Structural Determinants of Water Permeation through the Sodium-Galactose Transporter vSGLT

Joshua L. Adelman,[†] Ying Sheng,[†] Seungho Choe,[‡] Jeff Abramson,[§] Ernest M. Wright,[§] John M. Rosenberg,[†*] and Michael Grabe[†¶*]

[†]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania; [‡]School of Basic Science, College of Convergence, Daegu Gyeongbuk Institute of Science & Technology, Daegu, Korea; [§]Department of Physiology, University of California, Los Angeles, California; and [¶]Cardiovascular Research Institute, Department of Pharmaceutical Chemistry, University of California, San Francisco, California

ABSTRACT Sodium-glucose transporters (SGLTs) facilitate the movement of water across the cell membrane, playing a central role in cellular homeostasis. Here, we present a detailed analysis of the mechanism of water permeation through the inward-facing state of vSGLT based on nearly 10 μs of molecular dynamics simulations. These simulations reveal the transient formation of a continuous water channel through the transporter that permits water to permeate the protein. Trajectories in which spontaneous release of galactose is observed, as well as those in which galactose remains in the binding site, show that the permeation rate, although modulated by substrate occupancy, is not tightly coupled to substrate release. Using a, to our knowledge, novel channel-detection algorithm, we identify the key residues that control water flow through the transporter and show that solvent gating is regulated by side-chain motions in a small number of residues on the extracellular face. A sequence alignment reveals the presence of two insertion sites in mammalian SGLTs that flank these outer-gate residues. We hypothesize that the absence of these sites in vSGLT may account for the high water permeability values for vSGLT determined via simulation compared to the lower experimental estimates for mammalian SGLT1.

INTRODUCTION

The Na^+-glucose transporter (SGLT) is a member of a broad class of cotransporters that harness the energy stored in ionic gradients to drive the transport of substrates across membranes. Specifically, SGLTs are symporters that use sodium gradients to pump simple sugars into the cell (1,2). In addition to its primary role as a sugar transporter, it is well accepted that mammalian SGLTs, such as human SGLT isoform 1 (hSGLT1), are permeable to water (3–7). Additionally, intestinal water absorption is closely linked to the coupled absorption of salt and sugar, and this provides the basis for oral rehydration therapy (ORT). It is estimated that ORT has reduced the mortality of children with acute diarrhea from 4.6 million in 1980 to 1.5 million in 2000 (8). Although there is no dispute that glucose-stimulated water absorption is directly linked to Na^+/glucose cotransport by SGLTs, the exact nature of the coupling is controversial.

There is broad consensus that SGLTs facilitate passive water flow across membranes in response to osmotic gradients (5–7,9), and this property has been observed for several other transporters (10). Passive water flow is independent of the presence or absence of galactose, but it is sensitive to the conformational state of the transporter, since flow is blocked by inhibitors known to trap SGLTs in an outward-facing conformation (11). In addition to its role as a passive water channel, it has been observed that SGLT1 undergoing normal transporter turnover energized by a Na^+ or glucose gradient can move water from a region of low osmolyte concentration to a region of high osmolyte concentration. Two general classes of model have been developed to explain this observation: molecular-based models involving the cotransport of water (4,11–14), or an unstirred-boundary-layer model, which posits the existence of local osmotic gradients due to the build-up of sugar near the membrane (5,6,9,15).

Experiments supporting the water cotransport hypothesis show that each transport cycle of SGLT1 delivers 260 water molecules into the cell, along with two Na^+ and one sugar, in a tightly coupled manner (4). These models rely on the alternating-access mechanism (1,16), in which a water-filled cavity is first presented to the extracellular solution to bind ion and substrate (outward-facing state), and then closed to the extracellular solution and opened toward the interior of the cell (inward-facing state) via a rocking mechanism. The water-filled cavities contain a substantial number of water molecules, and the inward movement of ion and substrate moves these water molecules with them via a number of potential mechanisms nicely outlined by Zeuthen and MacAulay (17). Contrary to the cotransport models, the unstirred-boundary-layer hypothesis suggests that SGLTs are not able to actively pump water molecules, but rather that water movement occurs passively due to the establishment of a local osmolyte concentration gradient (5,6,9,15). In this model, SGLT activity increases the intracellular concentration of Na^+ and sugar at the membrane due to slow intracellular diffusion while at the same time depleting the extracellular concentration. Together, these changes set up an osmotic gradient that is favorable for water entry into the cell through a conformational state of the transporter, which is permeable.

Submitted October 31, 2013, and accepted for publication January 2, 2014.

*Correspondence: jmr@pitt.edu or michael.grabe@ucsf.edu

Editor: Gerhard Hummer.

0006-3495/14/03/1280/10 $2.00

http://dx.doi.org/10.1016/j.bpj.2014.01.006

In the absence of a high-resolution structure of hSGLT1, structures of the bacterial homolog from *Vibrio parahaemolyticus*, vSGLT, provide a useful model for beginning to understand the molecular underpinning of water transport in the broader sodium solute symporter families (SSS and NSS). vSGLT shares significant sequence identity with SGLT1 and retains even higher conservation among key residues involved in gating and ligand binding (2,18–23). Unfortunately, no experimental data exist for the water permeability of vSGLT due to its poor expression levels in *Xenopus* oocytes (24), thus preventing a direct comparison. However, all-atom molecular dynamics (MD) simulation coupled with the crystallographic structure of vSGLT provides a way of tracking the motions of the solvent, substrate, and transporter with the high temporal and spatial resolution required to provide a dynamical account of their coupling. Using this approach, our group performed, to our knowledge, the first simulations showing the existence of an aqueous pathway traversing vSGLT (14). This pathway allows water to permeate the inward-facing state. This observation was later confirmed by other groups (15,25,26).

In this study, we used trajectories collected from a series of MD simulations, each 500–1800 ns in length, to sample water flow through the inward-facing state of vSGLT before, during, and after substrate release. This ensemble of 21 simulations, which includes 10 spontaneous sugar escapes, permits a careful analysis of the role of both sugar and transporter motions in controlling water flow through the transporter. We observe the formation of transient water-conducting states in which residues in the outer gate of the transporter undergo a conformational transition to open a pore connecting the aqueous lumen of vSGLT to the extracellular bulk. Although sugar does not influence the passive water permeability of SGLT1, our simulations show that bound galactose occludes the water channel, reducing permeability. Fluctuations in the inward and outward flow of water give rise to net water movement over short periods of time, but this net movement occurs outward as much as inward, as we would expect at equilibrium, and it was poorly correlated with galactose exit from the binding site. Based on the location of key residues along the water permeation pathways and multiple sequence alignments, we propose that two large, conserved extracellular loops found in SGLT1 and SGLT 2 from higher organisms, but not in prokaryotic homologs, may explain differences between the permeabilities calculated from our simulations and those measured experimentally for mammalian SGLT1 (27).

METHODS

MD simulations

Four independent all-atom models of monomeric vSGLT embedded in a lipid bilayer were constructed based on the x-ray structure of the inward-occluded, galactose-bound state (PDB ID 3DH4) (21) using MODELLER (28) and the CHARMM-GUI Membrane Builder (29). The model was parameterized using the CHARMM22 force field (30) with CMAP corrections (31) for the protein, the CHARMM36 lipid force field (32), the CHARMM pyranose monosaccharides parameter set for the galactose (33), and the TIP3P explicit water model (34). Each model was minimized and equilibrated using NAMD 2.7 (35), and then coordinates and velocities from these runs were used to initiate simulations on the Anton special-purpose supercomputer (36).

The first model (M0) was equilibrated for 20 ns without restraints, during which time the sodium ion bound in the Na2 site exited and migrated into bulk solution before a 2.8 μs production simulation was initiated on Anton. For the remaining three models (M1–M3), the sodium ion remained bound throughout the 20 ns equilibration period. We then extended each simulation an additional 70 ns and selected snapshots of the simulation spaced 10 ns apart to use as starting points for 21 separate production runs on Anton. Upon initializing these simulations on Anton, both galactose and the Na2-site sodium ion were placed near their crystallographic poses. In nine of the simulations, the galactose had a root mean-square deviation (RMSD) of <1 Å from the 3DH4 coordinates, whereas in the remaining simulations, the RMSDs were between 2 and 7 Å. Fourteen of the simulations began with sodium within 2 Å of the crystallographic position, with the remainder exhibiting RMSDs of <3.5 Å.

We consider the resulting set of trajectories to be effectively independent, since the distribution of water molecules and the RMSD of the protein decorrelate on the 10 ns timescale. Each production trajectory was run until either both the sodium and galactose escaped into the bulk solvent phase or the simulation reached ~480 ns. Initially, coordinates were recorded every 100 ps. For the 10 simulations in which sugar exited the transporter, we resampled the exit event at a 1 ps interval to track the movement of individual water molecules with high fidelity. Additional details about the preparation of the systems and simulations are provided in the Supporting Material.

Water-flow calculations

Water molecules in the simulation could exchange between the intracellular and extracellular solutions by permeating through the transporter, or by crossing the upper and lower simulation boundaries along the z direction due to the periodicity of the simulation box. The latter mechanism is an artifact of the simulation method and is explicitly removed from the analysis, as described below. Crossings mediated by the protein-lipid interface were negligible and do not affect the overall statistics.

We confine our analysis to permeation events wherein water molecules pass directly through the transporter as they move from one bath to the other. After centering the sugar-binding site of the transporter at the origin, we divided the simulation box into three regions using two planes oriented parallel to the membrane in the xy plane. The planes were placed at ± 15 Å along the z axis, and all water molecules were initially assigned to a region based on the coordinates of the oxygen atom at the start of the trajectory. Time-dependent assignments for every water molecule were calculated, and we counted only permeation events that sequentially passed from one bulk phase region to another by way of the central region. Using this method makes it possible to distinguish events that move up through the transporter from those that move down through the transporter. The numbers of permeation events in each direction were averaged over 1 ns windows, and the net water flow through the transporter is defined as the cumulative sum of upward events minus downward events.

Water molecules cannot cross the entire transporter in <1 ps, so it is possible to unambiguously determine vSGLT-facilitated movement from one bulk to the other using our 1 ps sampling data. However, the distance that water can diffuse in 100 ps is significantly larger, especially in the bulk phase. At that temporal resolution of sampling, errors arise in correctly assigning the order in which a water molecule visits each region. We estimate that by downsampling our 1 ps data sets to 100 ps, statistics gathered

using the longer sampling interval undercount the true number of permeation events by ~6% after correcting for systematic errors caused by the periodic boundary (see the Supporting Material).

Path-based method for calculating pore size

We developed a, to our knowledge, novel method for determining the narrowest region along a water permeation pathway by using the atomic positions of water molecules recorded from MD trajectories. Identifying this region can be challenging for irregular three-dimensional pathways through a protein, especially when there are multiple pathways and cavities that bifurcate from the main channel and residues lining the channel are dynamic. To address this problem, we enumerate the number of distinct water chains that pass from the extracellular space to the intracellular solution through the transporter for each frame of a trajectory. A water chain, or path, is defined as a series of contiguous water molecules where the distance between adjacent oxygen atoms in the chain, d_c, is ≤ 3.2 Å.

First, we wrap the simulation cell for all trajectories such that the center of mass of the protein coincides with the origin. The entire system is then rotated to superimpose the instantaneous coordinates of the protein's nonhydrogen atoms onto the crystallographic reference structure. We then calculate the pairwise distances between the oxygen atoms of all water molecules separated by no more than d_c and within a cylinder 25 Å in radius with ends at $+35$ Å and -35 Å in the z-direction. These pairwise distances are used to construct an adjacency matrix describing the network of water molecules in and around the transporter.

Starting from the water whose oxygen atom is closest to (0, 0, -30) Å, Dijkstra's algorithm (37) is employed to calculate the shortest path through the network terminating above $z = 30$ Å. If a chain is found that successfully connects the two bulk phases through the transporter, all water molecules in the path are pruned from the network, and the shortest path on the subset of remaining water molecules is calculated. This iterative process is repeated until no additional pathways can be found. The resulting set of pathways are independent, as the set of water molecules within a path are unique to that path (i.e., the pathways are nonoverlapping). The number of water chains that connect the bulk regions is then a measure of the width of the channel at its narrowest point. Failure to find a pathway for a given frame indicates that the intra- and extracellular regions are separated by a dehydrated constriction zone.

Identification of residues that gate the water channel

The pathfinding algorithm described in the previous section provides not only an estimate of the size of the water pathway, but a direct way to identify the moieties that disrupt water flow through the transporter. For each frame that lacks a continuous water chain connecting the intra- and extracellular regions, we attempted to identify the closest water molecules on each side of the discontinuity to enumerate the intervening residues or substrate molecule that restricted water movement through that region. In these frames, by definition, oxygen atoms on either side of the discontinuity are separated by more than the 3.2 Å cutoff, d_c. To isolate these water molecules, the original cutoff distance is supplemented by a second cutoff that is increased gradually in 0.5 Å steps until a continuous pathway can be found. Those pairs of oxygen atoms whose distance is $>d_c$ but less than the second cutoff are penalized by a factor of $\alpha = 10$ during the calculation of the shortest path using Dijkstra's algorithm. This penalty term ensures that the method preferentially selects oxygen atoms separated by the shorter distance when available in building the path. Once the second cutoff is increased to a sufficient distance to create a continuous path through the discontinuity, all pairs of oxygens separated by $>d_c$ are recorded.

For each recorded oxygen pair, we then create a cylindrical volume formed by the line segment between the pair and extending outward 2 Å in radius from the axis. We record all nonhydrogen atoms belonging to either the protein or the galactose molecule within the volume. The corresponding residues play at least a transient role in gating water movement through the transporter by occluding the pathway.

Sequence alignment

The sequence of vSGLT (NCBI accession number AAF80602.1) was obtained from GenBank along with the corresponding sequences for SGLT1 and SGLT2 from human, rabbit, and rat (accession numbers P13866.1, P11170.1, P53790.1, P31639.1, P26430.1, P53792.1, respectively). The sequences of several prokaryotic homologs with known functional conservation to the SGLT family, including the proline transporter, PutP (accession number ZP_03034013.1), and the pantothenate permease, PanF (accession number P16256.2), both from *E. coli*, were obtained from the same source. A sequence alignment of the residues corresponding to the 10 core transmembrane helices in vSGLT, from Faham et al. (21), has been used previously to construct an experimentally validated homology model of hSGLT1 (23). We started with this alignment, which included the partial sequences of vSGLT, hSGLT1, and PutP and extended it using TM-Coffee (38) to include the SGLT isoforms from rabbit and rat, as well as PanF, for the complete sequences of each transporter.

RESULTS

Water permeation is correlated with the transient formation of a continuous water channel through the extracellular gate

The amount of water permeating vSGLT fluctuates significantly during our simulations, as the transporter appears to interconvert between putative water-conducting and nonconducting states. Although water readily moves through the transporter for periods of hundreds of nanoseconds (Fig. S2, *A–D*, in the Supporting Material), we also detect quiescent states with substantially less water permeation (Fig. S2, *E–H*). These changes are not mediated by large-scale conformational changes in vSGLT, as the RMSD of the core-transporter domain with respect to the starting x-ray structure remains small throughout the trajectory (see Fig. S3).

To understand the subtle structural rearrangements in the transporter that were controlling these changes in water flow, we attempted to identify the water pathways through the transporter. Initially, we used the programs HOLE (39) and 3V (40) to identify channels and cavities in static snapshots of the protein extracted from the simulation trajectories. Unfortunately, unlike many ion channels, the water pathways through vSGLT are convoluted and lack axial symmetry, and using these software packages failed to reliably identify pathways whose biophysical properties quantitatively correlated with the magnitude of water flux.

To overcome this issue, we developed a, to our knowledge, novel path-based algorithm to identify water chains through the protein directly from simulation snapshots, allowing us to calculate the effective radius of the water channel created by the protein as a function of time. The method first marks all water molecules near the transporter, and then traces out single water pathways from the bulk intracellular

solution to the bulk extracellular solution using Dijkstra's algorithm. Two water molecules are assumed to be in contact if their oxygen atoms are within $d_c = 3.2$ Å of each other. After a pathway is identified through the transporter, all waters in the path are removed and the process is repeated until a path can no longer be found. This occurs when a dehydrated zone appears in the transporter in which water molecules are separated by $>d_c$. The total number of identified pathways approximates the size of the water channel at its narrowest point.

Fig. 1 shows the total number of paths detected in 1 ns blocks together with the maximum instantaneous water influx and efflux. The number of nonoverlapping paths correlates well with the water transport, suggesting that water flow is strongly coupled to the size of the constriction zone along the water permeation pathway. There are, however, periods in which flow and path number are poorly correlated (see 100–150 ns in Fig. 1 *A* and 375–400 ns in Fig. 1 *B*). Closer qualitative inspection of the simulations revealed a primary channel encompassing most pathways (Fig. 1 *C*, *red path*) and a secondary, transient channel that branches from the primary channel (Fig. 1 *C*, *green path*). The secondary channel diverges from the primary channel passing between the interface of the extracellular half of TM10 above G432 and TM helices 11 and 12. This bifurcation occurs above the galactose binding site,

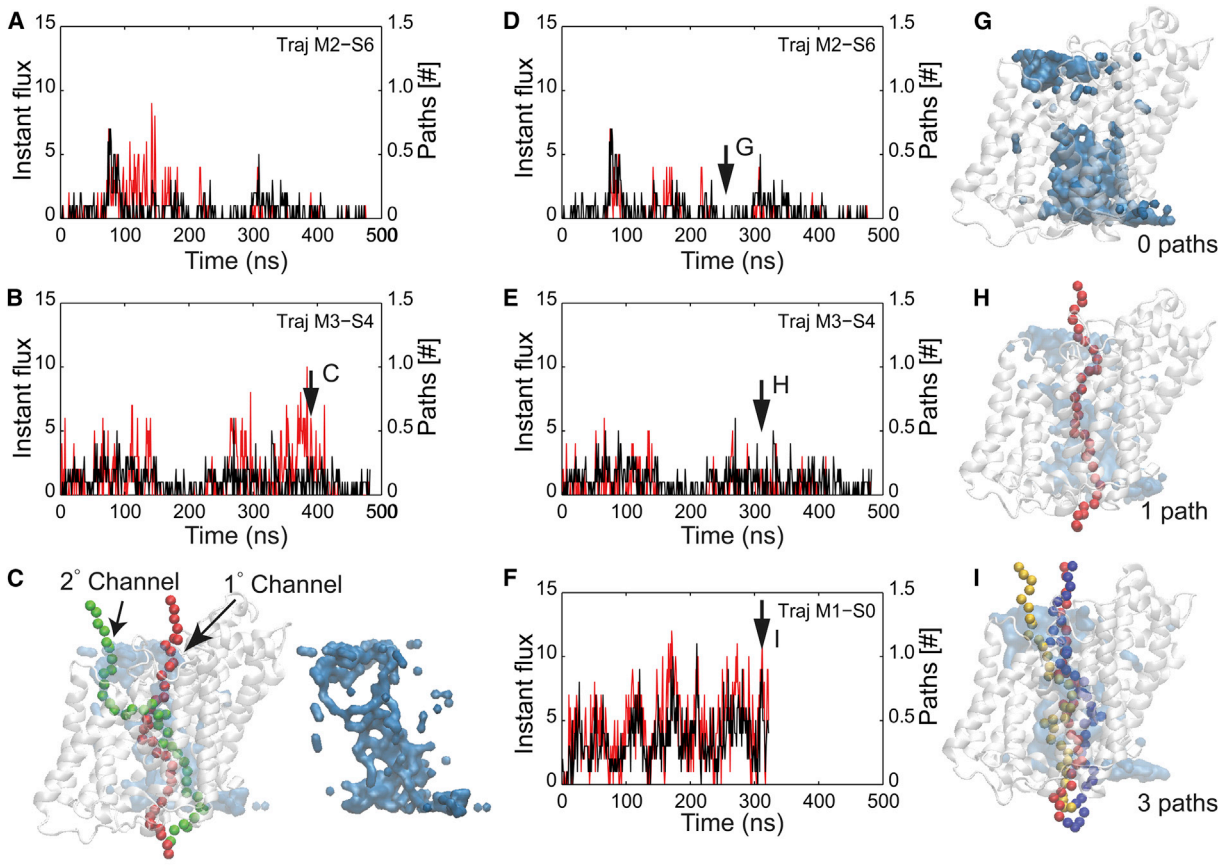

FIGURE 1 Instantaneous water flow and the identification of water channels. (*A* and *B*) Total number of independent water pathways (*red*) and maximum instantaneous water flow (*black*) versus time for two representative trajectories. Instantaneous water flux is defined as the number of upward or downward permeation events in 1 ns, and the maximum instantaneous water flux is the larger of these two values. The independent water pathways are the average number of paths identified from the lower bulk to the upper bulk averaged over 1 ns. The red and black curves show marginal agreement, but there are noticeable exceptions, for instance at 375 ns in *B*, where there are many paths through the transporter even though the water flux is low. The black arrow in *B* indicates the time from which the snapshot in *C* is taken. (*C*) Visualization of the water pathways when the instantaneous water flow and the total pathways are poorly correlated. The snapshot corresponds to the time indicated in *B* (*black arrow*). In the snapshot at left, vSGLT is white, and the molecular surface representation of the water is light blue. Two pathways are present: the main channel (*red*) and a secondary channel (*green*). Additional analysis revealed that water movement through the secondary channel is slow, and that the primary channel is responsible for the majority of the water flux. At right, the water surface is pictured alone to show clearly the trajectories of the two independent and contiguous pathways. (*D–F*) Removal of secondary-channel data results in excellent correlation between pathway numbers (*red*) and maximum instantaneous water flow (*black*). Pathways that ran through the secondary site (*C*, *green path*) were systematically removed, resulting in marked improvement in the correlations shown in *A* and *B*, as can be seen in *D* and *E*, respectively. Simulations represented in *D–F* show an increasing degree of channel openness, as well as an increasing degree of water flux. (G–I) Snapshots with zero (*G*), one (*H*), and three (*I*) independent pathways corresponding to the plots in *D–F*, respectively. The color scheme is as in *C*. In each snapshot, the oxygen atoms from independent pathways are represented by different colors. To see this figure in color, go online.

circumventing the extracellular gate. Water in the secondary channel is far less dynamic than that in the primary channel; consequently, the secondary channel has a significantly reduced permeability. When we removed contributions from the secondary channel from our calculation of the number of permeant paths, the discrepancies previously noted were largely ameliorated (Fig. 1, *D* and *E*).

Next, we used our method to determine which residues were responsible for occluding the aqueous channel. We started by selecting every frame from the trajectories in which no continuous paths were detected. Focusing on the constriction zones along these pathways, as described in Methods, we identified the residues that lie in the dehydrated region. We tallied the number of times that a particular residue was involved in blocking the pathway and plotted the probability of finding any given residue in any gap (Fig. 2 *A*). Picking a cutoff of 5% (*dashed line*), our analysis reveals that the side chains of residues S76, Y87, N245, D336, Q422, F424, Q425, and Q428, along with the bound galactose, are responsible for interrupting the continuous water pathway through the transporter. A subset of these residues (Y87, F424, and Q428) were recently identified as lining the constriction zone based on direct observation of simulations (25), and Y87 and F424 were proposed to form part of the extracellular gate along with M73, based on visualization of the static x-ray structure (21). Using grand canonical Monte Carlo simulations of the transporter, Sasseville et al. also identified Y87, Q425, and Q428, as well as A259, F479, and M483, as components of a constriction zone that acted as a barrier to water permeation through vSGLT (15). It is not surprising that some of these amino acids have bulky hydrophobic side chains capable of occluding and drying pathways in particular rotameric states, whereas others are charged or polar and form interactions between adjacent TM helices to hold the outer gate shut.

We highlighted the primary gating residues on the vSGLT structure in a high-water-flux configuration (Fig. 2 *B*) and two low-water-flux configurations (Fig. 2, *C* and *D*). As can be seen from the lower row of structures in Fig. 2, *B–D*, the identified residues all localize to the outer-gate region above the galactose binding site on the extracellular side of the protein. In addition, the residues form an open annulus when the water flux is high (Fig. 2 *B*) and a completely closed mass when the water flux is low (Fig. 2 *D*). Interestingly, we realized that the residues were sometimes open, as in Fig. 2 *C*, despite a very low water flux. In these cases, the galactose molecule occludes the water channel, preventing water flow (see Fig. 2 *C*).

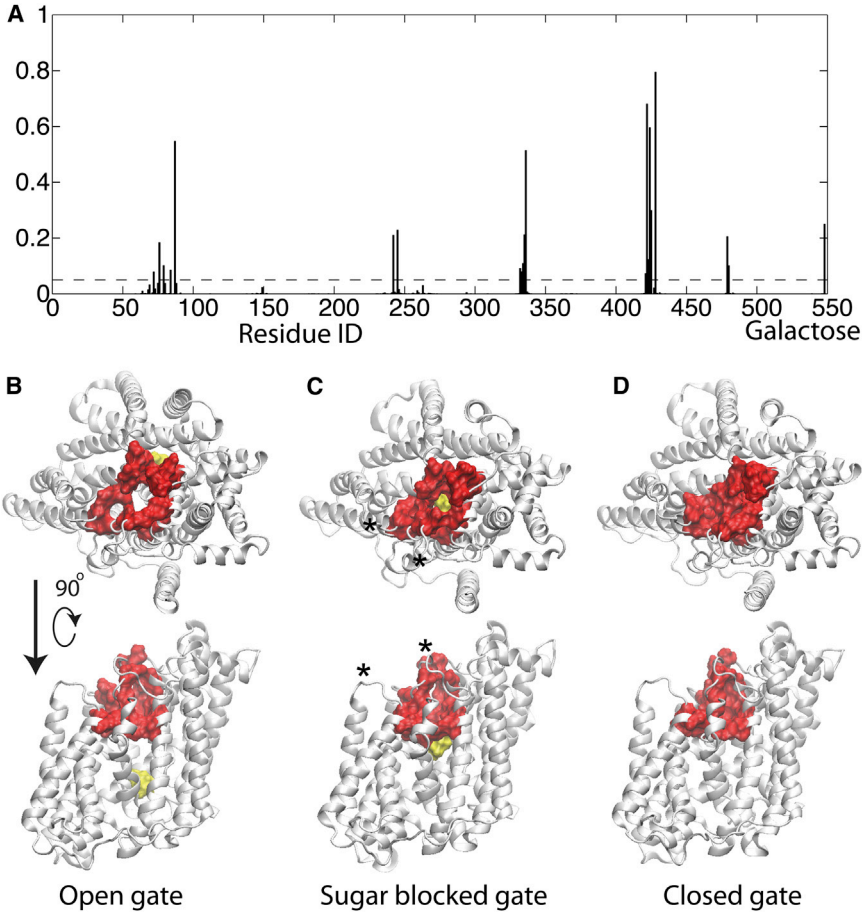

FIGURE 2 Identification of residues that control the number of water pathways. (*A*) Probability of a protein residue or galactose being identified in a gap region. Probability is calculated as the number of total observations divided by the total frames involving water gaps. The *x* axis is the protein residue number in vSGLT with galactose (1–547), with the galactose being the final value. The dashed line corresponds to a 5% (0.05) probability. (*B–D*) All residues observed in >5% of gaps were highlighted (*red*) on the vSGLT structure (*white*) in high water flux (*B*) and in two low-water-flux states (*C* and *D*). Galactose is yellow. Each structure is shown from the extracellular space (*upper*) and from the membrane (*lower*). In *B*, the residues form an annulus with an open center. The side view shows that all residues are localized to the extracellular side of the transporter. In *C*, the residues form an open annulus, but with the sugar plugging the hole, which explains why no water pathways are found in this snapshot. The asterisks indicate the positions of the insertion sites described in Fig 4 and the Discussion. In *D*, it can be seen that the outer residues all come together to occlude the outer gate and prevent water pathways through the channel. All of the molecular images were rendered using VMD (61). To see this figure in color, go online.

We also examined the role of galactose in controlling water flow through the transporter. As shown in Fig. S4, when the sugar has an RMSD of <2 Å with respect to the crystallographic pose, the average water flow is nearly four times smaller than when the sugar is in bulk solution. The fluctuations in flow are also dramatically suppressed when sugar occupies the binding site. This is consistent with the ability of the gating residues to occlude the permeation pathway in the absence of substrate. A similar dependence of water flow on sugar occupancy was previously observed in simulations carried out by the Tajkhorshid lab (25).

Diffusional and osmotic water permeabilities of vSGLT

Additional insight into the biophysical nature of the permeation mechanism was obtained from a comparison of the osmotic permeability coefficient, p_f, and the coefficient of diffusional water exchange, p_d. We estimated these quantities directly from our equilibrium simulations using the theory developed by Zhu et al. (41), as done in our previous work with vSGLT (14). Water movement in the system must be accurately tracked to provide high-fidelity estimates of the water-diffusion and osmotic-permeability coefficients. All simulations were recorded at either 1 ps or 100 ps intervals. The 100 ps sampling introduces some uncertainty in the tracking of water movement, as described both in Methods and in the Supporting Material. Our analysis indicates, however, that the error associated with the calculation of p_d from the trajectories sampled at 100 ps intervals is only 6%. We therefore used our entire collection of trajectories, with an aggregate total simulation time of 6.5 μs, to calculate $p_d = 5.1 \times 10^{-14}$ cm^3/s with a 95% confidence interval of $p_d = (4.6, 5.7) \times 10^{-14}$ cm^3/s.

The osmotic permeability coefficient, however, is much more sensitive to the sampling interval, so accurate estimates could only be made using our 1 ps sampling data. Using this smaller subset of the data, we determined that $p_f = 2.7 \times 10^{-13}$ cm^3/s with a 95% confidence interval of $p_f = (2.2, 3.1) \times 10^{-13}$ cm^3/s. The discrepancy between the values of p_f reported in this study and in our previous work (14) and the much smaller value presented in Li et al. (25) are discussed in the Supporting Material.

Experimentally, p_f measured in SGLT1 from rabbit is 4.5 $\times 10^{-16}$ cm^3/s (27), which is 600 times smaller than our reported value. The p_f and p_d values reported here are consistent with results from previous simulations, where the high p_f compared to the experimental value was originally noted (14). We will return to this topic in the Discussion. The ratio p_f/p_d is related to the structure of water in the channel (42) and is a measure of the degree of cooperativity in the transport process, with values close to 1 indicative of large pores that lack cooperativity and values >10 indicative of a high degree of cooperativity in transport, as observed for single-file transport in aquaporins (41). Our simulations indicate this ratio for vSGLT is $p_f/p_d = 5.3$, in good agreement with other transporters and water channels (43).

Directional water flow is not correlated with substrate release

Previous analysis of a single vSGLT simulation suggested that the inward release of galactose to the intracellular space is coupled to the inward cotransport of water (14). The simulation showed that 70–80 water molecules entered the intracellular space during the time that galactose left the central binding pocket and exited the protein through a narrow intracellular cavity. It was hypothesized that the galactose acted as a plug in the tight confines of the exit cavity, causing water to move inward as the galactose unbound. To determine whether water and galactose movement are correlated, we examined the net water flow and transient fluctuations from eight independent galactose unbinding simulations and 11 other simulations in which galactose remained bound to the transporter.

Fig. 3 shows a representative subset of nine of these trajectories, where the position of the galactose relative to the center of mass of the sugar binding site is plotted along with the cumulative net water flow through the transporter. We observe transient directional flow through the transporter that does not appear to correlate with substrate release. Although there are trajectories in which water moves in the same direction as the galactose during exit, consistent with our previous observation (14) (Fig. 3, *C*, *E*, *F*, and *G*), there are nearly as many instances where water moves against the direction of galactose exit (Fig. 3, *A* and *B*) or fails to flow at all (Fig. 3 *D*). Further, there are two trajectories in which galactose remains bound during the entire simulation, yet the water fluctuations are comparable in magnitude to values observed during galactose exit (Fig. 3, *H* and *I*). These results suggest that the directional flux of water observed by Choe et al. (14). represented a transient fluctuation that was coincident with galactose exit, rather than tight coupling to the movement of galactose.

DISCUSSION

We have used a series of long, 500- to 1800-ns MD simulations, starting from a sugar and sodium-bound inward-facing state, to probe the structural basis of water movement through vSGLT. A major advance of this study is our observation of multiple independent events of sugar unbinding from the intracellular space, where previously we had observed only a single unbinding event (14). Thus, this analysis provides a statistical description of solute-coupled water flow, which has allowed us to refine our initial hypothesis. In addition, we have significantly more simulation of

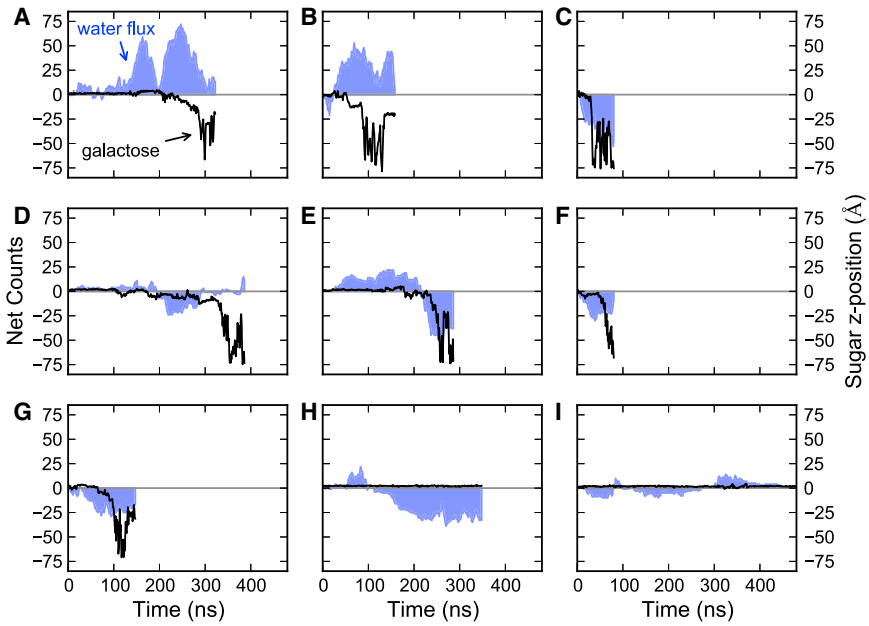

FIGURE 3 Water flow and movement of galactose as a function of time. (*A–I*) For each trajectory, the cumulative net flow of water (*blue-shaded areas*), defined as the difference between cumulative efflux and influx through the transporter, is shown alongside the position of the galactose relative to the sugar binding site projected onto the *z* axis (*black*). Net efflux through the transporter corresponds to positive values, whereas net influx results in negative values. Panels *A-G* show examples of simulations in which galactose was released into the bulk intracellular solution. In panels *H* and *I*, the sugar is stable in the binding site for the entire simulation. To see this figure in color, go online.

the substrate-bound conformation, as well as outer-gate conformations, allowing us to discern the roles of binding-site occupancy and protein conformations in modulating water permeation.

For the inward-facing conformation, we find that the magnitude of the passive water flow is controlled by small structural rearrangements in the outer gate, as well as the occupancy of galactose in the binding site. Although the conformation of vSGLT corresponding to the crystal structure (21) is nearly impermeable to water, due to the tight packing of side chains in the gate region, the thermalized system permits water to leak through the outer gate during transient excursions to more permissive conformations. To properly account for the aqueous pathways, we developed a believed novel procedure for detecting continuous water chains connecting the two bulk solvent phases in the simulation cell. The hole that forms in the outer gate varies in size from the diameter of a single-file chain of waters, reminiscent of aquaporins (44), to a large pore several water molecules in diameter. We found that the number of water paths closely matched the magnitude of the water flow, acting as a surrogate for the effective radius of the pore, despite the complicated geometry of the gate. For this reason, we believe that our method will be widely applicable for the detection of narrow water channels with complex geometry in other systems.

These longer simulations confirm our previous finding that water permeates vSGLT through the sugar-binding site. Estimates of the diffusional and osmotic permeabilities are also consistent with our previous measurements (14). The data presented here, however, do not support a mechanism of water transport in which the exiting sugar forces water in the lumen of the transporter into the intracellular solution. Instead, the net water movement exhibits large, apparently stochastic, fluctuations that are not strongly correlated with the motion of the exiting substrate. This finding indicates that the hyperosmotic (12) and Brownian piston (14) models of active water pumping are most likely not correct for vSGLT. Water is able to move past galactose while it is in the binding site and the intracellular exit tunnel, which explains why the inward movement of galactose is not coupled to water motion. We still believe that the piston mechanism could rectify water flow in a more tightly constricted exit tunnel or during different stages of the transport process. In addition, active pumping mechanisms that rely on water being expelled in a unidirectional manner by large-scale protein conformational changes (17) cannot be addressed by this work, since we did not observe major rearrangements. Such changes occur on the millisecond timescale for alternating access and are hard to capture by simulation (45).

In agreement with our previous simulations of vSGLT, we find that the osmotic water permeability is nearly 600–900 times greater (14) than that reported for mammalian homologs (7). It is possible that this difference arises due to deficiencies in the underlying force field or because experiments average the water flow over the entire transport cycle and our simulations are limited to the inward-facing state only. For instance, if the inward-facing state is the only conformation that allows water permeation, and if the fraction of time spent in this state is 1/1000, then there would be excellent agreement between our vSGLT simulations and experiments on mammalian homologs. However, kinetic models of hSGLT1 transport suggest that a significant amount of time is spent in the inward-facing state (46), pointing to the possibility that there are structural differences between the prokaryotic and eukaryotic transporters.

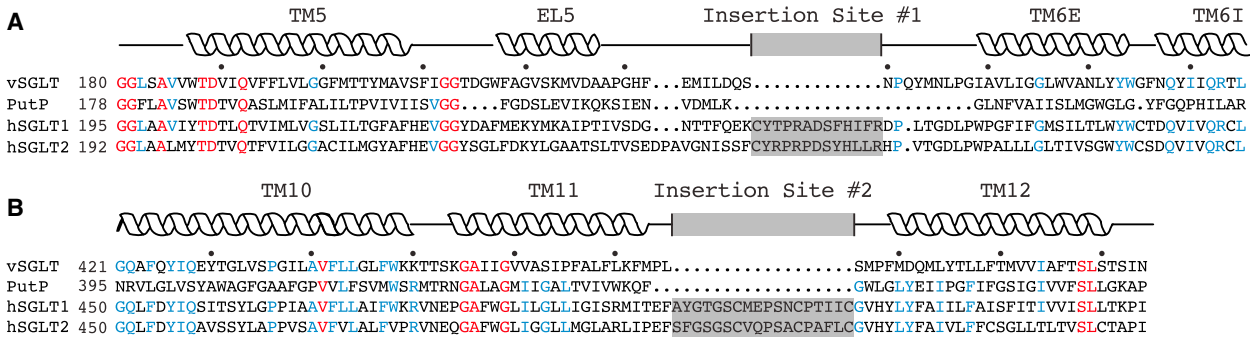

FIGURE 4 Insertion sites in mammalian transporters near the extracellular gate. Alignment of the primary sequences of vSGLT, PutP, hSGLT1, and hSGLT2 showing the regions around the two insertion sites found in mammalian transporters, which are absent in the bacterial homologs. Secondary-structure elements are shown above the sequences based on the structure of vSGLT. Strictly conserved and highly conserved residues are shown in red and blue, respectively. Insertion sites 1 (*A*) and 2 (*B*) are shown as gray boxes. To see this figure in color, go online.

To investigate this possibility, we performed a multiple-sequence alignment between vSGLT and several of its closely related family members. The resulting alignment revealed two extracellular loops that are present in the mammalian transporters but notably absent in vSGLT and other prokaryotic homologs. These insertion sites and the conserved flanking transmembrane helices are shown in Fig. 4 for vSGLT, the proline permease from *E. coli*, PutP, hSGLT1, and hSGLT2. Insertion site 1 (Fig. 4 *A*) is 13 residues long and is positioned in the extracellular loop connecting TM5 and TM6, between S239 and N240 in vSGLT (C255–R267 in hSGLT1). The second insertion site (Fig. 4 *B*) results in an additional 18 residues inserted in the loop between TM11 and TM12, and it is flanked by residues L475 and S476 (A505 and C522 in hSGLT1). It is predicted, based on the alignment, that both of these insertion sites will lie proximal to the extracellular gate and the residues that control water permeation through the inward-facing state of vSGLT. In Fig 2C, the asterisks indicate where the loops would be inserted on vSGLT. These loops, absent in vSGLT, might further occlude the water pathway to the extracellular solution in the mammalian transporter by forming a steric block. Alternatively, the loops may help maintain the outer gate in a closed, water-tight conformation without sterically hindering the water pathway directly. We believe that the hypothesis that vSGLT possesses an intrinsically leakier outer gate may also explain why experiments predict that galactose binding fails to alter water flow, whereas simulations suggest that it does. If movement of water through a tightly sealed outer gate is rate-limiting in mammalian transporters, then the presence or absence of galactose occluding the permeation pathway would not be detectable with such a small water flow.

Both loops contain a number of highly conserved residues that are present in both hSGLT1 and hSGLT2, as well as isoforms from rat and rabbit. Among the conserved residues, C255 at the N-terminal end of insertion site 1 and C511 near the center of insertion site 2 are hypothesized to form a disulfide bridge (47,48). Such a link would orient the loop corresponding to insertion site 2, toward the opening of the water permeation pathway that we have identified, consistent with the proposed proximity of C255 and C511 to the sugar binding site (48). Although these differences are suggestive, any effect of sequence differences on the functional behavior of the transporters would require experimental validation. Nonetheless, these protein elements may be critical in controlling the flow of material across the membrane and may play a role in regulating key steps in the alternating-access cycle.

SUPPORTING MATERIAL

Four figures, simulation details and analysis, Supporting Discussion, and references (49–60) are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00067-8.

We thank Shannon McCullough and Ambika Ramesh for their initial analysis on this project. Simulations were performed on the Anton special-purpose supercomputer provided by the National Resource for Biomedical Supercomputing (NRBSC), the Pittsburgh Supercomputing Center (PSC), and the Biomedical Technology Research Center for Multiscale Modeling of Biological Systems (MMBioS) through grant P41GM103712-S1 from the National Institutes of Health and grant PSCA00015P (M.G.). The Anton machine at NRBSC/PSC was generously made available by D. E. Shaw Research.

This work was supported by National Institutes of Health (NIH) grants GM089740-01A1 (M.G. and J.M.R.), T32-DK061296 (J.L.A.), DK19567 (E.M.W.), and GM078844 (J.A.). The work of S.C. was supported by a start-up fund from the Daegu Gyeongbuk Institute of Science and Technology.

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. Wright, E. M., D. D. F. Loo, and B. A. Hirayama. 2011. Biology of human sodium glucose transporters. *Physiol. Rev.* 91:733–794.

3. Loike, J. D., S. Hickman, …, J. Fischbarg. 1996. Sodium-glucose cotransporters display sodium- and phlorizin-dependent water permeability. *Am. J. Physiol.* 271:C1774–C1779.

4. 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.
5. Duquette, P.-P., P. Bissonnette, and J.-Y. Lapointe. 2001. Local osmotic gradients drive the water flux associated with Na^+/glucose cotransport. *Proc. Natl. Acad. Sci. USA.* 98:3796–3801.
6. Gagnon, M. P., P. Bissonnette, …, J.-Y. Lapointe. 2004. Glucose accumulation can account for the initial water flux triggered by Na^+/glucose cotransport. *Biophys. J.* 86:125–133.
7. Loo, D. D., B. A. Hirayama, …, E. M. Wright. 1999. Passive water and ion transport by cotransporters. *J. Physiol.* 518:195–202.
8. Victora, C. G., J. Bryce, …, R. Monasch. 2000. Reducing deaths from diarrhoea through oral rehydration therapy. *Bull. World Health Organ.* 78:1246–1255.
9. Charron, F. M., M. G. Blanchard, and J.-Y. Lapointe. 2006. Intracellular hypertonicity is responsible for water flux associated with Na^+/glucose cotransport. *Biophys. J.* 90:3546–3554.
10. MacAulay, N., S. Hamann, and T. Zeuthen. 2004. Water transport in the brain: role of cotransporters. *Neuroscience.* 129:1031–1044.
11. Loo, D. D., E. M. Wright, and T. Zeuthen. 2002. Water pumps. *J. Physiol.* 542:53–60.
12. Naftalin, R. J. 2008. Osmotic water transport with glucose in GLUT2 and SGLT. *Biophys. J.* 94:3912–3923.
13. Zeuthen, T. 2010. Water-transporting proteins. *J. Membr. Biol.* 234: 57–73.
14. Choe, S., J. M. Rosenberg, …, M. Grabe. 2010. Water permeation through the sodium-dependent galactose cotransporter vSGLT. *Biophys. J.* 99:L56–L58.
15. Sasseville, L. J., J. E. Cuervo, …, S. Y. Noskov. 2011. The structural pathway for water permeation through sodium-glucose cotransporters. *Biophys. J.* 101:1887–1895.
16. 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.
17. Zeuthen, T., and N. MacAulay. 2012. Transport of water against its concentration gradient: fact or fiction? *Wiley Interdiscip. Rev. Membr. Transp. Signal.* 1:373–381.
18. Turk, E., O. Kim, …, E. M. Wright. 2000. Molecular characterization of *Vibrio parahaemolyticus* vSGLT: a model for sodium-coupled sugar cotransporters. *J. Biol. Chem.* 275:25711–25716.
19. Xie, Z., E. Turk, and E. M. Wright. 2000. Characterization of the *Vibrio parahaemolyticus* Na^+/glucose cotransporter. A bacterial member of the sodium/glucose transporter (SGLT) family. *J. Biol. Chem.* 275: 25959–25964.
20. Veenstra, M., S. Lanza, …, E. M. Wright. 2004. Local conformational changes in the Vibrio Na^+/galactose cotransporter. *Biochemistry.* 43:3620–3627.
21. 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.
22. Watanabe, A., S. Choe, …, J. Abramson. 2010. The mechanism of sodium and substrate release from the binding pocket of vSGLT. *Nature.* 468:988–991.
23. Sala-Rabanal, M., B. A. Hirayama, …, E. M. Wright. 2012. Bridging the gap between structure and kinetics of human SGLT1. *Am. J. Physiol. Cell Physiol.* 302:C1293–C1305.
24. Leung, D. W., E. Turk, …, E. M. Wright. 2002. Functional expression of the *Vibrio parahaemolyticus* Na^+/galactose (vSGLT) cotransporter in *Xenopus laevis* oocytes. *J. Membr. Biol.* 187:65–70.
25. Li, J., S. A. Shaikh, …, E. Tajkhorshid. 2013. Transient formation of water-conducting states in membrane transporters. *Proc. Natl. Acad. Sci. USA.* 110:7696–7701.
26. Bisha, I., A. Laio, …, J. Sgrignani. 2013. A candidate ion-retaining state in the inward-facing conformation of sodium/galactose symporter: clues from atomistic simulations. *J. Chem. Theory Comput.* 9:1240–1246.
27. Zampighi, G. A., M. Kreman, …, E. M. Wright. 1995. A method for determining the unitary functional capacity of cloned channels and transporters expressed in *Xenopus laevis* oocytes. *J. Membr. Biol.* 148:65–78.
28. Sali, A., and T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815.
29. Jo, S., T. Kim, and W. Im. 2007. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS ONE.* 2:e880.
30. MacKerell, Jr., A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.
31. Mackerell, Jr., A. D., M. Feig, and C. L. Brooks, 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.* 25:1400–1415.
32. Klauda, J. B., R. M. Venable, …, R. W. Pastor. 2010. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B.* 114:7830–7843.
33. Guvench, O., S. N. Greene, …, A. D. Mackerell, Jr. 2008. Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* 29:2543–2564.
34. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
35. Phillips, J. C., R. Braun, …, K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
36. Shaw, D. E., R. O. Dror, …, B. Towles. 2009. Millisecond-scale molecular dynamics simulations on Anton. *Proc. Conf. High Perform. Comput. Network. Storage Anal., Portland, OR.* 2009:39.
37. Dijkstra, E. 1959. A note on two problems in connexion with graphs. *Numer. Math.* 1:269–271.
38. Chang, J.-M., P. Di Tommaso, …, C. Notredame. 2012. Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. *BMC Bioinformatics.* 13 (*Suppl 4*):S1.
39. Smart, O. S., J. G. Neduvelil, …, M. S. Sansom. 1996. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* 14:354–360, 376.
40. Voss, N. R., and M. Gerstein. 2010. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* 38 (Web Server issue):W555–W562.
41. Zhu, F., E. Tajkhorshid, and K. Schulten. 2004. Collective diffusion model for water permeation through microscopic channels. *Phys. Rev. Lett.* 93:224501.
42. Portella, G., and B. L. de Groot. 2009. Determinants of water permeability through nanoscopic hydrophilic channels. *Biophys. J.* 96:925–938.
43. Finkelstein, A. 1987. Water Movement through Lipid Bilayers, Pores, and Plasma Membranes: Theory and Reality. Wiley, New York.
44. Tajkhorshid, E., P. Nollert, …, K. Schulten. 2002. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science.* 296:525–530.
45. Adelman, J. L., A. L. Dale, …, M. Grabe. 2011. Simulations of the alternating access mechanism of the sodium symporter Mhp1. *Biophys. J.* 101:2399–2407.
46. 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.
47. Gagnon, D. G., P. Bissonnette, and J.-Y. Lapointe. 2006. Identification of a disulfide bridge linking the fourth and the seventh extracellular loops of the Na^+/glucose cotransporter. *J. Gen. Physiol.* 127:145–158.

48. Gagnon, D. G., C. Frindel, and J.-Y. Lapointe. 2007. Voltage-clamp fluorometry in the local environment of the C255–C511 disulfide bridge of the Na^+/glucose cotransporter. *Biophys. J.* 92:2403–2411.

49. Scargle, J. D., J. P. Norris, …, J. Chiang. 2013. Studies in astronomical time series analysis. vi. bayesian block representations. *Astrophys. J.* 764:167.

50. Tuckerman, M., B. J. Berne, and G. J. Martyna. 1992. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* 97:1990–2001.

51. Chen, V. B., W. B. Arendall, 3rd, …, D. C. Richardson. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66:12–21.

52. Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* 1:54–75.

53. Zhu, F., and G. Hummer. 2010. Pore opening and closing of a pentameric ligand-gated ion channel. *Proc. Natl. Acad. Sci. USA.* 107:19814–19819.

54. Kräutler, V., W. F. van Gunsteren, and P. H. Hünenberger. 2001. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* 22:501–508.

55. Jensen, M. Ø., and O. G. Mouritsen. 2006. Single-channel water permeabilities of *Escherichia coli* aquaporins AqpZ and GlpF. *Biophys. J.* 90:2270–2284.

56. Hashido, M., M. Ikeguchi, and A. Kidera. 2005. Comparative simulations of aquaporin family: AQP1, AQPZ, AQP0 and GlpF. *FEBS Lett.* 579:5549–5552.

57. Shan, Y., J. L. Klepeis, …, D. E. Shaw. 2005. Gaussian split Ewald: a fast Ewald mesh method for molecular simulation. *J. Chem. Phys.* 122:54101.

58. Aksimentiev, A., and K. Schulten. 2005. Imaging α-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys. J.* 88:3745–3761.

59. Lomize, M. A., A. L. Lomize, …, H. I. Mosberg. 2006. OPM: orientations of proteins in membranes database. *Bioinformatics.* 22:623–625.

60. Vanderplas, J., A. Connolly, …, A. Gray. 2012. Introduction to astroMl: machine learning for astrophysics. *Proc. Conf. Intell. Data Understand.* (*CIDU*), Boulder, CO 2012:47–54.

61. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.

Supporting Materials:

Structural determinants of water permeation through the sodium-galactose transporter vSGLT

Joshua L. Adelman[1], Ying Sheng[1], Seungho Choe[2], Jeffrey Abramson[3], Ernest Wright[3], John M. Rosenberg[1] and Michael Grabe[1,4]

[1] *Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA*
[2] *School of Basic Science, College of Convergence, Daegu Gyeongbuk Institute of Science & Technology, Daegu, Korea*
[3] *Department of Physiology, University of California, Los Angeles, Los Angeles, CA*
[4] *Cardiovascular Research Institute, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA*

Contents

1. System Preparation

Atomistic models of monomeric vSGLT embedded in a lipid bilayer were constructed using chain A from the x-ray structure deposited in the RCSB Protein Data Bank (PDB ID: 3DH4) (1). The first 52 residues of vSGLT including the unassigned section of the first transmembrane helix (TM-1) and the unresolved loop connecting it to TM1, were removed. These residues do not constitute part of the core structure of the transporter and the corresponding TM is not conserved among superfamily members. The MODELLER software package (2), was then used to mutate any non-wild type residues back to their WT sequence and rebuild the side chains of residues K124, R273, K454 and K547, which were missing in the 3DH4 structure. Additionally, the loop connecting TM4 and 5 (residues 179 to 184) was rebuilt using MODELLER's loop modeling routine.

Two independent sets of models were constructed that differed only in the random seed used during loop modeling and the number of flanking residues on either side of the missing loop that were permitted to move. One model was chosen from each of the two independent ensembles of putative structures based on screening using the MODELLER DOPE score and analysis with MolProbity (3). These two final models were used to construct the full systems described below.

The protein along with the galactose substrate and crystallographic sodium, were oriented with respect to the z-axis using OPM (Orientations of Proteins in Membranes) (4) and then inserted into POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylethanolamine) membrane using the CHARMM-GUI Membrane Builder (5). The embedded protein with membrane was then solvated in a rectangular box with dimensions 97 x 97 x 97 $Å^3$ containing approximately 91,000 atoms. The system was neutralized using sodium and chloride counterions, which were added in sufficient quantity to approximate a physiological concentration of 150 mM.

2. Simulations

2.1 Equilibration

Each system was prepared using NAMD (6) to obtain well-equilibrated starting conformations for the production runs on the Anton special-purpose supercomputer (7). Simulations were carried out using the CHARMM22 parameter set (8) for the protein with CMAP corrections (9). The membrane was parameterized using the CHARMM36 lipid forcefield (10) and galactose parameters were taken from the CHARMM force field for pyranose monosaccharides (11). Explicit water molecules were included using the three-site TIP3P model (12).

The systems were minimized using 10,000 steps of the conjugate gradient method followed by gradual heating from 1 K to 310 K at a rate of 1 K every 400 fs using temperature reassignment. During the heating phase the dynamics were carried in the constant volume ensemble with all water molecules, galactose, Na^+, and heavy backbone and sidechain atoms restrained using a harmonic potential with a 10.0 kcal/mol/$Å^2$ force constant. Subsequent dynamics were performed with a Langevin piston barostat with a 200 fs piston period and 100 fs piston decay constant to maintain a constant pressure of 1 atm. Temperature was maintained at 310 K using Langevin dynamics with a 0.5 ps^{-1} damping coefficient. Restraints on the water molecules and side-chain heavy atoms were then removed over the next 2.1 ns. Finally, restraints on the backbone, Na^+ and galactose were removed over another 1.8 ns. During the entire equilibration protocol, an external force was

applied to all water molecules outside of the transporter to prevent aberrant hydration of the protein-membrane interface. All further dynamics were simulated in the absence of external restraints.

Bond lengths between hydrogen and heavy atoms were constrained using the SHAKE algorithm and water was held rigid using the SETTLE algorithm. This permitted us to use a discretized time step of 2 fs. The Particle Mesh Ewald summation was employed for electrostatics with a grid spacing of less than 1 *Å* along each dimension. Both van der Waals and electrostatic interactions were smoothly switched off between 8 and 10 *Å*.

2.2 Production simulations on Anton

Productions simulations were run with Anton software version 2.4.1 (7). Coordinates and velocities for each setup prepared in NAMD were converted to the Anton specified format. The same force field parameters were used in both the preparatory and production runs. Additional parameters for the Anton runs were obtained using the default outputs of the `guess_chem`, `refine_sigma` and `subboxer` pre-processor scripts to tune the cut-off distances, electrostatic settings and spatial domain decomposition. Bonds between hydrogen and heavy atoms were constrained to their equilibrium length using the M-SHAKE algorithm (13). Long-range electrostatic interactions were treated using the Gaussian split Ewald (GSE) method with a 64x64x64 mesh (14). The equations of motion were integrated using the RESPA multiple time-step method (15) with time steps of 6.0 fs for the long-range electrostatic interactions and 2.0 fs for all other interactions. Simulations were run at constant temperature (300 K) and pressure (1 atm). Constant pressure was maintained using a semi-isotropic Berendsen barostat with a relaxation time of 2 ps and temperature was controlled with the corresponding Berendsen thermostat with $\tau = 1$ ps.

3. Analysis

3.1 Calculation of permeability and diffusion coefficients

The diffusion permeability, p_d, was computed using the linear flux equation

$$J_t = p_d\,(t_o - t_i)\,. \tag{1}$$

Here, J_t (mol/s) is the flux of tracked water molecules permeating from one side of the membrane to the other and t_i and t_o are the inner and outer tracer concentrations respectively. The flux of water is calculated directly from counting permeation events through the transporter as described in the main text. In this study the channel region extended 30 Å in the z-direction, which is longer than the definition employed in Ref. (16), where the channel boundaries were defined at $\pm$ 7.5 Å from the center of the transporter. Using the larger boundary separation has the effect of decreasing the number of observed permeation events by approximately a factor of two, but ensures that waters must make a complete transition from one bulk region to the other and minimizes recrossing events. Given a molar concentration of water, $c_w \sim 0.055$ mol/cm^3, the diffusion permeability is then calculated from Eq. 1 as

$$p_d = \frac{\text{counts}_{\text{up}} + \text{counts}_{\text{down}}}{2 c_w T_{\text{total}} N_A}, \tag{2}$$

where N_A is Avogadro's Number and T_{total} is the total time of the simulation.

In the absence of an osmotic concentration or pressure gradient across the lipid membrane in our simulations, we estimate the osmotic permeability, p_f using the theoretical framework proposed by Zhu et al. (17). Briefly, this calculation is based on calculating the total incremental change in the position of all water molecules in the channel at time t

$$dn(t) = \sum_{i=1}^{M} \frac{z_i(t+dt) - z_i(t)}{L}, \tag{3}$$

where L is the length of the channel and $z_i(t)$ is the z component of the position of the ith water molecule's oxygen atom. After numerically integrating Eq. 3, the osmotic permeability is calculated from a linear regression of

$$\left\langle n(t)^2 \right\rangle = 2\frac{p_f}{v_w} t, \tag{4}$$

where v_w is the average volume of a single water molecule (18 cm^3/mol/N$_A$ $\approx 3\times10^{-23}$ cm^3) and the angular brackets denote an ensemble average over trajectory segments. In this study, trajectory segments were obtained by dividing the time series of $n(t)$ into non-overlapping windows 300 ps in length. Uncertainties in the calculated value of p_f were computed using a Monte Carlo bootstrapping procedure. The value of $n(t)$ for each trajectory segments was first averaged in blocks of 80 consecutive windows, and then a bootstrap sample was created from the block averages using 10,000 samples drawn at random with replacement. This sample was then used to determine the 95% confidence interval for the calculated value of p_f. A similar procedure was used to calculate the confidence interval for p_d.

The water channel in vSGLT is not aligned with the z-axis of the simulation box and is instead curved as it winds from the extracellular to intracellular side of the transporter. We therefore modify the protocol for estimating p_f as follows. After centering the protein at the origin of the simulation box, the water molecules were wrapped into the central periodic image; the protein was then aligned onto a reference conformation. Subsequently, to define the permeation path, the center-of-mass of water molecules within cylindrical slabs 0.5 Å in thickness, and with a radius of 20 Å from the center-of-mass of the transporter, were calculated for $-15 < z < 15$ Å. A smooth continuous path, φ, is generated by fitting

$$\varphi_\alpha(\lambda) = \varphi_0 + (\varphi_N - \varphi_0)\,\lambda + \sum_{i=1}^{N_{\text{dim}}} \left[\sigma_{i,0}\sin(\pi\lambda) + \sigma_{i,1}\sin(2\pi\lambda)\right]\cdot\hat{e}_i, \tag{5}$$

through the center-of-mass of water molecules within each slab by varying λ over the range $[0,1]$ (18). Here, $N_{\mathrm{dim}} = 3$, $\hat{e}_i$ is the unit vector of the i^{th} dimension and $\sigma_{i,j}$ are the coefficients of 2 sinusoidal basis functions in each dimension. The parameters $\sigma_{i,j}$ and λ_α are selected to minimize the difference between the center-of-mass calculated for each slab, α and the parameterized curve,

$$\chi^2 = \sum_{\alpha=0}^{N_{slab}} |\varphi_\alpha(\lambda_\alpha) - \varphi_\alpha|^2 . \tag{6}$$

Equation 3 is then reformulated in terms of the displacement along the curve, rather than a displacement along z, as

$$dn(t) = \sum_{i=1}^{M} \frac{s_i(\lambda_i(t), \lambda_i(t+dt))}{L} \tag{7}$$

where s is the contour length between $\lambda_i(t)$ and $\lambda_i(t+dt)$ along $\varphi(\lambda)$ and L is the contour length of the curved path traversing the entire channel. The projection $\lambda_i(t)$ minimizes the distance between the position of the i$^{\text{th}}$ water molecule in the channel at time t and the curve. Results using the projection obtained by identifying the point on the curve with the same z value of the oxygen atom at each time, as in (16), yielded similar results.

As with the calculation of p_d, defining the channel as the volume within the transporter between $-15 < z < 15$ Å decreases p_f by approximately a factor of 2 compared to the value obtained in Ref. (16) where the channel boundaries were between $-7.5 < z < 7.5$ Å. We find that, empirically, both permeability values scale with L in a similar manner between $L = 15$ and 30 Å. As such, the ratio of p_f/p_d for this system is largely independent of L over this range.

3.2 Estimating errors in permeation count statistics from low temporal resolution trajectories

The ability to accurately enumerate the number of water permeation events crossing the transporter is limited by the temporal resolution with which the system coordinates were recorded. At a 1 ps sampling frequency, each water molecule in the system only moves a small fraction of the total length of the transporter. This allows us to unambiguously demarcate when a specific water molecules moves from one side of the membrane to the other via the lumen of the transporter. At decreasing temporal resolution of the trajectory data, the distance that a water molecule can diffuse between observations approaches the length of the passage through vSGLT as well as the distance between the water-lipid interface and the periodic boundary. When those distances become comparable, a growing subpopulation of the crossing events will not be detected by our counting algorithm, leading to under counting of the true directional flux of water. A distribution of the crossing times for waters that move through the transporter through the central region, which is bounded by two planes separated by 30 Å along the z-axis, is shown in Fig. S1.

In addition to errors arising from *false negatives*, low temporal resolution of the trajectory data can result in *false positive* counts, wherein water molecules that move from one bulk region to the other through the periodic boundary are counted as crossing events due to transient excursions into the central region of the simulation box. While it is impossible to prevent under counting due to false negatives, counts arising from the second class of errors can be eliminated from our statistics. This is done by detecting the large jumps (> 50 Å) in the z position of a water between adjacent time steps due to a periodic boundary crossing that lead to an otherwise spurious crossing event.

To estimate the error in the counting statistics from using original trajectory data sampled at 100 ps intervals, we examine the subset of the trajectories that were re-sampled at 1 ps. Crossing statistics from the 1 ps data are considered to be a true measure of the number of permeation events in the simulation. We then down sample those trajectories, taking the coordinates of every hundredth frame and recalculate the number of permeation events while applying our filtering method to remove the false positive events. The time-dependent directional flux using the 1 ps data and the down sampled 100 ps data are shown in Fig. S2. The difference between the total number of counts using the same data sampled at different temporal resolutions is approximately 6 % and is consistent with the small fraction of the total crossing events that take on the order of the 100 ps observation interval.

4. Discussion

4.1 Discrepancies between reported osmotic permeabilities for vSGLT in the literature

The osmotic permeability, p_f, reported here (2.7×10^{-13} cm^3/s) is consistent with the value we reported previously (4.1×10^{-13} cm^3/s) based on a different set of simulations of the same inward-facing state of vSGLT. Meanwhile, the Tajkhorshid laboratory has reported a p_f value for simulations on the same conformation of the protein, which is much smaller (4.75×10^{-15} cm^3/s) (19). This discrepancy arises from a difference in the calculation of the net crossing events, $n(t)$, appearing in Eq. 4. As described above, we integrate the differential form of the collective variable (Eq. 7) to obtain $n(t)$ as originally described in Ref. 17. In this formalism, $n(t)$ is a function of all water molecules within the channel at time t and $t+dt$. Conversely, Li et al. (19) calculate $n(t)$ from the cumulative sum of the discrete efflux and influx permeation events (Dr. Emad Tajkhorshid, personal communication). Using this later method, $n(t)$ depends only on the dynamics of water molecules that fully permeate the transporter. On long timescales both methods provide qualitatively similar time series for $n(t)$, as can be seen by comparing Figures 2B and S3A in Ref. 16; however, on short timescales these two methods give different results. Specifically, $n(t)$ calculated as in Ref. (19) changes in integer steps, while $n(t)$ calculated from the collective coordinate varies continuously. When the averaging time is shorter than or comparable to the mean time

between permeation events, the discrete event method often results in $n(t)^2$ values that are nearly constant and close to zero. As a result, a linear regression fit of the the average slope of $n(t)^2$ calculated from the discrete permeation counts systematically yields a smaller estimate of p_f than the collective coordinate method. We note, however, that Eq. 4 is only applicable when t is much longer than the velocity correlation time of $n(t)$ (17). Analyzing the data presented here with the discrete event method employed in Ref. (19) also produces a much smaller value of $p_f = 9.5\times10^{-15}$ cm^3/s, suggesting that the water dynamics underlying the two studies are similar. Nonetheless, we believe that the collective coordinate method developed by Zhu et al. (17), and used here, is preferred as it has been validated against nonequilibrium simulations in the presence of a chemical potential difference (17) and experimental measurements (20–22).

References

[1] Faham, S., A. Watanabe, G. M. Besserer, D. Cascio, A. Specht, B. A. Hirayama, E. M. Wright, and J. Abramson, 2008. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+/sugar symport. *Science* 321:810–4.

[2] Sali, A., and T. L. Blundell, 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815.

[3] Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12–21.

[4] Lomize, M. A., A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, 2006. OPM: orientations of proteins in membranes database. *Bioinformatics* 22:623–5.

[5] Jo, S., T. Kim, and W. Im, 2007. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS One* 2:e880.

[6] Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, 2005. Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–802.

[7] Shaw, D. E., R. O. Dror, J. K. Salmon, J. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, et al., 2009. Millisecond-scale molecular dynamics simulations on Anton. *In* Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. ACM, 39.

[8] MacKerell, A. D., D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. a. Ha, et al., 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B* 102:3586–3616.

[9] Mackerell, A. D., Jr, M. Feig, and C. L. Brooks, 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* 25:1400–15.

[10] Klauda, J. B., R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, Jr, and R. W. Pastor, 2010. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* 114:7830–43.

[11] Guvench, O., S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor, and A. D. Mackerell, Jr, 2008. Additive empirical force field for hexopyranose monosaccharides. *J Comput Chem* 29:2543–64.

[12] Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* 79:926.

[13] Kräutler, V., W. F. van Gunsteren, and P. H. Hünenberger, 2001. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry* 22:501–508.

[14] Shan, Y., J. L. Klepeis, M. P. Eastwood, R. O. Dror, and D. E. Shaw, 2005. Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J Chem Phys* 122:54101.

[15] Tuckerman, M., B. J. Berne, and G. J. Martyna, 1992. Reversible multiple time scale molecular dynamics. *The Journal of chemical physics* 97:1990.

[16] Choe, S., J. M. Rosenberg, J. Abramson, E. M. Wright, and M. Grabe, 2010. Water permeation through the sodium-dependent galactose cotransporter vSGLT. *Biophys J* 99:L56–8.

[17] Zhu, F., E. Tajkhorshid, and K. Schulten, 2004. Collective diffusion model for water permeation through microscopic channels. *Phys Rev Lett* 93:224501.

[18] Zhu, F., and G. Hummer, 2010. Pore opening and closing of a pentameric ligand-gated ion channel. *Proc Natl Acad Sci U S A* 107:19814–9.

vSGLT	hSGLT1
S76	T90
Y87	F101
N245	G272
D336	N363
Q422	Q451
F424	F453
Q425	D454
Q428	Q457

Table S1. Residues involved in modulating water flow through the outer-gate of vSGLT and the corresponding residues in hSGLT1 based on the sequence alignment in (1)

[19] Li, J., S. A. Shaikh, G. Enkavi, P.-C. Wen, Z. Huang, and E. Tajkhorshid, 2013. Transient formation of water-conducting states in membrane transporters. *Proc Natl Acad Sci U S A* 110:7696–701.

[20] Aksimentiev, A., and K. Schulten, 2005. Imaging alpha-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys J* 88:3745–61.

[21] Hashido, M., M. Ikeguchi, and A. Kidera, 2005. Comparative simulations of aquaporin family: AQP1, AQPZ, AQP0 and GlpF. *FEBS Lett* 579:5549–52.

[22] Jensen, M. Ø., and O. G. Mouritsen, 2006. Single-channel water permeabilities of Escherichia coli aquaporins AqpZ and GlpF. *Biophys J* 90:2270–84.

[23] Scargle, J. D., J. P. Norris, B. Jackson, and J. Chiang, 2013. Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations. *The Astrophysical Journal* 764:167.

[24] Vanderplas, J., A. Connolly, Ž. Ivezić, and A. Gray, 2012. Introduction to astroML: Machine learning for astrophysics. *In* Conference on Intelligent Data Understanding (CIDU). 47 –54.

[25] Efron, B., and R. Tibshirani, 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 54–75.

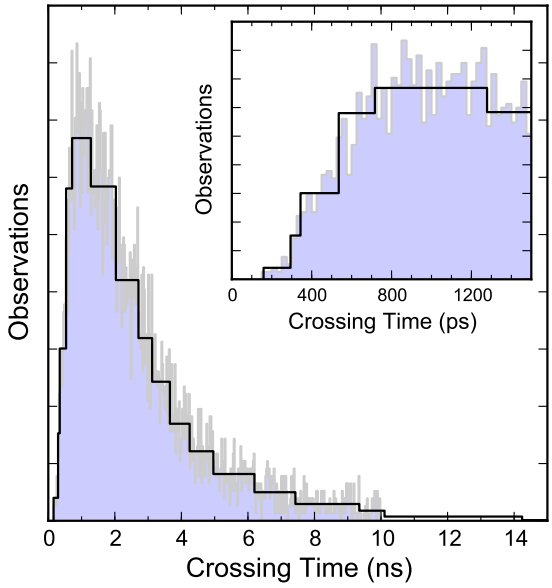

Figure S1. Distribution of crossing times for waters moving between the two planes defining the central region of the simulation box, which contains the lipid membrane and transporter. The crossing times are first partitioned using a uniform bin size of 25 ps (blue) and then using a set of variable bins determined by the Bayesian blocks approach (black) (23, 24). (Inset) The same data plotted over crossing times less than 1.5 ns.

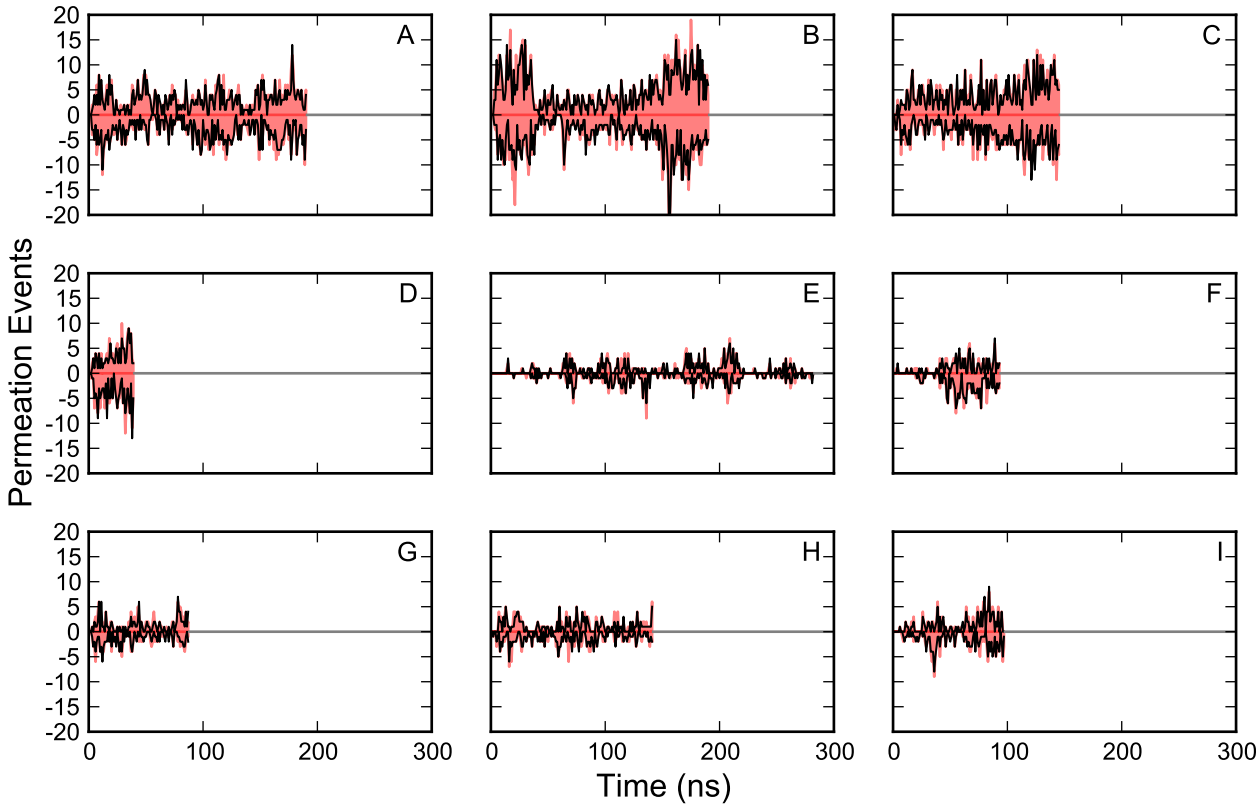

Figure S2. Comparison of time-dependent direction fluxes of water calculated from trajectories sampled at a 1 ps interval (red) and the same data down sampled to 100 ps between frames (black). Counts arising from efflux of water through the transporter is shown with positive values, and influx as negative values.

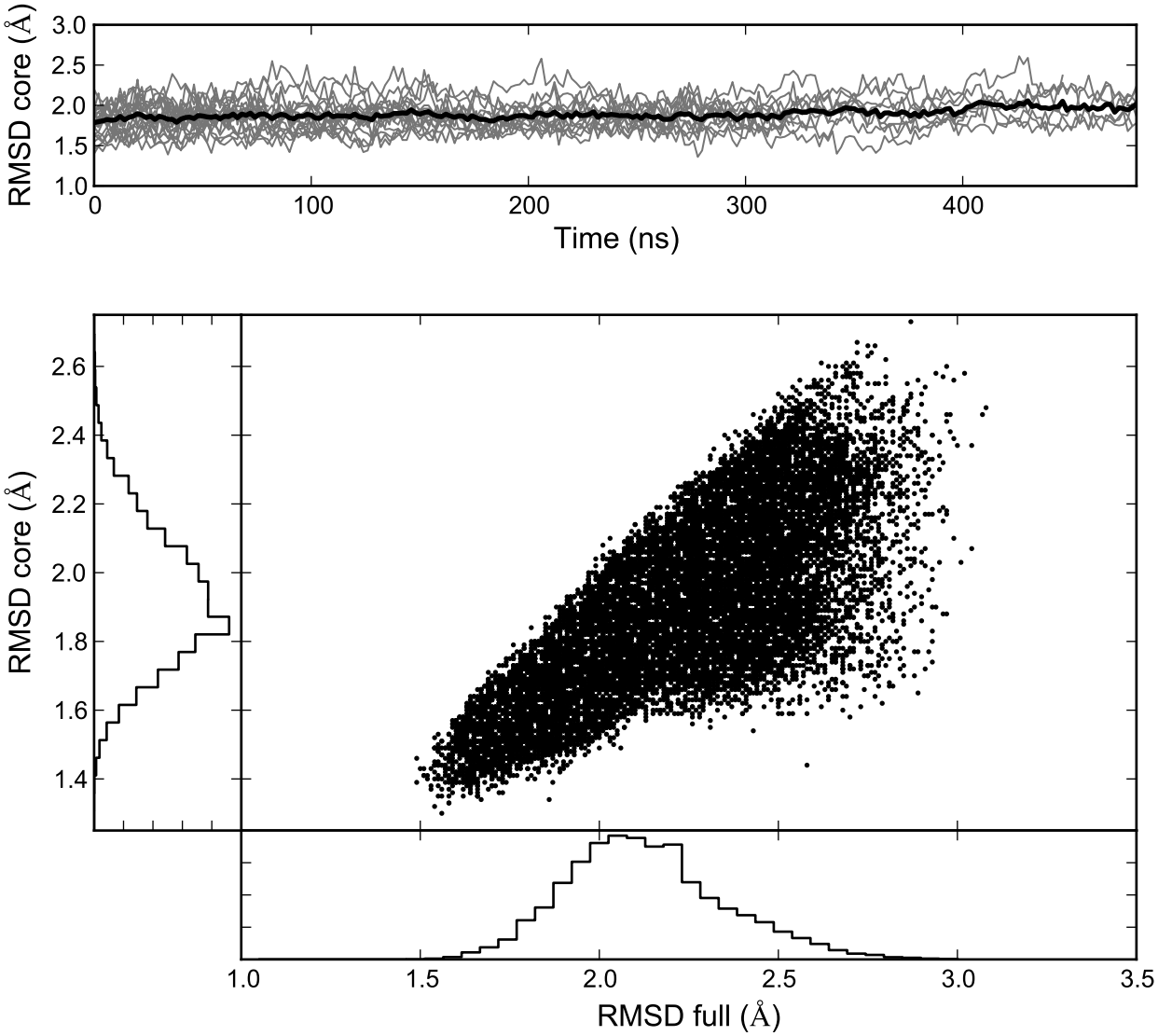

Figure S3. Cα RMSD of vSGLT. (Upper) Time series of the Cα RMSD from the crystallographic structure (PDB: 3DH4) of the ten core transmembrane helices for the 21 independent trajectories. The individual traces for each simulation and the average RMSD for the ensemble are shown in gray and black, respectively. (Lower) The Cα RMSD for core 10 TMs and the full transporter for all conformations visited in the aggregate ensemble. The probability distribution for each RMSD metric is shown adjacent to the central scatter plot, along its respective axis.

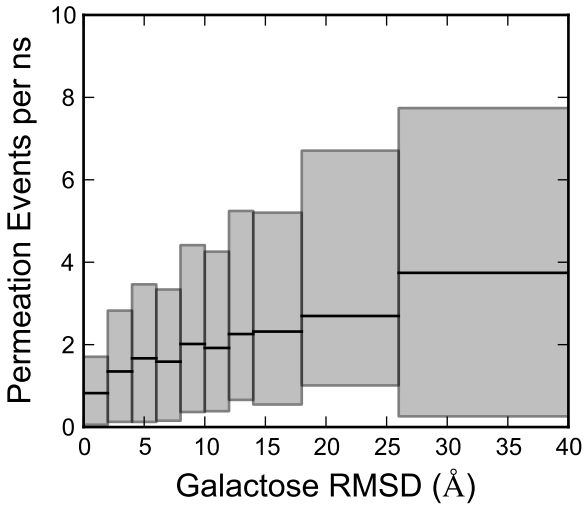

Figure S4. Galactose position dependent flow of water through vSGLT. Time series of the average inward and outward flow were calculated along with the corresponding RMSD of the galactose from its crystallographic pose after superimposing the transporter. After averaging the both quantities in 2 ns blocks, the blocks were partitioned into bins based on the mean RMSD in the block. The average flow in each bin is shown as a black line with the grey box denoting the 95 % confidence interval calculated using Monte Carlo bootstrapping (25).