Check for updates

Aquaglyceroporin AQP7's affinity for its substrate glycerol. Have we reached convergence in the computed values of glycerol-aquaglyceroporin affinity?†

Michael Falato, Ruth Chan and Liao Y. Chen [ID] *

AQP7 is one of the four human aquaglyceroporins that facilitate glycerol transport across the cell membrane, a biophysical process that is essential in human physiology. Therefore, it is interesting to compute AQP7's affinity for its substrate (glycerol) with reasonable certainty to compare with the experimental data suggesting high affinity in contrast with most computational studies predicting low affinity. In this study aimed at computing the AQP7-glycerol affinity with high confidence, we implemented a direct computation of the affinity from unbiased equilibrium molecular dynamics (MD) simulations of three all-atom systems constituted with 0.16 M, 4.32 M, and 10.23 M atoms, respectively. These three sets of simulations manifested a fundamental physics law that the intrinsic fluctuations of pressure in a system are inversely proportional to the system size (the number of atoms in it). These simulations showed that the computed values of glycerol-AQP7 affinity are dependent upon the system size (the inverse affinity estimations were, respectively, 47.3 mM, 1.6 mM, and 0.92 mM for the three model systems). In this, we obtained a lower bound for the AQP7-glycerol affinity (an upper bound for the dissociation constant). Namely, the AQP7-glycerol affinity is stronger than 1087/M (the dissociation constant is less than 0.92 mM). Additionally, we conducted hyper steered MD (hSMD) simulations to map out the Gibbs free-energy profile. From the free-energy profile, we produced an independent computation of the AQP7-glycerol dissociation constant being approximately 0.18 mM.

Introduction

Aquaglyceroporins (AQGPs) are a subfamily of aquaporin (AQP) proteins[1–3] responsible for facilitated diffusion of glycerol and some other small neutral solutes across the cell membrane along the solute concentration gradient.[4] They also conduct water transport down the osmotic gradient. Among the 13 human AQPs, the AQGP subfamily consists of AQPs 3, 7, 9, and 10. AQGPs are fundamental to many physiological processes. For example, pancreatic AQP7 is involved in insulin secretion; all AQGPs participate in fat metabolism. Therefore, AQGPs are investigated as drug targets for metabolic diseases.[5]

Among the many experimental and theoretical-computational investigations of aquaglyceroporins, one fundamental question remains: does an AQGP have affinity for its substrate glycerol? In functional characterization experiments in 1994,[6] *Escherichia coli* aquaglyceroporin GlpF was shown to facilitate unsaturable uptake of glycerol up to 200 mM

into *Xenopus oocytes*, suggesting that GlpF has very low affinity for its substrate glycerol. In a series of functional experiments from 2008 to 2014,[7–9] human aquaglyceroporins AQP7, AQP9, and AQP10 were shown to conduct saturated transport of glycerol with Michaelis constants around 10 μM, indicating that human AQGPs have high affinities for glycerol. In the crystal structures available to date (GlpF in 2000,[10] *Plasmodium falciparum* PfAQP in 2008,[11] AQP10 in 2018,[12] and AQP7 in 2020 (ref. 13–15)), glycerol molecules were found inside the AQGP channel and near the channel openings on both the intracellular (IC) and the extracellular (EC) sides, showing that all four AQGPs have affinities for glycerol. If we insisted that unsaturated transport precludes high affinity, these experimental data would suggest inconsistency. However, in an *in silico–in vitro* study[16] of glycerol uptake into human erythrocytes through AQP3,[17] it was shown that an AQGP (having high affinity for its substrate glycerol) can conduct glycerol transport that is unsaturated up to 400 mM. The transport pathway for unsaturated transport through a high affinity facilitator protein was shown to involve two glycerol molecules next to each other both bound inside an AQP3 channel (one at the high affinity site and one at a low affinity site) for the transport of one glycerol molecule across the cell membrane.[16] It is the substrate–

Department of Physics, University of Texas at San Antonio, San Antonio, Texas 78249, USA. E-mail: Liao.Y.Chen@gmail.com

† Electronic supplementary information (ESI) available: Two movies and 10 additional figures. See DOI: 10.1039/d1ra07367b

substrate interactions (mostly repulsion due to steric exclusion) inside a single-file channel that make it easy for two glycerol molecules cooperatively to move one substrate molecule across the AQGP channel *via* the high affinity site.

On the theoretical-computational side, the predicted affinities of AQGPs were derived from the computed free-energy profiles – the PMF curves (the potential of mean force as a function of an order parameter, namely, the Gibbs free energy of the system when the chosen degrees of freedom are set to a given set of values). The predictions are dependent upon the methods of computation used in a given study. For example, the estimated values of the glycerol-GlpF affinity range from <1/M (from the PMF curve of ref. 18 and 19) to >10^3/M (from the PMF curve of ref. 20). Currently, the estimations of the glycerol-AQP7 affinity stand at <1/M (from the PMF curves of ref. 13 and 15) in contrast with the experimental data of ref. 7 showing high affinity ∼10^5/M. All these point to the need for further theoretical-computational studies of AQGP-glycerol affinities.

In this research, we aim to reach computational convergence on glycerol-AQP7 affinity. We first carried out direct computations of the glycerol-AQP7 affinity by estimating the probability p_b of glycerol binding inside an AQP7 channel for a given glycerol concentration c_G. The dissociation constant $k_D = c_G(1 - p_b)/p_b$ is the inverse of the glycerol-AQP7 affinity. Running equilibrium molecular dynamics (MD) without any biases or constraints on three systems ranging from approximately 0.2 M atoms to 10 M atoms in sizes, we observed a convergence toward high AQP7-glycerol affinity. We also examined the intrinsic fluctuations of the model systems (Fig. 1). We found that the pressure fluctuations were inversely proportional to the system size as expected based on statistical thermodynamics.[21] In a system (consisting of 0.2 M atoms) typical in the current literature, the root mean squared pressure fluctuations >100 bar in the simulation of a system under a constant pressure of 1.0 bar (see, *e.g.*, NAMD User's Guide https://www.ks.uiuc.edu/Research/namd/2.14/ug/node39.html). In another word, the model system is subject to constant agitations of an artificial sonicator in inverse proportion to the system size. These agitations are expected to loosen the binding between a protein and its substrate and thus to reduce the apparent affinity (*i.e.*, the computed value of the glycerol-AQGP affinity). Our simulations of various system sizes showed that the computed values of glycerol-AQP7 affinity were strongly dependent upon the system size and that convergence of computational studies points to strong affinity between an AQGP and its substrate instead of weak affinity observed in small simulations. Seeking an independent confirmation of strong AQGP-glycerol affinity, we also determined the glycerol-AQP7 affinity from the PMF curve that was computed from a large set of hyper steered MD (hSMD) simulations.

Methods

The parameters, the coordinates, and the scripts for setting up the model systems, running the simulations, and analyzing the data are available at Harvard Dataverse.[22]
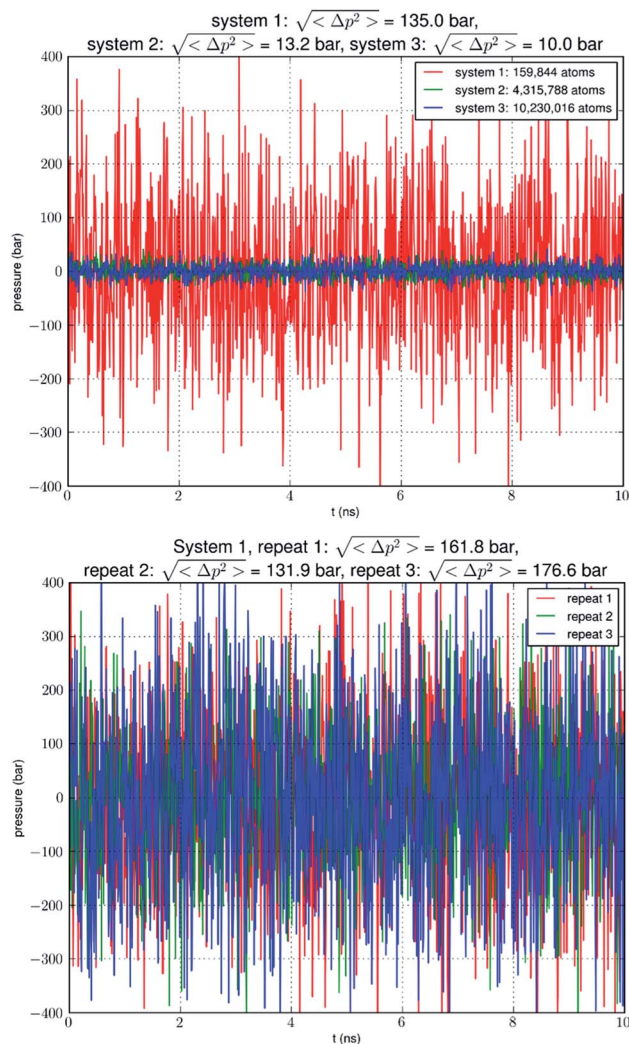

Fig. 1 Pressure fluctuation in an NPT simulation of a small system (159 844 atoms) containing a single AQP7 tetramer, a large system (4 315 788 atoms) with 27 AQP7 tetramers, or a huge system (10 230 016 atoms) with 64 AQP7 tetramers. The pressure fluctuation is approximately proportional to the inverse of the system size in terms of atom numbers. Three repeated simulations of system 1 are also shown.

Model system setup and simulation parameters

Following the well-tested steps in the literature, we employed CHARMM-GUI[23–25] to build an all-atom model of an AQP7 tetramer embedded in a 117 Å × 117 Å patch of membrane (lipid bilayer consisting of 193 phosphatidylethanolamine/POPE, 119 phosphatidylcholine/POPC, and 80 cholesterol/CHL1 molecules). The AQP7 coordinates were taken from ref. 13 (PDB: 6QZI). The positioning of the AQP7 tetramer was determined by matching the hydrophobic side surface with the lipid tails and aligning the channel axes perpendicular to the membrane. The AQP7-membrane complex was sandwiched between two layers of TIP3P waters, each of which was approximately 30 Å thick. The system was then neutralized and salinated with Na$^+$ and Cl$^-$ ions to a salt concentration of 150 mM. Glycerol was added to the system to 50 mM in
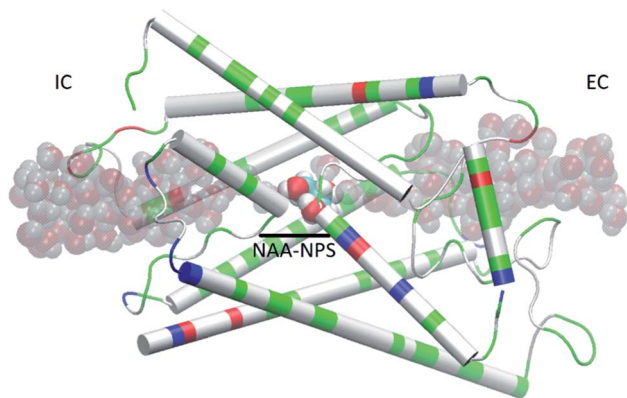
Fig. 2 AQP7 monomer channel with a glycerol molecule (large spheres colored by atoms: C, cyan; O, red; H, white) at the central binding site near the NAA/NPS motifs. The whole monomer protein is shown as cartoons colored by residue types (positively charged, blue; negatively charged, red; hydrophilic, green; hydrophobic, white). The water molecules inside and near the channel are shown in shadowy spheres colored by atoms (O, red; H, white). All molecular graphics in this paper were rendered with VMD.[35]

concentration. The system so constructed consists of a single AQP7 tetramer (four monomer channels) constituted with 159 844 atoms, which is referred to as SysI (shown in ESI, Fig. S1†). We employed NAMD 2.13 and 3.0 (ref. 26 and 27) as the MD engines. We used CHARMM36 parameters[28–30] for inter- and intra-molecular interactions. We followed the literature's standard steps to equilibrate the system.[15,31–33] Then we ran unbiased MD for 2000 ns (namely, 8000 monomer·ns) with constant pressure at 1.0 bar (Nose–Hoover barostat) and constant temperature at 303.15 K (Langevin thermostat). The Langevin damping coefficient was chosen to be 1/ps. The periodic boundary conditions were applied to all three dimensions. The particle mesh Ewald (PME) was used for the long-range electrostatic interactions (grid level: $128 \times 128 \times 128$). The time step was 2.0 fs. The cut-off for long-range interactions was set to 10 Å with a switching distance of 9 Å. The last 500 ns (2000 monomer·ns) of the trajectory was used in the computation of the glycerol-AQP7 affinity. We varied the Nose–Hoover barostat parameters and the cut-off distances to ascertain that the large pressure fluctuations are not an accidental consequence of the aforementioned choice of parameters which are typical in the literature.

We replicated SysI 26 times to obtain 27 copies of SysI. With appropriate translations of these copies, we formed SysII consisting of 27 AQP7 tetramers (illustrated in ESI, Fig. S2†).

Unbiased MD was run for 15 000 monomer·ns for this large SysII with identical parameters used for SysI except that the PME was implemented on a grid of $384 \times 384 \times 384$. The last 5000 monomer·ns were used in the computation of the glycerol-AQP7 affinity. Likewise, we replicated SysI 63 times to form SysIII consisting of 64 AQP7 tetramers (illustrated in ESI, Fig. S3†). We ran unbiased MD on SysIII (with PME grid of $512 \times 512 \times 512$) for 15 000 monomer·ns and used the last 5000 monomer·ns in the computation of the glycerol-AQP7 affinity.

Direct computation of AQP7-glycerol affinity

We used the part of an MD trajectory when the system is fully equilibrated to compute the probability p_b for an AQP7 channel being occupied with a glycerol molecule (being inside the single-file region of the channel, 7.1 Å to the IC/EC side from the NAA/NPS motifs illustrated in Fig. 2). Based on the equilibrium kinetics, $p_b = c_G/(c_G + k_D)$ with c_G being the glycerol concentration, we computed the dissociation constant from the binding probability: $k_D = c_G(1 - p_b)/p_b$.

Computing the Gibbs free-energy profile and the affinity

We conducted 2100 ns hSMD of SysI (illustrated in Fig. 2) to compute the PMF along the glycerol transport path through an AQP7 channel across the membrane. We followed the multi-sectional protocol detailed in ref. 34. We defined the forward direction as along the z-axis pointing from the intracellular side to the extracellular side. We divided the entire glycerol transport path across the membrane from $z = -28$ Å to $z = 22$ Å into 50 evenly divided sections. From the central binding site ($z = -1$ Å, shown in Fig. 2) to the EC side ($z \geq 22$ Å), the center-of-mass z-degree of freedom of glycerol was steered at a speed of 0.25 Å ns^{-1} for 4 ns over one section for a z-displacement of 1.0 Å to sample a forward path over that section. At the end of each section, the z-coordinate of the glycerol center-of-mass was fixed (or, technically, pulled at a speed of 0.0 Å ns^{-1}) while the system was equilibrated for 10 ns. From the end of the 10 ns equilibration, the z-coordinate of the glycerol center-of-mass was pulled for 4 ns for a z-displacement of −1.0 Å to sample a reverse path. From the binding site ($z = -1$ Å) to the IC side ($z \leq -28$ Å), the center-of-mass z-degree of freedom of glycerol was steered for 4 ns for a z-displacement of −1.0 Å to sample a reverse path over one section. At the end of that section, the z-coordinate of the glycerol center-of-mass was fixed while the system was equilibrated for 10 ns. From the end of the 10 ns equilibration, the z-coordinate of the glycerol center-of-mass was pulled for 4 ns for a z-displacement of +1.0 Å to sample a forward path. In this way, section by section, we sampled a set

Table 1 Three repeats of 100 ns equilibrium MD runs of SysI

Repeat	Cutoff	Switching	Langevin piston period	Langevin piston decay	$\sqrt{\langle \Delta p^2 \rangle}$
0	10 Å	9 Å	50 fs	25 fs	135.0 bar
1	10 Å	9 Å	60 fs	30 fs	161.8 bar
2	12 Å	10 Å	50 fs	25 fs	131.9 bar
3	12 Å	10 Å	60 fs	30 fs	176.6 bar

of four forward paths and four reverse paths in each of the 50 sections (28 sections from the central binding site to the IC side and 22 sections from the central binding site to the EC side) along the entire transport path between the IC and the EC sides. The force acting on the glycerol center-of-mass was recorded along the forward and the reverse pulling paths for computing the PMF along the entire transport path from the IC side to the central binding site and then to the EC side. The PMF was computed from the work along the forward paths and the work along the reverse paths (ESI, Fig. S6†) *via* the Brownian-dynamics fluctuation-dissipation theorem.[34]

Following the standard literature (*e.g.*,[36]), one can relate the binding affinity (inverse of the dissociation constant k_{Di}) at the i-th binding site to the PMF difference in 3 dimensions (3D) and the two partial partitions as follows:

$$c_0/k_{Di} = \exp[-\Delta W_i/RT]Z_i/Z_\infty. \tag{1}$$

Here ΔW_i is the PMF at the i-th binding site minus the PMF in the dissociated state when glycerol is far away from the protein. R is the gas constant. T is the absolute temperature. Z_i is the partial partition of glycerol in the i-th bound state which can be computed by sampling the fluctuations in 3 degrees of freedom of the glycerol center of mass and invoking the Gaussian approximation for the fluctuations in the bound state.[37] $Z_\infty = 1/c_0$ is the corresponding partial partition in the dissociated state with $c_0 = 1$ M being the standard concentration.

Results and discussion

Large pressure fluctuations in small simulation systems

Our simulation of a small system SysI showed that pressure fluctuations were very large in simulations of small model systems. To verify that this was not accidental, we repeated the simulation of SysI three times with different parameters. The parameters are tabulated in Table 1. The pressure fluctuations during the last 10 ns of the 100 ns MD runs are shown in Fig. 1. The root mean squared pressure fluctuations are shown in Fig. 1 and Table 1. These results are not accidental in our study but fully in line with the current literature (*e.g.*, NAMD User's Guide https://www.ks.uiuc.edu/Research/namd/2.14/ug/node39.html).

The long process of equilibration in a glycerol-AQP7 system

We conducted an MD run (under constant temperature and constant pressure, NPT) for 2000 ns to fully equilibrate SysI consisting of one AQP7 tetramer (four AQP7 monomer channels) constituted with 159 844 atoms (ESI, Fig. S1†). We computed the root mean squared deviation (RMSD) from the crystal structure for each of the four protein monomers, which are shown in Fig. 3. The RMSD being 2 to 2.5 Å from the crystal structure confirms the quality of the high-resolution structure of ref. 13 representing the AQP7-glycerol chemistry under equilibrium conditions. The small but significant spikes in the RMSD curves corresponds to the events of binding/dissociating of glycerol into/from the AQP7 channel, in line with the concept
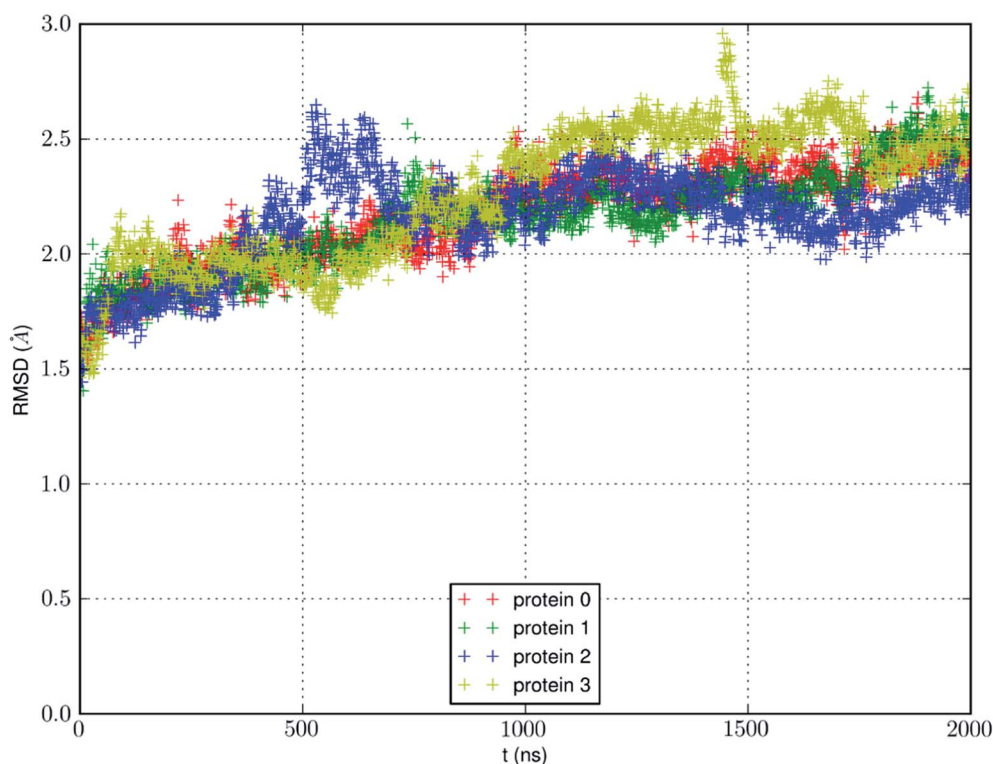

Fig. 3 RMSD from the crystal structure of the protein monomers during the MD run of a system with a single AQP7 tetramer (4 monomer channel proteins) for 2000 ns (*i.e.*, 8000 monomer·ns).

of induced fit in a glycerol-GlpF complex[10] (and glycerol-aquaglyceroporin complexes, in general).

We also learnt from Fig. 3 that equilibrium was not reached until after 1500 ns (*i.e.*, 6000 monomer·ns). Only during the last 500 ns, we observed multiple events of glycerol moving into and out of the AQP7 channel as illustrated in Movie 1.† Therefore, only the last 500 ns (2000 monomer·ns) of the MD trajectory should be used in the statistical analyses of the system. The earlier part of the trajectory represents a transient process toward equilibrium which is inevitably dependent upon the initial conditions of the model system. Multiple monomer·μs simulations are necessary for significant sampling of glycerol-AQP7 kinetics. This indicates that glycerol-AQP7 interactions are strong rather than weak. What interactions are responsible for the strong AQP7-glycerol affinity? First, the hydrogen bonds between glycerol and the channel lining residues of AQP7 and the hydrogen bonds between glycerol and the water molecules inside the channel. When a glycerol resides inside the channel near the NAA-NPS motifs (Fig. 2), it forms 2 hydrogen bonds with the surrounding AQP7 residues and 2 hydrogen bonds with the 2 water molecules (one on each side). More importantly, when a glycerol is away from the protein and fully surrounded by water molecules, it forms 6 hydrogen bonds with the surrounding water molecules, but it interrupts 10 water–water hydrogen bonds because it displaces 4 water molecules. Second,

the van der Waals (vdW) energy between glycerol and AQP7 is estimated to be −14.8 kcal mol⁻¹ when glycerol resides inside the protein near the NAA-NPS motifs where there is sufficient room for glycerol. When glycerol is away from the protein, the vdW energy between glycerol and water is estimated to be −4.9 kcal mol⁻¹. All these factors combine to give rise to the strong AQP7-glycerol affinity.

Small simulations suggest low glycerol-AQP7 affinity

Analyzing the MD trajectory of SysI, a system consisting of one AQP7 tetramer in the presence of 50 mM glycerol, we counted one or more glycerol molecules residing inside a monomer channel as glycerol being within 7.1 Å from the NAA/NPS motifs located in the central part of the AQP7 channel. The probability of a channel being occupied by one or more glycerol molecules is shown in Fig. 4 along with the probability of a channel being occupied by two glycerol molecules. From the last 500 ns (2000 monomer·ns), we computed the probability of an AQP7 channel being occupied by glycerol $p_b^I = 0.514 \pm 0.058$ leading to a computed value of the glycerol-AQP7 dissociation constant $k_D^I = 47.3$ mM. The computed affinity is not high, far from the experimentally measured value of 0.01 mM.[7]

Does the large discrepancy from the *in vitro* data mean that *in silico* studies cannot be quantitatively accurate at all? Where does this large discrepancy come from? Our model system
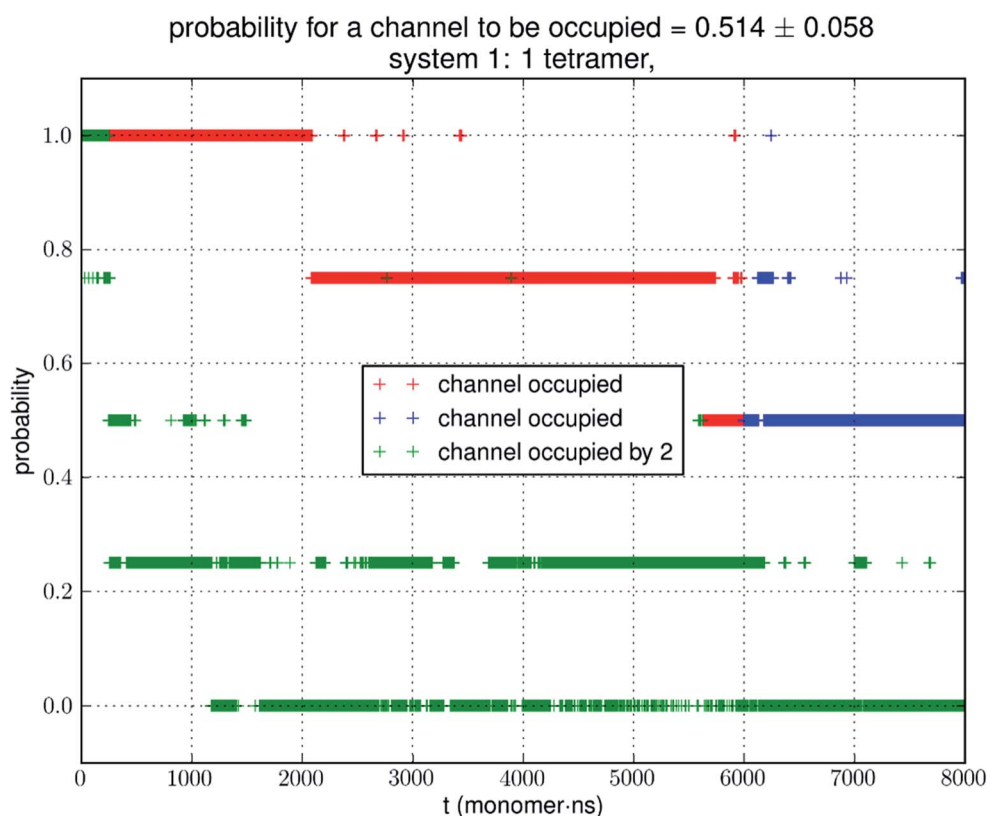

Fig. 4 Glycerol binding characteristics of one tetramer in a typical simulation. The last 500 ns of the trajectory (*i.e.*, 2000 monomer·ns of dynamics, colored in blue) was used in the statistical calculation of the probability. During the first 1500 ns shown in red, the system has not reached equilibrium. A channel is considered occupied when one or more glycerol molecules are within 7.1 Å from the NAA/NPS motifs.

(SysI) is typical of the current literature.[15] We used the standard CHARMM force field parameters. We did not employ any biases in the MD simulation that can generate artifacts. However, during the NPT run for a constant pressure of 1.0 bar, the model system was actually subjected to the mechanic pressure that fluctuated between ± 400 bar (Fig. 1). This pressure fluctuation is inevitable in any simulation because it is intrinsic to any system that is smaller than the thermodynamic limit. The mean square fluctuation of pressure is inversely proportional to the system volume (thus the number of atoms constituting the model system).[21] In light of all this, it is only logical to build larger systems to ascertain whether or not the pressure fluctuations caused the glycerol-AQP7 affinity to appear weak.

Larger simulations yield greater estimates of the glycerol-AQP7 affinity

In Fig. 5, we show the results of 15 000 monomer·ns simulations of two larger systems, SysII consisting of 4.3 M atoms and SysIII consisting of 10.2 M atoms. The pressure fluctuations of these two systems (shown in Fig. 1) were significantly smaller than SysI. The mean square fluctuations were approximately in inverse proportion to the system size (the number of atoms) as expected from statistical thermodynamics.[21] Using the same criterium as for SysI, when one or more glycerol molecules are within 7.1 Å of the NAA/NPS motifs of an AQP7 monomer, that AQP7 channel is counted as being occupied. For a given time interval, SysII and SysIII have many more events of glycerol binding to and dissociating from an AQP7 channel than SysI. Naturally, with larger simulations, we have better statistics in addition to the fact that we have much smaller artifactitious fluctuations in pressure. Taking the last 5000 monomer·ns of the MD trajectories into the statistical calculations, we obtained the probability for an AQP7 channel being occupied by glycerol, $p_b^{II} = 0.969 \pm 0.004$ for SysII and $p_b^{III} = 0.982 \pm 0.003$ for SysIII. Correspondingly, the computed values of the glycerol-AQP7 dissociation constant were $k_D^{II} = 1.6$ mM for SysII and $k_D^{III} = 0.92$ mM for SysIII. Considering the computed value for SysI, $k_D^I = 47.3$ mM, we observed the convergence toward higher affinities (lower k_D values) in larger model systems. There is a strong correlation between the computed k_D values and the artifactitious pressure fluctuations that are inevitable in any computational studies. Ideally, one can build a large enough system whose pressure fluctuation is much less than 1.0 bar for NPT runs under a constant pressure of 1.0 bar, which is still infeasible with today's computing power. However, our study of SysI, SysII, and SysIII together showed that the glycerol-AQP7 affinity is indeed high as one would expect for a facilitator protein with its substrate. It is emphasized here that the afore-presented computations are directly from unbiased equilibrium MD simulations. As long as the parameters are accurate for the intra- and inter-molecular interactions, the conclusion of high glycerol-AQP7 affinity should be valid, free from artifacts that may be present in biased MD simulations.
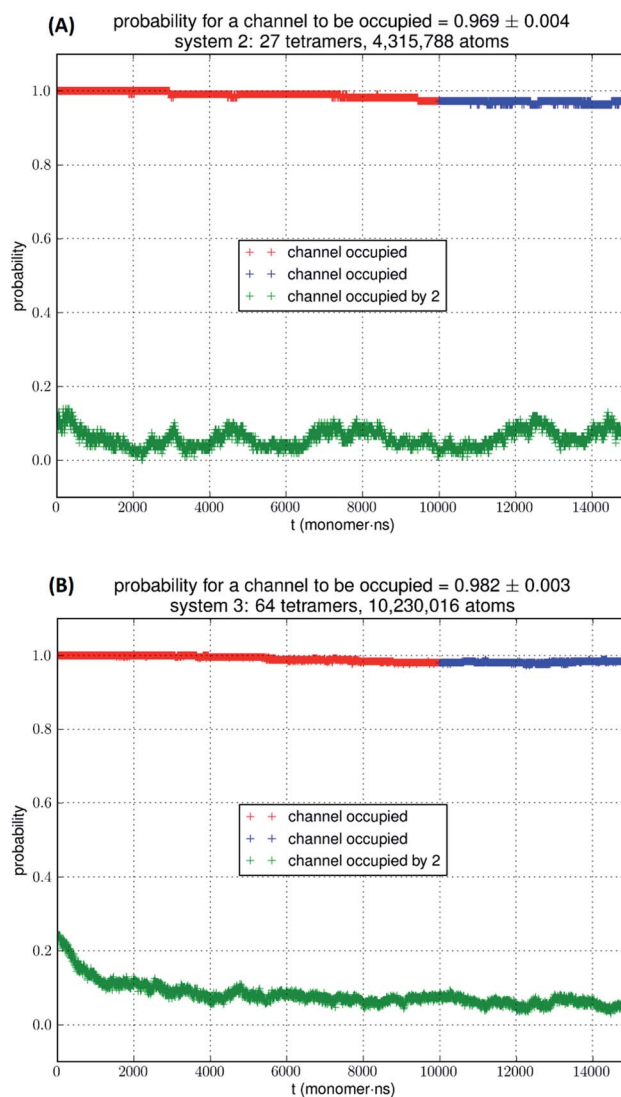

Fig. 5 (A) SysII, 27 tetramers in a large simulation. (B) SysIII, 64 tetramers in a huge simulation. The last 5000 monomer·ns of the trajectory (colored in blue) was used in the statistical calculation of the probability. More details are shown ESI, Fig. S4 and S5.†

Affinity from the Gibbs free-energy profile

Fig. 6 shows the PMF throughout the AQP7 channel as a function of the z-coordinate of the glycerol's center of mass. The PMF was computed from hSMD sampling of glycerol transport through AQP7 as illustrated in Movie 2.† It represents the Gibbs free energy of the system when a glycerol molecule is located at a given location. The reference level of the PMF was chosen at the bulk level on either the EC or the IC side. The two bulk levels must be equal for neutral solute transport across the cell membrane which is not an actively driven process but a facilitated passive process of diffusion down the concentration gradient. The PMF curve leveling off to zero on both the EC side and the IC side in Fig. 6 indicates the accuracy of our computation. Inside the protein channel, the PMF presents a deep well ($\Delta W_0 = -9.2$ kcal mol^{-1}) near the NAA/NPS motifs (around $z \sim -1$), which is a binding site for glycerol (site 0). On the EC side,
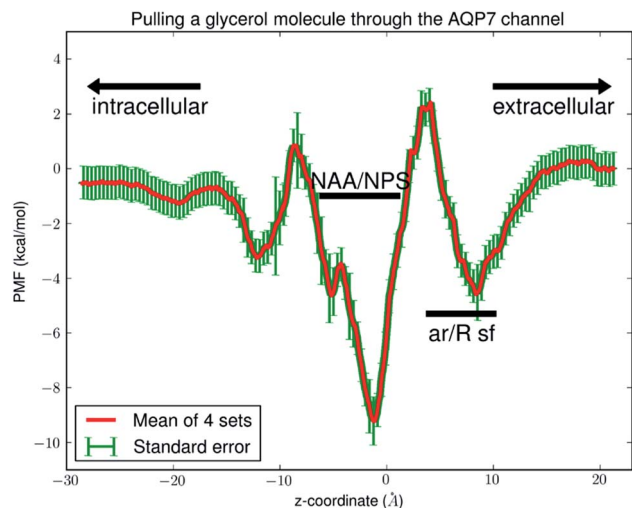
Fig. 6 PMF of glycerol throughout the AQP7 channel. The coordinates are set so that the center of the membrane is located at $z \sim 0$ Å. In the single-file region (-11 Å $< z < 9$ Å), the PMF is one dimensional. In the IC ($z < -11$ Å) or EC ($z > 9$ Å) side of the channel, the PMF is three dimensional. The three PMF wells (binding sites) are located at: site 0, $z \sim -1$ Å; site 1, $z \sim 9$ Å; site 2, $z \sim -11$ Å.

near the aromatic/Arginine (ar/R) selectivity filter (sf), there is another binding site (site 1) where the PMF has a local minimum ($\Delta W_1 = -4.7$ kcal mol^{-1}). The third binding site (site 2) is located on the IC side of the NAA/NPS where the PMF is $\Delta W_2 = -3.3$ kcal mol^{-1}. The PMF well depth is the main factor to determine the affinity (the inverse dissociation constant) at a given binding site. $1/k_D = f_0 \exp[-\Delta W_0/RT]$ for the central binding site. The other factors involved in the determination of the affinities are the fluctuations (shown in ESI, Fig. S7–S10†) which were computed straightforwardly from the equilibrium MD runs with the Gaussian approximation. Combining the fluctuations and the PMF well depth, we obtained the dissociation constants as follows: $k_D = 0.18$ mM for the central binding site. This independent computation of the AQP7-glycerol affinity from hSMD simulations supports our direct computation from equilibrium MD simulations ($k_D < 0.92$ mM).

It is interesting to note that AQP7 and GlpF are similar in channel pore radius:[13] the widest part of the channel is around the NAA/NPS motifs of AQP7 (Fig. 2) and the NPA motifs of GlpF,[10] respectively. The narrowest part is near the ar/R sf. The PMF curve shown in Fig. 6 clearly reflects these characteristics in similarity to the PMF of GlpF.[20] At site 0, near the NAA/NPS motifs, there is sufficient room to accommodate a glycerol there and thus no conformational frustrations exist for the AQP7 residues or the glycerol. The vdW interactions between them are all attractive. At site 1, near the ar/R sf, both the glycerol and the pore residues are frustrated in their conformations for the induced fit.[10] Likewise but in a lesser degree, there are conformational frustrations when glycerol passes through the IC side of the channel. All these point to the importance of vdW interactions in the glycerol-aquaglyceroporin affinity.[20,38]

Conclusions

This study illustrates a fundamental issue in computational chemistry that begs for reexamination: are the computed values of binding affinities or some other characteristics sensitive to the large pressure fluctuations of small model systems? Based on the unbiased MD simulations of a typically sized system and two very large systems, we observed that larger pressure fluctuations in smaller systems cause the glycerol-aquaglyceroporin affinity to appear lower. Beyond the consequence of the artifactitious pressure fluctuations, the computed values of the glycerol-AQP7 dissociation constant indicate high affinity of an aquaglyceroporin for its substrate, which is in agreement with the *in vitro* data on AQP7.

Data availability

The Dataset (parameters, coordinates, scripts, *etc.*) to replicate this study is available at Harvard Dataverse.[22]

Author contributions

MF and LYC did the computational work; LYC conceptualized the research and wrote the paper; all participated in analyzing the data and editing the manuscript.

Grant support

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 P. Agre, L. S. King, M. Yasui, W. B. Guggino, O. P. Ottersen, Y. Fujiyoshi, A. Engel and S. Nielsen, *J. Physiol.*, 2002, **542**, 3–16.

2 G. Benga, *Mol. Aspects Med.*, 2012, **33**, 514–517.

3 P. Agre, M. Bonhivers and M. J. Borgnia, *J. Biol. Chem.*, 1998, **273**, 14659–14662.

4 A. Engel and H. Stahlberg, in *International Review of Cytology*, ed. W. D. S. Thomas Zeuthen, Academic Press, Cambridge, MA, 2002, vol. 215, pp. 75–104.

5 G. Calamita, J. Perret and C. Delporte, *Front. Physiol.*, 2018, **9**, 851.

6 C. Maurel, J. Reizer, J. I. Schroeder, M. J. Chrispeels and M. H. Saier, *J. Biol. Chem.*, 1994, **269**, 11869–11872.

7 T. Katano, Y. Ito, K. Ohta, T. Yasujima, K. Inoue and H. Yuasa, *Drug Metab. Pharmacokinet.*, 2014, **29**, 244–248.

8 M. Ishii, K. Ohta, T. Katano, K. Urano, J. Watanabe, A. Miyamoto, K. Inoue and H. Yuasa, *Cell. Physiol. Biochem.*, 2011, **27**, 749–756.

9 Y. Ohgusu, K.-y. Ohta, M. Ishii, T. Katano, K. Urano, J. Watanabe, K. Inoue and H. Yuasa, *Drug Metab. Pharmacokinet.*, 2008, **23**, 279–284.

10 D. Fu, A. Libson, L. J. W. Miercke, C. Weitzman, P. Nollert, J. Krucinski and R. M. Stroud, *Science*, 2000, **290**, 481–486.

11 Z. E. Newby, J. O'Connell 3rd, Y. Robles-Colmenares, S. Khademi, L. J. Miercke and R. M. Stroud, *Nat. Struct. Mol. Biol.*, 2008, **15**, 619–625.

12 K. Gotfryd, A. F. Mósca, J. W. Missel, S. F. Truelsen, K. Wang, M. Spulber, S. Krabbe, C. Hélix-Nielsen, U. Laforenza, G. Soveral, P. A. Pedersen and P. Gourdon, *Nat. Commun.*, 2018, **9**, 4749.

13 S. W. de Maré, R. Venskutonytė, S. Eltschkner, B. L. de Groot and K. Lindkvist-Petersson, *Structure*, 2020, **28**, 215–222.

14 L. Zhang, D. Yao, Y. Xia, F. Zhou, Q. Zhang, Q. Wang, A. Qin, J. Zhao, D. Li, Y. Li, L. Zhou and Y. Cao, *Sci. Bull.*, 2021, **66**, 1550–1558.

15 F. J. Moss, P. Mahinthichaichan, D. T. Lodowski, T. Kowatz, E. Tajkhorshid, A. Engel, W. F. Boron and A. Vahedi-Faridi, *Front. Physiol.*, 2020, **11**, 728.

16 R. A. Rodriguez, R. Chan, H. Liang and L. Y. Chen, *RSC Adv.*, 2020, **10**, 34203–34214.

17 N. Roudier, J.-M. Verbavatz, C. Maurel, P. Ripoche and F. Tacnet, *J. Biol. Chem.*, 1998, **273**, 8407–8412.

18 M. Ø. Jensen, S. Park, E. Tajkhorshid and K. Schulten, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 6731–6736.

19 J. S. Hub and B. L. de Groot, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 1198–1203.

20 L. Y. Chen, *Biochim. Biophys. Acta, Biomembr.*, 2013, **1828**, 1786–1793.

21 L. D. Landau and E. M. Lifshitz, *Statistical Physics, Part 1*, Pergamon Press, Tarrytown, 1985.

22 L. Y. Chen, *Harvard Dataverse*, 2021, DOI: DOI: 10.7910/DVN/RCZG6V.

23 S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.*, 2008, **29**, 1859–1865.

24 J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, A. D. MacKerell, J. B. Klauda and W. Im, *J. Chem. Theory Comput.*, 2016, **12**, 405–413.

25 J. Lee, D. S. Patel, J. Ståhle, S.-J. Park, N. R. Kern, S. Kim, J. Lee, X. Cheng, M. A. Valvano, O. Holst, Y. A. Knirel, Y. Qi, S. Jo, J. B. Klauda, G. Widmalm and W. Im, *J. Chem. Theory Comput.*, 2019, **15**, 775–786.

26 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

27 J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot and E. Tajkhorshid, *J. Chem. Phys.*, 2020, **153**, 044130.

28 R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell, *J. Chem. Theory Comput.*, 2012, **8**, 3257–3273.

29 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, *J. Comput. Chem.*, 2010, **31**, 671–690.

30 J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell and R. W. Pastor, *J. Phys. Chem. B*, 2010, **114**, 7830–7843.

31 S. Padhi and U. D. Priyakumar, *Vitam. Horm.*, 2020, **112**, 47–70.

32 L. S. M. Neumann, A. H. S. Dias and M. S. Skaf, *J. Phys. Chem. B*, 2020, **124**, 5825–5836.

33 J. A. Freites, K. L. Németh-Cahalan, J. E. Hall and D. J. Tobias, *Biochim. Biophys. Acta, Biomembr.*, 2019, **1861**, 988–996.

34 L. Y. Chen, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6176–6183.

35 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

36 H.-J. Woo and B. Roux, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6825–6830.

37 R. A. Rodriguez, L. Yu and L. Y. Chen, *J. Chem. Theory Comput.*, 2015, **11**, 4427–4438.

38 R. A. Rodriguez, H. Liang, L. Y. Chen, G. Plascencia-Villa and G. Perry, *Biochim. Biophys. Acta, Biomembr.*, 2019, **1861**, 768–775.