Supporting Information: Chemically Selective Transport in a Cross-linked H_{II} Phase Lyotropic Liquid Crystal Membrane

Benjamin J. Coscia Michael R. Shirts

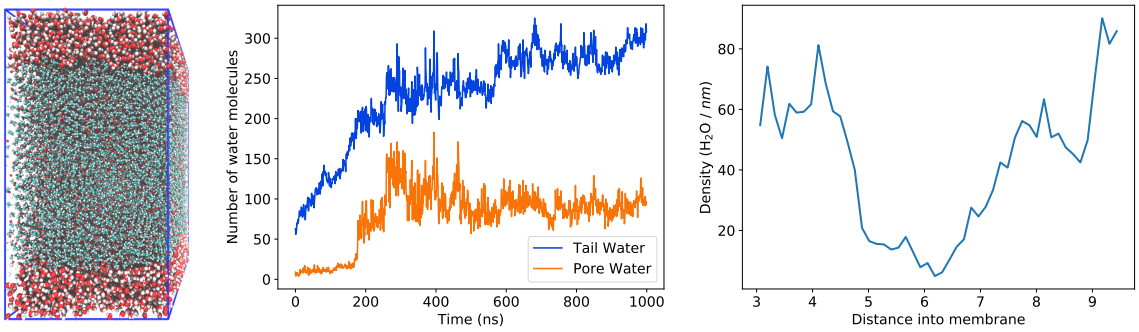
May 8, 2019

S1 Setup and analysis scripts

All python and bash scripts used to set up systems and conduct post-simulation trajectory analysis are available online at `https://github.com/shirtsgroup/LLC_Membranes`. Documentation for the `LLC_Membranes` repository is available at `https://llc-membranes.readthedocs.io/en/latest/`. Table S1 provides more detail about specific scripts used for each type of analysis performed in the main text.

S2 Water content equilibration

We initially attempted to equilibrate our system with water by allowing water molecules to naturally penetrate the membrane from a water bath separating periodic images of the system in the z-direction (see Figure S1a). We allowed a dry, previously equilibrated system to further equilibrate in coexistence with a 3 nm-thick (in the z-direction) layer of water. Water readily enters the tail region where the density of monomers is low. About 3 times more water molecules occupy the tail region after 1000 ns of equilibration (see Figure S1b). Although the water level in the pore appears to plateau in this system, it is clear that equilibration of this system is kinetically limited since water does not fill the pores uniformly (see Figure S1c). The density of water along the pore axis, averaged over the last 50 nanoseconds of simulation, is close to zero at the membrane center. Therefore, we required a different equilibration technique in order to overcome the kinetic limitation.



(a) (b) (c)

Figure S1: (a) Using an equilibrated dry configuration, we inserted a 3 nm thick layer of water between periodic copies of the system in the z-direction. (b) Water slowly enters the membrane. Most water enters the tail region where the density of monomers is lowest. Water entering the pore plateaus after 500 ns. (c) Although the water content of the pore appears equilibrated in (b), the density of water throughout the pores is not uniform, with almost no water close to the pore center. Note that the density is not shown below 3 nm or above 9.5 nm, because the the system enters the water layer at those points.

Script Name	Section	Description
/setup/param.sh	2.1	Parameterize liquid crystal monomers and solutes with GAFF
/setup/solvation_equilibration.py	2.2	Add water to the pores and tails in order to achieve a specific total water content and ratio of water molecules in each region, then equilibrate the solvated system.
/setup/input.py	2.2	Create GROMACS topology and .mdp files
/setup/xlink.py	2.2	Iteratively cross-link a configuration
/setup/place_solutes_pores.py	2.2	Place a desired number of solutes in the pore center, equally space in the z-direction.
/analysis/msd.py	2.3	Calculate the mean squared displacement of residues
/analysis/radius.py	2.4	Calculate the maximum end-to-end distance of a solute over a trajectory
/timeseries/forecast_ctrw.py	2.6	Construct dwell time and hop length distributions
/analysis/rdf.py	2.7, 2.11	Calculate the cylindrical radial distribution function of a solute or membrane component with respect to the pore centers.
/analysis/hbonds.py	2.8	Identify hydrogen bonds based on geometric criteria
/analysis/coordination_number.py	2.9	Calculate number of molecules or atoms within a cutoff distance of another type of molecule or atom.
/analysis/lifetime.py	2.10	Calculate association lifetimes
/analysis/ztrace.py	3.2.2	Plot the center of mass z-coordinate of an atom or molecule versus time which is colored according to its distance from the pore center.

Table S1: The first column provides the names of the python scripts available in the LLC_Membranes GitHub repository that were used for system setup and post-simulation trajectory analysis. Paths preceding script names are relative to the LLC_Membranes/LLC_Membranes directory. The second columns lists the section in the main text where the output or usage of the script is first described. The third column gives a brief description of the purpose of each script.

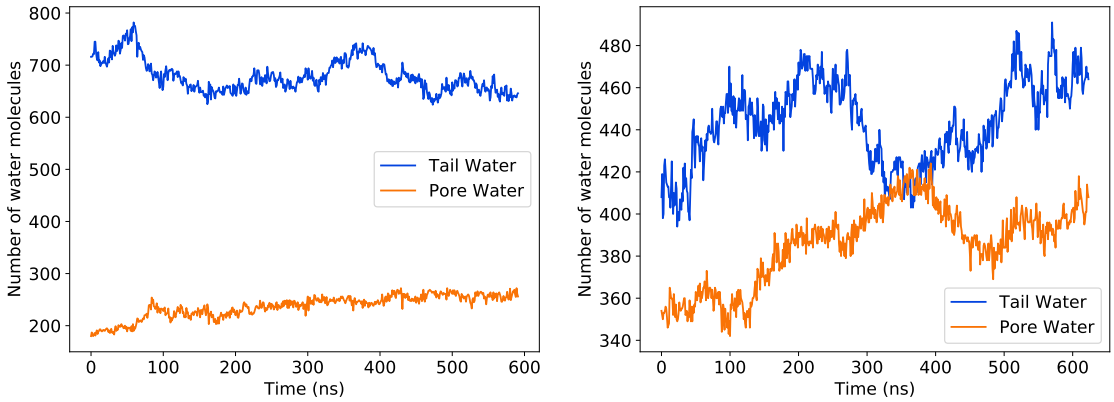
We equilibrated 4 systems where we initially placed water in the pores and in the tails in addition to a water reservoir between periodic images. We filled the pores with water by running `gmx solvate` on our initial configuration and removing any water molecules placed in the tail region. The GROMACS command `gmx solvate` places water molecules in proximity to other atoms based on their van der Waals radii and therefore does a decent job of preventing equilibration issues. This also means that the initial pore radius dictates the water content of the pore. We can however, put arbitrary amounts of water into the tail region. We chose to test systems with initial pore radii of 5, 6, 7 and 8 Å with tail and pore water compositions given in Table S2.

Systems appear to be most stable when there is more water in the pores than tails. In systems started with more water in the tails (Figures S2a and S2b), the pore water content tends to increase over time, while that of the tails decreases or stays stable. Filling the pores with water is likely a very long process since it requires monomers to make space for water molecules. Systems started with a higher pore water content (Figures S2c and S2d) tend to plateau relatively quickly, with about one third of the water staying in the tails. We used this ratio in order to construct the initial configurations used for the studies in the main text. Clearly, a more complicated methodology is needed in order to predict the equilibrium water content of a given LLC membrane. However, getting the value exactly right is not required for our study. Instead, we

Pore Radius (Å)	wt % water tails	wt % water pores
5	5.67	1.09
6	2.88	2.38
7	1.91	4.12
8	2.78	6.00

Table S2: We chose a diverse combination of initial pore and tail water contents in order to study its effect on equilibrium water content.

can observe mechanisms as a function of water content in each region.



(a) r = 5 Å

(b) r = 6 Å

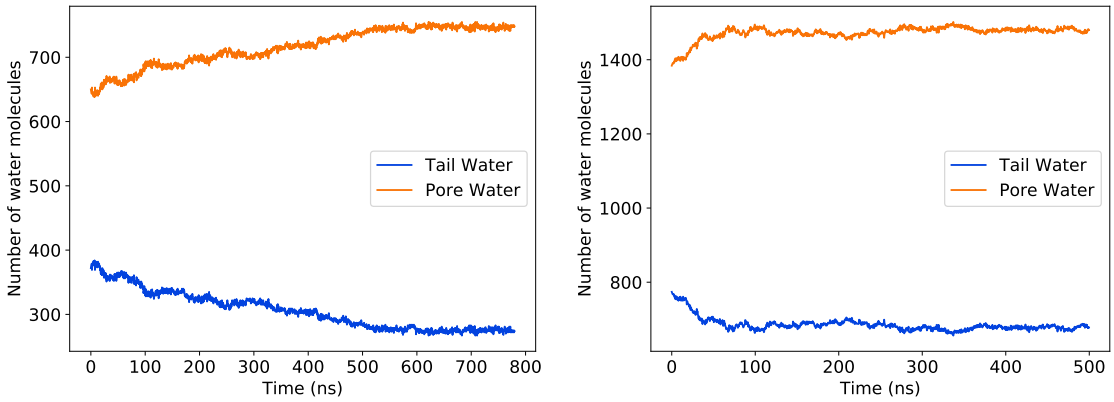
(c) r = 7 Å

(d) r = 8 Å

Figure S2: There is likely more water in the pore region than in the tail region. When we create configurations with more water in the tails, equilibration is slow. When configurations start with more water in the pores, the water content in each region equilibrates quickly.

Since the equilibrium water content is unclear based on the previous simulations, we elected to choose and study systems with two different water contents. We removed the water reservoir and allowed the pore and tail water contents to equilibrate with 5 and 10 wt % total water. We placed one third of the total water needed in the tails, based on the intuition gained in Figure S2. We considered the water content equilibrated once the water contents plateaued. The 5 wt% system did not plateau until $\sim$ 600 ns (Figure S3a) while the 10 wt % water system equilibrated within the first 100 ns of simulation (Figure S3b). The pores contain 72 % and 69 % of the total water in the 5 and 10 wt % systems respectively.

We cross-linked the equilibrated solvated systems, then allowed them to equilibrate further for 100 ns.

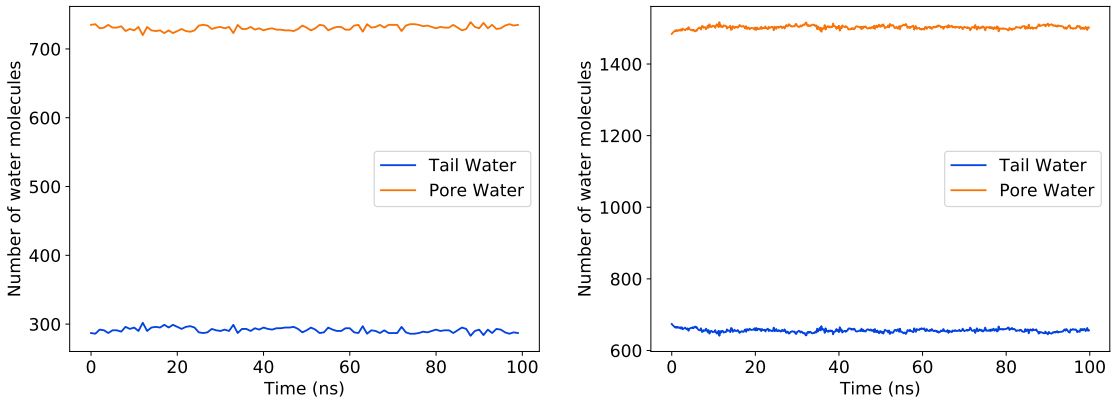


(a) 5 wt %

(b) 10 wt %

Figure S3: We created solvated systems with one third of the total water initially placed in the tail region. (a) With 5 wt % total water, the water content equilibrates after 600 ns, with $\sim$ 72 % of the total water in the pores. (b) With 10 wt % total water, the water content equilibrates after 100 ns, with $\sim$ 69 % of the total water in the pores.

The water contents in each region does not change significantly in either case.



(a) 5 wt%

(b) 10 wt%

Figure S4: The water content in the tail and pore region is not affected by cross-linking

S3 Solute Interaction

We chose to model 6 solutes in each pore because there was a low degree of interaction between solutes which gave us a sufficient number of independent trajectories to observe and analyze. There are a negligible number of occurrences where the center of a given solute came within 3.5 Å of another same-solute center of mass.

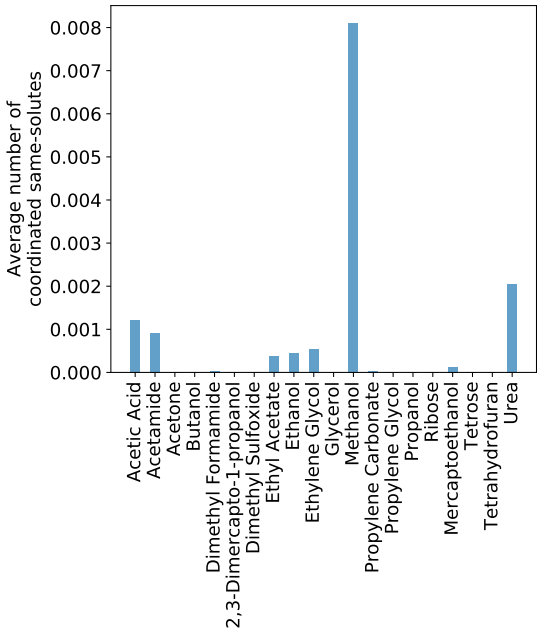
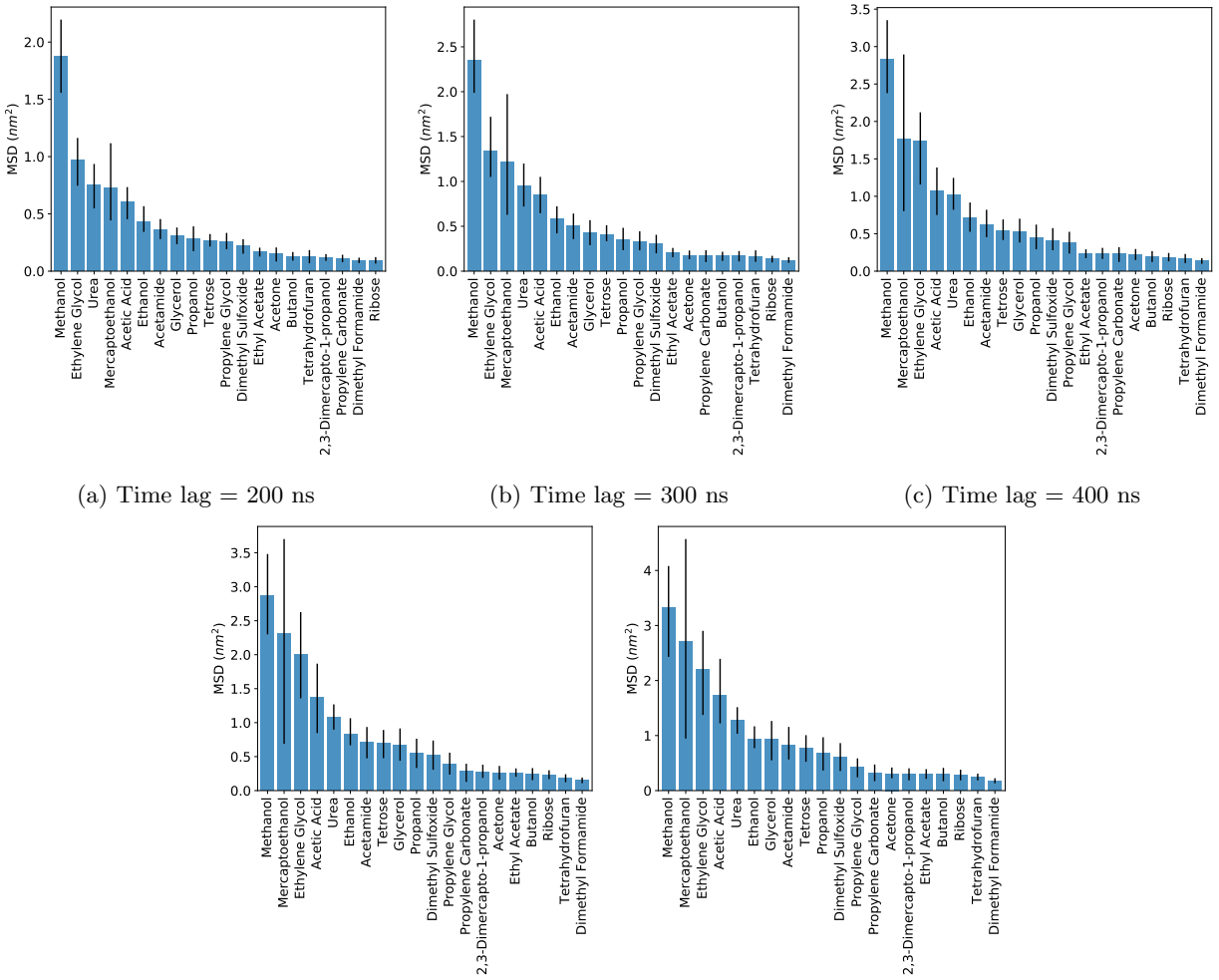


Figure S5: Interactions between same-solutes are negligible when systems are built with 6 solutes per pore. Even methanol, which interacts with other same-solute molecules the most, does so less than 0.05 % of the time.

S4 Solute MSD Ranking Time Lag Sensitivity

We chose to report the MSD of our solutes after a 400 ns time lag. This choice does not have a significant influence on the ranking of the solute MSDs. Although there is minor re-ranking among solutes with similar MSDs, it does not affect any of the conclusions drawn in the main text.



(a) Time lag = 200 ns

(b) Time lag = 300 ns

(c) Time lag = 400 ns

(d) Time lag = 500 ns

(e) Time lag = 600 ns

Figure S6: The choice of time lag used to report solute MSDs does not influence the trends reported in the main text. We observe minor re-ranking of solutes with similar MSDs and note that longer time lags generally lead to large error bars on the MSDs, as expected.

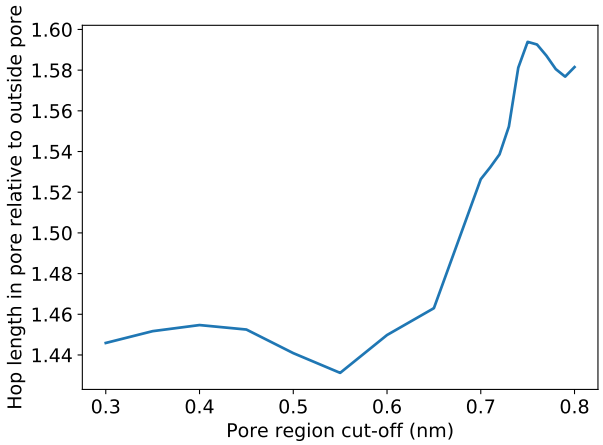


Figure S7: Defining the cut-off 0.75 nm from the pore center maximizes the difference in hop lengths between solutes in and out of the pore. We varied the location of the cut-off defining the edge of the pore region in increments of 0.05 from 0.3 to 0.7. We varied the cut-off in increments of 0.01 from 0.7 to 0.8 in order to increase our resolution about the maximum.

S5 Pore Region Definition

We performed a sensitivity analysis in order to determine the radial cut-off that maximizes the difference in hop lengths in and out of the pore region. We varied the location of the cut-off defining the edge of the pore region in increments of 0.05 from 0.3 to 0.7. We varied the cut-off in increments of 0.01 from 0.7 to 0.8 in order to increase our resolution about the maximum (see Figure S7). Defining the cut-off 0.75 nm from the pore center maximizes the difference in hop lengths between solutes in and out of the pore.

S6 Hydrogen Bond Detection Sensitivity

In Section 2.8 of the main text, we established geometric criteria for the identification of hydrogen bond interactions. Namely, if the distance between the donor, D, and acceptor, A, atoms is less than 3.5 Å and the angle formed by D–H...A is less than 30°, then we consider the interaction to be a consequence of hydrogen bonding.

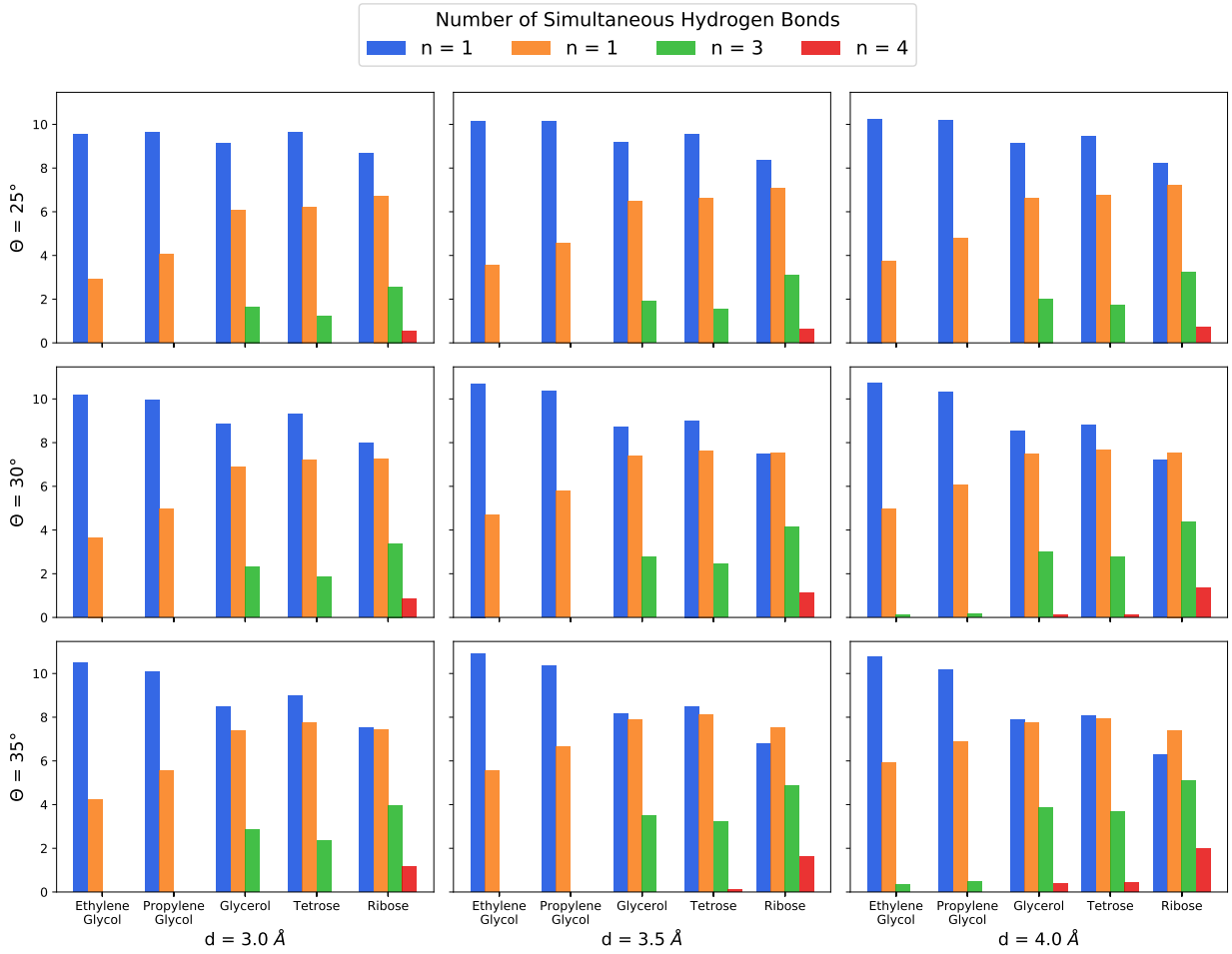


Figure S8: We tested the sensitivity of our hydrogen bond detection to the chosen geometric criteria for hydrogen bonding. We varied the distance cut-off between 3.0 and 4.0 Å, and varied the angle cut-off between 25 and 35 °.

We tested the sensitivity of this criteria to ensure that our conclusions were not a strong function of the chosen criteria. In Figure S8 we plotted the results of the same calculation performed to create Figure 11cc of the main text. We varied the distance cut-off between 3.0 and 4.0 Å, and varied the angle cut-off between 25 and 35 °.

As the distance and angle cut-offs are increased, there is a slight increase in the total number of hydrogen bonds, which is primarily manifested by an increased number of multiple-hydrogen-bond interactions. Using our most lenient set of criteria (d = 4.0 Å and θ = 35 °), glycerol, tetrose and ribose all show instances where they simultaneously hydrogen bond with 4 head group atoms. In the strictest case (d = 3.0 Å and θ = 25 °), ribose is the only solute with a detected instance of a quadruple hydrogen bond.

We can draw the same conclusions from any of the plots in Figure S8:

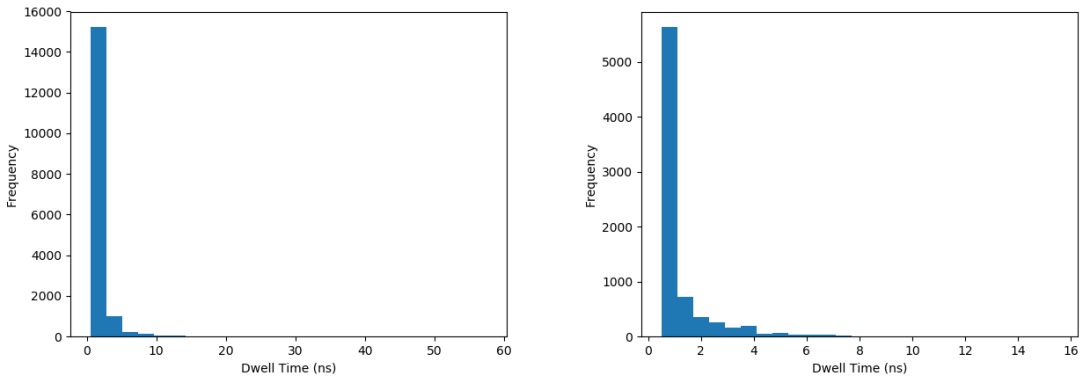
1. An increased number of hydroxyl groups results in an increased number of hydrogen bond interactions.

2. The number of multiple-hydrogen-bond interactions increases.

3. On average, propylene glcyol participates in more hydrogen bond interactions than ethylene glycol.

S7 Lifetime Distributions

The distribution of hydrogen bond lifetimes and association lifetimes for all solutes appears to be power law or exponentially distributed. Example distributions generated from ethylene glycol are shown in Figure S9.



(a) (b)

Figure S9: The distribution of hydrogen bond lifetimes and association lifetimes of ethylene glycol both appear to be power law or exponentially distributed. We do not attempt to distinguish between the type of distribution and instead report the 95% confidence interval of association lifetimes.

S8 Pore Splines

We captured the apparent z-dependence of the pore center locations by constructing splines running through each pore (See Figure S10). Each spline consists of 10 points, equally spaced in the z-direction, whose (x, y) coordinates are defined based on the center of mass of all head groups closest, in z, to the given point.

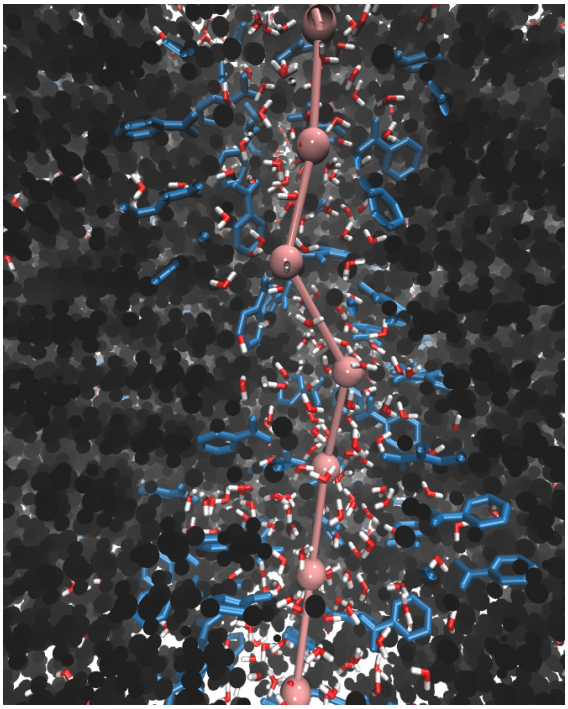


Figure S10: We traced the center of each pore as a function of z using a spline (pink line). We constructed the spline in each pore using 10 points (pink spheres) whose positions we defined based on the center of mass of the head groups in closest proximity to the spline point in the z-direction.

We calculated the tortuosity, τ, of the pores by calculating the ratio $\frac{L}{Z}$ where L is the length of the spline and Z is the length of the unit cell in the z-direction. The average tortuosity of each pore is 1.03 ± 0.01 and 1.07 ± 0.02 in the 5 and 10 wt% water systems respectively.