 480 ns. (*C*–*E*) Coordination of a sodium ion placed at the Na1′-binding site in the absence (*C*) and presence (*D* and *E*) of a water molecule. Distances are calculated between the ion and a side-chain hydroxyl (OH), amide (OD), backbone oxygen (O), or the average position of the carbon atoms in the phenyl ring of F380 (Ring) (*D*) or a specific water molecule (*E*). Data in *D* and *E* are distributions from three replica trajectories with >200-ns total simulation time.

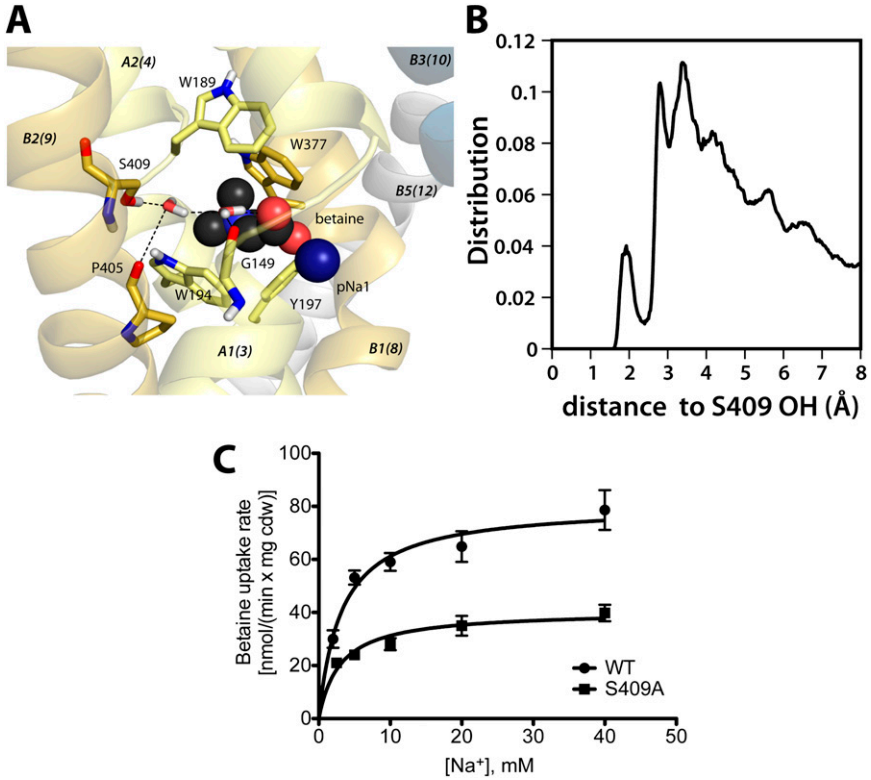

Fig. S8. Analysis of S409 as a putative sodium-binding residue. (*A*) Snapshot from simulation of BetP inward-occluded trimers, in which waters enter the bundle core and hydrate the S409 side chain. (*B*) Distribution of water molecules around the side-chain oxygen atom of S409 in each of three protomers in three replica MD simulations. The peak at 2 Å reflects direct interaction with a water molecule. (*C*) Sodium dependence of betaine uptake by BetP S409A compared with that of WT BetP. Uptake rates in nmol · min^{-1}·mg^{-1} cdw were measured as a function of the external sodium concentration in *E. coli* MKH13 cells expressing WT BetP or the S409A mutants. Each point shows the average of at least three independent experiments. Error bars represent SD.

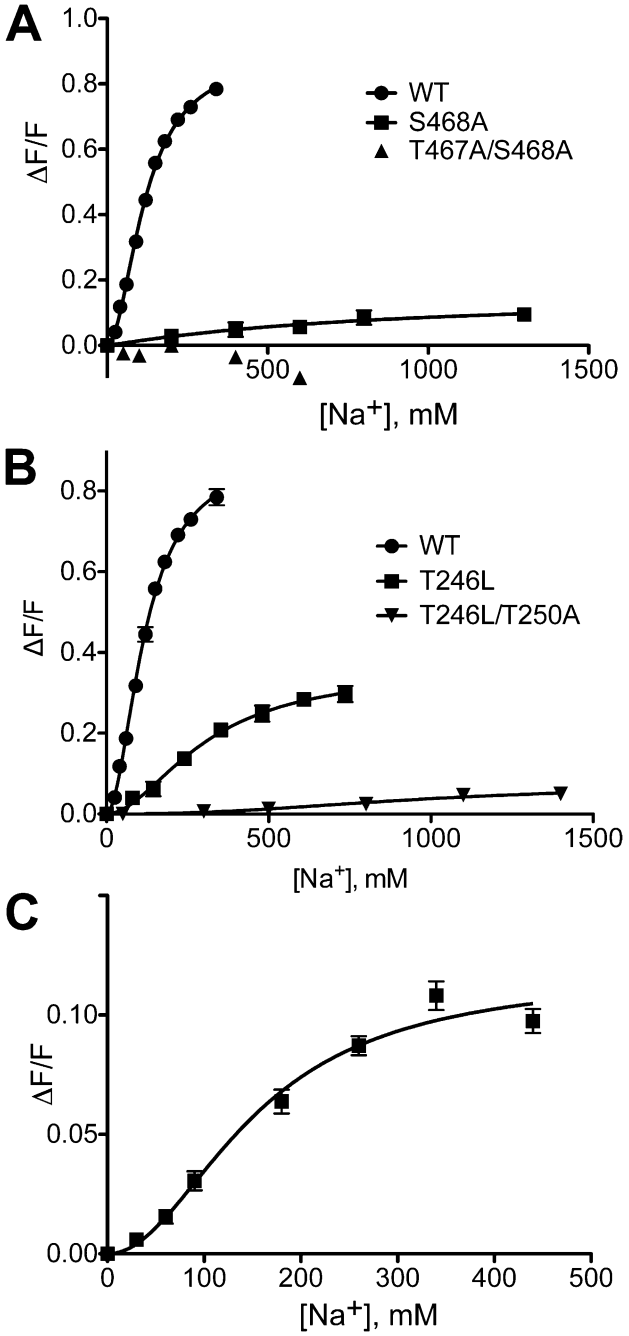

Fig. S9. Cooperativity in the sodium dependence of Trp fluorescence in BetP. (*A* and *B*) Mutants of Na2 (*A*) and Na1′ (*B*) shown at a large scale; see Fig. 9 for details. (*C*) The monomeric W101A/T351A BetP mutant.

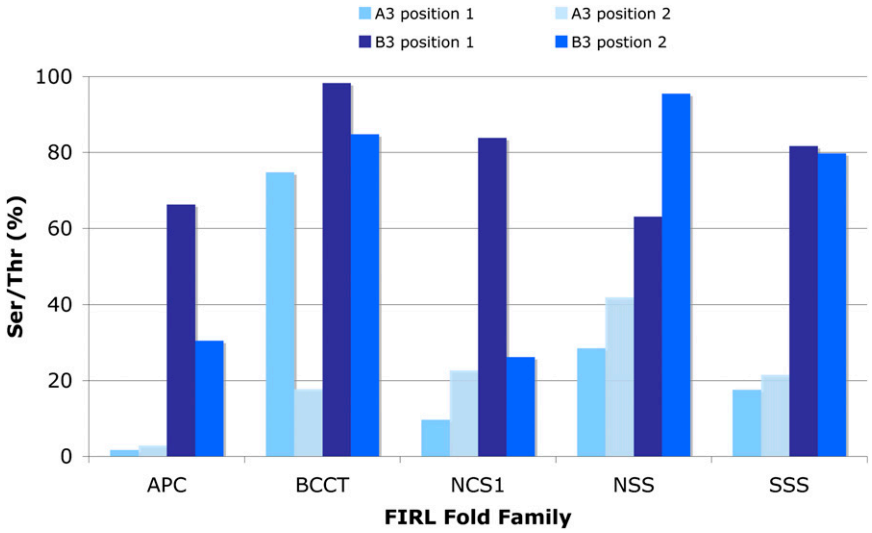

Fig. S10. Conservation of hydroxylic side chains in FIRL-fold transporters. The percentage composition of serine and threonine amino acids was calculated at key positions (see Fig. 2) in multiple-sequence alignments of five different families for which representative structures are known. Sequences were collected using a PSI-BLAST search (*SI Materials and Methods*). Positions 1 and 2 in helix A3 (lighter blue bars) correspond to BetP residues T246 and T250, respectively. Positions 1 and 2 in helix B3 correspond to BetP residues T467 and S468, respectively (darker blue bars).

Fig. S11. Sequence alignment of functionally characterized BCC transporters. Amino acid sequences are of BetP; LcoP; EctP from *Corynebacterium glutamicum* (CORGL); BetH from *Halobacillus trueperi* (HATR); BetL from *Listeria monocytogenes* (LISMN); BetM and EctM from *Marinococcus halophilus* (MARHA); BetS

Legend continued on following page

from *Sinorhizobium meliloti* (RHIML); BetT from *Pseudomonas syringae* (PSESM) and *Aphanothece halophytica* (APHHA); BetT and BetU from *Escherichia coli* (ECOLI or ECOBR); ButA from *Tetragenococcus halophile* (TETHA); EctT from *Virgibaccilus pantothenticus* (VIRPA); CudT from *Staphylococcus xylosus* (STAXY); DddT from *Psychrobacter sp*. (PSYSP); OpuD from *Bacillus subtilis* (BACSU); and CaiT from *Proteus mirabilis* (PROMH). Only CaiT is not sodium dependent. Helical residues in BetP are indicated by gray bars above the sequence. Positions of residues mentioned in the main text are marked with arrows; see legend to Fig. 2 for more details. The alignment was generated using PRALINE (1).

1. Pirovano W, Feenstra KA, Heringa J (2008) PRALINE: A strategy for improved multiple alignment of transmembrane proteins. *Bioinformatics* 24(4):492–497.

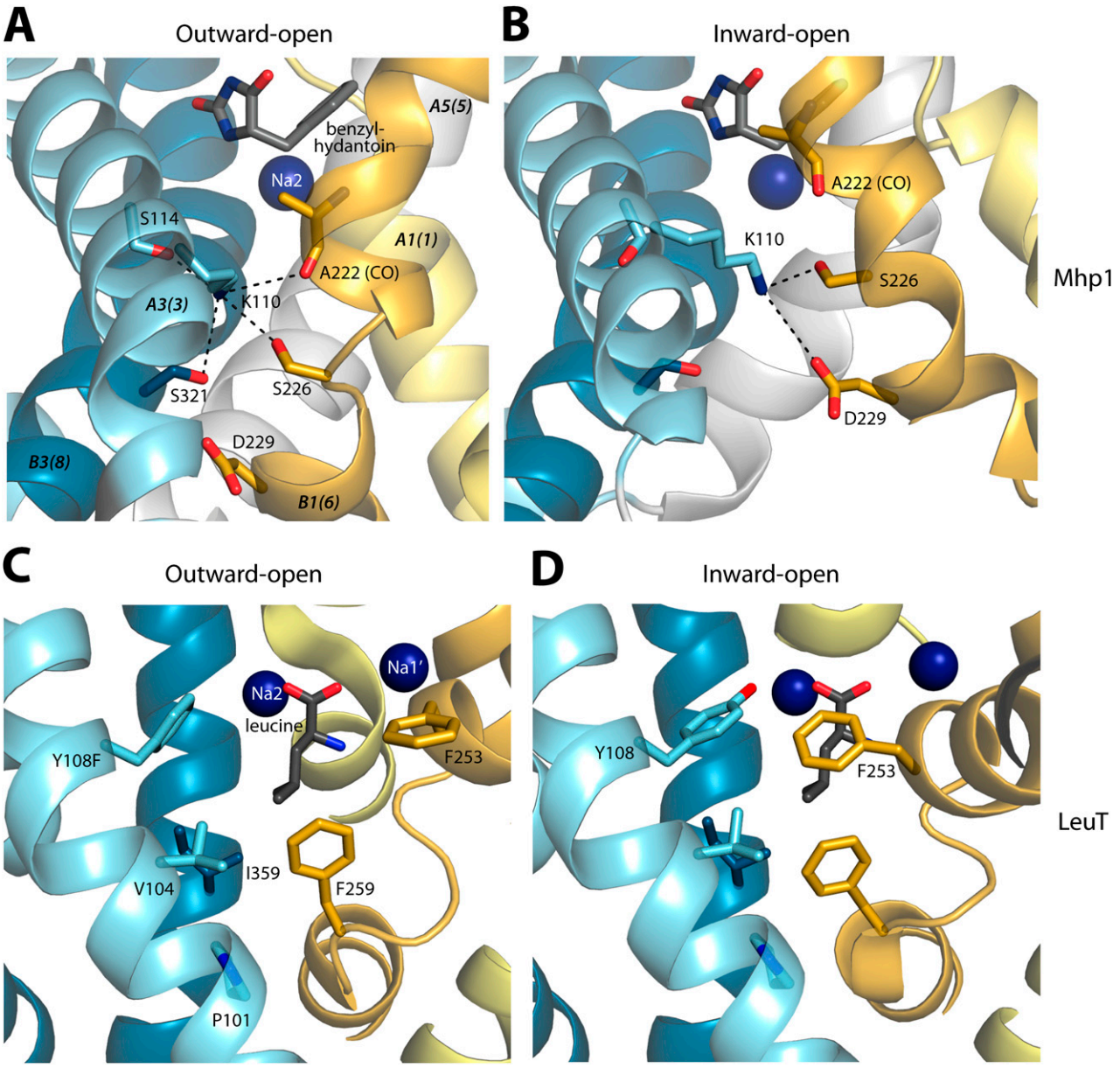

Fig. S12. Comparison of structures of two FIRL-fold transporters, Mhp1 and LeuT, in outward-open and inward-open conformations, focusing on the location equivalent to that of Na1′ in BetP. Helices are colored according to Fig. 1 and area viewed from within the membrane with the periplasm toward the top of the page. Selected residues are shown as sticks and labeled. (*A* and B) Structures of Mhp1 in outward-open (*A*) and inward-open (*B*) conformations, from PDB ID 2JLN and 2X79, respectively, after superposition of the hash domains (A3, A4, B3, and B4). The substrate benzylhydantoin (gray sticks) and one sodium ion (blue sphere) from an outward-occluded conformation (PDB entry 2JLO) are included for reference. (*C* and *D*) Structures of (*C*) WT LeuT, in outward-open conformation, from PDB ID 3TT1 and (*D*) a Na2-deletion mutant of LeuT bound to an antibody fragment in an inward-open conformation, from PDB ID 3TT3, after superposition of the hash domains. Bound leucine (gray sticks) and two sodium ions (blue spheres) from an outward-occluded conformation (PDB ID 2A65) are shown for reference. The position approximately equivalent to that of Na1′ in BetP is between residues P101, V104, F259, and I359 of LeuT.

Table S1. Residues in actual or proposed sodium-binding sites in BetP

Site\helix	A1 (3)	A3 (5)	A5 (7)	B1 (8)	B2 (9)	B3 (10)	B4 (11)	Other
pNa1 (1)	A148 (O) M150 (O)					*T467 (OH)* *S468 (OH)*		Betaine-CO_2^-
pNa2 (1)	A147 (O)		S306 (O) S306 (OH) M310 (S)					
–	G149 (O)				P405 (O) S409 (OG)			Betaine-CO_2^-
Na1′		T246 (O) T246 (OH) T250 (OH)		F380 (ring)			*T499 (OH)*	Water
Na2	A147 (O) M150 (O)					F464 (O) T467 (OH) S468 (OH)		

Residues in italics indicate possible indirect coordination, e.g., through a water molecule. Helix numbering (column titles) reflects the position in the FIRL-fold, and the TM helix numbering in BetP (in parentheses; see Fig. S1).

Table S2. Primer sequences

Oligonucleotide	Sequence 5′ → 3′
W101A	GAC AAT CTC GGC GCG GCC TTC ATT TTG
T246L	GCG ATT ATC GCC CTC GTA TTC GGC ACG
T250A	GTA TTC GGC GCG GCA TGT TCC CTT GGC
S306A	GGA ATC CAG TAC CTC GCC AAC GCC AAC ATG
M310A	CTC TCC AAC GCC AAC GCG GTT CTG GCA GCT CTG
T351A	CAA ATG GCA GGC CGC GCT GCC ATG AGT GCC GAC
S376A	GCA TGG TGG ATC GCT TGG TCA CCA TTC
F380A	ATC TCT TGG TCA CCA GCC GTA GGA ATG TTC TTG
S409A	GTC CCA GCA GGT GTG GCC ACC GTA TGG TTC TCC
F464A	TTG CTG GGT ACT GCC TTC ATT ACC TCT GCT
T467A	GGT ACT TTC TTC ATT GCC TCT GCT GAC TCT GCT
S468A	ACT TTC TTC ATT ACC GCT GCT GAC TCT GCT TCC
T467A/S468A	GGT ACT TTC TTC ATT GCC GCT GCT GAC TCT GCT TCC

Exchanged codons are underlined. Only sense primers are listed.