Supporting Information

Molecular Dynamics Simulations of Selective Metabolite Transport across the Propanediol Bacterial Microcompartment Shell

Jiyong Park,[1,†] Sunny Chun,[2] Thomas A. Bobik,[3] Kendall N. Houk,[1] Todd O. Yeates[1,2,4,]

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, [2]Molecular Biology Institute, University of California, Los Angeles, [3]Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, [4]UCLA-DOE Institute for Genomics and Proteomics, [†]Center for Catalytic Hydrocarbon Functionalization, Institute for Basic Science (IBS), Daejeon, 34141, Republic of Korea.

I. Permeability through a pore in the presence of a free energy barrier

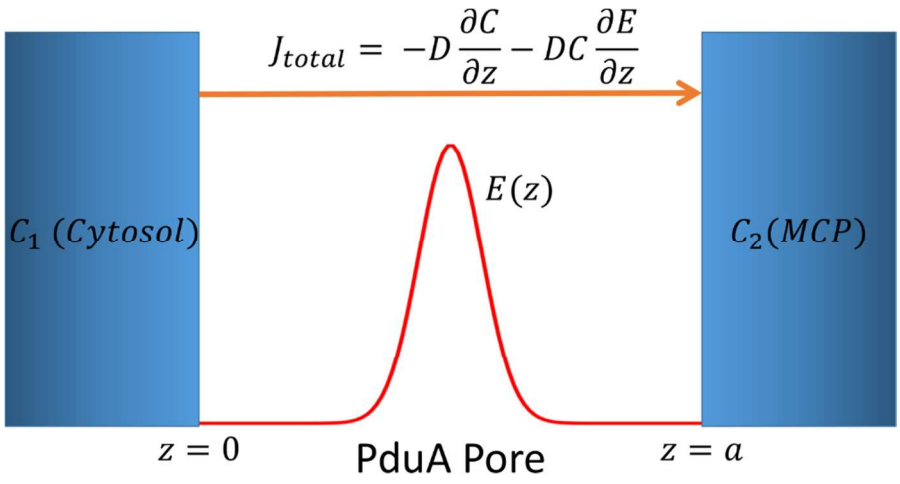

Figure S1. Mathematical description of the selective permeation through the PduA pore. We assume the initial substrate or the intermediate has constant concentrations at both the outside (C_1) and the inside (C_2) of the bacterial microcompartment (BMC). Following Fick's law of diffusion, the total flux (J_{total}) is the sum of the contributions from the concentration gradient ($-D\frac{\partial C}{\partial z}$) and the external energy gradient ($-DC\frac{\partial E}{\partial z}$).

We consider the diffusion of a small molecule through the PduA pore where Fick's laws of diffusion are obeyed. In this system, 1-D diffusion of the small molecule along an axis (z) through the PduA pore is considered (Figure S1). By Fick's 1st law, in the absence of an external force, the diffusion of the molecules in a concentration gradient gives rise to a net flux (J):

$$J_1 = -D\frac{\partial C}{\partial z}$$

, where D is the diffusion coefficient. In the presence of an external energy term, E, an additional transport velocity is caused by the external force:

$$v_E = -\frac{1}{f}\frac{\partial E}{\partial z}$$

where f is the frictional coefficient and is inversely related to D by the Einstein-Smoluchowski relation ($D = k_B T/f$), such that

$$v_E = -D\frac{\partial E}{\partial z}$$

for E expressed in units of $k_B T$. The transport velocity due to the energy term gives rise to a flux,

$$J_2 = v_E C = -DC\frac{\partial E}{\partial z}.$$

Thus, in the case of molecular diffusion on a non-uniform energy landscape, the total flux can be written as

$$J_{total} = J_1 + J_2 = -D\frac{\partial C}{\partial z} - DC\frac{\partial E}{\partial z}$$

, which is one form of the Smoluchowski equation for diffusion across an energy barrier.[1]

We consider the relative permeation at a steady state, where the two substrates are in the same constant concentrations at both sides of the PduA pore. This is a hypothetical situation where both the initial and the intermediate substrates are consumed at the same rate inside the MCP. At steady state, the total flux is constant over time ($\frac{\partial J_{total}}{\partial t} = 0$) and we can consider the spatial dependence of the flux. We are left to solve an ordinary differential equation of the form

$$-DC' - DCE' = c$$

, where the primes indicate the derivative with respect to position.

Let b stand in for the constant $-c/D$, so

$$C' + CE' = b.$$

For the one-dimensional system considered along z, where $z = 0$ on the lumenal side of the bacterial microcompartment pore, the general solution to this first-order linear differential equation is

$$C(z) = be^{-E(z)}\int_0^z e^{E(z)}dz.$$

We drop the additive constant from the integrated expression under the assumption that $C(0) = 0$. Then, expressing the permeability p as a ratio of flux J to concentration difference ΔC from positions z=0 to z=a,

$$p = \frac{-J}{C(a)}$$

Noting from substitutions above that the constant flux is equal to $-bD$, substituting the equation for C into the equation above and taking $E(a)$ to be zero, gives the following expression for permeability

$$p = \frac{D}{\int_0^a e^{E(z)}dz}$$

A similar method is used by Bauer and Nadler (2006) to derive flux of particles that are non-interacting with a channel, interacting with a channel, or blocked from passage through a channel.[2]

Thus, the permeabilities of the small molecules 1,2-propanediol and propionaldehyde through the PduA pore are proportional to expressions of the form $\frac{1}{\int_0^a e^{E(z)}dz}$, where the energy profiles are different for the two molecules, and those energy profiles are obtained in the present work from MD-based enhanced sampling simulations.

II. Lowest energy conformations of the metabolites

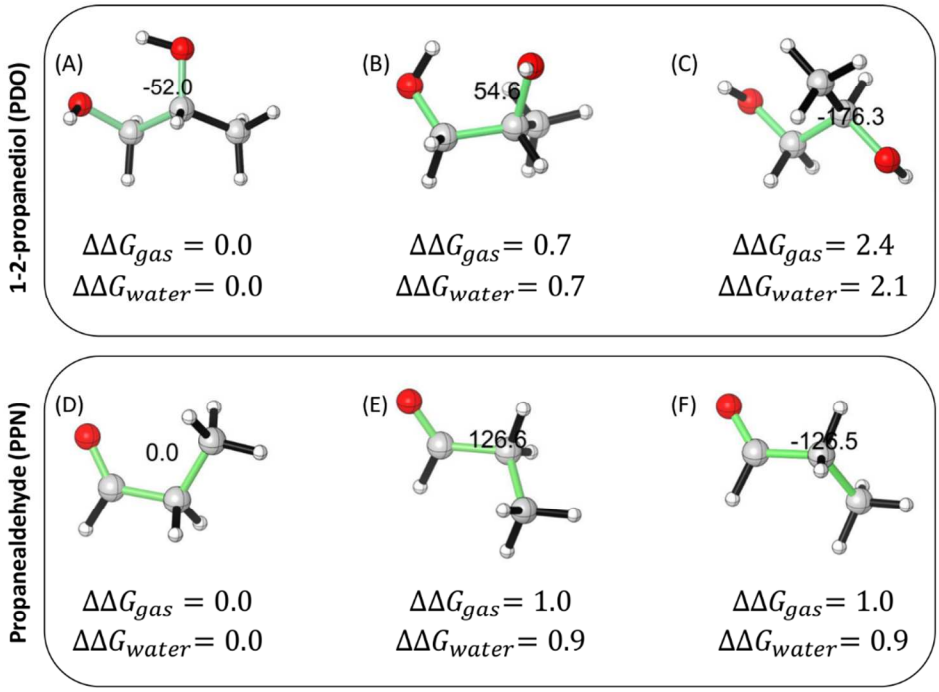

Figure S2. Three local energy minimum conformations of 1-2-propanediol (A-C) and propionaldehyde (D-F). The gas phase optimized structures at the B3LYP/6-31(d) level theory are shown. The free energy ($\Delta\Delta G_{gas}$ and $\Delta\Delta G_{water}$) of each conformation relative to the lowest energy conformation is shown. The energy unit is in kcal/mol. The three bonds defining the dihedral angle scanned (OH-C-C-OH of PDO and OH-C-C-C of PPN) are highlighted in green. The unit of angle is in degrees.

The lowest energy conformation of each metabolite was identified using conformational scanning analysis. For each metabolite, three local minimum conformations are optimized at B3LYP/6-31(d) level theory as shown in Fig. S2. The three local minimum conformations were identified by scanning a dihedral angle: θ(OH-C-C-OH) of PDO and θ(OH-C-C-C) of PPN. For PDO, the lowest energy conformation is 0.7 kcal/mol lower in energy than the second lowest conformation. Likewise, the lowest

energy conformation of PPN is 1.0 kcal/mol lower in energy than the second lowest energy conformation. The results show that the lowest energy conformation is more populated by > three times than the second lowest conformation, that justifies the use of force field parameters derived from the lowest energy conformations. The influences of the solvation environment were studied: the gas phase optimized structures are reoptimized at the B3LYP/6-31(d)/IEFPCM(water) level theory that accounts for the solvation environment. The reoptimized structures differ from the corresponding gas phase conformations by less than 0.1 Å of RMSD (data not shown). The finding indicated that the lowest energy conformation of each metabolite is not affected by the presence the solvation environment.

III. Normalization of the free energy profile computed from the metadynamics MD simulations

The collective variables used for the metadynamics simulations were: the radial distance (r) from the pore axis and the axial coordinate (z). At a radial coordinate r, the differential volume element in polar coordinates, namely $2\pi r\mathrm{d}r\mathrm{d}\theta\mathrm{d}z$, had to be considered in the final calculation of free energies. If the appropriate correction is not applied, a radial dependence on energy arises in the absence of forces, and this was confirmed in our analysis (data not shown). Specifically, the free energy difference between points (r_2, z_2) and (r_1, z_1) must take into account the ratio between the differential volume elements at r_2 and r_1, which is $2\pi r_2\mathrm{d}r\mathrm{d}\theta\mathrm{d}z/2\pi r_1\mathrm{d}r\mathrm{d}\theta\mathrm{d}z$, or simply r_2/r_1. More succinctly, this takes into account the Jacobian of the coordinate transformation between cylindrical and Cartesian coordinates. Expressed as a ΔG,

$$\Delta G = G(r_2, z_2) - G(r_1, z_1) = -k_B T ln\left(\frac{e^{-(U(r_2,z_2)-U(r_1,z_1))/k_B T}}{r_2/r_1}\right)$$

Where (r_1, z_1) can be chosen as a reference point remote from the energetic effects of the protein.

The apparent divergence of the expression above at r=0 is mitigated by the discretized sampling of the collective variables in the MD simulations. The radial coordinate (r) was sampled every 1.5 Å, beginning at r=0.75 Å (to cover the region at the center), and the axial coordinate (z) was recorded every 1.0 Å.

IV. Convergence of metadynamics MD simulations

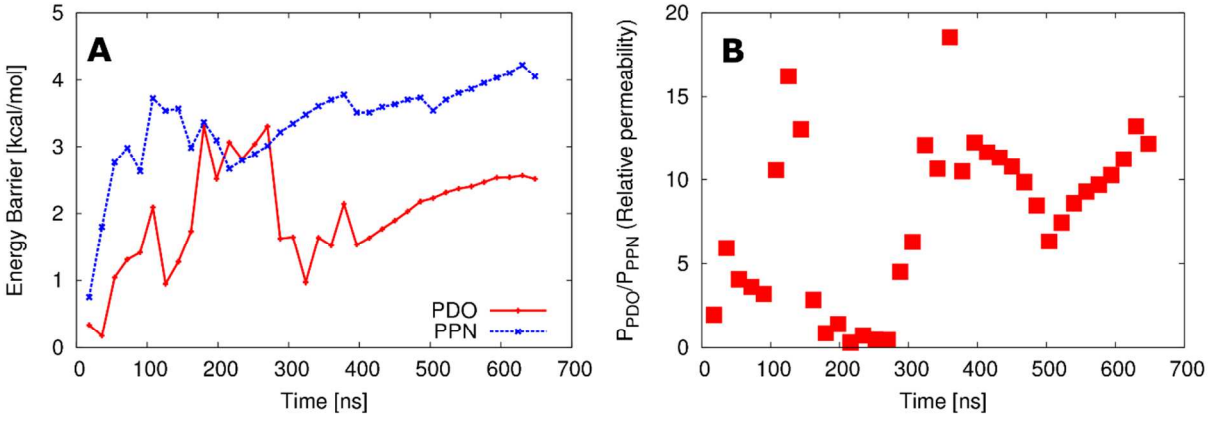

Figure S4. (A) The computed energy barrier heights and (B) the relative permeabilities from the metadynamics MD simulations are plotted over time.

The energy barrier height from the metadynamics MD simulations are plotted in Fig. S4 (A). The barrier height is defined by two values: the maximum near the pore ($|z| < 5.0$ Å) and the average free energy away from the pore (20 Å $\leq |z| \leq$ 25 Å). As the barrier height fluctuates over time, we estimated the statistical error of the computed barrier height by computing the standard deviation of the barrier height over the last 180 ns duration. The error in barrier height of the PDO simulations is 0.13 kcal/mol and that of the PPN simulations is 0.20 kcal/mol. Relative permeabilities of PDO and PPN are computed as a function of simulation time (Fig. S4 (B)). The permeabilities of both ligands (P_{PDO} and P_{PPN}) are defined by Eq. 2. The relative permeability (P_{PDO}/P_{PPN}) fluctuates significantly until 350 ns, and later converges to 9.7 ± 1.8, when averaged over the last 180 ns.

V. Statistical error of REUS MD simulations

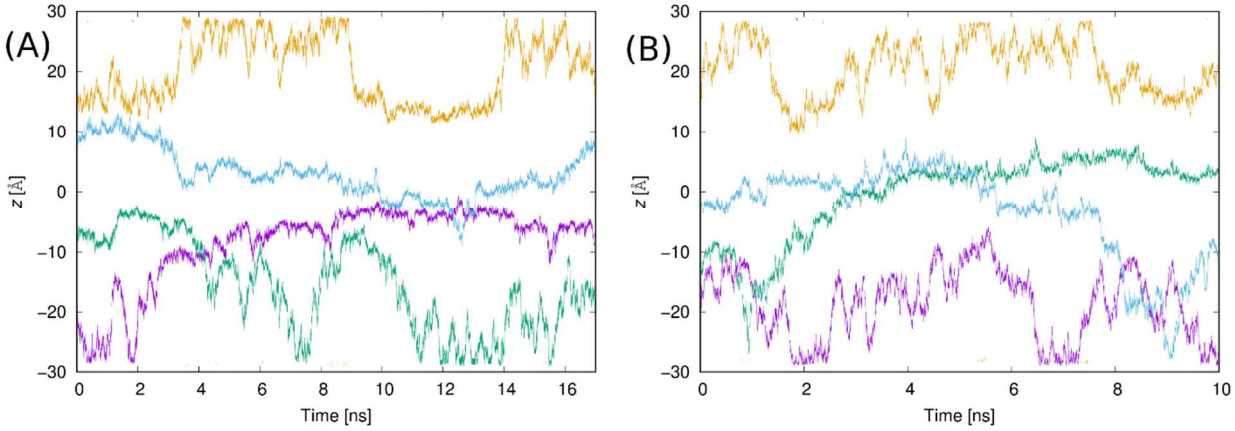

Figure S5. Sampled z-coordinates from REUS MD simulations: (A) PduA+PDO and (B) PduA+PPN. Four representative trajectories for the simulations of each small molecule are shown.

We used the weighted histogram analysis method (WHAM)[3, 4] to compute the Gibbs free energy profile from the REUS MD simulations. The statistical uncertainty of the computed free energy values were estimated using the bootstrap analysis method.[5] We repeated random resampling with replacement 200 times to compute the standard error of the free energy at each window. As shown in Fig S5 (A) and (B), two successive values of the sampled z-coordinates from an individual replica can be correlated. In order to produce uncorrelated resampled data, two data points separated beyond the correlated time interval needs to be sampled. We computed a normalized autocorrelation function ($C_Z(t)$) of the z-coordinates sampled from individual replicas[6]:

$$C_z(t) = \frac{\langle z(\tau + t)z(\tau)\rangle_\tau - z_0^2}{\sigma_z^2},$$

where $z(t)$ is the axial coordinate of a replica at *t*, z_0 is averaged energy, and σ_z^2 is the variance of energy. The axial coordinates are computed from the trajectories of individual replicas. The autocorrelation function is then fitted to an exponential function:

$$C_z(t) = \exp\left(-\frac{t}{t_{corr}}\right),$$

to give the energy correlation time (t_{corr}). The most extended correlation time is 23.0 ps, and that value was used as the correlation time interval in the bootstrap analysis. The estimated errors are depicted in Figure 3 of the main text.

The barrier heights (E_{height}) and the corresponding variances ($Var(E_{height})$) are analyzed:

$$E_{height} = \Delta G(z = 0) - \Delta G(away),$$

$$Std(E_{height}) = \sqrt{Var(\Delta G(z = 0)) + Var(\Delta G(away))}$$

, where $\Delta G(\mathrm{z} = 0)$ is the potential of mean force (PMF) at the center of the pore, $\Delta G(away)$ is the averaged PMF away from the pore (20 Å < |z| < 25 Å), and $Var(\Delta G)$ and $Std(E_{height})$ are the variance and the standard deviation of the energy values. We report the standard deviations as the uncertainty of the computed energy barrier heights.

VI. Diffusion coefficients of the metabolites in the aqueous medium and in the constricted pore region

We compared relative permeabilities of the PDO and PPN molecules under the assumption that the two metabolites have comparable diffusion constant in the aqueous medium and near the center of the constricted pore. We confirmed the assumption by computing the diffusion constants in the aqueous medium using unconstrained MD simulations. In addition, we calculated the diffusion constants in the pore based on a formula proposed by Hummer.[7] The computed diffusion coefficients show that the two metabolites share near identical diffusion constants in the bulk medium and at the pore. Of note, the diffusion constants at the center of the pore are about half those in the bulk medium.

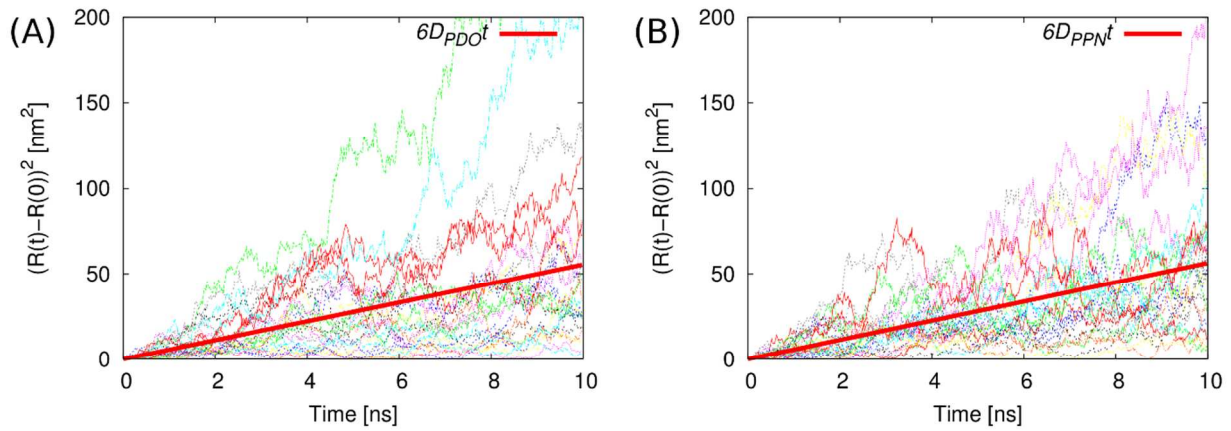

Figure S6. Diffusion of the metabolites in aqueous medium: (A) PDO and (B) PPN. Squared deviations of the center-of-mass coordinates from the initial positions are plotted.

Figure S6 shows the squared distances from the initial positions of the two metabolites observed from the unconstrained MD simulations. For each metabolite, the unconstrained molecular trajectories were sampled for 10 ns at 300 K. The simulations were repeated 24 times to provide an ensemble of trajectories for each metabolite. The averaged squared distances are proportional to the diffusion constant (D) in the bulk media by the Einstein relation:[6]

$$6Dt = \langle R(t)R(0)\rangle$$

, where t is the time, $R(t)$ is the center-of-mass coordinates of the metabolite, and $\langle\cdots\rangle$ represents ensemble average over many realizations. The Einstein relation was least-square fitted to the ensemble of MD trajectories to give the diffusion constant of each metabolite. The computed diffusion coefficient of the PDO molecule is $0.92 \pm 0.01 \times 10^{-5}$ cm^2/s and of the PPN molecule is $0.94 \pm 0.01 \times 10^{-5}$ cm^2/s. Here the standard errors of the least-square fittings are understood as the uncertainty of the estimation. The computed diffusion coefficients are in good agreement with the experimental observations, which are 1.0×10^{-5} cm^2/s for the PDO and 1.16×10^{-5} cm^2/s for the PPN.

We computed the diffusion coefficient of each metabolite at the center of the PduA pore. Due to geometrical restraints, the metabolites may diffuse through the pore with different diffusion rates relative to the bulk medium. We analyzed the restrained MD simulations (Figure 4B) to compute the local diffusion coefficient. In the restrained MD simulations, the center-of-mass positions of the two metabolites were harmonically restrained at the center of the PduA pore. Without the restraints, the diffusive behaviors of the metabolites are difficult to observe. This is because the center of the pore is energetically uphill in energy as shown in Figures 2 and 3. According to Hummer,[7] the local diffusion coefficient under restraints is expressed:

$$D_{Local} = \frac{Var(R)}{\tau_R}$$

, where D_{Local} is the local diffusion coefficient at the pore, $Var(R)$ is the variance of the center-of-mass coordinates of the metabolite under the restraints, and τ_R is the correlation time that is defined as below.

$$\tau_R = \frac{\int_0^\infty C_R(t)\, d\mathrm{t}}{Var(R)}$$

, where $C_R(t)$ is the normalized autocorrelation function of the center-of-mass positions. The entire duration (100 ns) of the MD simulations were split into 20 ns overlapping windows. For each window, we computed the autocorrelation time and consequently the local diffusion coefficients. Finally, the computed diffusion coefficients were averaged over the 80 windows. The resultant diffusion coefficients are $0.49 \pm 0.20 \times 10^{-5}$ cm^2/s for the PDO and $0.47 \pm 0.17 \times 10^{-5}$ cm^2/s for the PPN. The standard deviations of the averages are used as the errors of the estimates.

Table S1. Optimized coordinates and atomic partial charges of the metabolites

Molecule	**Atom**	**X***	**Y**	**Z**	**RESP partial charge****
1-2-propanediol	**O**	-1.942	-0.054	-0.002	-0.663329
	C	-0.71	-0.748	-0.227	0.17897
	H	-0.741	-1.766	0.188	0.019313
	H	-0.612	-0.82	-1.315	0.019313
	C	0.465	0.049	0.342	0.449042
	O	0.412	1.378	-0.157	-0.686685
	H	0.365	0.058	1.446	-0.04614
	C	1.816	-0.548	-0.025	-0.381434
	H	1.927	-1.558	0.386	0.099838
	H	1.929	-0.595	-1.114	0.099838
	H	2.622	0.077	0.372	0.099838
	H	-2.144	-0.115	0.946	0.406932

	H	-0.535	1.604	-0.178	0.404505
propionaldehyde	**C**	1.44	-0.506	0	-0.212956
	C	0.549	0.734	0	0.121244
	C	-0.931	0.425	0	0.471457
	O	-1.404	-0.69	0	-0.503403
	H	-1.595	1.321	0	-0.022998
	H	1.246	-1.125	0.881	0.059357
	H	2.497	-0.225	0	0.059357
	H	1.246	-1.125	-0.881	0.059357
	H	0.747	1.377	0.872	-0.015707
	H	0.747	1.377	-0.872	-0.015707

***: Unit of coordinates is angstrom; **: Unit of atomic partial charge is electron**

Table S2. Computed diffusion constants from the MD simulations. The unit of the constants is cm^2/s

Metabolite	**Diffusion constant in aqueous medium**	**Diffusion constant at the PduA pore**
Propanediol (PDO)	$0.92 \pm 0.01 \times 10^{-5}$	$0.49 \pm 0.20 \times 10^{-5}$
Propionaldehyde (PPN)	$0.94 \pm 0.01 \times 10^{-5}$	$0.47 \pm 0.17 \times 10^{-5}$

References

1. Ray, D. S., Notes on Brownian Motion and Related Phenomena. *arxiv.org*, **1999,** arXiv:physics/9903033.
2. Bauer, W. R.; Nadler, W., Molecular Transport Through Channels and Pores: Effects of In-channel Interactions and Blocking. *Proc Natl Acad Sci U S A* **2006,** *103*, 11446-51.
3. Grossfield, A., WHAM: The Weighted Histogram Analysis Method, version 2.0.9. University of Rochester: Rochester, NY, 2002.
4. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992,** *13*, 1011-1021.
5. Efron, B.; Tibshirani, R., Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* **1986,** *1*, 54--77.
6. Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*. Clarendon Press: 1989.

7. Hummer, G., Position-dependent Diffusion Coefficients and Free Energies from Bayesian Analysis of Equilibrium and Replica Molecular Dynamics Simulations. *New Journal of Physics* **2005,** *7*, 34.