

Simulations of the Alternating Access Mechanism of the Sodium Symporter Mhp1

Joshua L. Adelman,[†] Amy L. Dale,[†] Matthew C. Zwier,[‡] Divesh Bhatt,[§] Lillian T. Chong,[‡] Daniel M. Zuckerman,[§] and Michael Grabe^{†*}

[†]Department of Biological Sciences, [‡]Department of Chemistry, and [§]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania

ABSTRACT Sodium coupled cotransporters of the five-helix inverted repeat (5HIR) superfamily use an alternating access mechanism to transport a myriad of small molecules across the cell membrane. One of the primary steps in this mechanism is the conformational transition from a state poised to bind extracellular substrates to a state that is competent to deliver substrate to the cytoplasm. Here, we construct a coarse-grained model of the 5HIR benzylhydantoin transporter Mhp1 that incorporates experimental structures of the outward- and inward-open states to investigate the mechanism of this conformational change. Using the weighted ensemble path-sampling method, we rigorously sample the outward- to inward-facing transition path ensemble. The transition path ensemble reveals a heterogeneous set of pathways connecting the two states and identifies two modes of transport: one consistent with a strict alternating access mechanism and another where decoupling of the inner and outer gates causes the transient formation of a continuous permeation pathway through the transporter. We also show that the conformational switch between the outward- and inward-open states results from rigid body motions of the hash motif relative to the substrate bundle, supporting the rocking bundle hypothesis. Finally, our methodology provides the groundwork for more chemically detailed investigations of the alternating mechanism.

INTRODUCTION

Cells use membrane transporters to acquire small molecules such as sugars, essential amino acids, and neurotransmitters from the environment. A large class of transporters has evolved to couple the inward movement of sodium ions, Na⁺, down their concentration gradient with the inward movement of substrates. These transporters are referred to as sodium-dependent cotransporters, or sodium symporters, and the free energy provided by the downhill movement of Na⁺ allows the substrates to be concentrated in the cytoplasm (1–3). In general, transport is thought to occur via an alternating access mechanism in which ion and substrate tightly bind to an outward-facing conformation of the transporter followed by a conformational change that closes the extracellular binding cavity and exposes the substrate and ion to an intracellular cavity (4). The conformational change to the inward-facing state is thought to weaken substrate and ion binding to the transporter such that they are easily released to the cytoplasm (Fig. 1). The apo state is then poised to reset to the outward-facing conformation to start the cycle again. Although this hypothetical transport mechanism is reasonable, the molecular details of the cycle are only now becoming clear and the dynamics of the process are not well understood.

Recently, there have been a number of structures solved for members in the five-helix inverted repeat (5HIR) superfamily of sodium-dependent cotransporters (5–8). Superfamily members have a variable number of α -helical transmembrane

(TM) segments, but all share a common core of 10 TM segments numbered TM1–TM10. Substrate binds near the center of the transporter between prominent kinks in TM1 and its symmetry-related pair, TM6. As the name suggests, the first five helices form an inverted repeat of the second five. Remarkably, some superfamily members share < 10% sequence identity, yet the core domains superpose with a root mean-squared deviation (RMSD) < 4.0 Å (1).

Superfamily members have been solved at different stages along the transport cycle lending insight into how the molecular architecture uses the energy stored in the Na⁺ gradient to pump the substrate into the cytoplasm. There are five primary states that the transporters take on: outward open, outward occluded, occluded, inward occluded, and inward open. Outward-facing states have a large hydrophilic cavity penetrating the cotransporter from the extracellular space to the substrate binding site, while a hydrophobic stretch of protein occludes substrate and ion escape to the cytoplasm (Fig. 1 A). Meanwhile, the inward-open state has a hydrophilic cavity penetrating the protein from the cytoplasm and an impassible stretch of tightly packed protein that prevents release to the extracellular space (Fig. 1 B). In the presence of bound substrates another critical feature of the 5HIR superfamily becomes clear—a handful of hydrophobic residues block substrate exit to the cytoplasm for structures in the inward-facing state or exit to the extracellular space for outward-facing structures. For this reason, these configurations are termed the inward-occluded state and outward-occluded state, respectively. These blocking residues have been termed the thin gates while the thick stretches of packed protein

Submitted May 26, 2011, and accepted for publication September 15, 2011.

*Correspondence: mdgrabe@pitt.edu

Editor: Benoit Roux.

© 2011 by the Biophysical Society
0006-3495/11/11/2399/9 \$2.00

doi: 10.1016/j.bpj.2011.09.061

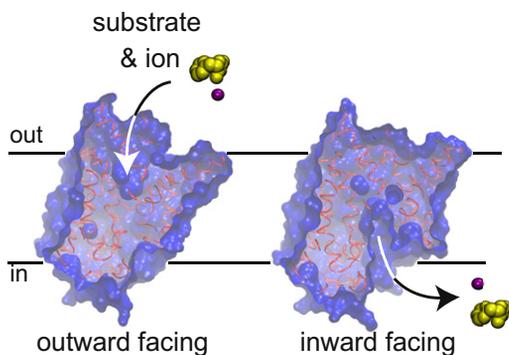


FIGURE 1 Alternating access and the structures of Mhp1. The outward-open x-ray structure of Mhp1 (PDBID: 2JLN) is pictured on the left and the inward-open structure (PDBID: 2X79) is shown on the right. Sodium ion (purple) and hydantoin molecule (yellow) are shown for reference to indicate that the outward-open state binds cargo from the extracellular space (*out*) and the inward-open state has delivered the cargo to the cytoplasm (*in*).

are referred to as the thick gates (3). Finally, the occluded state possesses thick barriers blocking substrate escape to the cytoplasm or the extracellular space (8).

Although these x-ray structures are an invaluable starting point for understanding the transport mechanism, there are open questions concerning the dynamics of the process. Most importantly, there is little known about how transporters move from an outward-facing state to an inward-facing state. Does this transition occur in a highly concerted fashion like the movement of cogs in a watch, or are there many paths between these states with only loose coupling between all of the parts of the transporter? To address these questions, we used molecular dynamics (MD) to simulate the outward-to-inward transition of a coarse-grained (CG) model of the sodium-dependent benzylhydantoin transporter Mhp1. This is the only 5HIR member whose structure has been solved in the outward-open state (7) and inward-open state (9). Mhp1 mediates the uptake of nucleobases and related metabolites as part of a metabolic salvage pathway in *Microbacterium liquefaciens* and is a member of the nucleobase cation symporter-1 family of transporters (10).

Typical transport cycles occur on the millisecond to second timescale (11–14), which is well beyond the range of current fully atomistic membrane protein simulations that are on the order of hundreds of nanoseconds to microseconds. A common strategy to bridge the transport timescale is to use steered dynamics to force the protein into a different conformation or occupancy state (15,16). However, applying external forces to the system can bias the trajectories through unphysical pathways. Instead, we combined three alternative approaches to bridge these timescales: an efficient path sampling method that does not bias the dynamics, graphical processing unit accelerated dynamics, and a coarse-grained model of the transporter.

This simulation protocol has allowed us to generate thousands of unbiased outward-to-inward transitions, providing

us with a statistical sampling of the path ensemble in which we observe distinct structural pathways from the outward-open state to the inward-open state. Along the primary pathway, the outer gate closes before the inner gate opens, with trajectories passing through an occluded intermediate structure. Additionally, we predict a second noncanonical pathway in which both gates are open at the same time. Along this second pathway, a continuous path is created from the extracellular space to the intracellular space, and this may be an important property of 5HIR superfamily members, which are known to allow water permeation (12,17,18). This study serves as an initial point for probing the molecular nature of alternating access in other 5HIR superfamily members, and lays the foundation for subsequent studies with increased chemical accuracy and the inclusion of the substrate and energizing ions.

MATERIALS AND METHODS

Multistate structural model

To characterize the conformational transition from the outward- to inward-facing conformation, we constructed a CG multistate structure-based model of Mhp1. We employed a strategy based on a class of $\bar{G}\bar{O}$ -like models that incorporate information about the native structures in the CG force field (19,20). In these models, pairs of residues that interact in the experimental structure are stabilized by an attractive term in the potential energy, while nonnative interactions between CG sites are ignored. Additionally, bonded terms are treated using a set of standard molecular mechanics terms to constrain the local geometry of the protein. Although originally devised to study protein folding to a single native conformation, $\bar{G}\bar{O}$ models have subsequently been adapted to investigate switching between metastable states. One approach is to use a global mixing procedure to create an energy potential with two stable basins corresponding to the reference states of the protein (21–23). This procedure makes the transitions strongly cooperative because interactions arising from one reference state globally suppress interactions arising from the other. Here, we used an alternative approach, which allows a heterogeneous set of contacts from each reference state to form simultaneously (24–26). Mixing of contact formations in this manner allows the protein to locally deform and can give rise to metastable intermediates and heterogeneous pathways connecting the reference states. This flexibility may more accurately capture the motions of real proteins. Specifically, we used a united residue Double- $\bar{G}\bar{O}$ model with Gaussian contacts (26). The reference states were generated by selecting the C^α atoms from the Mhp1 x-ray structures in the outward-facing conformation (PDBID: 2JLN) and inward-facing conformation (PDBID: 2X79). The specific functional form and parameterization of the potential energy are given in the [Supporting Material](#).

Path sampling

Weighted ensemble (WE) simulation is an enhanced sampling method that tracks the evolution of a statistical ensemble of trajectories in conformation space. WE belongs to a large class of path sampling techniques including transition path sampling (27), transition interface sampling (28), forward flux sampling (29), milestone sampling (30), and nonequilibrium umbrella sampling (31) that have been developed to study systems with slow kinetics.

The conformational space separating the outward- and inward-facing states was partitioned into nonoverlapping bins using a set of progress coordinates. A group of N simulations, each with weight $1/N$, were initiated from the outward-open conformation using stochastic MD with distinct

random seeds. After a fixed time interval τ , the simulations were suspended, and for each replica, the full coordinates were projected onto the progress coordinate space to determine its assignment to one of the bins. We then applied a statistically rigorous procedure to maintain a constant number of replicas, M , in each populated bin. In bins containing fewer than M replicas, trajectories were split, and the weights were divided equally among the copies. Replicas in bins with more than M replicas were culled, and the weights of the terminated replicas were redistributed among the surviving trajectories in the bin. Iterating this procedure produces an ensemble of trajectories that begin in an initial state A and progress toward a target state B in an unbiased manner, while maintaining the correct statistical weight of all simulations. A detailed description of the theory and practical implementation of WE path sampling can be found elsewhere (32–34).

WE can be adapted to calculate the steady-state ensemble and reaction rate between two states (35,36). For a given set of initial and target states, the steady state may be obtained by instantaneously removing trajectories that enter the target state and placing them back in the starting state. Once a steady-state probability current into the target state is obtained, the rate constant k is given by the average of the probability flux into the target state I_B

$$k = \langle I_B \rangle \approx \left\langle \frac{P_B(\tau N_\tau)}{\tau} \right\rangle, \quad (1)$$

where τ is the weighted ensemble propagation/resampling time step and $P_B(\tau N_\tau)$ is the total weight of all replicas within the target state at time τN_τ immediately before recycling.

Initially, all of the probability resides in the initial state, and therefore, there is a finite relaxation time required for the probability to approach the steady-state distribution. The presence of one or more metastable intermediates along the transition pathway can significantly slow this relaxation. To eliminate the potentially long relaxation time, we reweighted the probability in each bin to approximate the steady-state probability distribution using the method in (35) and detailed in the Supporting Material.

Progress coordinates

We selected four progress coordinates that track conformational rearrangements in the transporter and enhance sampling between the outward- and inward-facing states. The choice of progress coordinates only effects the efficiency of the weighted ensemble method, but does not alter the results in the asymptotic limit of infinite sampling. The first coordinate is the distance-RMSD (dRMSD) to the target state (inward-open structure) defined for an instantaneous conformation of the CG model as

$$Q_{\text{dRMSD}} = \sqrt{\frac{2}{N \times (N - 1)} \sum_{i < j} (d_{i,j} - d_{i,j}^0)^2}, \quad (2)$$

where $d_{i,j}$ and $d_{i,j}^0$ are the instantaneous and reference distances between residues i and j , respectively. The remaining three progress coordinates track the movements of the inner gate, outer gate, and Na^+ -site gate as defined by Shimamura et al. (9). Gate distances are measured between the centers of mass of two groups of C^α atoms in each of the gate motifs. The extracellular gate opening distance, Q_{EC} , is defined as the distance between the TM9/10 loop (residues N360, T361, and F362) and residue I47 on the substrate bundle. The intracellular gate opening distance, Q_{IC} , measures the distance between I161, T162, and F163 on TM5 and D229, I230, and V231 on TM6. Finally, Q_{Na} is the distance between A309, S312, and T313 on TM8 and A38 and I41 on TM1. The values of each order parameter in the outward- and inward-open states as well as the rectilinear definitions of the initial and target states are shown in Table S1.

Simulation details

The multistate structural model was implemented in a custom simulation package built using the graphical processing unit-accelerated OpenMM package (37,38) and the PyOpenMM Python wrapper. The dynamics were propagated using a Langevin integrator at 300 K with a friction coefficient of 90 ps^{-1} and a timestep of 10 fs. Only the repulsive term of the potential was implemented as a nonbonded interaction, and it was truncated without switching at 5 Å. All other nonbonded interactions were modeled using the CustomBondForce class in OpenMM. For all nonbonded interactions, 1–4 exclusions were applied. The velocities of all replicas were initialized from a Maxwell-Boltzmann distribution.

RESULTS

Description and calibration of the model

To characterize the ensemble of dynamical transitions between the outward- and inward-facing conformations of Mhp1, we constructed a multistate coarse-grained molecular model based on the experimental structures of the transporter in two reference configurations (Fig. 2).

We initiated twelve 100 ns brute-force simulations from each of the reference conformations, totaling $2.4 \mu\text{s}$, to assess the stability of the model. Interestingly, both starting

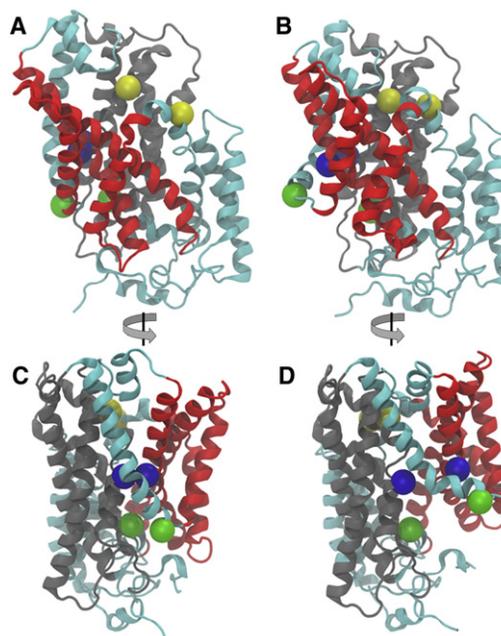


FIGURE 2 Conformational change between the inward-facing and outward-facing states. (A and C) Ribbon diagram of the outward-facing state of Mhp1 (PDB ID: 2JLN). The figure in panel A is rotated 90 degrees about the long axis of the protein compared to the figure in panel C. The centers of mass of the two groups used to define each of the gate distance progress coordinates are shown as spheres (yellow: extracellular gate; green: intracellular gate; blue: Na^+ -site gate). It can be seen that the extracellular gate is open and the intracellular gate is closed. The hash motif is red, substrate bundle is gray, and the flexible and noncore helices are cyan. (B and D) Model of target inward-open state of Mhp1 (PDB ID: 2X79). The coloring scheme and geometries are the same as in panels A and C. For all four panels, Mhp1 is oriented so that the cytoplasmic side is at the bottom.

structures quickly transitioned to a strongly metastable intermediate in which both the extra- and intracellular gates are closed. This intermediate is structurally similar to the known outward-occluded conformation of Mhp1 (7), yet we did not include any information about the outward-occluded structure in the potential energy function. Many 5HIR superfamily members have been crystallized in occluded states suggesting they are stable intermediates in the transport cycle (5,7,8). A quick closure of the gates occurs because the CG energy potential stabilizes native contacts in the reference structures, and many pairs of residues throughout both the intra- and extracellular gates form contacts in the closed, but not open states; this energetically favors closed gate conformations. Nonetheless, we expected the outward-open and inward-open states to be metastable on such short timescales because these configurations are stable in the crystals, so we recalibrated the potential energy function to make each reference state metastable by making a subset of the contacts bistable subject to a longer cutoff distance (see [Supporting Material](#)). This extended cutoff distance was tuned to roughly match the fluctuations of the intra- and extracellular gate distances observed in all-atom equilibrium MD simulations of Mhp1 in each reference state (9). Using an extended cutoff of 14 Å, simulations initiated from each reference structure maintain stable RMSD values of 1.7 ± 0.5 Å and 1.6 ± 0.4 Å from the starting coordinates for the outward- and inward-facing structures, respectively. Both RMSD measures and residue RMSF profiles for our multistate CG model are similar to those observed for all-atom simulations of Mhp1 (9) indicating that our current model captures the correct degree of protein flexibility and provides a partial validation of its correctness.

Brute force simulations were then used to quantify the fluctuations of the progress coordinates around the initial and final conformations. On the basis of these fluctuations, we demarcated native basins around the reference states that define when a trajectory enters or exits these states. In our subsequent analysis of the transition path ensemble, we exclude the finite time that trajectories dwell in the initial basin. The distribution for each progress coordinate is shown in [Fig. S1](#) in the [Supporting Material](#). As expected, the brute force simulations do not undergo transitions within 100 ns.

The outward-to-inward transition rate rapidly converges

In an attempt to observe the outward-to-inward transition, we initiated dynamic simulations from 30 replicas of the outward-facing conformation pausing at 20 ps intervals, referred to as τ , to split and cull trajectories in a manner that preserves the true steady-state probabilities according to the WE method. Near 50 τ , we observed that the first trajectories reach the target inward-facing state as defined

by the four progress coordinates (see [Table S1](#)). Although the flux of probability into the target state over the first 200 τ is increasing, indicating that we are able to simulate the transition, the arrival rate does not reach steady state in this time frame. Because it is likely that the initial fast, low-weight trajectories are not representative of the true steady-state path ensemble, we applied a reweighting procedure (35) based on the dynamics observed in the first 200 τ to hasten the convergence of the WE simulations. Readjusting the weight of the replicas does not perturb the underlying dynamics of the system, but as shown in [Fig. 3](#) the procedure abolishes the slow transient in the arrival flux into the target state. After reweighting, the flux into the target state is stationary for the remainder of the WE simulation, strongly indicating that the system is at steady state. A second independent WE simulation, also reweighted at 200 τ , produced a nearly identical steady-state flux suggesting that the convergence properties of the simulation are correct and that the ensemble of pathways has been appropriately sampled (also see [Fig. 4](#) for additional support of this claim). Ignoring the first 400 τ , we calculated the rate of the outward-to-inward transition to be 23.9 ± 1.7 ms⁻¹ using [Eq. 1](#). Both independent data sets were used in the calculation, and the error represents the 95% confidence interval calculated using Monte Carlo Bootstrapping (32,39). Related 5HIR family members are reported to undergo a much slower outward-to-inward transition at a rate of 50 s⁻¹ (14); however, the rate of this transition for Mhp1 has not been experimentally determined. Because our model lacks chemical detail and neglects nonnative interactions, it is unlikely that our computed rate is a true estimate of the experimental rate. It is important to note that we use the convergence of the rate primarily as an

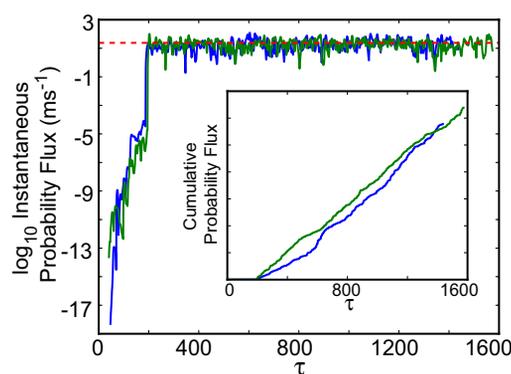


FIGURE 3 Convergence of probability flux into the target state. The probability flux into the target state is shown for each of the two independent WE simulations. Fluxes were window-averaged with a 10 τ window size. After ~ 200 τ , each system was reweighted resulting in a rapid convergence to steady state. The average flux into the target state resulting from aggregating results from both simulations is shown as the dashed line for reference. (*Inset*) Cumulative flux into the target state as a function of simulation time τ plotted on a linear scale. For both simulations, the fluxes reach a linear regime after reweighting.

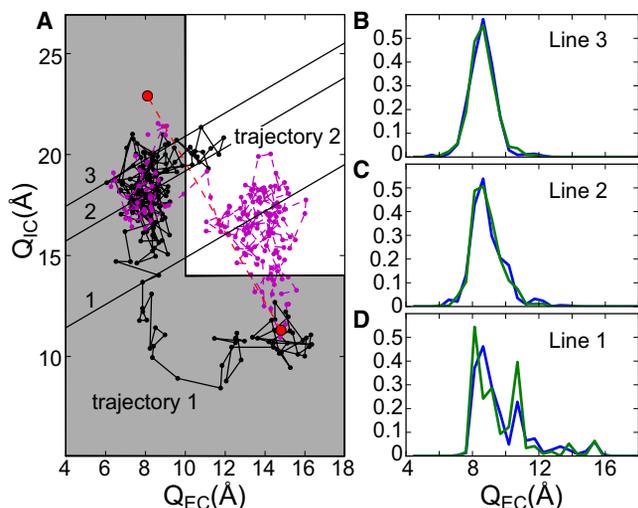


FIGURE 4 Structural distribution of the path ensemble from the outward-open to inward-open conformations of Mhp1 from weighted ensemble simulations. (A) Two representative pathways projected onto the intra- and extracellular gate opening distances (Q_{IC} and Q_{EC} , respectively). The two paths illustrate different mechanisms contained within the path ensemble. Trajectory 1 resides almost exclusively in the shaded region corresponding to extracellular gate closure preceding intracellular gate opening, whereas for trajectory 2 both gates are open simultaneously for part of the transition. (B–D) The ensemble distribution of Q_{EC} values along the dividing lines in A for each of the two WE simulations. The histograms are computed from the last crossing point of each trajectory on the line. The similarities between the two distributions in panels B–D serve as additional verification of the convergence of the path ensemble.

indicator that the simulations have reached steady state and not to predict a theoretical value to compare with experiment. We have not tuned the model to reproduce a specific transition rate, and in general, unless care has been given to parameterizing the barrier height between states, CG models will not reproduce experimentally measured rates (see the [Supporting Material](#) for further discussion).

Not all paths obey a strict alternating access model

The alternating access mechanism, as envisioned by Jardetzky (4), requires at least two states for operation: one in which a molecular cavity is open to one side of the membrane, say the outward-open state, and a second where the cavity is open to the opposite side, say the inward-open state. Jardetzky did not suggest how the transporter would interconvert between these two states, but it is commonly believed that one cavity should close before the second cavity opens, because there is the potential for uncoupling the movement of the energizing ions and the transported substrate. If this is indeed how transporters operate, then a critical state in the transition is an intermediate occluded state in which both the thick inner and thick outer gates are simultaneously closed (3). Comparison of the high-resolution structures of Mhp1 in the outward-open, substrate

free, and outward-occluded, substrate bound conformations (7) as well as spectroscopic measurements on the LeuT outer gate closing in the presence of leucine (40) have demonstrated that the occluded state represents a dominant intermediate along the transport cycle. Furthermore, the high-resolution structure of the glycine betaine transporter BetP reveals a substrate binding site that is inaccessible to the solvent on both sides and was interpreted as an occluded intermediate (8). Nonetheless, equilibrium simulations of Mhp1 indicate that the extracellular gate is highly mobile even in the outward-occluded conformation (9), and simulations of vSGLT show that flexibility of the extracellular gate even persists in the inward-occluded state because the gate moves enough to allow water to permeate the substrate binding site (18). Thus, it is important to determine the coordination between the inner and outer gates.

The WE simulations of the CG model provide a rigorous sampling of the transition path ensemble connecting the outward-open and inward-open conformations, allowing us to directly investigate the coordination of the extra- and intracellular gates during transport. For each reactive trajectory that makes a successful transition between the outward-open and inward-open conformation, the coordinates of the system were projected onto the gate opening distances Q_{EC} and Q_{IC} . The distance Q_{Na} closely tracks Q_{IC} so we excluded it from our analysis. How successful trajectories progress through the Q_{EC} , Q_{IC} conformational space is informative of the transport mechanism.

To classify whether a trajectory strictly adheres to the canonical alternating access mechanism in which the outer gate closes before the inner gate opens, we defined a region of conformational space in which at most only one gate is open at a time using the equilibrium gate fluctuations around the initial and target states as a criteria for closure. Although the true openness of each gate depends on the full atomistic configuration, we believe that these single distances provide a good approximation of the degree of openness given the CG nature of our model. The region corresponding to strict alternating access is shaded in Fig. 4 A and is defined as those conformations in which $Q_{EC} < 10.0 \text{ \AA} \vee Q_{IC} < 14.0 \text{ \AA}$. For two representative trajectories shown in Fig. 4 A, we can see that trajectory 1 adheres to the canonical mechanism while trajectory 2 enters a region in which both gates are open at the same time. The heterogeneity of the transition is evident in the distribution of Q_{EC} values perpendicular to the line joining the inward- and outward-facing states in Fig. 4 D (line 1 in panel A). The larger peak near 8 Å adheres to a strict alternating access mechanism, like trajectory 1, and represents 68% of the transitions. Meanwhile, the shoulder to the right at 11 Å is representative of trajectory 2 in which the outer gate stays open as the inner gate opens. The histograms show that there is a heterogeneous path ensemble in which the majority of trajectories follow a canonical pathway, but a nonnegligible population of the trajectories does not adhere to a strict

alternating access mechanism. The Q_{EC} distributions along lines closer to the inward-facing state shift into the region where both gates are closed, increasing to 92% at line 2 and 97% at line 3. The focusing of the trajectories indicates that the extracellular gate is uncoupled near the outward-open conformation of Mhp1, but is more constrained in its motions as the conformational transition approaches the inward-open state in agreement with all-atom simulations of Mhp1 (9).

The hash and substrate motifs act as a moving bundle and stationary scaffold

A static comparison of the high-resolution structures of Mhp1 suggests that during the transport cycle the hash motif (TMs 3,4,8,9) and the substrate bundle (TMs 1,2,6,7) move relative to each other as semirigid domains (9). The resulting model of the conformational changes driving transport resembles the rocking bundle mechanism first proposed by Forrest et al. based on symmetry of the inverted helical motifs in LeuT (41,42). In the galactose transporter vSGLT, superposition of the equivalent subdomains suggests that the inward-occluded to inward-open transition also occurs by similar rigid-body movements (43).

Although the rocking bundle seems reasonable based on end point structures, it is quite possible that specific rearrangements, such as the bending of a TM segment or the separation of closely packed helices, are required for the outward-to-inward transition to take place. We critically probed the rocking bundle hypothesis by examining the hash motif and substrate bundle during transitions from the outward-open to the inward-open conformation of Mhp1 to determine if the domains are rigid throughout the conformational change. For each successful trajectory, the RMSD of the hash motif and substrate bundle were calculated after optimal superposition of either 1) the instantaneous coordinates of the hash motif onto the hash residues of the reference start state or 2) the instantaneous coordinates of the substrate bundle onto the substrate bundle residues in the reference start state. All RMSDs and superpositions were calculated using the fast QCP method (44,45). We then determined the distribution of observed RMSDs for each domain, $P(\text{RMSD}) = \sum_i^N w_i p_i(\text{RMSD})$, where p_i is the distribution of RMSDs in trajectory i and w_i is the normalized weight of the trajectory. The resulting distributions are shown in Fig. 5. Fig. 5 A shows the distributions for each domain when it is first superposed onto itself. If a given domain remains absolutely rigid during the course of the dynamics, we would expect to see a sharply peaked distribution near the origin. In fact that is what we observe in panel A; the mean RMSD of the substrate bundle is 1.2 Å, whereas the hash motif has slightly more internal flexibility with a mean RMSD of 1.6 Å. Meanwhile, when we superpose the substrate bundle (hash motif) onto the substrate bundle (hash motif) of the initial structure and

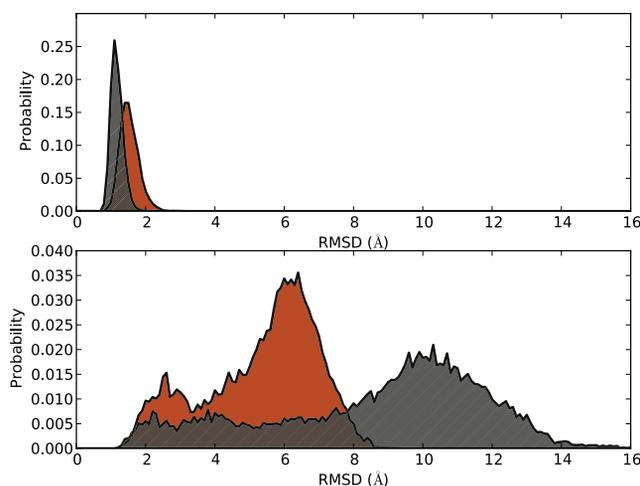


FIGURE 5 Rigid body motions of domains. (A) The distribution of RMSD values for the hash motif and substrate bundle superposed onto themselves. Snapshots along each productive transition were collected every 100 ps, and the hash motif was superposed onto the hash motif of the starting structure and the RMSD recorded. We repeated this process for the substrate bundle. The distributions are similar when superpositions are carried out onto the target state structure. Values for the hash motif are shown in red, and values for the substrate bundle are shown in gray with hatching. (B) The distribution of RMSD values for one domain after superposition of the other. Superpositions were carried out as in panel A, but then RMSD values were computed for the nonsuperposed domain. Values for the hash motif are shown in red, and values for the substrate bundle are shown in gray with hatching.

calculate the RMSD of the hash motif (substrate bundle), we see a very different distribution. As shown in Fig. 5 B, the opposing unaligned domain deviates substantially from the reference position, resulting in a broad distribution of RMSDs. Taken together, these distributions indicate that both the hash motif and substrate bundle remain relatively rigid throughout the entire transition and the conformational change is a result of their movement relative to one another.

Time ordering of critical conformational transitions

It is possible that the 5HIR architecture encodes a very specific mechanical escapement required for movement between the dominant states. To address this possibility, we systematically explored the transition path ensemble to attempt to identify regions of the transporter whose movements are ordered in time. Such analysis, if successful, has the potential to identify latch regions that must first move before other portions of the protein can undergo the outward-to-inward transition. We carried out this analysis at the finest level of detail using individual residue contacts; however, such analysis presents us with an overwhelming amount of information. To synthesize this large data set, we developed a variation of the contact appearance order metric originally developed to investigate the mechanism

of protein folding using discrete space $G\bar{o}$ models (46). For each transition, we determined the order in which residue-residue contacts in the target state were permanently formed and contacts in the initial state were last permanently broken. This metric does not resolve the unproductive or transient formation or loss of contacts; instead it emphasizes conformational changes in the transporter that are persistent. For each successful trajectory we recorded the time when each *B*-state contact last stably formed, and likewise, the time that each *A*-state contact permanently dissociated. These association/dissociation event times were then temporally ordered for each transition using the fractional ranking method, and a distribution was calculated for each contact. This distribution gives the probability that the contact between residues *i* and *j* permanently associates/dissociates *N*th among all contacts.

In this analysis, we considered only contacts whose separation distances, $\Delta d_{i,j}^{A,B} = |d_{i,j}^A - d_{i,j}^B|$, change by $> 3 \text{ \AA}$ between the outward- and inward-open states. Loss of *A*-state contacts were identified when $d_{i,j} > d_{i,j}^A + \min\{1.7, \Delta d_{i,j}^{A,B}/2\} \text{ \AA}$. Similarly we considered *B*-state contacts to be formed when $d_{i,j} < d_{i,j}^B + \min\{1.7, \Delta d_{i,j}^{A,B}/2\} \text{ \AA}$. Furthermore, $d_{i,j}$ was averaged over a 10τ moving window to suppress transient fluctuations.

The distributions for gain of *B*-state contacts and loss of *A*-state contacts are shown in the upper and lower panels of Fig. 6, respectively. Although all contacts were considered simultaneously during the initial rank ordering, we separated the two classes of contacts for clarity. Additionally, in both panels of Fig. 6, the first contacts to form or break were sorted starting from the bottom to the top.

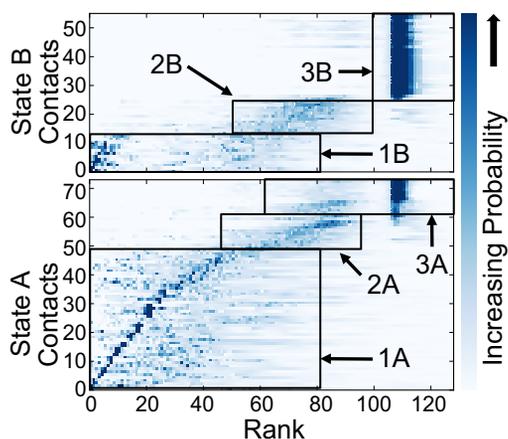


FIGURE 6 Time ordering of critical events. The temporal ordering histograms for the loss of State *A* contacts (*lower panel*) and gain of State *B* contacts (*upper panel*) are shown as density maps. Darker colors correspond to a higher probability for a contact to form/dissociate at a given rank. In both panels, the first contacts to form (*upper*) or break (*lower*) were sorted starting from the bottom to the top. Groups of contacts have been boxed off depending on whether they break/form early (1), at an intermediate time (2), or late (3).

The time ordering of the loss of State *A* contacts and the gain of State *B* contacts can be divided into three distinct groups indicated by the boxes in Fig. 6. Initially, TM9 and 10 of the extracellular gate close making contacts with TM1 and TM6 in the substrate bundle (group 1B), which is consistent with our analysis in Fig. 4. However, the distributions for these contacts are bimodal because the extracellular gate sometimes closes after a rigid body rotation of the hash motif about TM 3 that causes TM 8 to lose contact with the substrate bundle (group 1A). As the rigid body rotation of the hash motif completes, TM5 bends to further open the intracellular gate. This bending results in a loss of contacts between the intracellular half of TM5 and intracellular residues in TM1 and 8 (group 2A). TM5 bending also results in the subsequent formation of a different set of contacts between TM5 and TM1 and 8 (group 2B), further opening the intracellular cavity. The transition ends when the small extracellular helix connecting TM7 to TM8 fully converts to the target state (groups 3A and 3B). Interestingly, for a fraction of transitions the movement of the extracellular helix is concomitant with the bending of TM5 as can be seen by comparing groups 2A and 3A.

DISCUSSION

The recently determined x-ray structures of 5HIR transporters have provided a wealth of information concerning the ion-coupled transport of small molecules across the cell membrane. However, a deeper understanding of the transport process requires a detailed picture of how the proteins move between their primary stable states. Simulation can provide a high resolution method for exploring these transitions, but the long time required for major protein conformational changes to occur makes brute force simulations and fully atomistic simulations computationally intractable. Here, we developed a multistate, structure-based model that captures important aspects of Mhp1's topology and flexibility, while coarse-graining some atomic details to allow us to simulate the transition in a reasonable time frame. These simulations do not directly address how the binding of substrate and ion drive the conformational change; instead we focus on the coordination of the intracellular and extracellular gates and global rearrangement of the transporter during the transition. We used this model to investigate the primary substrate transport step that moves the transporter from the outward- to the inward-facing conformation.

Our analysis of the unbiased ensemble of reactive trajectories reveals that two distinct pathways connect the outward- and inward-open states. Along the dominant pathway, the extracellular gate closes before the intracellular gate opens producing an occluded intermediate. This predicted occluded state has been experimentally observed (7), but we did not use this structure in building our energy potential. Although the initial alternating access model (4)

did not include an occluded state, subsequent work has advanced the idea that this state is important to the transport process (1,3,47). In a smaller, but nonnegligible subset of the pathways, the extracellular gate remains partially open even as the intracellular gate begins to open. Nonetheless, as shown in Fig. 4, the ensemble of pathways merges into a single reactive channel as they approach the inward-open state. This merging may effectively restrict the size of a putative membrane-spanning pore through Mhp1 during transport. A transiently open, but partially restricted, channel through the protein might allow selective permeation of water, ions, and other small molecules, while preventing escape of the substrate to the extracellular space. Uncoupled ion and water permeation through active transporters have been experimentally measured (12,17,48) and observed in simulation (18), and water permeation might arise either from incompletely occluded states similar to those observed in our simulations or alternative permeation pathways.

Additionally, our coarse-grained simulations support a rocking bundle mechanism, in which the hash motif and substrate bundle move as rigid domains relative to each other during the transition from the outward-open to inward-open conformation. This rigid body rotation of the hash motif occurs by pivoting along TM3, and the entire motion is accompanied by a bending of TM5. TM10 is the symmetry-related partner of TM5, and while TM10 bends to open the extracellular gate, this bending is not strictly coupled to the rocking bundle transition.

Our methodology and analysis provide a framework for understanding the conformational transitions that occur in transport proteins. At the core of our approach is the ability of the weighted ensemble method to provide an efficient way of extracting the path ensemble without introducing bias in the dynamics. Convergence of the path ensemble allowed us to quantitatively analyze the conformational change, providing a statistical picture of the process that captures the heterogeneity of pathways and permits us to quantify each pathway's relative importance to the transition. Although little functional data are available for Mhp1, we believe that our results will be applicable to other 5HIR transporters, and therefore, some of our claims can be tested in members that are more amenable to experimentation. For instance, if the hash motif and the substrate bundle are truly rigid throughout the transition, then cross-linking pairs of residues within the same domain should have little effect on transport, whereas cross-linking residue pairs between domains should destroy transport. Moreover, several emerging experimental techniques such as single molecule fluorescence resonance energy transfer and double electron-electron resonance have been used to investigate other 5HIR family members (40,49) and can potentially be used to observe subpopulations that adopt different conformations along the transition pathway in response to changes in substrate occupancy.

One of the goals to understanding sodium-coupled transport is to elucidate how substrate and ion binding drive the alternating access process. Related to this, the order of binding and protein mediated allosteric coupling between the binding sites is of extreme importance. The coarse-grained nature of our current simulations does not allow us to directly answer these questions since the ion and substrate were not explicitly included. Nonetheless, we imagine that the gross conformational features described here will be present in more detailed models of the alternating access mechanism. Importantly, our framework is easily generalized to varying levels of atomistic detail, other steps of the transport cycle, or different complex biomolecular systems that undergo conformational transitions on timescales inaccessible via long brute-force simulations.

SUPPORTING MATERIAL

Additional information about the steady-state simulations, rate calculation and multistate structural model, along with a table, a figure, and references are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(11\)01243-4](http://www.biophysj.org/biophysj/supplemental/S0006-3495(11)01243-4).

We thank Seungho Choe, Jeff Abramson, and Ernest Wright for ongoing discussions concerning the alternating access mechanism. We also thank Peter Eastman, Randy Radmer, John Chodera and Vijay Pande for assistance with the GPU implementation. We thank Kenneth Jordan, Akila Gothandaraman and the University of Pittsburgh Center for Simulation and Modeling for computational resources.

This work was supported by a grant from FundScience and a National Institutes of Health (NIH) postdoctoral fellowship (T32-DK061296) to J.L.A., an NIH grant (R01GM89740-1A1) to M.G., a National Science Foundation CAREER Award (MCB-0845216) to L.T.C., and an NIH grant (R01GM070987) to D.M.Z. M.C.Z. was supported by an Arts and Sciences Fellowship from the University of Pittsburgh, and M.G. is an Alfred P. Sloan Research Fellow.

REFERENCES

1. Abramson, J., and E. M. Wright. 2009. Structure and function of Na(+)-symporters with inverted repeats. *Curr. Opin. Struct. Biol.* 19:425–432.
2. Forrest, L. R., R. Krämer, and C. Ziegler. 2011. The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta.* 1807:167–188.
3. Krishnamurthy, H., C. L. Piscitelli, and E. Gouaux. 2009. Unlocking the molecular secrets of sodium-coupled transporters. *Nature.* 459:347–355.
4. Jardetzky, O. 1966. Simple allosteric model for membrane pumps. *Nature.* 211:969–970.
5. Yamashita, A., S. K. Singh, ..., E. Gouaux. 2005. Crystal structure of a bacterial homologue of Na⁺/Cl⁻-dependent neurotransmitter transporters. *Nature.* 437:215–223.
6. Faham, S., A. Watanabe, ..., J. Abramson. 2008. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na⁺/sugar symport. *Science.* 321:810–814.
7. Weyand, S., T. Shimamura, ..., A. D. Cameron. 2008. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science.* 322:709–713.

8. Ressler, S., A. C. Terwisscha van Scheltinga, ..., C. Ziegler. 2009. Molecular basis of transport and regulation in the Na⁺/betaine symporter BetP. *Nature*. 458:47–52.
9. Shimamura, T., S. Weyand, ..., A. D. Cameron. 2010. Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science*. 328:470–473.
10. Suzuki, S., and P. J. F. Henderson. 2006. The hydantoin transport protein from *Microbacterium liquefaciens*. *J. Bacteriol.* 188:3329–3336.
11. Mackenzie, B., D. D. Loo, ..., E. M. Wright. 1996. Mechanisms of the human intestinal H⁺-coupled oligopeptide transporter hPEPT1. *J. Biol. Chem.* 271:5430–5437.
12. Loo, D. D., T. Zeuthen, ..., E. M. Wright. 1996. Cotransport of water by the Na⁺/glucose cotransporter. *Proc. Natl. Acad. Sci. USA*. 93:13367–13370.
13. Jung, H., S. Tebbe, ..., K. Jung. 1998. Unidirectional reconstitution and characterization of purified Na⁺/proline transporter of *Escherichia coli*. *Biochemistry*. 37:11083–11088.
14. Loo, D. D. F., B. A. Hirayama, ..., E. M. Wright. 2006. Conformational dynamics of hSGLT1 during Na⁺/glucose cotransport. *J. Gen. Physiol.* 128:701–720.
15. Shaikh, S. A., and E. Tajkhorshid. 2010. Modeling and dynamics of the inward-facing state of a Na⁺/Cl⁻ dependent neurotransmitter transporter homologue. *PLOS Comput. Biol.* 6:e1000905.
16. Shi, L., M. Quick, ..., J. A. Javitch. 2008. The mechanism of a neurotransmitter:sodium symporter—Inward release of Na⁺ and substrate is triggered by substrate in a second binding site. *Mol. Cell*. 30:667–677.
17. Cao, Y., S. Mager, and H. A. Lester. 1997. H⁺ permeation and pH regulation at a mammalian serotonin transporter. *J. Neurosci.* 17:2257–2266.
18. Choe, S., J. M. Rosenberg, ..., M. Grabe. 2010. Water permeation through the sodium-dependent galactose cotransporter vSGLT. *Biophys. J.* 99:L56–L58.
19. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.
20. Taketomi, H., Y. Ueda, and N. Gö. 1975. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* 7:445–459.
21. Hummer, G., L. Pratt, and A. García. 1997. Multistate Gaussian model for electrostatic solvation free energies. *J. Am. Chem. Soc.* 119:8523–8527.
22. Okazaki, K.-i., N. Koga, ..., P. G. Wolynes. 2006. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA*. 103:11844–11849.
23. Yang, S., and B. Roux. 2008. Src kinase conformational activation: thermodynamics, pathways, and mechanisms. *PLOS Comput. Biol.* 4:e1000047.
24. Zuckerman, D. M. 2004. Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. *J. Phys. Chem. B*. 108:5127–5137.
25. Takagi, F., and M. Kikuchi. 2007. Structural change and nucleotide dissociation of Myosin motor domain: dual go model simulation. *Biophys. J.* 93:3820–3827.
26. Lammert, H., A. Schug, and J. N. Onuchic. 2009. Robustness and generalization of structure-based models for protein folding and function. *Proteins*. 77:881–891.
27. Bolhuis, P. G., D. Chandler, ..., P. L. Geissler. 2002. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* 53:291–318.
28. Van Erp, T., D. Moroni, and P. Bolhuis. 2003. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* 118:7762–7774.
29. Allen, R. J., D. Frenkel, and P. R. ten Wolde. 2006. Simulating rare events in equilibrium or nonequilibrium stochastic systems. *J. Chem. Phys.* 124:024102.
30. Faradjian, A. K., and R. Elber. 2004. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* 120:10880–10889.
31. Warmflash, A., P. Bhimalapuram, and A. R. Dinner. 2007. Umbrella sampling for nonequilibrium processes. *J. Chem. Phys.* 127:154112.
32. Huber, G. A., and S. Kim. 1996. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* 70:97–110.
33. Zhang, B. W., D. Jasnow, and D. M. Zuckerman. 2007. Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. *Proc. Natl. Acad. Sci. USA*. 104:18043–18048.
34. Zhang, B. W., D. Jasnow, and D. M. Zuckerman. 2010. The “weighted ensemble” path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. *J. Chem. Phys.* 132:054107.
35. Bhatt, D., B. W. Zhang, and D. M. Zuckerman. 2010. Steady-state simulations using weighted ensemble path sampling. *J. Chem. Phys.* 133:014110.
36. Zwier, M., J. Kaus, and L. Chong. 2011. Efficient explicit-solvent molecular dynamics simulations of molecular association kinetics: methane/methane, Na⁺/Cl⁻, methane/benzene, and K⁺/18-crown-6 ether. *J. Chem. Theory Comput.* 7:1189–1197.
37. Friedrichs, M. S., P. Eastman, ..., V. S. Pande. 2009. Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.* 30:864–872.
38. Eastman, P., and V. S. Pande. 2010. Efficient nonbonded interactions for molecular dynamics on a graphics processing unit. *J. Comput. Chem.* 31:1268–1272.
39. Press, W. H., S. A. Teukolsky, ..., B. P. Flannery. 2007. Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press, New York.
40. Claxton, D. P., M. Quick, ..., H. S. McHaourab. 2010. Ion/substrate-dependent conformational dynamics of a bacterial homolog of neurotransmitter:sodium symporters. *Nat. Struct. Mol. Biol.* 17:822–829.
41. Forrest, L. R., Y.-W. Zhang, ..., G. Rudnick. 2008. Mechanism for alternating access in neurotransmitter transporters. *Proc. Natl. Acad. Sci. USA*. 105:10338–10343.
42. Forrest, L. R., and G. Rudnick. 2009. The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology (Bethesda)*. 24:377–386.
43. Watanabe, A., S. Choe, ..., J. Abramson. 2010. The mechanism of sodium and substrate release from the binding pocket of vSGLT. *Nature*. 468:988–991.
44. Theobald, D. L. 2005. Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. *Acta Crystallogr. A*. 61:478–480.
45. Liu, P., D. K. Agrafiotis, and D. L. Theobald. 2010. Fast determination of the optimal rotational matrix for macromolecular superpositions. *J. Comput. Chem.* 31:1561–1563.
46. Gin, B. C., J. P. Garrahan, and P. L. Geissler. 2009. The limited role of nonnative contacts in the folding pathways of a lattice protein. *J. Mol. Biol.* 392:1303–1314.
47. Csanády, L., and J. A. Mindell. 2008. The twain shall meet: channels, transporters and things between. Meeting on membrane transport in flux: the ambiguous interface between channels and pumps. *EMBO Rep.* 9:960–965.
48. Hirayama, B. A., D. D. Loo, and E. M. Wright. 1994. Protons drive sugar transport through the Na⁺/glucose cotransporter (SGLT1). *J. Biol. Chem.* 269:21407–21410.
49. Zhao, Y., D. S. Terry, ..., J. A. Javitch. 2011. Substrate-modulated gating dynamics in a Na⁺-coupled neurotransmitter transporter homologue. *Nature*. 474:109–113.