Biophysical Journal, Volume 99

Supporting Material

Water Permeation Through the Sodium-Dependent Galactose Cotransporter vSGLT

Seungho Choe, John M Rosenberg, Jeff Abramson, Ernest M Wright, and Michael Grabe

Water Permeation Through the Sodium-Dependent Galactose Cotransporter vSGLT

Supporting Material

Water Permeation Through the Sodium-Dependent Galactose Cotransporter vSGLT

Seungho Choe[†], John M. Rosenberg[†‡], Jeff Abramson[§], Ernest M. Wright[§], and Michael Grabe[†‡]

[†]Department of Biological Sciences, and [‡]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania, 15260, USA; and [§]Department of Physiology, David Geffen School of Medicine at University of California, Los Angeles, California, 90095-1751, USA

System Setup for Molecular Dynamics Simulations

We extracted chain A of the recent vSGLT (PDB ID 3DH4) x-ray structure (1) and oriented the protein with respect to the membrane using the Orientation of Proteins in Membranes (OPM) algorithm (2). The structure was then inserted into a pre-equilibrated and solvated 1-palmitoyl-2-oleoyl phosphatidylcholine (POPC) bilayer using the CHARMM-GUI (3). The first 20 residues of the first transmembrane segment (called TM-1) are not assigned, and the loop from TM-1 to the second transmembrane segment (called TM1) is missing. Therefore, since TM-1 is not part of the 10 conserved helices in the Na^+-dependent co-transporter family (1, 4), we removed it and the TM-1 to TM1 loop. Our simulation includes residues 53-547. Next, we used the loop modeling routine in MODELLER (5) to build in 6 missing residues between TM4 and 5. During system construction, 36 sodium and 36 chloride ions were added to solution to approximate a physiological concentration of 150 mM. The final system was electroneutral in a hexagonal box approximately $96 \times 96 \times 84$ $Å^3$ consisting of about 63,000 atoms.

NAMD (6) was used to carry out the simulations using the CHARMM27 parameter set (with CMAP corrections (7)) and the TIP3P water model (8). The TIP3P model has three interaction sites, corresponding to the three atoms of a water molecule (two positively charged hydrogen atoms and one negative oxygen atom). TIP3P is a natural water model to use for our simulations since the CHARMM force field was developed hand in hand with TIP3P. The self-diffusion coefficient of TIP3P is 2-3 times larger than experimental values (9). This overestimate implies that permeability values calculated from our simulations are most likely artificially high, which helps to reconcile some, but certainly not all, of the differences between our high computational estimates and the true experimental values. MATLAB (Mathworks, Inc., Natick, Ma) and VMD (10) were used for visualization and analysis. Initial minimization was carried out using the conjugate gradient method followed by gradual heating to 310 K using Langevin dynamics with a 10 ps^{-1} damping coefficient. The first 300 ps of dynamics were carried out in the NVT ensemble and all water molecules, galactose, Na^+, and heavy backbone and side-chain atoms were constrained in a harmonic potential with a 10.0 kcal/mol/$Å^2$ force constant. We then switched to the NPT ensemble, and the restraints on the water molecules and side-chain heavy atoms were removed in 5 steps over the next 1.5 ns. This was followed by the removal of the backbone, Na^+ and galactose restraints in 6 discrete steps over the next 1.8 ns. Finally, the system was run restraint free for an additional 10 ns before carrying out the production run. Pressure was maintained at 1 atm with the Langevin piston method employing a 200 fs piston period and 100 fs piston decay constant. We used a discretized time step of 2 fs, and trajectories were saved every 0.1 ps for analysis. Bond lengths between hydrogen and heavy atoms were constrained using the SHAKE algorithm (11). Van der Waals interactions were cut-off at 10 Å, and the Particle Mesh Ewald summation was employed for electrostatics.

The Role of Water In Mediating Substrate-Protein Interactions

Galactose is poised in the vSGLT structure where it makes intimate contact with several binding site residues (1). Notably, the structure shows that Y263 on TM6 is the primary residue making up the inner hydrophobic gate blocking exit to the intracellular space. In our simulation, Y263 adopts a new rotamer conformation at 52 ns (from $\chi_1 \sim -180°$ as seen in the x-ray structure to $\chi_1 \sim -50°$) that widens the exit pathway allowing galactose to exit via the hydrophilic cavity. During the course of the simulation we observe that water enters the site primarily from the intracellular space. We calculated the average interaction energies between the binding site residues and the galactose during unbinding to understand the role of water in this event (see Supplemental Table 1). We considered two time periods: one from 0-52 ns during which Y263 adopts the conformation seen in the x-ray structure (first column) and 52-105 ns during which Y263 flips (second column). The interaction energy between galactose and binding site residues E88, N260, and Q428 all decrease indicating a weaker interaction with the protein. Q428 on TM10 shows the largest change since it stops hydrogen bonding with galactose and starts bonding to Y263 after the rotamer change at 52 ns. Meanwhile, the decreased interaction with the protein is

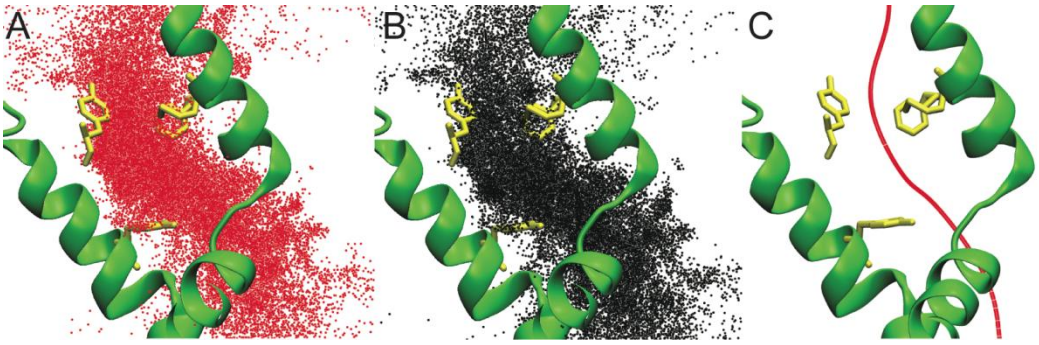
compensated for by an increased interaction energy with water due to increased solvation (~10 kcal/mol). Note that while the inner gate, composed of Y263, opens from 52 to 105 ns, the water flux through vSGLT is nearly constant from 0 to 100 ns (see Figure 2A in the main text). Only after ~100 ns does the water flux through vSGLT increase as shown in Figure 2A. Galactose begins to enter the bulk intracellular space between 105-110 ns, and it is at this point that it unblocks the water permeation pathway. We believe that this is the primary reason why permeability increases after 100 ns.

SUPPLEMENTAL TABLE 1 Galactose interaction energies [kcal/mol]

	0 – 52 ns	52 – 105 ns
entire protein	-57 ± 0.4	-49 ± 23
Q69	-1 ± 2	-1 ± 2
E88	-26 ± 6	-18 ± 17
S91	-3 ± 3	-3 ± 3
N260	-5 ± 1	-2 ± 2
K294	-4 ± 6	-4 ± 7
Q428	-5 ± 3	-1 ± 1
N64	-1 ± 1	-3 ± 3
E68	1 ± 1	-3 ± 4
water	-21 ± 7	-31 ± 13
bound Na^+	-0 ± 0.3	Not applicable

WATER PERMEATION OCCURS VIA THE GALACTOSE BINDING SITE

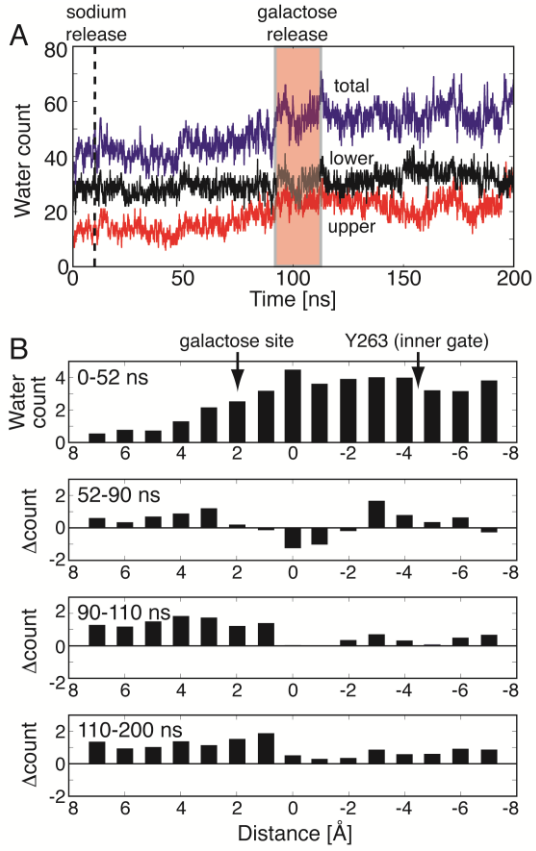
The simulation system is periodic in the z direction, and this gives rise to rapid exchange of waters between the intracellular and extracellular baths at the upper and lower simulation boundaries, ± 42 Å. We used the history of each water molecule, sampled every 10 ps, to discern events that pass from one bath to the other by going through the membrane mid-plane or the simulation boundaries. For our initial analysis, all water molecules were initially labeled as being in the upper bath, lower bath, or mid-plane region if the z coordinate of the oxygen atom was greater than 7.5 Å, lower than -7.5 Å, or between both values, respectively. When a water entered a new region, it was relabeled, and those that passed through the mid-plane region from one bath to the other were counted as a permeation event. Inspection of the successful permeation trajectories in Supplemental Figure 1 showed that all water molecules moved through the galactose binding site. Panels A and B show the superposition of all successful outward trajectories (red dots) and inward trajectories (black dots), respectively. There is no discernable difference between outward and inward moving events. Next, we refined the definition of the water channel by using all the successful events in Supplemental Figure 1A,B. We calculated the center of mass of the water molecules in 1 Å slabs along the z-axis of the transporter. These points formed a continuous pathway. Along this path, we used a 10 Å radius in the x-y plane to define the channel and identify permeating water molecules, and we used this for the analysis of p_d and p_f below.



SUPPLEMENTAL FIGURE 1 Inward movement is defined to be negative. (*A and B*) Superposition of successful outward water trajectories (red dots in A) and inward trajectories (black dots in B) over the static vSGLT structure from the inset of Figure 1 of the main text. Dots represent the position of the oxygen atom. (*C*) Path through vSGLT used to carry out permeability coefficient calculations. A polynomial spline was fit to the center of mass of the water positions in A and B using MATLAB built in functions. Water movement was projected onto this curve to determine dn(t) as described below.

CHANGES IN THE PORE STRUCTURE DURING THE SIMULATION

We wanted to look more closely at the protein conformational changes that vSGLT undergoes during our simulation. First, in Supplemental Figure 2A, we calculated the total number of water molecules in the pore as a function of time (blue curve). As above, waters in the pore had z-values between ±7.5 Å. To distinguish changes in the upper gate and the lower gate, we also determined the number of water molecules in the upper part of pore (z = +7.5 to 0 Å, red curve) and the lower pore (z = 0 to -7.5 Å, black curve). The water in the lower pore is nearly constant over the 200 ns simulation, while the number of waters in the upper pore increase from 14 to 23 during galactose exit. This increase is mirrored in the total water count (blue curve). Galactose occupies the space of about 9 water molecules, so this increase is primarily due to water filling the void left by galactose. In panel B, we calculated the distribution of water molecules along the pore averaged over 4 important time periods: 0-52 ns (before the Y263 undergoes a rotamer conformation), 52-90 ns (Y263 has adopted a rotamer compatible with galactose exit, but galactose is still in the binding site), 90-110 ns (the time of galactose exit), and 110-200 ns (the remainder of the simulation after galactose exits). Fifteen uniformly spaced bins were constructed along the pathway shown in Supplemental Figure 1C, and the average number of water molecules in each slice was calculated. For reference, arrows indicate the center of mass of the galactose binding site and the inner gate formed by Y263. The bottom 3 panels show the difference (Δ count) compared to the first panel. There are several features that can be gleaned from these panels. First, the top panel shows that there is much more water in the lower half of the channel than the upper half. This is consistent with vSGLT being an inward-facing state. Second, the bottom two panels show an overall increase in water at all points along the pore, especially in the upper part of the pore from +7.5 to 0 Å. This is due to the exit of galactose as well as a slight widening of the pore. In particular, our simulations show that the extracellular gate, composed of residues M73, Y87 and F424, has a degree of flexibility, which is important for water permeation.



SUPPLEMENTAL FIGURE 2 The spatial and temporal distribution of water in the pore. (*A*) The total water in the pore over time. We split the pore into two regions: the upper pore, which includes the extracellular gate (z = +7.5 to 0 Å), and the inner pore, which includes the intracellular gate (z = 0 to -7.5 Å). The red curve is the total number of water molecules in the upper, the black curve the total in the lower pore, and the blue curve is the sum in the entire pore. Na^+ escape is indicated by a dashed line and galactose release occurs during the time span

highlighted by the shaded box. (*B*) The average number of water molecules distributed along the pore over 4 time frames. The top panel shows the number of water molecules from 0-52 ns before Y263 adopts the conformation that allows galactose exit. The pore was split into 1 Å thick slices along the pathway identified in Supplemental Figure 1C. The position of Y263 and the center of mass of the galactose binding site are indicated by arrows. The bottom three panels show changes in the average number of water molecules compared to the first panel: 52-90 ns – Y263 is in a position to allow galactose exit, 90-110 ns – galactose exits, and 110-200 ns – the transporter is galactose free.

TIME SCALE OF GALACTOSE EXIT

It is interesting to ask why galactose undergoes such a long unidirectional random walk out of the binding site and into the bulk water on what appears to be a short timescale. Energetic analysis of galactose exit shows that galactose is only weakly bound to the inward-occluded conformation and that the potential of mean force for unbinding does not have any large barriers (data not shown). The movement of the inner gate, Y263, is a critical molecular event required for galactose to initially exit the site. After galactose moves 6-8 Å to pass the gate, Y263 adopts a conformation that prohibits the return of the galactose to the binding site. Therefore, galactose movement is rectified by conformational changes in the protein. This helps to explain the unidirectional flow of galactose and water during the 90-110 ns time frame.

To better gauge if the observed escape time of 20 ns is characteristic of a natural diffusive process, we estimate the time for galactose to diffuse 17 Å along the exit pathway using the 1-dimensional diffusion law: $\tau = R^2/(2D)$, where D is the diffusion coefficient of galactose in the protein. Based on our simulations from 0 to 110 ns, the diffusion coefficient is 0.023×10^{-9} m^2/s along the exit pathway. For reference, the experimental value in bulk solution is 1×10^{-9} m^2/s (12), about 40 times higher, which is expected due to the lack of protein-substrate interactions in water. We then estimate the exit time as:

$$\tau = \frac{R^2}{2D} = \frac{(17\times10^{-10}m)^2}{2\times0.023\times10^{-9}m^2/s} \approx 60\ ns. \qquad 1$$

Thus, based on the actual diffusion coefficient of galactose in the transporter, the observed exit time is faster, but of the same order of magnitude as simple estimates based on diffusion. Therefore, we feel that the simulation is representative of a typical naturally occurring event.

GALACTOSE ACTS AS A SLOW MOVING WALL

If galactose exit forcefully expels water from the intracellular vestibule of vSGLT, this may be an activated process that requires energy. Here we use simple analysis to estimate the magnitude of this energy. About 75 water molecules are pushed out of the intracellular cavity during exit. Assuming that they form a sphere, the radius, r, is given by

$$r = \left(\frac{3}{4\pi}\cdot N_{water}\cdot\rho_{water}\right)^{1/3} \approx 0.8\ nm, \qquad 2$$

where the density of 55 molar water is about $\rho_{water} = 0.03$ nm^3/mol. Analysis of the simulation indicates that galactose exits at a rate of 1 Å/ns, or 0.1 m/s. Using this observation in conjunction with Stokes' formula, the drag force on the expelled water is:

$$F_{drag} = 6\pi\eta r v = 6\pi\cdot\left(0.7\times10^{-3}\frac{kg}{m\cdot s}\right)\cdot\left(8\times10^{-10}m\right)\cdot\left(0.1\frac{m}{s}\right) = 1.1\ pN, \qquad 3$$

where η is the viscosity of water at 310 K (13). Integrating the force along the length of the exit pathway (about 2 nm), we arrive at the energy dissipated due to correlations created in the water:

$$W = F_{drag}\times d = 1.1\ pN\times 2\ nm = 2.2\ pN\cdot nm \approx 0.5\ k_BT. \qquad 4$$

This energy value is smaller than a thermal fluctuation indicating that it is a reversible process. In other terms, the time scale of galactose exit is much slower than the diffusive motion of water.

CALCULATION OF PERMEABILITY AND DIFFUSION COEFFICIENTS

In the presence of an osmotically active molecule, water will flow through from the low concentration side to the high concentration side if water channels are present. If the inner concentration is c_i (mol/cm^3) and the outer concentration c_o (mol/cm^3), then, in the absence of hydrostatic pressure and in dilute concentrations, the water flux is linearly proportional to the concentration difference of osmotically active particles:

$$J_w = p_f(c_o - c_i), \tag{5}$$

where J_w (mol/s) is the flux of water, and the constant of proportionality, p_f (cm^3/s), defines the osmotic permeability of the system (14). Another closely related phenomenon is the diffusion of tagged water molecules from one side of the channel to the other. This can easily be accomplished in a molecular simulation by labeling all waters on one side of a membrane and tracking their movement through the pore, and it can be achieved experimentally by using radioactive tracer assays to track isotopes specifically added to one side of the membrane. Again, the flux of tracer particles obeys a linear flux equation:

$$J_t = p_d(t_o - t_i), \tag{6}$$

where J_t (mol/s) is the tracer flux, t_i and t_o are the respective inner and outer tracer concentrations, and p_d (cm^3/s) defines the diffusion permeability. Note that the inner and outer indices in Eqs. 5 and 6 are reversed since water flows *toward* the region of high solute concentration, but tracers diffuse *away* from the region of high tracer concentration. For an excellent discussion of p_d and p_f please refer to the work of Zhu *et al.* (15).

The diffusion permeability, p_d, can be computed directly from the permeation counts in Figure 2 of the main text by using Eq. 6. If all intracellular waters are tagged, t_i then becomes the molar concentration of water ~ 0.055 mol/cm^3. By the end of the 200 ns simulation 823 outward crossing events were recorded and 805 inward events giving an average of 814. Using this value, Eq. 6 can be solved to give $p_d = (814/200\ \text{ns}^{-1})/(0.055\ \text{mol/cm}^3)/N_A = 1.2\times10^{-13}$ cm^3/s, where N_A is Avogadro's Number. When the length used to define the channel region, L, was increased from 15 Å to 20 Å, the number of total outward and inward permeation events decreased to 658 and 635, respectively. This leads to 20% reduction in p_d with very little change to the net number of crossings. Since L = 15 Å is beyond the hydrophobic gates and inner hydrophilic vestibule, we carried out all of our calculations with this value.

Since our simulation is at equilibrium, there is no osmotic concentration gradient or pressure gradient across the membrane, which makes it impossible to calculate p_f directly from Eq. 5. Fortunately, Zhu and co-workers provided an elegant solution to this problem when they showed that p_f is related to change in the square of the number of permeation events determined from an equilibrium simulation (16). The following discussion follows directly from their work. Consider the incremental change of all water molecules in the pore during a time step dt:

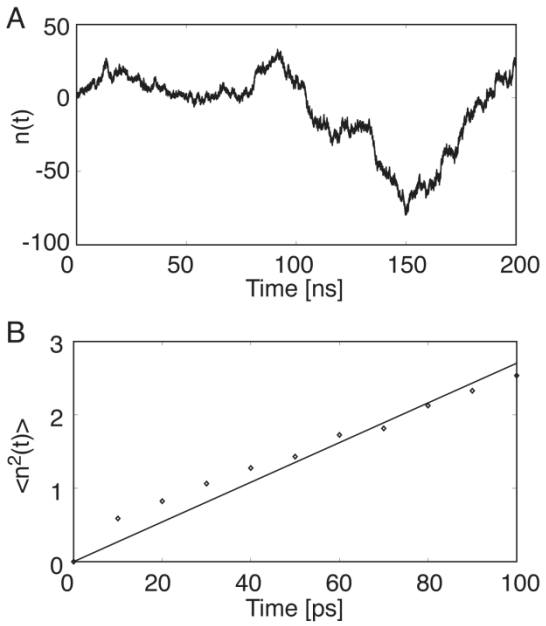
$$dn(t) = \sum_{i=1}^{M} \frac{z_i(t+dt) - z_i(t)}{L}, \tag{7}$$

where $z_i(t)$ is the z component of the oxygen atom of the i^{th} water molecule in the channel at time t and $z_i(t+dt)$ is the value a time dt later. In our case, the water channel through the transporter is curved as shown in Supplemental Figure 1C. So we project the incremental movement of each water molecule onto this curve. The position at time t is projected to point p_1 and the position at t+dt is projected to point p_2. Then in Eq. 7, L is the total contour length along the path, and the incremental distance $z_i(t+dt) - z_i(t)$ is replaced by the contour integral from p_1 to p_2. Projections were carried out by identifying the point on the contour with the same z value of the oxygen atom at each time. Contour integrals were computed numerically using MATLAB. The sum is over all M water molecules in the channel as defined in the last section with the additional requirement that the water be in the pore at time t and t+dt. From this definition it follows that n(t) is the integral of dn(t) from time 0 to a later time t with the condition that n(0) = 0. With this definition, when a single water passes from the bottom to the top of the channel n = +1 and the value is -1 for inward permeation events. Thus, this metric corresponds to the net number of transported water molecules at time t. It can be shown that n(t) is related to the water diffusion coefficient through the channel, D_n, and therefore p_f, according to:

$$\left\langle n^2(t) \right\rangle = 2D_n t = 2\frac{p_f}{v_W} t, \tag{8}$$

where the brackets denote an ensemble average over many trajectories and v_W is the average volume of a single water molecule (18 cm^3/mol/$N_A \cong 3\times10^{-23}$ cm^3). To compute this quantity, we first squared each value in the trace of n(t) in Supplemental Figure 3A to obtain $n^2(t)$. Next, we split the trace into 2000 consecutive 100 ps windows, and each window was shifted to the origin along the time axes and then translated along the y-axis so that the beginning of the trace was at

zero. From this point, we averaged all n^2 values at each time point to obtain the diamonds in Supplemental Figure 3B. Next, we carried out a linear regression to give the best fit (solid line) with the constraint that the line pass through the origin. The slope is 27 counts^2/ns, which gives an estimate of the osmotic permeability of $p_f = 4.1\times10^{-13}$ cm^3/s using Eq. 8.



SUPPLEMENTAL FIGURE 3 Calculation of the permeability coefficient p_f from an equilibrium simulation. (*A*) Fractional movement n(t) during the 200 ns simulation. (*B*) The ensemble average of $n^2(t)$ (diamonds) calculated from 0 to 100 ps computed from the trace in panel A. The solid line is the linear regression whose slope determines p_f.

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* 321:810-814.

2. Lomize, M. A., A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg. 2006. OPM: orientations of proteins in membranes database. *Bioinformatics* 22:623-625.

3. Jo, S., T. Kim, V. G. Iyer, and W. Im. 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29:1859-1865.

4. Weyand, S., T. Shimamura, S. Yajima, S. i. Suzuki, O. Mirza, K. Krusong, E. P. Carpenter, N. G. Rutherford, J. M. Hadden, J. O'Reilly, P. Ma, M. Saidijam, S. G. Patching, R. J. Hope, H. T. Norbertczak, P. C. J. Roach, S. Iwata, P. J. F. Henderson, and A. D. Cameron. 2008. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* 322:709-713.

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

6. Kale, L., R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. 1999. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comp. Phys.* 151:283-312.

7. Mackerell, A. D., Jr. 2004. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* 25:1584-1604.

8. 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. *J. Chem. Phys.* 79:926-935.

9. Spoel, D. v. d., P. J. v. Maaren, and H. J. C. Berendsen. 1998. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *J. Chem. Phys.* 108:10220-10230.

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

11. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. *J. Comp. Phys.* 23:327-341.

12. Mogi, N., E. Sugai, Y. Fuse, and T. Funazukuri. 2007. Infinite dilution binary diffusion coefficients for six sugars at 0.1 MPa and temperatures from (273.2 to 353.2) K. *J. Chem. Eng. Data.* 52:40-43.

13. Chemical Rubber Co. 53rd edition, editor R.C. Weast, Handbook of Chemistry and Physics. Chemical Rubber Co., Cleveland.

14. Finkelstein, A. 1987. Water movement through lipid bilayers, pores, and plasma membranes. John Wiley & Sons, New York.

15. Zhu, F., E. Tajkhorshid, and K. Schulten. 2004. Theory and simulation of water permeation in aquaporin-1. *Biophys. J.* 86:50-57.

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