

**Understanding Structure and Selectivity in Nanostructured
Lyotropic Liquid Crystal Membranes using Molecular
Simulations**

by

Benjamin Joseph Coscia

B.A., University of Connecticut, 2014

M.S., University of Colorado Boulder, 2017

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Department of Chemical and Biological Engineering

2020

This thesis entitled:
Understanding Structure and Selectivity in Nanostructured Lyotropic Liquid Crystal Membranes
using Molecular Simulations
written by Benjamin Joseph Coscia
has been approved for the Department of Chemical and Biological Engineering

Prof. Michael Shirts

Prof. Richard Noble

Prof. Douglas Gin

Prof. Matthew Glaser

Prof. Robert Davis

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Coscia, Benjamin Joseph (Ph.D., Chemical Engineering)

Understanding Structure and Selectivity in Nanostructured Lyotropic Liquid Crystal Membranes
using Molecular Simulations

Thesis directed by Prof. Michael Shirts

The ability to perform highly selective aqueous separations at low cost is beneficial to a variety of chemical industries

Periodic, nanostructured porous polymer membranes made from the cross-linked inverted hexagonal phase of self-assembled lyotropic liquid crystals (LLCs) are a promising class of materials for selective separations.

The uniform size and complex chemical topology of the pores formed by self-assembled amphiphilic molecules such as liquid crystals make them promising candidates for selective separations.

In this work, we investigate an experimentally characterized LLC polymer membrane using atomistic molecular modeling. We

First, we gained a deep understanding of nanoscopic pore structure in order to make sure we got the chemical environment in the pores correct.

In particular, we compare simulated X-ray diffraction (XRD) patterns with experimental XRD data to quantify and understand the differences between simulation and experiment.

Finally, we explore the composition and structure of the nanopores and reveal that there exists a composition gradient rather than an abrupt partition between the hydrophilic and hydrophobic regions. A caveat is that the time scales of the dynamics are extremely long for this system, resulting in simulated structures that appear too ordered, thus requiring careful examination of the metastable states observed in order to draw any conclusions.

We observe transport of water, sodium ions and 20 small polar solutes within the pores of a lyotropic liquid crystal (LLC) membrane using atomistic molecular simulations. We find that the transport of a species is dependent not only on molecular size, but on chemical functionality as

well. We observe that all solutes perform intermittent hops between lengthy periods of entrapment. Three different trapping mechanisms are responsible for this behavior. First, solutes that drift out of the pore can become entangled among the dense monomer tails. Second, solutes can donate hydrogen bonds to the monomer head groups. Third, solutes can coordinate with sodium counter ions. The degree to which a solute is affected by each mechanism is dependent on the chemical functionality of the solute.

Mathematically modeling complex transport phenomena at the molecular level can be a powerful tool for identifying transport mechanisms and predicting macroscopic properties. We use two different stochastic time series models, parameterized from long molecular dynamics (MD) simulation trajectories of a cross-linked H_{II} phase lyotropic liquid crystal (LLC) membrane, in order to predict solute mean squared displacements (MSDs) and solute flux, and thus solute selectivity, in macroscopic length pores.

First, using anomalous diffusion theory, we show how solute dynamics can be modeled as a fractional diffusion process subordinate to a continuous time random walk. From the MD simulations, we parameterize the distribution of dwell times, hop lengths between dwells and correlation between hops. We explore two variations of the anomalous diffusion modeling approach. The first variation applies a single set of parameters to the solute displacements and the second applies two sets of parameters based on the solute's radial distance from the closest pore center.

Next, we generalize Markov state models, treating the configurational states of the system as a Markov process where each state has distinct transport properties. For each state and transition between states, we parameterize the distribution and temporal correlation structure of positional fluctuations as a means of characterization and to allow us to predict solute MSDs. We show that both models reasonably reproduce the MSDs calculated from MD simulations. However, qualitative differences between MD and Markov state dependent model-generated trajectories may limit its usefulness.

Finally, we demonstrate how one can use these models to estimate flux of a solute across a macroscopic-length pore and, based on those quantities, the membrane's selectivity towards each

solute. This work helps to connect microscopic chemically-dependent solute motions that do not follow simple diffusive behavior with macroscopic membrane performance.

Dedication

To all of the fluffy kitties.

Acknowledgements

I would like to acknowledge and thank a number of people whose support and guidance has not gone unnoticed. First, I thank my advisor, Michael Shirts, who has been instrumental to my growth as a competent and confident researcher, pushing me to do high quality science. I thank the entire Shirts group for their feedback and encouragement as well as for making our lab an enjoyable to work. I thank Richard Noble and Douglas Gin for their guidance on past and present experimental research on LLC membranes. I also thank Michael McGrath, Gregory Dwulet and Sarah Dischinger, members of the Gin and Noble groups, for providing invaluable experimental perspective. I thank Matthew Glaser for his contributions to my understanding of liquid crystal systems. Our first paper would not have been complete without his insight. I thank Joseph Yelk for his effort creating the X-ray diffraction code on which our first paper relied. I thank Christopher Calderon for his guidance on applying Bayesian non-parametric techniques to solute time series.

I would like to acknowledge my financial support: The ACS Petroleum Research Fund grant #59814-ND7 and the Graduate Assistance in Areas of National Need (GAANN) fellowship which is funded by the U.S. Department of Education.

Finally, I would like to acknowledge my high performance computing resources: Molecular simulations were performed using the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC). This work also utilized the RMACC Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and

ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University.

Contents

Chapter

1	Introduction	1
1.1	Membranes for Selective Separations	1
1.1.1	How Small Molecule Membrane Separations Work	1
1.1.2	Separation Applications	1
1.2	Competing Membrane Technologies	1
1.2.1	Amorphous Membranes	1
1.2.2	Nanostructured Membranes	2
1.3	Cross-linked Self Assembled Liquid Crystal Membranes for Selective Aqueous Separations	2
1.3.1	The H _{II} Phase	2
1.3.2	The Q _I Phase	2
1.4	Atomistic Molecular Simulation of LLC Membranes	2
2	Understanding Nanoscale Membrane Structure	4
2.1	Introduction	4
2.2	Methods	12
2.2.1	Source code	12
2.2.2	Monomer Parameterization	13
2.2.3	Unit Cell Preparation	13

2.2.4	Monomer Placement	14
2.2.5	Equilibration	15
2.2.6	Equilibrium Calculations	16
2.2.7	Simplified Systems	21
2.2.8	<i>Ionic Conductivity Calculations</i>	22
2.2.9	Cross-linking	22
2.3	Results and Discussion	23
2.3.1	Density of Monomers Around Pores	23
2.3.2	Simulated XRD Comparison to 2D-WAXS Data	25
2.3.3	Atomistic Structure of the Pore Columns	44
2.3.4	Slow Dynamics	46
2.3.5	Ionic Conductivity Measurements	48
2.3.6	Effect of Cross-linking	49
2.4	Conclusions	49
3	Chemically Selective Transport in the H_{II} Phase	53
3.1	Introduction	53
3.2	Methods	58
3.2.1	Molecule Parameterization	58
3.2.2	System Setup	59
3.2.3	Mean Squared Displacement	61
3.2.4	Molecular Size Determination	62
3.2.5	The Stokes-Einstein Relationship	63
3.2.6	Hop Detection	64
3.2.7	Time Spent in Pore Region	65
3.2.8	Identification and Analysis of Hydrogen Bonds	65
3.2.9	Coordination Number	66

3.2.10 Association Lifetimes	66
3.2.11 Radial Distribution Functions	67
3.3 Results and Discussion	68
3.3.1 Structure of Membrane Constituents	68
3.3.2 Mechanisms Governing Small Solute Transport	71
3.3.3 Transport of Simple Alcohols	79
3.3.4 Transport of Diols, Triols and Sugars	80
3.3.5 Transport of Ketones and Amides	83
3.3.6 Transport of Thiols	85
3.3.7 Solutes Unable to Donate Hydrogen Bonds	86
3.4 Conclusion	87
Bibliography	91
Appendix	
A Test	99
A.1 Setup and analysis scripts	99
A.2 Further details regarding monomer parameterization	99
A.3 Attempted Self-assembly	103
A.4 Monomer build procedure	105
A.5 Number of monomers per column	106
A.6 Initial Configuration Dependence	107
A.6.1 Initial pore spacing	107
A.6.2 Initial pore radius	108
A.6.3 Initial stacking distance between monomers	110
A.7 Placement of monomer head groups	113

A.8 Equilibration Details	113
A.9 Calculation of pore-to-pore spacing statistics	116
A.10 2D Small Angle X-ray Scattering Data	118
A.11 Component radial density functions	119
A.12 Ionic Conductivity	120
A.13 Cross-linking algorithm details	120
A.14 The nematic order parameter	124
A.15 Pore radius as a function of monomers per column	127
A.16 Noise in simulated diffraction patterns	128
A.17 The effect of imposed tail tilt	129
A.18 Tail organization	129
A.19 Correlation Functions	132
A.20 Structure factors of simplified systems	133
A.21 Ensemble of ordered sandwiched configurations	136
A.22 Carboxylate Dihedrals	138
A.23 All solvated systems	139
A.24 The role of water in the appearance of R-double	140
A.25 Diffusion of monomers	142
A.26 Autocorrelation of monomer dihedrals	142
A.27 Influence of Cross-linking on physical properties	144
A.28 Setup and analysis scripts	144
A.29 Water content equilibration	146
A.30 Solute Interaction	150
A.31 Solute MSD Ranking Time Lag Sensitivity	150
A.32 Pore Region Definition	152
A.33 Hydrogen Bond Detection Sensitivity	153
A.34 Lifetime Distributions	154

A.35 Pore Splines	156
-----------------------------	-----

Tables

Table

2.1 For 5 column-per-pore systems, the equilibrated vertical stacking distance, d_{equil} , of monomers in the disordered basin is approximately 0.6 Å higher than their initial configuration stacking distance, d , while d_{equil} is smaller than d for disordered basin systems, but 0.9 Å higher than experiment. The correlation length is larger for systems in the ordered basin.	25
2.2 The simulated XRD patterns of the systems tested, normalized so that the average intensity of R-alkanes in each pattern equals 1, show R-pores and R- π reflections that are significantly higher than experiment and R-spots reflections that are slightly lower than experiment. R-double does not appear in any simulated patterns, and thus has no measurable intensity. In terms of the locations of the reflections, R-pores, R-alkanes and R-spots appear at $ \mathbf{q} $ values that are close to experiment, while R- π appears at a significantly lower $ \mathbf{q} $ value than experiment.	52
2.3 Deviation of the positions of the center of mass of head groups from their average positions (thermal disorder) as well as their idealized positions (quenched disorder) to use as simulated disorder in our model system.	52
3.1 Names of solutes along with their molecular structures and the abbreviations which we use in this paper. Colors correspond to solute groupings and are used in many plots.	57

A.1 The first column provides the names of the python scripts available in the <code>llcsim</code> GitHub repository that were used for system setup and post-simulation trajectory analysis. Paths preceding script names are relative to the <code>llcsim</code> root directory.	102
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.	
A.2 The average pore radii of systems built with an initial pore radius of 2.5 Å and 5 Å equilibrate to values that agree within error. If the pore radius is too large, slit pores may form.	110
A.3 The pore spacing measured in two ways agree within uncertainty. In the main text, we report values calculated by measuring the average distance between pore centers. Alternatively, one can estimate the pore spacing as half of the <i>x</i> and <i>y</i> box vectors. Both methods agree within uncertainty, but measuring the distance between pore centers does a better job of showing the spread of pore-to-pore distances.	117
A.4 The diffusion constants for all systems are on the order of $10^{-14} m^2/s$	142
A.5 The first column provides the names of the python scripts available in the <code>LLC_Membranes</code> GitHub repository that were used for system setup and post-simulation trajectory analysis. Paths preceding script names are relative to the <code>LLC_Membranes/LLC_Membranes</code> 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.	145
A.6 We chose a diverse combination of initial pore and tail water contents in order to study its effect on equilibrium water content.	147

Figures

Figure

- | | |
|---|---|
| 2.1 (a) The LLC monomer Na-GA3C11 (b) rendered atomistically (c) exhibits wedge-like character. (d) Monomers stack on top of each other to create columns with short range order, then assemble into pores with hydrophilic head groups (blue) facing towards the pore center. (e) The pores assemble into hexagonally packed columnar mesophases. | 7 |
| 2.2 (a) Previous understanding of the pores are essentially speculations based on limited chemical and experimental data. (b) We use detailed molecular modeling in this paper in order to appropriately model the pore's complex architecture, which is crucial to understanding the mechanism of solute transport. In both pictures, the head group region is colored blue and the tail region is colored cyan. | 9 |

- 2.3 (a) 2D-WAXS gives details about repeating features on the order of angstroms. Experimentalists have explained each of the 5 major reflections present as follows: (R- π) Aromatic head groups $\pi - \pi$ stack 3.7 Å apart. (R-double) Monomers arrange vertically in a 2_1 helix. (R-alkanes) Alkane chain tails pack 4.5 Å apart. (R-spots) Monomer tails are tilted with respect to the membrane plane. (R-pores) As derived from SAXS, the pores are spaced 4.12 nm apart and pack hexagonally (b) (Reproduced from Ref.[1]) The repeat spacing in the 1D-SAXS scattering pattern is characteristic of hexagonal packing. The leading peak, q^* , represents the distance between the d_{100} planes, which translates into a distance between pore centers of 4.12 nm. 11
- 2.4 (a) We measured the intensity of R-alkanes by calculating the average intensity within the region bounded by $|\mathbf{q}| = 1.4 \text{ \AA}^{-1}$ and 1.57 \AA^{-1} (between 4.0 and 4.5 Å in real space). We excluded intensities within $\pm 30^\circ$ of the meridional axis defined by $q_r = 0$, since the simulated spectra overlap with R- π in those regions in all cases. (b) We measured the intensity of R- π and R-double based on the peaks of the q_z cross-section of the diffraction pattern. (c) We measured the intensity of R-spots by averaging the peak heights produced by radially integrating the patterns within the R-alkanes region. We took the intensity of R-spots as the average of the peak intensities near 37 and 143°. (d) We measured the intensity of R-pores by measuring the height of the d_{100} peak (the leading peak closest to $q_r=0$) of the cross-section of the diffraction patterns along the q_r axis at $q_z=0$ 19
- 2.5 Systems with 5 columns-per-pore have equilibrated pore spacings closest to the experimental value of 4.12 nm. The equilibrated pore spacing of the model increases as the number of columns in each pore increases. 24

- 2.7 (a) R-spots increases in intensity when the temperature of the system is lowered to 280 K. (b) We measured the average angle made between each monomer alkane tail and the membrane plane. The average tilt angle (dashed line) is near -2° which is far from the 37° tilt angle previously used to explain R-spots. (c) To isolate the main cause of R-spots, we removed all atoms from the trajectory except for carbon atoms that constitute the tails. The simulated XRD pattern of the tails-only trajectory still shows R-spots. (d) Since the tails stay relatively flat, we plotted the center of mass of the first four carbon atoms of each tail originating from the head groups (for example, green colored centroids in the plot surround the red centroid in hexagonal fashion). Visually, the packing looks hexagonal. (e) We hypothesize that R-spots is the result of ordered tail packing. Defining the membrane plane to be 0° , we measured the angles between each COM and its nearest neighbor COMs for the equilibrated sandwiched configuration simulated at 280 K. Peaks appear in the distribution at -60° , 0° and 60° . (f) The Fourier transform of a hexagonally packed grid of points defined by the angles in (e) shows intensity at the same locations where we expect to see R-spots, as well as intensity along the q_z axis where R- π would appear. 31
- 2.8 The maximum intensity of R- π generated from simulations of the ordered parallel displaced configuration is 3 times larger than experiment. The q_r cross-section of R- π ((a) and (b)) is qualitatively different between experiment (a) and simulation (b). We fit Lorentzian profiles to each peak and the FWHM (full width at half maximum) of the simulated pattern agrees with experiment within error. However the fit to the simulated data is affected by the three sharp peaks which appear near $q_r = 0$. The q_z cross-sections of R- π ((c) and (d)) are qualitatively similar. Each fits a Lorentzian profile well, however the FWHM of the simulated cross-section (d) is 48% larger than the experimental cross-section (c). Additionally, the peak of the simulated q_z cross-section is located at a lower q_z value than experiment. 32

2.9 As we increase column independence using the f parameter, the intensity of R- π decreases and the q_r cross-section of the simulated diffraction pattern becomes more smooth. (a) When we place all columns at the same reference z coordinate, $f = 0$, the intensity of R- π is 18.3 and is characterized by sharp Bragg-like peaks. (b) When columns have a moderate amount of independence, $f = 0.75$, the intensity of R- π decreases 5-fold to 3.4. The edges of the peak are beginning to smooth out, but sharp peaks still exist. (c) When columns are completely independent, $f = 1$, the intensity of R- π decreases 9-fold to 2.1 and the cross-section is relatively smooth.	35
2.10 The q_r section of R- π is much smoother than Figure 2.8b when we generate simulated XRD patterns from an ensemble of 40 independent simulations, suggesting that much of the extra initial intensity may be due to the starting conditions. The maximum intensity of R- π is $3\times$ higher in the sandwiched configuration (a) than in the parallel displaced configuration (b).	37
2.11 (a) The distributions of the head group COM deviations from their idealized positions generated using pooled data from all frames of each independent configuration are symmetric, implying that there is equal probability for a head group to displace in the positive or negative z , r or θ -direction. (b) The COM position of a given head group is displaced randomly upon quenching. The ECDF generated from the pooled distributions (black) in (a) agree with the means of the ECDFs generated from each COM (red). The red error bars are larger than the black error bars since there is a wider distribution of mean COM deviations than the mean deviations of distributions sampled from the full distribution.	38

2.12 (a) The standard deviation of the distribution of quenched disorder from the first 5 ns of the main ordered parallel displaced system studied in this paper (black dashed line) is in agreement with the distribution of quenched disorder standard deviations calculated from the ensemble of simulations (histogram). (b) The mean values of r and θ from the first 5 ns of the main system trajectory (black dashed line) is in agreement with the distribution of mean values calculated from each simulation in the ensemble (histogram). The mean values of z are necessarily 0 so they are not plotted. (c) The radial densities of tail atoms, head group atoms and sodium atoms calculated from the first 5 ns of the main system simulation trajectory (black lines) and from the ensemble of trajectories (all other lines) look qualitatively similar.	40
2.13 (a) When monomer head groups are rotated with respect to vertically adjacent monomers (top), R-double is visible while the heavy atoms of the head groups are held in place with position restraints (middle). R-double fades once the position restraints are released. (b) When monomers are non-uniformly spaced (top), R-double appears if all heavy atoms of the head groups are held in place with position restraints (middle). R-double quickly fades once the position restraints are released (bottom). (c) When we add 1 wt% water to the parallel displaced configuration in the ordered basin (top) R-double is not initially present during the restrained portion of equilibration (middle). After 200 ns of equilibration, R-double becomes visible and persists for another 200 ns (bottom).	43
2.14 (a) The structure of the head groups is responsible for the appearance of R-double. When we remove head groups from the trajectory, the simulated diffraction pattern no longer shows R-double. (b) Monomer head groups above or below each other that hydrogen bond with a shared water molecule are drawn closer together in the z -direction. Blue monomers were stacked above orange monomers in the initial configuration.	44

3.2 The radial densities of various monomer components paint a picture of the pore topology where the pore centers of hydrated systems are primarily, but not exclusively, composed of water and sodium ions. The monomer groups labeled in each plot correspond to the color-coded monomer pictured in the upper right corner of the figure. RDFs show the average number of atoms located at a given distance from the pore center normalized by the volume of the annular bin to which they belong. The vertical dashed black lines are positioned so that they intersect with the maximum head group density. (a) In the dry system, the density of head groups and sodium ions are highest 0.15 to 0.4 nm from the pore center. There is an appreciable density of head groups crowding the pore center. (b) In the 5 wt% system, monomer head groups retreat radially in order to make room for water molecules at the pore center. (c) Monomers in the 10 wt% system retreat an additional 0.2 nm to make room for more water.	69
3.3 Looking down at the <i>xy</i> plane of the cross-linked LLC membranes, the lowest density regions of both membranes are in the pore center and in the distal tail region, furthest from the pore center (a & c). Water concentration is highest in regions of low monomer density (b & d).	70
3.4 (a) The <i>z</i> -direction MSD of water in the 10 wt% water system is about 51 times higher than water in the 5 wt% water system. Water appears to exhibit Brownian motion in the 10 wt% system due to its linear MSD. (b) The MSD of sodium in the 10 wt% water system is about 49 times higher than sodium in the 5 wt% water system.	71
3.5 The MSDs of solutes in the 5 wt% water system (a) are significantly smaller than those of the solutes in the 10 wt% water system (b). The MSDs are not a monotonic function of molecular size (c and d). A significant number of solute MSDs fall below the theoretical lines predicted by the Stokes-Einstein equation and Gierer and Wirtz' corrected Stokes-Einstein equation.	74

- 3.6 All solutes show subdiffusive transport behavior inside the membrane's nanopores, similar to that exhibited by ethanol. (a) The z -coordinate trace of 3 representative ethanol COMs shows clear periods of entrapment separated by hops. In general, the longest dwell times occur when solutes are situated far from the pore center and more frequent hops occur when solutes are close to the pore center. (b) The time-averaged MSD of ethanol is sublinear which suggests transport is governed by an anomalous subdiffusion process. 75
- 3.7 (a) Hops made in the pore region of the 10 wt% water system are, on average, 59% larger than those made outside the pore region. The trend in hop lengths is similar to the trend in MSDs shown in Figure 3.5b implying that solutes which make consistently larger hops have higher MSDs. The fraction of time spent by a solute in the pore region (b) does not necessarily lead to more frequent hopping (c). For example, ribose spends the largest fraction of time in the pore region, yet hops the fifth least frequently. 76
- 3.8 (a) Solutes capable of donating hydrogen bonds to monomer head groups do so to varying degrees. The reported percentages represent unique solute-monomer hydrogen bonds. Individual solutes that hydrogen bond with multiple head groups simultaneously are only counted once. (b) The lifetime of individual hydrogen bonds appears correlated to the percentage of solutes involved in hydrogen bond interactions. Hydrogen bond lifetimes tend to be longer for solutes that hydrogen bond frequently. Note that solutes incapable of donating hydrogen bonds are omitted from this figure. 78
- 3.9 (a) Solutes, especially those with carbonyl groups, spend a significant fraction of time coordinated to sodium ions. (b) The length of time a solute-sodium pairs spends associated tends to be higher for pairs that associate more frequently. 79

3.12 (a) The MSD of ketones and amides decreases and the deviation from Stokes-Einstein predicted behavior increases with molecular size. (b) The radial density near the pore center ($r = 0$) decreases with decreasing solute MSD. Peaks in the RDFs of urea, acetamide and acetone that appear 0.2 - 0.4 nm from the pore center are likely due to coordination with sodium ions. (c) The amides hydrogen bond with water far less than acetic acid, however they tend to coordinate with sodium ions more frequently. Acetone coordinates with sodium with same frequency as acetic acid.	84
3.13 (a) The RDF of mercaptoethanol is similar to ethylene glycol except for its higher density in the tail region and consequently lower density in the pore region. (b) 2,3-dimercapto-1-propanol is densest near the head groups, unlike glycerol whose density is very high close to the pore center. (c) Overall, dimethylsulfoxide has a higher density than acetone within the pore region which may in part explain its marginally larger MSD.	85
3.14 (a) The MSDs of solutes that can only receive hydrogen bonds are significantly lower than expected. (b) The solutes' radial density is surprisingly high in the pore region, but is balanced by an appreciable amount of solute trapped in the tails. (c) The low MSDs exhibited by each of these solutes is due to a combination of entrapment within the tail region and a high degree of coordination with sodium ions.	88
A.1 Atomistic representation of the monomer Na-GA3C11. White atoms represent hydrogen, cyan atoms represent carbon, red atoms represent oxygen and the blue atom is sodium.	101
A.2 (a) We created a box of isotropically packed monomers and allowed it to simulate for 100 ns using isotropic and anisotropic pressure coupling. (b) The nematic order parameter hovers close to zero for the duration of the simulation meaning the system maintains its isotropic alignment.	104

- A.3 (a) We parameterized a single monomer and annealed it to produce a low-energy configuration. (b) We assembled monomers into columns by stacking them on top of each other. (c) We duplicated each column and rotated them to create hydrophilic pore centers. We chose to stack twenty monomers into each column. (d) We duplicated the pores and placed them into a monoclinic unit cell with hexagonal symmetry. 105
- A.4 (a) The correlation functions generated from ordered basin sandwiched configurations are nearly the same when the z -dimension of the system is doubled. (b) Oscillations persist throughout the length of the correlation function for the sandwiched system in the ordered basin built with 40 monomers per column. The full system is ca. 17 nm tall so the full correlation function covers half this length since periodicity forces the remaining length to be its mirror image. 106
- A.5 (a) In a perfect hexagonal array, there are 5 distances that should be exactly equal. As pictured in this diagram the distance from 1-2, 1-3, 1-4, 2-4, and 3-4 are the same. All plots in (b)-(f) are the average of these distances and represent the pore spacing during the restrained portion of the dry equilibration procedure. Every 50 ps (0.05 ns) we reduce the force constant on the position restraints according to the sequence: 1000000, 3162, 56, 8, 3, 2, 1, 0 kJ mol⁻¹ nm⁻². When we chose an initial pore spacing below experimental value (b) or at the experimental value (c), there is an abrupt change in pore spacing when the position restraints are reduced to 56 kJ mol⁻¹ nm⁻². When we initially position pores 45 Å apart (d) the pore spacing remains relatively stable. When we initially place pores 50 Å apart (e), the pore spacing decreases nearly linearly once the restraints are reduced to 56 kJ mol⁻¹ nm⁻². When we initially place pores 55 Å apart (f), so that monomers do not intersect with adjacent pores, and position restraints are reduced to 56 kJ mol⁻¹ nm⁻², the pore spacing changes erratically before stabilizing when force constants are reduced below 3 kJ mol⁻¹ nm⁻². 109

A.6 When creating an initial configuration, we define the pore radius, R, based on the distance of the carbonyl carbon from the pore's central axis.	111
A.7 A system that was built with an initial pore radius of 8 Å equilibrates to a structure that exhibits both cylindrical and slit-like pores. As pictured here, sodium ions are colored blue, carbon atoms in the aromatic ring of the head group are colored orange and all else is colored cyan.	111
A.8 When layers are initially stacked 10 Å apart and the system is equilibrated using the dry equilibration procedure, large vacuum gaps form as the monomers attempt to fill space.	112
A.9 (a) Sandwiched benzene dimers stack 3.8 Å apart. (b) Parallel-Displaced benzene dimers stack 3.4 Å vertically and 1.6 Å horizontally apart. (c) T-shaped benzene dimers stack 5.0 Å apart. (d) Monomers stacked in the sandwiched configuration (e) Monomers stacked in the parallel-displaced configuration.	114
A.10 Head groups are likely stacked in a configuration between sandwiched and parallel displaced. (a) In our parallel displaced configurations with 5 columns per pore, we placed head groups so the angle, θ , between vectors extending from vertically stacked head groups to the pore center equals $36^\circ ((360^\circ / ncol) / 2$, where $ncol$ is the number of columns per pore). (b) In the true configuration, θ is likely distributed between 0° and 36° . When $\theta=0^\circ$, the system architecture is the same as a sandwiched configuration.	115

- A.11 Two dimensional small angle X-ray scattering experiments show reflections along the q_r axis at $q_z=0$. The spacing between the reflections is characteristic of a hexagonal phase and is directly related to the distance between pores. The reflections are arced because the pores are not perfectly aligned. A perfectly aligned system would show dots at each peak location. An isotropically aligned system would show concentric rings about the origin which pass through each peak location. The d_{100} and d_{110} peaks (the first two pairs of reflections from origin located at $|\mathbf{q}| \approx 0.18 \text{ \AA}^{-1}$ and $|\mathbf{q}| \approx 0.31 \text{ \AA}^{-1}$) are not visible in the 2D WAXS pattern. The edges of the d_{200} reflection (the third pair of reflections located at $|\mathbf{q}| \approx 0.36 \text{ \AA}^{-1}$) are slightly visible in the 2D WAXS pattern, so we used them to approximate the experimental intensity of R-pores. 118
- A.12 In all cases, the component radial distribution functions exhibit a composition gradient transitioning from the hydrophilic to the hydrophobic regions. Systems with monomers initially stacked 3.7 Å apart in the parallel displaced (a), and sandwiched (b) configurations both show the highest concentrations of ions and head groups away from the pore center. Systems with layers initially spaced 5 Å apart (c) and (d), both show the highest concentrations of ions and head groups near the pore center, implying a more uniform, disordered pore. 119
- A.13 The groups used for radial distribution calculations. Red atoms are in the tails group. Green atoms are in the head group region. The blue atom is sodium. 121
- A.14 Looking down onto the xy plane of the membrane, we binned the radial distance of all atoms in chosen groups from pore centers. The bins are defined by the annulus bounded by concentric circles centered at the pore centers, as shown. We normalize by dividing the count of atoms within each bin annulus by its volume. 121

A.15 The Collective Diffusion model and the Nernst-Einstein relation yield agreeing values of ionic conductivity for both types of ordered basin systems simulated. Both methods estimate the value of the ionic conductivity to be an order of magnitude higher than the experimental value. There is more noise in the collective diffusion model because there is inherently less data that can be used for its calculation. For that reason, we use only the Nernst-Einstein relationship in the main text.	122
A.16 Terminal vinyl groups present on each monomer tail react with free radical initiators to create monomers with terminal vinyl radicals. Vinyl radicals react with the vinyl groups of other monomers in order to propagate crosslinking.	122
A.17 We calculate the nematic order parameter by measuring the angle between \mathbf{v} and the z -axis.	124
A.18 The distribution of angles between nematic director vector (see Figure A.17) and the z -axis averaged over the equilibrated portion of each trajectory. The ordered sandwiched configuration has a higher nematic order parameter than the ordered parallel displaced configuration. The disordered sandwiched configuration has a higher nematic order parameter than the disordered parallel displaced configuration.	126
A.19 The pore radius increases as we increase the number of columns per pore.	127
A.20 The magnitude of noise, especially along $q_r=0$, decreases as the number of independent configurations (N) increases.	128
A.21 As we reduce the force constants applied to the end of each tail of an initially tilted monomer configuration, the average tilt angle with respect to the membrane plane decreases until it is nearly negligible.	129
A.22 We divided the tail into three regions: tail-fronts (blue), tail-middles (green) and tail-ends (red).	130

- A.23 Monomer tails pack together hexagonally. (a) The center of mass of each tail-front (See Figure A.22) is visualized as a blue sphere. The red sphere highlights an example of a tail-front center of mass with its nearest neighbors (green spheres) surrounding it in a hexagonal pattern. (b) We calculated the angle between each tail-front center of mass and its nearest neighbor tail-front center of masses and plotted the distribution. There are distinct peaks ca. -60° , 0° and 60° which is indicative of hexagonal packing. (c) We repeated the same procedure outlined in (a) and (b) with the tail-middles. There is still a high degree of order. (d) We repeated the procedure again with the tail-ends. There is far less order at the ends of the tails where there is more space to fill. 131
- A.24 1D correlation functions of the center of masses of aromatic head groups, $g(z)$, show decaying oscillatory behavior and have consistent correlation lengths with experiment. We calculated the correlation length by fitting a decaying exponential function (Equation 2.2 of the main text) to the peaks of $g(z)$. The correlation length is longer for ordered basin systems (Table 2.1 of the main text). We did not attempt to calculate the correlation length for (d) because there are no clear peaks. We assume that this correlation length is less than the vertical distance between monomers. . . 132
- A.25 When we average all z -slices of the 3D correlation function, the correlation length of the ordered sandwich configuration nearly triples compared to that shown in Figure A.24a. 133
- A.26 We can decrease the maximum intensity of $R-\pi$ by increasing thermal noise in the z -direction; however, the q_r cross-section profile does not change. (a) The intensity of $R-\pi$ drops precipitously as we increase σ_z . We can decrease the intensity of $R-\pi$, relative to the background intensity, relative to systems with simulation levels of disorder (blue) by a factor of 3 when we increase σ_z by 12% (green) and by a factor of 15 when we increase σ_z by 27% (red). (b) Although the maximum intensity of $R-\pi$ decreases with increasing σ_z , the peak shape remains the same. 134

- A.27 Increasing the amount of disorder in the xy plane decreases the FWHM of the q_r cross-section of R- π . When we cut the amount of r and θ disorder in half (Noise factor = 0.5), the FWHM increases by 88%. When we double the disorder (Noise factor = 2), the FWHM decreases by 51% 135
- A.28 (a) The distributions of the head group COM deviations from their idealized positions generated using pooled data from all frames of each independent configuration are symmetric, implying that there is equal probability for a head group to displace in the positive or negative z , r or θ -direction. (b) The COM position of a given head group is displaced randomly upon quenching. The ECDF generated from the pooled distributions (black) in (a) agree with the means of the ECDFs generated from each COM (red). The red error bars are larger than the black error bars since there is a wider distribution of mean COM deviations than the mean deviations of distributions sampled from the full distribution. 136
- A.29 (a) The standard deviation of the distribution of quenched disorder from the first 5 ns of the main ordered sandwiched system studied in this paper (black dashed line) is in agreement with the distribution of quenched disorder standard deviations calculated from the ensemble of simulations (histogram). (b) The mean values of r and θ from the first 5 ns of the main system trajectory (black dashed line) is in agreement with the distribution of mean values calculated from each simulation in the ensemble (histogram). The mean values of z are necessarily 0 so they are not plotted. (c) The radial densities of tail atoms, head group atoms and sodium atoms calculated from the first 5 ns of the main system simulation trajectory (black lines) and from the ensemble of trajectories (all other lines) look qualitatively similar. . . . 137

- A.30 A potential energy diagram for the dihedral angle about the carbon-carbon bond of a carboxylate group attached to an aromatic ring, using parameters from GAFF (blue), Rakitin (orange) and Borkman (green). It is energetically favorable for the carboxylate group to stay in plane with the aromatic ring. There is a large energy penalty as the dihedral becomes antiparallel to the plane of the aromatic ring. 138
- A.31 Adding 1 wt% water to the ordered basin dry systems causes an increase in the intensity R-spots. R-double becomes visible in the parallel displaced configuration. Additional amounts of water increase discrepancies with experiment 139
- A.32 When water is removed from the trajectory which produces R-double, R-double is still present since it is the result of uneven spacing between head groups. 140
- A.33 (a) We group the z -coordinate of the center of mass of monomer head groups into 20 bins. Peaks in the distributions indicate a shared hydrogen bonded water molecule between a monomer head group and one in the bin vertically above it. The distance between the peaks roughly corresponds to the distance between pairs of monomers. The distance between peaks in pore 1 (indicated by arrows) is equal to two times the π -stacking distance. In this scenario, monomers are no longer evenly spaced in the z -direction since they are pulled closer together by hydrogen bonds. (b) We performed discrete Fourier transforms on each distribution in (a) In Pore 1, there is periodicity every 8.3 \AA , twice the π -stacking distance, as indicated by the intensity at 0.76 \AA^{-1} . In this case, the center of mass of two adjacent pairs is separated by twice the average monomer stacking distance, while the four monomers involved are unequally spaced. These conditions seem likely to give rise to R-double. 141
- A.34 We measured the autocorrelation functions of the dihedrals highlighted in black. We only included dihedrals associated with the outer tails in our calculation since the position of the center tail leads to fundamentally different interactions and dynamics. 143

- A.35 The outer tail ether dihedrals become decorrelated on a reasonable timescale when we raise the system temperature. (a) The dihedrals do not become decorrelated for ca. 100 ns when the system is equilibrated at 300 K. (b) The ether dihedrals become decorrelated after 11 ns when the temperature of the system is increased to 500 K. . 143
- A.36 There are minor differences between the physical properties of a crosslinked (xlink) system versus an uncrosslinked system. The sandwiched configuration (S) exhibits a smaller decrease in its ionic conductivity, while the parallel displaced (PD) configuration remains constant. The pore spacing of both systems decreases upon crosslinking. The vertical distance between stacked monomers increases in the sandwiched configuration and decreases in the parallel displaced configuration. The correlation length decreases in both configurations. 144
- A.37 (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. 146
- A.38 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. 148
- A.39 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. 149

A.40 The water content in the tail and pore region is not affected by cross-linking	149
A.41 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.	150
A.42 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.	151
A.43 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.	152
A.44 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 °.	153
A.45 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.	155
A.46 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.	156

Chapter 1

Introduction

1.1 Membranes for Selective Separations

- Membranes are useful separation tools. Cover historical usage
- Some discussion of competing separation technologies.
- They are more efficient and have a smaller footprint than thermal distillation.

1.1.1 How Small Molecule Membrane Separations Work

- Maybe briefly mention ultrafiltration, microfiltration
- Overview of reverse osmosis and nanofiltration.
- Separation mechanisms (pore flow versus solution diffusion)

1.1.2 Membrane Separation Applications

- Desalination
- Separation of Organic Micropollutants
- Recovery of Valuable Dissolved Species
- Breathable Barriers

1.2 Competing Membrane Technologies

1.2.1 Amorphous Membranes

- Mostly about the industry standard, thin film composite polyamide membranes.

1.2.2 Nanostructured Membranes

- Graphene
- Carbon nanotubes
- zeolites
- MOFs (?)

1.3 Cross-linked Self Assembled Liquid Crystal Membranes for Selective Aqueous Separations

- A less studied class of membranes, LLCs, have the potential to outperform competing membranes.

1.3.1 The H_{II} Phase

- Ideal geometry for transport
- Go through history
- Highlight issues with synthesis
- Talk about alignment approaches by Xunda Feng.

1.3.2 The Q_I Phase

- Easier to synthesize
- History of monomers used

1.4 Atomistic Molecular Simulation of LLC Membranes

- Why MD?
- Why HII phase?
- What work has been done already

Questions to be answered:

- (1) How do we build and equilibrate these systems?
- (2) What is the nanoscopic structure?
- (3) What factors affect solute transport?
- (4) Can we estimate macroscopic properties?
- (5) How can we learn mechanisms with minimal human intervention?

Chapter 2

Understanding Nanoscale Membrane Structure

2.1 Introduction

More highly selective nanoporous membranes would be extremely useful for performing complex aqueous separations with seawater and various types of wastewater. For example, Sodium chloride and boron in seawater [2] and organic micropollutants found in municipal and industrial wastewaters [3] represent just a few of the diverse contaminants of water sources. By efficiently separating contaminants from feed solutions with highly selective membranes, it is possible to reduce the number of required membrane passes and post-treatment steps needed for a given filtration process [4], thus lowering energy requirements. Additionally, one can also extract valuable resources from the feed streams. For example, flowback water produced during hydraulic fracturing of shale formations contains dissolved species such as acetate whose extraction has economic value [5].

Reverse osmosis (RO) and nanofiltration (NF) are two prevailing membrane filtration processes that can be used to separate solutes on the order of 1 nm in size and smaller, including ions. However, commercially available amorphous RO membranes and porous NF membranes lack uniformity, which severely impacts their selectivity [6]. Although scalable, their fabrication involves the spontaneous assembly of polymers into disordered structures that offer separation pathways that are tortuous and polydisperse in size [4]. Tortuosity increases the effective length that a solute must travel, while pore size polydispersity limits membrane selectivity.

Ordered, nanostructured porous membrane materials have become an increasingly popular research area for aqueous separation applications because they offer the ability to control pore

architecture at the molecular scale, thereby permitting the design of solute-specific separation membranes [7]. However, their development has been limited by the ability to synthesize and scale various technologies. Graphene sheets are atomically thick, which results in excellent water permeability but defects during manufacturing severely impact selectivity [8]. Molecular dynamics (MD) simulations of carbon nanotubes have shown promise [7], but synthetic techniques are unable to achieve scalable alignment and pore monodispersity [9, 10]. Consequently, there is a need for new materials that could provide scalable nanostructured porous membranes.

Preliminary evidence has shown that cross-linked lyotropic liquid crystal (LLC) membranes can be produced at moderate scale and may be capable of performing highly selective separations. LLCs are amphiphilic molecules that have the ability to self-assemble into porous nanostructures [11] and can be cross-linked to create mechanically strong membrane films with periodic pores on the order of 1 nm in diameter [12]. Unlike the pores in most commercial NF membranes, polymeric LLC membrane pores are uniform in size because they are formed by self-assembly. Since LLC polymer membranes lack an appreciable pore size distribution, they inherently exhibit high selectivity due to their strict molecular weight cut-off (MWCO) [12]. Additionally, the LLC monomers examined in this paper are salts, and therefore lead to Donnan exclusion of ions in solution. The membrane gains a net surface charge when counterions from the head groups that line the pore walls escape into the feed solution in an effort to balance the gradients of concentration and electric potential [13].

The feasibility of nanostructured LLC polymer membranes for selective separations has been demonstrated using LLC monomers that form the type I bicontinuous cubic (Q_I) [14, 15, 16] and the inverted hexagonal (H_{II}) [12] phases. When separating organic solutes from NaCl, Q_I -phase membrane filtration experiments have shown selectivity 2–3 times higher than commercial RO and 6–12 times higher than commercial NF membranes [5]. When separating a series of various sized dyes, the H_{II} -phase membrane showed complete rejection of dyes bigger than 1.2 nm in size [12].

The H_{II} -phase pore geometry (Figure 2.1e) has a higher theoretical capacity for transport than the Q_I phase. The H_{II} phase forms at room temperature in the presence of ca. 10 wt% water

and consists of hexagonally packed, hydrophilic pore columns[11]. In the absence of water, neat monomer will form the same hexagonal columnar structure which, in the literature, has been referred to as the Col_h thermotropic phase[17]. Q_I-phase membranes consist of a tortuous network of three dimensionally interconnected pores that prevent optimal through-plane transport. In contrast, the densely packed, non-tortuous and uniform sized pores of H_{II}-phase membranes represent the ideal geometry for achieving high solute flux[18]. However, the hexagonally packed LC domains of the H_{II}-phase generally form mutually unaligned domains, which hurts membrane permeability. This domain scale misalignment had inhibited further development of this technology, and research efforts were focused on the Q_I phase, whose geometry does not require alignment [19].

However, recently researchers have learned how to macroscopically align the hexagonal domains which has revived research into H_{II}-phase LLC polymer membranes. In 2014, Feng et al. showed that one can align Col_h domains, created by the “dry” monomer Na-GA3C11, using a magnetic field with subsequent cross-linking to lock the structure in place[17]. In 2016, Feng et al. showed that one could also obtain the same result by confining the neat monomer between PDMS or glass substrates since hexagonal mesophases preferentially anchor perpendicular to both surfaces[1]. Current experimental efforts are focused on extending the method to the H_{II} phase and characterizing the performance of these newly aligned systems.

Our current understanding of the molecular details of resulting LLC polymer membranes' nanostructure is not sufficient to be able to precisely design them for specific separations. Dischinger et al. attempted to use an empirical model that correlates the physiochemical properties of the counterion used in a Q_I-phase LLC membrane to solute rejection[20]. Although their model showed some qualitative agreement with experiment, the quality of fit of their model was limited due to complex solute-membrane interactions that could not easily be modeled. Additionally, they observed an unexpected discrepancy in the relationship between uncharged solute rejection and water permeability, which will require a more in-depth knowledge of the difference between solute and solvent transport.

Over the past 20 years, H_{II}-phase LLC polymer membrane studies have been limited primarily

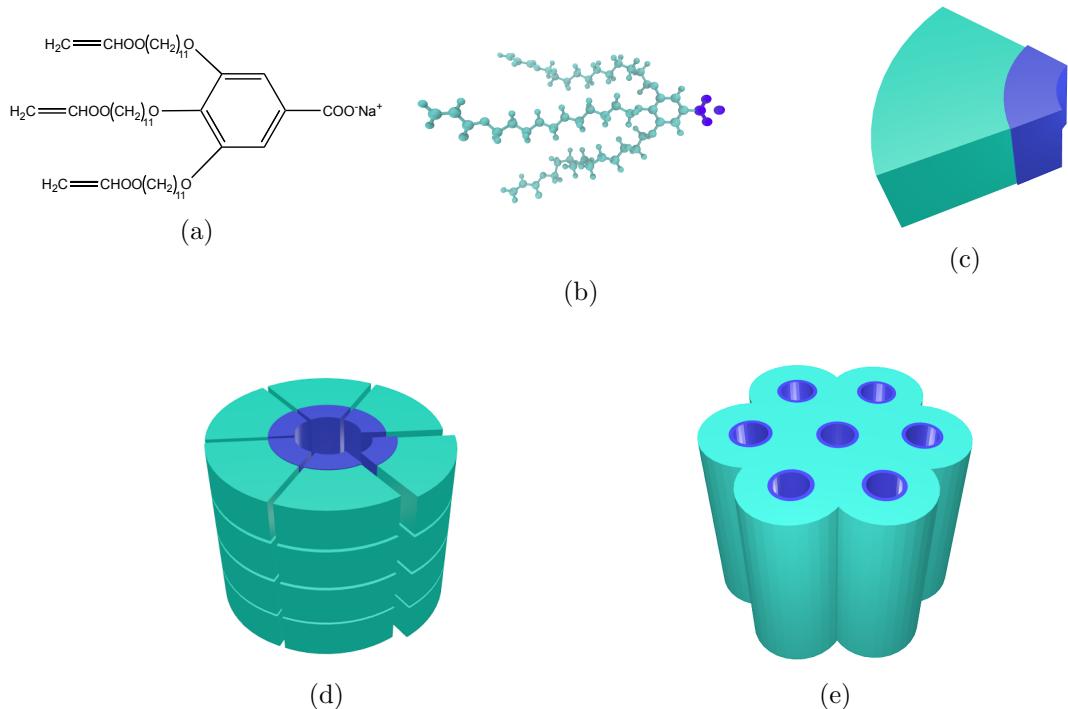


Figure 2.1: (a) The LLC monomer Na-GA3C11 (b) rendered atomistically (c) exhibits wedge-like character. (d) Monomers stack on top of each other to create columns with short range order, then assemble into pores with hydrophilic head groups (blue) facing towards the pore center. (e) The pores assemble into hexagonally packed columnar mesophases.

to the Na-GA3C11 monomer with some characterization done after minor structural modifications. For example, Resel et al. varied the length of the monomer tails and the counterion used and observed its effect on pore spacing [21]. In a later study of rejection performance, it was shown that membranes formed by cross-linked Na-GA3C11 in the H_{II} phase cannot separate solutes less than 1.2 nm in diameter because the pores are too large [12]. We do not yet understand how to controllably reduce the effective pore size or how to tune the chemical environment in the nanopores of this or related materials for small molecule separations. The only source of predictive modeling for LLC systems have been macroscopic models that likely do not adequately describe transport at these length scales [14]. Modeling with molecular detail could provide sufficient information about the mechanisms and chemical features to better inform experimental design of similar nanostructured membranes.

A molecular-level understanding of LLC polymer membrane structure, enabled by molecular dynamics (MD) simulations, will enhance our understanding of the role of monomer structure in solute transport and provide guidelines to reduce the large chemical space available to design monomers for creation of separation-specific membranes. It is important that we model the system with a nanoscopic pore structure that is maximally consistent with experimental data in order to best approximate the chemical environment experienced by solutes in future simulated transport studies. Atomistic MD simulations can provide the required level of detail (Figure 2.2b), assuming the force fields are sufficiently accurate. With such an atomistic model, we can directly observe molecular-level solute transport and suggest governing mechanisms. We can also observe how the choice of head group interacts with solutes of interest. In addition, we can interchange counterions which may influence both the pore size and the strength of the Donnan potential.

In this study, we achieve a more realistic atomistic description of hexagonal LLC polymer membranes than, to our knowledge, has ever previously been created, and explore what new structural information can be gained and what structure hypotheses are supported by this model. We validate the results using as much experimental information as possible. We are most interested in reproducing the conclusions about structure drawn from small-angle X-ray scattering (SAXS)

and wide-angle X-ray scattering (WAXS) experiments, as well as in matching ionic conductivity measurements [1].

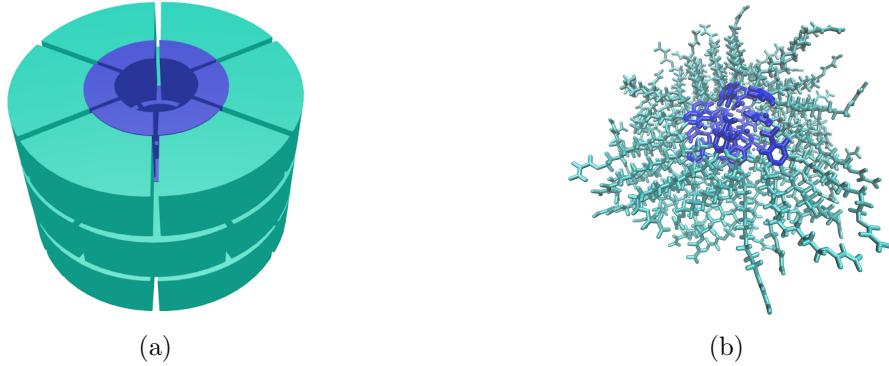


Figure 2.2: (a) Previous understanding of the pores are essentially speculations based on limited chemical and experimental data. (b) We use detailed molecular modeling in this paper in order to appropriately model the pore's complex architecture, which is crucial to understanding the mechanism of solute transport. In both pictures, the head group region is colored blue and the tail region is colored cyan.

In this paper, we perform molecular modeling of the Col_h thermotropic (i.e., solvent-free) assembly formed by Na-GA3C11. Compared to the lyotropic (i.e, solvent-containing) H_{II} phase, the Col_h phase is a simpler starting point. The system is not assembled in aqueous phase which allowed us to simulate longer timescales by omitting the solvent in the simulations, and there exists detailed experimental characterization of the fully aligned state, including 2D-WAXS patterns (Figure 2.3a) that are useful for reconstructing structural data.

There are five major features of interest present in the 2D experimental WAXS pattern shown in Figure 2.3a.

- (1) $R\text{-}\pi$: The location of the first is at $q_z = 1.7 \text{ \AA}^{-1}$, corresponding to a real space separation of 3.7 Å. Previous work attributes this reflection to $\pi\text{-}\pi$ stacking between aromatic rings in the direction perpendicular to the membrane plane, or z -axis [17]. For simplicity, we will refer to this reflection as R- π .

- (2) *R-double*: A weak intensity line, located at exactly half the q_z value of R- π ($q_z = 0.85 \text{ \AA}^{-1}$), corresponds to real space periodicity of 7.4 Å. Since this reflection corresponds to double the spacing of R- π in real space, we will refer to it as R-double. R-double has been previously interpreted as 2_1 helical ordering of aromatic rings along the z -axis[17].
- (3) *R-alkanes*: A low intensity ring located at $|\mathbf{q}| = 1.4 \text{ \AA}^{-1}$ marks the third major reflection of interest. The real space separation corresponds to 4.5 Å which is characteristic of the average spacing between packed alkane chains [22]. We will call this reflection R-alkanes.
- (4) *R-spots*: Within R-alkanes, are four spots of higher relative intensity. Accordingly, we name these reflections R-spots. The location of all spots is $\sim 37^\circ$ from the q_r axis in their respective quadrants. In many liquid crystal systems, such spots are explained as the result of alkane chains tilted with respect to the membrane plane[23].
- (5) *R-pores*: The final significant feature corresponds to the spacing and symmetry of the pore columns. The feature, which we named R-pores, is characterized by reflections along the equatorial axis defined by $q_z = 0$. The spacing between dots is indicative of the hexagonal symmetry of the packed pores. We observe the same information with higher resolution by looking at the same system's 1D-SAXS pattern (Fig. 2.3b). The location of the leading SAXS peak (closest to $q_r = 0$) is related to the distance between pores.

Despite having structural data, there is still information which experiment cannot definitively answer. Specifically, we want to know:

- (1) What is the density of monomers that pack around each hydrophilic core?

Authors often describe this and similar systems as being made up of stacked monomer layers. A simple molecular simulation study of a similar molecule suggested that there are 4 monomers in each layer. Their estimation is based on a simulated system containing only 16 total monomers which likely does not sufficiently model the chemical environment

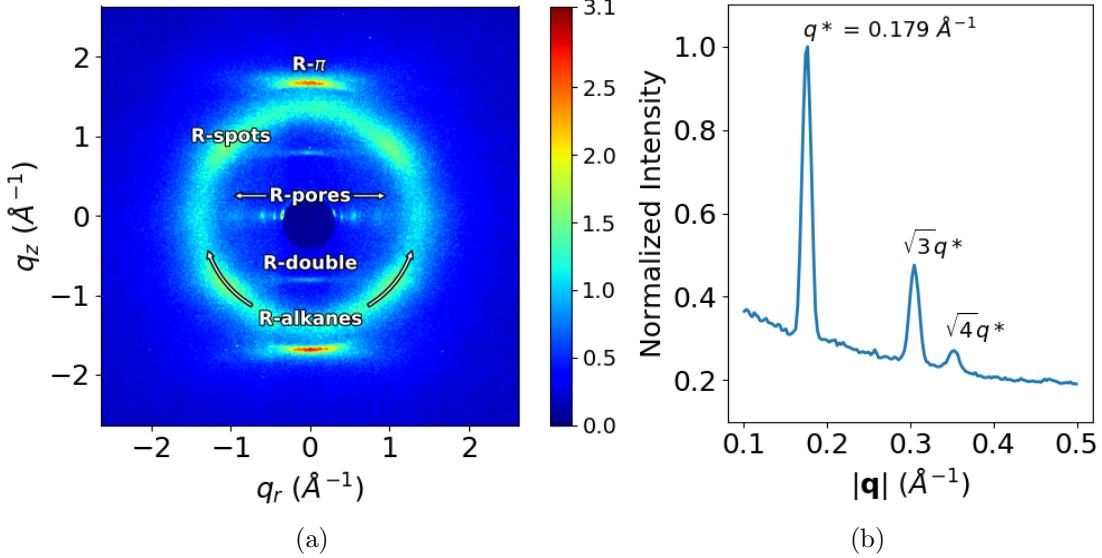


Figure 2.3: (a) 2D-WAXS gives details about repeating features on the order of angstroms. Experimentalists have explained each of the 5 major reflections present as follows: (R- π) Aromatic head groups $\pi - \pi$ stack 3.7 Å apart. (R-double) Monomers arrange vertically in a 2_1 helix. (R-alkanes) Alkane chain tails pack 4.5 Å apart. (R-spots) Monomer tails are tilted with respect to the membrane plane. (R-pores) As derived from SAXS, the pores are spaced 4.12 nm apart and pack hexagonally (b) (Reproduced from Ref.[1]) The repeat spacing in the 1D-SAXS scattering pattern is characteristic of hexagonal packing. The leading peak, q^* , represents the distance between the d_{100} planes, which translates into a distance between pore centers of 4.12 nm.

present in the real system. [24]. A separate calculation based on the volume of the liquid crystal monomers proposes that there are seven monomers in each layer [21].

We are careful to avoid the term “layers” since many liquid crystalline systems have long range order in 1 or 2 spatial dimensions and short range order in the other dimensions. [25] In the system we are studying, there are long-range 2D correlations in the hexagonal array of pores (xy plane) and short range z -direction correlations within stacked columns of monomers. In this study, we use atomistic molecular modeling to study how the system’s structure is affected by the density of these monomer columns that surround each pore’s hydrophilic core.

- (2) What structural motif best matches experimental 2D-WAXS patterns?

On the short timescales accessible to MD (even the 100’s of nanoseconds of simulation per-

formed here are short compared to experimental timescales), we observe distinct metastable configurations which depend on starting configuration. We simulated X-ray diffraction (XRD) patterns of our system and compared them to experimental 2D WAXS patterns (Figure 2.3a) so that we ensure our model creates a nanoscopic chemical environment maximally consistent with experiment within the constraints of our force field. Using this approach, we are able to confirm some previous interpretations of the WAXS pattern and refute others.

- (3) Is it necessary to include any water in order to appropriately model the Col_h phase?

While the Col_h phase is described as dry, it is likely that small amounts of ambient water are absorbed into the system. The hydrogen-bonding network formed by the water may play a role in structuring the pore. We used simulated XRD patterns to uncover any meaningful structural difference between a “dry” and a “wet” system.

- (4) What is the detailed atomistic structure of the pores?

The limited picture that experiment provides tells us that there are hexagonally packed, hydrophilic regions where transport is likely to occur. One may instinctively imagine these regions as tube-like pathways with well-defined boundaries. We explored the composition of the pores, the partition between the hydrophilic and hydrophobic regions, and its sensitivity to initial configuration, including both dry and wet systems.

2.2 Methods

2.2.1 Source code

Python scripts used to set up systems and conduct post-simulation trajectory analysis are available online at <https://github.com/bencoscia/l1csm>. The python scripts used to simulated XRD patterns are publicly available online at <https://github.com/joeyelk/MD-Structure-Factor>. Additional details are available in Section

A.1 of the Supporting Information (SI).

2.2.2 Monomer Parameterization

We parameterized the interaction potential for the liquid crystal monomer Na-GA3C11 using the Generalized AMBER Force Field (GAFF) [26] with the Antechamber package [27] provided with AmberTools16 [28]. We chose GAFF because it has been parameterized for use with organic molecules. Although researchers have optimized force fields, including GAFF, for use with liquid crystals, they are often tuned in order to reproduce one or more experimental observables for specific types of liquid crystal molecules [29, 30]. Na-GA3C11 does not have a structure characteristic of more commonly studied liquid crystal monomers so we chose to stick with a widely used force field. If we can reproduce experimental trends using a single ‘out-of-the-box’ force field, then there will be no need to complicate our studies with customized parameter sets. We assigned atomic charges using the am1bccsym method of `molcharge` shipped with QUACPAC from Openeye Scientific Software. We ran all molecular dynamics simulations using GROMACS 2016 [31, 32, 33, 34].

We generated an ensemble of characteristic, low-energy vacuum monomer configurations by applying a simulated annealing process to a parameterized monomer. We cooled monomers from 1000 K to 50 K over 10 ns. We randomly pulled a low energy configuration from the trajectory then reassigned charges using `molcharge`. Using the new charges, we annealed the monomer system again and pulled a random monomer configuration from the trajectory which we used for full system construction. Section A.2 of the SI provides further detail of the parameterization process.

2.2.3 Unit Cell Preparation

The time scale for self-assembly of monomers into the hexagonal phase is unknown and likely outside of a reasonable length for an atomistic simulation, requiring a more efficient way to build the system. Previous work has shown that a united-atom model of an LLC can self-assemble into the H_{II} phase configuration in ∼1000 ns [35]. We attempted atomistic self-assembly by packing monomers into a box using Packmol [36]. Simulations of greater than 100 ns show no indicators of progress

towards an ordered system (see Section A.3 of the SI). To bypass the slow self-assembly process, we used automated procedures to assemble monomers into a structure close to one of a number of hypothesized equilibrium configurations (See Figure A.3 of the SI). These initial structures may result in equilibration into slowly-interconverting metastable states, an issue we will address later in the paper.

A typical simulation volume contains four pores in a monoclinic unit cell, the smallest unit cell that maintains hexagonal symmetry when extended periodically. Each pore is made of columns of stacked monomers with periodic continuity along the pore axis, avoiding any edge effects and creating an infinite length pore ideal for studying transport. We prefer a small number of stacked monomers in order to reduce computational cost and to allow us to look at longer timescales. Ultimately, we chose to build a system with 20 monomers per column in order to minimize finite size effects with reasonable computational expense and to obtain sufficient resolution when simulating XRD patterns (see further discussion in Section A.5 of the SI).

2.2.4 Monomer Placement

When constructing an initial configuration, there are a number of variables which require careful consideration while placing monomers. We find that the equilibrium configuration is sensitive to some while insensitive to others. For example, we find the starting pore radius, defined as the distance of a chosen head group carbon from the pore's central axis, does not influence the equilibrium structure if one chooses a reasonable value. The initial distance between pores, within a wide range, also has little effect on the equilibrated structure. However, one should not start them too close or there will be high energy repulsions during early equilibration. We chose a pore radius of 0.5 nm and an initial pore spacing of 4.5 nm, $\sim 10\%$ larger than the experimental value of 4.12 nm for our initial configurations. A sensitivity analysis of both parameters is presented in the SI, Section A.6. The distance between vertically stacked monomers, the xy position of monomers with respect to vertically adjacent monomers, and the number of columns-per-pore do influence the equilibrated structure and require further justification for their choices. We rely on experimental

data to inform them.

We chose the vertical spacing between monomers for the initial configuration based on $R-\pi$ and then allowed the system to readjust during equilibration. We rotated each monomer so the plane of its aromatic head group would be coplanar with the xy plane. We explored three different initial monomer spacings. The first is exactly equal to $R-\pi$ with monomers placed so aromatic rings stack 3.7 Å apart in the z -direction. We also extensively explored a second system with an initial spacing of 5 Å. We briefly explored a third system with an initial spacing of 10 Å. However this largest spacing yields non-physical behavior which is detailed in the SI, Section A.6.3.

We chose the relative orientation between vertically adjacent monomers in each column based on clues from diffraction data as well as the various known stacking modes of benzene and substituted benzene rings: sandwiched, parallel displaced and T-shaped [37]. We ruled out the T-shaped configuration because its ~5 Å equilibrium stacking distance [37] is inconsistent with $R-\pi$. It is also infeasible for the monomers to orient in the T-shaped conformation because of the bulky tail groups. We explored the system's preference towards the sandwiched vs. parallel displaced stacking modes in some detail. Both have reported stacking distances near the $R-\pi$ value of 3.7 Å. Head groups in our sandwiched initial configuration stack directly on top of each other while head groups in the parallel displaced initial configuration stack with an offset of $180^\circ/ncol$ where $ncol$ is the number of columns-per-pore. See Figure A.9 of the SI for a detailed illustration of the initial configurations in each mode.

The number of columns-per-pore is unknown, as stated in Question (1). We tested configurations constructed with a varied number of columns-per-pore. We built systems in the sandwiched and parallel displaced configurations with 4, 5, 6, 7 and 8 columns-per-pore.

2.2.5 Equilibration

We developed equilibration schemes for creating dry (i.e. thermotropic LC phase) and wet (i.e. LLC phase) configurations. Both schemes start with an initial configuration generated according to the previous guidelines. For wet systems, we added the desired concentration of water to the

initial configuration and carried out equilibration in the same way as the dry systems. First, we fixed monomer head groups in place using position restraints with a force constant of 10^6 kJ mol $^{-1}$ nm $^{-2}$. We gradually released the position restraints by decreasing the force constants over a series of NVT simulations. We allowed the resulting unrestrained structure to equilibrate for 5 ns in the NPT ensemble with pressure controlled by the Berendsen barostat, followed by NPT equilibration simulations run for at least 400 ns using the Parrinello-Rahman barostat. More equilibration details are given in Section A.8 of the SI.

2.2.6 Equilibrium Calculations

2.2.6.1 Determining Equilibration Time

Using equilibrated structures, we carried out various calculations to characterize the system. We defined the point at which a system is equilibrated based on when the distance between pores stopped changing. We determined when the distances stopped changing by applying the statistical test of Chodera [38], implemented as `pymbar.timeseries.detectEquilibration`, to the time series. Typically, the pore-to-pore distance equilibrated between 200 and 350 ns. We used data collected after equilibration to do all subsequent analysis.

2.2.6.2 Calculation of Pore Spacing

To calculate the equilibrated pore spacing, we measured the distance between pore centers. We located the pore centers by averaging the coordinates of sodium ions in their respective pores. We generated pore spacing statistics using the bootstrapping technique (See Section A.9 of the SI). The pore spacings calculated in this way are consistent with one half of the x and y box vectors (See Table A.3 in the SI). However, our method for their calculation does a better job capturing the spread of pore-to-pore distances.

2.2.6.3 Generation of Simulated X-ray Diffraction Patterns

We generated simulated XRD patterns based on atomic coordinates in order to make a direct experimental comparison. We modeled all atomic coordinates as Gaussian spheres of electron density whose maximum and width are defined by each atom's atomic number and electronic radius respectively. A three dimensional Fourier transform of the array of electron density results in a three dimensional structure factor which represents the unit cell in reciprocal space. The experimental WAXS measurement was made using a vertically aligned film whose pores were oriented perpendicular to the direction of the incident X-ray beam. Although the pores are vertically aligned, the crystalline domains are still misaligned with respect to the xy plane. To account for this, we averaged 2D slices of the structure factor at all angles about $\mathbf{q} = (0, 0, z)$.

We normalized all diffraction patterns relative to R-alkanes. We believe that the alkane-alkane density, averaged over all angles, is the feature most likely to be replicated between experiment and simulation, as atomistic alkane force field parameters are relatively well-studied [26]. Other features are dependent on system ordering which is likely to have some dependence on initial configuration. We calculated the average intensity within R-alkanes of the experimental pattern, I_{avg} , and divided all intensities by this value. In this way, the average intensity of R-alkanes was set equal to 1. When calculating I_{avg} , we excluded intensities within $\pm 30^\circ$ of the meridional axis defined by $q_r = 0$, since the simulated patterns differ from experiment in those regions in all cases (See Figure 2.4a). Specifically, in contrast to the experimental WAXS pattern, R- π , as it appears in the simulated diffraction patterns, intersects with R-alkanes (See Section 2.3.2.4). We set an upper bound on the colorbar by multiplying I_{avg} by a scaling factor, f . Intensities that appear in the patterns $\geq f \times I_{avg}$ are colored uniformly. We applied the same scaling method to the simulated patterns. We chose a scaling factor of $f = 3.1$ in order to visibly display all features in all patterns.

We reported the intensities of R- π and R-double by recording the maximum values of the peaks of the q_z cross-section of the experimental pattern at $q_r = 0$ (Figure 2.4b). In the case of the experimental pattern, the peak heights are not perfectly symmetrical, so we report the average of

the two heights.

We measured the intensity of R-spots by averaging the peak heights produced by radially integrating the patterns within the R-alkanes region (Figure 2.4c). In some patterns, the spots are not easily discernable due to obstruction by other features. In that case, we report the average intensity at the intersection of R-alkanes with the q_z value of R-double since that is where we expect it to appear based on experiment. Since R-double does not appear in our simulated patterns, we estimate where it should appear as half of the q_z value of R- π .

We measured the intensity of R-pores based on the intensity of the d_{100} peak (the leading peak closest to $q_r = 0$) of the q_r cross-section of the patterns at $q_z=0$ (Figure 2.4d). The beamstop covers most of the small angle reflections in the experimental 2D-WAXS pattern. In order to compare the simulated intensity of R-pores to experiment, we used the q_r cross-section of 2D-SAXS which was generated from the same sample (Figure A.11 of the SI). Since the d_{200} peak is partially exposed in the experimental 2D-WAXS pattern, we normalized the experimental 2D-SAXS cross-section by matching the intensity of the d_{200} peak between it and the experimental 2D-WAXS cross-section. It is possible that the peak is not fully captured in the 2D-WAXS pattern and that we have underestimated the intensity of R-pores in the experimental pattern.

2.2.6.4 Pair Distribution Functions and Correlation Length

The normalized pair distribution function, $g(\mathbf{r})$, describes the probability of finding a pair of particles separated by \mathbf{r} ,

$$g(\mathbf{r}) = \frac{1}{\rho N} \left\langle \sum_{i=1}^N \sum_{j \neq i}^N \delta(\mathbf{r} + \mathbf{r}_j - \mathbf{r}_i) \right\rangle \quad (2.1)$$

where ρ is the average number density of particles and $\delta(\mathbf{r})$ is the Dirac delta function[39]. We applied Equation 2.1 in three dimensions and then extracted one dimensional distribution functions using slices of the grid along the appropriate axis.

We measured the one dimensional pair distribution function, $g(z)$, between centers of masses (COMs) of aromatic head group rings along the z -axis (perpendicular to the membrane plane). We averaged all 1D z -directional slices of the full 3D correlation function within 2.1 Å of $(x, y) = (0, 0)$.

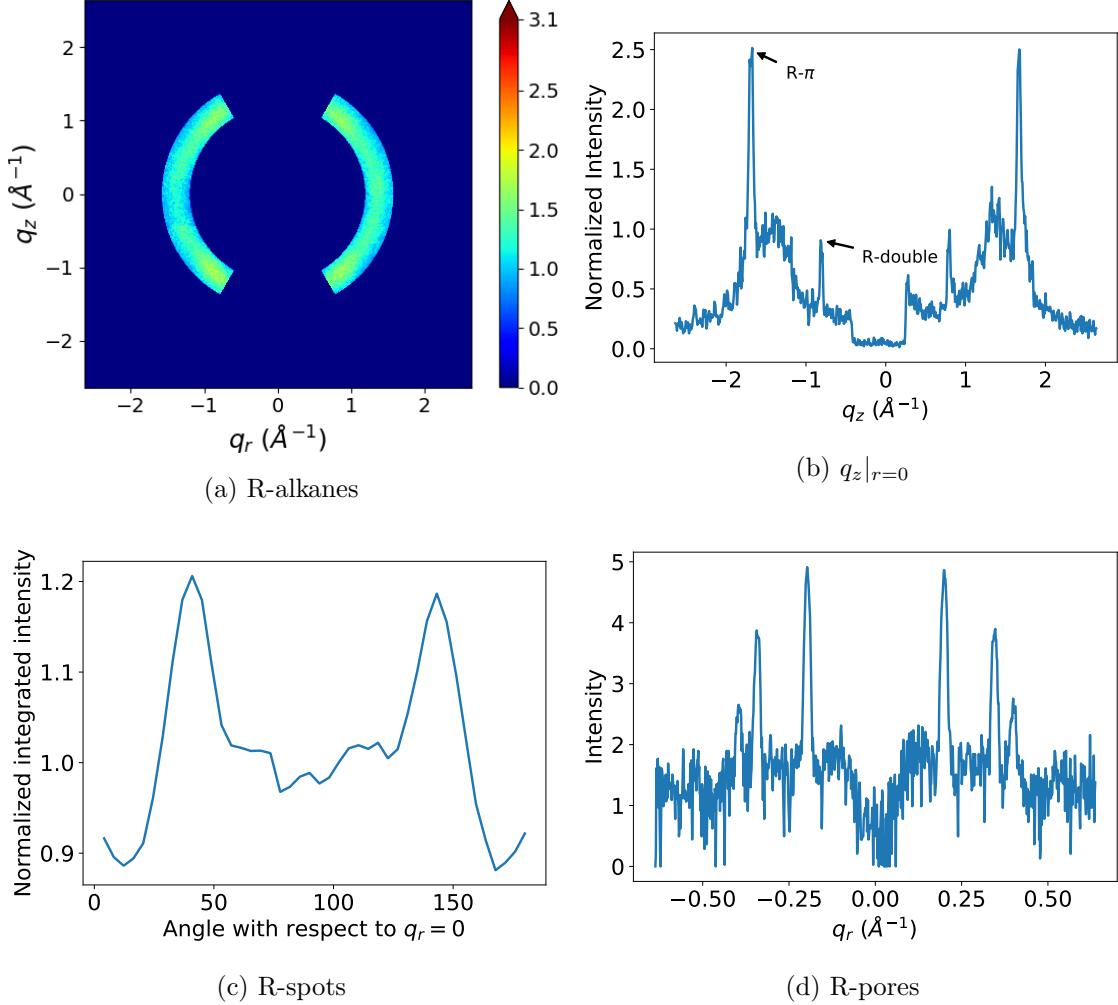


Figure 2.4: (a) We measured the intensity of R-alkanes by calculating the average intensity within the region bounded by $|\mathbf{q}| = 1.4 \text{ \AA}^{-1}$ and 1.57 \AA^{-1} (between 4.0 and 4.5 \AA in real space). We excluded intensities within $\pm 30^\circ$ of the meridional axis defined by $q_r = 0$, since the simulated spectra overlap with R- π in those regions in all cases. (b) We measured the intensity of R- π and R-double based on the peaks of the q_z cross-section of the diffraction pattern. (c) We measured the intensity of R-spots by averaging the peak heights produced by radially integrating the patterns within the R-alkanes region. We took the intensity of R-spots as the average of the peak intensities near 37 and 143° . (d) We measured the intensity of R-pores by measuring the height of the d_{100} peak (the leading peak closest to $q_r=0$) of the cross-section of the diffraction patterns along the q_r axis at $q_z=0$.

We chose 2.1 \AA as a crude approximation of the radius of the phenyl ring plane. We calculated the radius as the sum of the longest C-C distance within a phenyl ring (2.8 \AA) and two times the carbon atom electronic radius (0.7 \AA) [40]. In this study, $g(z)$ is characterized by an oscillatory function

with a period equal to the average distance between stacked monomers, and an amplitude that decays exponentially (see Section 2.3.2.4). The rate of decay is related to the correlation length, L, between monomer head groups. We estimated L by fitting the peaks of $g(z)$, using the python package `scipy.optimize.curve_fit` [41], to a decaying exponential function of the form:

$$Ae^{-z/L} \quad (2.2)$$

where A is a fitting parameter for amplitude, z is the independent variable of $g(z)$, and L is the fit correlation length. We calculated the error in the estimated value of L as the square root of the diagonal entry of the covariance matrix of optimized fit parameters.

We also used $g(z)$ to calculate the equilibrated vertical stacking distance between monomers, d_{equil} . We fit a decaying sinusoidal function (using `scipy.optimize.curve_fit`) to $g(z)$ of the form:

$$1 - A \cos\left(\frac{2\pi}{d_{equil}}z + B\right) e^{-z/L} \quad (2.3)$$

where A and B are fit parameters for the function's amplitude and phase shift respectively. This function could be used in place of Equation 2.2, however it does not consistently fit the peaks of $g(z)$ from parallel displaced configurations well enough to extract a reliable value of L.

2.2.6.5 Radial Distribution Functions

We explored the pores' compositions by measuring the average number densities of various monomer components as a function of distance from the pore centers. We looked at the average number density of sodium ions, aromatic rings and carbon atoms making up the monomer tails. We binned the radial distance of all atoms in each group from the pore centers, then normalized by the volume of the annulus defined by the bin edges and the z box vector (See Figure A.14 of the SI).

2.2.7 Simplified Systems

In order to gain a deeper understanding of discrepancies between the experimental and simulated R- π reflection, we used a simplified model where we represent each monomer as a single point scatterer located at the COM of its head group. In order to make a system approximately similar to our equilibrated simulation geometries, we created 4 hexagonally packed pore regions spaced 42.5 Å apart, each with 5 columns of scatterers spaced 4.4 Å apart in the z -direction. We built the system 4 times taller in the z -direction (80 scatterers per column) in order to access higher q_z resolution.

We gave the simplified models the same amount of disorder present in the atomistic simulations. We observe two sources of disorder in our atomistic simulations: thermal motion of atoms during the simulation and quenched disorder created by rapid structural rearrangement during early equilibration that is largely locked into place for the remainder of the simulation, even at the 100 ns time scale. We measured thermal disorder by calculating the standard deviation of the distribution of head group COM positions from their average positions. We measured quenched disorder by calculating the standard deviation of the distribution of head group COMs from their idealized average positions. In the z -direction, we measured the deviation of the head group COMs from an equally spaced column of head groups. In the xy plane, we calculated the standard deviation in radial position from the pore center and the angular deviation from equally spaced points surrounding the pore center. This method for calculating quenched disorder inherently includes the convoluted thermal disorder.

We approximated interactions between particles by correlating the z -distance between points within each column of point scatterers. We placed points in each column by drawing random samples from a multivariate normal distribution defined by the mean positions of an equally spaced column of points. We gave the distribution at each point along the column the same standard deviation, σ_z , and we defined a covariance matrix such that the covariance, v , of the distance, d , between scatterers decays exponentially from v according to the equation $ve^{-d/L}$, where L is the

correlation length. Unless noted otherwise, we used the experimental correlation length of 9.0 Å.

We calculated averaged R- π profiles by simulating the diffraction pattern of trajectories consisting of 1000 independent simple system configurations generated with point scatterer placement based on the quenched disorder seen in atomistic simulations. The quenched disorder observed in the atomistic simulations is up to 8 times greater in magnitude than thermal disorder (See Table 2.3 in Section 2.3.2.4). This method for simulating time-averaged R- π profiles assumes that the timescales for large scale rearrangements of the system are much longer than what we can feasibly simulate with MD.

2.2.8 Ionic Conductivity Calculations

We calculated ionic conductivity using the Nernst-Einstein relationship, which relates the DC ionic conductivity, σ , to ion diffusivity, D , concentration, C , ion charge, q , the Boltzmann constant, k_b , and absolute temperature, T :

$$\sigma = \frac{q^2 C D}{k_b T} \quad (2.4)$$

We measured sodium ion diffusion coefficients by calculating the slope of the linear region of the z -direction mean square displacement curve as indicated by the Einstein relation [42]. We visualized the MSD plot to determine where to begin and end a linear fit. We measured ion concentration with respect to the volume of the entire unit cell. More details are provided in the SI, Section A.12.

2.2.9 Cross-linking

In order to better match the experimental membrane synthesis process, we created a cross-linking algorithm that one can apply to equilibrated structures. The primary purpose of cross-linking experimentally is to create a mechanically robust membrane. For that reason, we are not concerned with replicating the kinetics of the reaction, but instead emphasize understanding how much and in what way cross-linking modulates the system's structure.

We based our cross-linking algorithm on the known reaction mechanism (Figure A.16 of the

SI). The reaction takes place at the terminal vinyl groups on each alkane tail. The procedure is carried out iteratively. Each iteration, the algorithm chooses carbons to cross-link based on the distance between eligible carbon pairs. The algorithm then updates the topology with the new bonds and atom types, energy minimizes the system and runs a short simulation before selecting the next group of eligible carbons atoms.

2.3 Results and Discussion

2.3.1 Density of Monomers Around Pores

Our simulations best support a model built with 5 monomer columns-per-pore based on the measured equilibrated pore-to-pore distances. To identify the density of monomer columns around each pore, addressing Question 1 in the introduction, we ran simulations of systems created with 4–8 columns-per-pore. We built systems in both the parallel displaced and sandwiched configurations and equilibrated them according to the dry equilibration procedure. We tested all systems with an initial vertical monomer spacing, d , of 3.7 Å in accordance with R- π . We tested 4 additional systems with monomers initially spaced 5 Å apart vertically (see Section A.6.1 of the SI for more details on sensitivity to initial monomer spacing). We considered the pore-to-pore spacing to be equilibrated as defined in Section 2.2.6.1. Figure 2.5 shows the equilibrated pore-to-pore distances for all systems tested.

All systems tested, although equilibrated from the perspective of the metrics used here, are frozen in metastable basins. Not all make physical sense or fit the experimental profile that we are trying to match. In the limit of infinite simulation time, all systems will in theory converge to a single equilibrium configuration, but that time is far beyond the 100’s of nanoseconds simulated here. For simplicity, we group the systems studied here into the ordered and disordered basins. What we find is that any system equilibrated with $d = 3.7$ Å can generally be characterized as being in a more ordered basin, and any system equilibrated with $d = 5.0$ Å can be characterized as being in a more disordered basin. The extra space between stacked monomers in the disordered basin

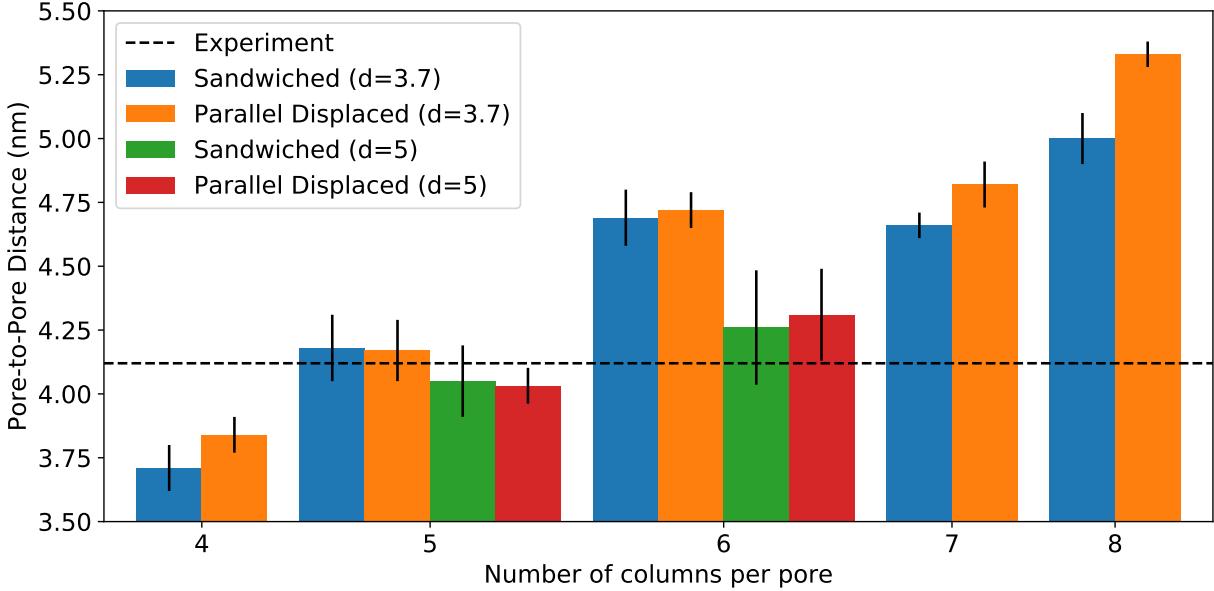


Figure 2.5: Systems with 5 columns-per-pore have equilibrated pore spacings closest to the experimental value of 4.12 nm. The equilibrated pore spacing of the model increases as the number of columns in each pore increases.

systems gives the monomer head groups more rotational freedom. We quantified the ordering of the head groups using the nematic order parameter (see Section A.14 for details of the calculation). Disordered basin systems have a lower nematic order parameter (meaning they are more disordered) than ordered basin systems. Generally, when monomers are started further apart, they stay further apart than systems where monomers are started closer together (see Table 2.1). This increased equilibrated stacking distance of monomers in disordered basin systems increases their discrepancy with experimental stacking distances derived from WAXS relative to ordered basin systems (See Section 2.3.2.4).

Systems built with 5 columns around each pore equilibrate to a pore spacing that is most consistent with the experimental value of 4.12 nm derived from SAXS measurements (Figure 2.3b). Ordered basin systems built with 4 columns-per-pore equilibrate to an average pore spacing 0.25 nm lower than experiment. Ordered basin systems built with 6 columns-per-pore, have an equilibrated pore spacing ca. 0.50 nm higher than experiment. Monomers in disordered basins systems stack

too far apart, therefore they are not suitable candidate structures. Although their pore-to-pore spacing is in close agreement with experiment, the 6 column-per-pore disordered sandwiched and disordered parallel displaced configurations stack ~ 4.87 and 4.94 \AA apart respectively, which is $\sim 1.2 \text{ \AA}$ further apart than suggested by experiment. The monomer columns in these systems are effectively stretched in the z -direction which allows them and the pores to pack closer together on the xy plane. 5 column-per-pore systems stack, at a maximum, 0.9 \AA further apart than experiment (see Table 2.1). The remainder of this discussion will therefore focus on the analysis of systems built with 5 columns-per-pore.

The number of columns-per-pore dictates the density of interaction sites within each pore and the pore radius. A higher number of interaction sites will surely play a role in transport of molecules that have an affinity for the monomer head groups. The pore radius increases with the number of columns-per-pore (See Figure A.19). This radius is likely related to the size of the molecules which can be excluded by this type of membrane. One may be able to control LLC membrane pore size by selecting monomers that are stable in configurations with more columns-per-pore.

System	$d (\text{\AA})$	$d_{equil} (\text{\AA})$	Correlation Length (\AA)
Sandwiched	3.7	4.27 ± 0.03	4.2 ± 0.8
Parallel Displaced	3.7	4.33 ± 0.04	14.5 ± 1.3
Sandwiched	5.0	4.48 ± 0.07	3.2 ± 0.9
Parallel Displaced	5.0	4.60 ± 0.08	$< d_{equil}$
Experiment	–	3.70	9 ± 1

Table 2.1: For 5 column-per-pore systems, the equilibrated vertical stacking distance, d_{equil} , of monomers in the disordered basin is approximately 0.6 \AA higher than their initial configuration stacking distance, d , while d_{equil} is smaller than d for disordered basin systems, but 0.9 \AA higher than experiment. The correlation length is larger for systems in the ordered basin.

2.3.2 Simulated XRD Comparison to 2D-WAXS Data

We can better understand the structure of the system by simulating XRD patterns produced from equilibrated MD trajectories (see Section 2.2.6.3) and comparing these simulated patterns to experiment, addressing Question 2 in the Introduction. We tested systems built with 5 columns-per-

pore in the parallel displaced and sandwiched configurations at 300 K in the ordered and disordered basins, as those were the most consistent with the most unambiguous XRD features, the pore-to-pore spacing and the stacking distance, as described in Section 2.3.1. We generated simulated patterns using the locally equilibrated portion of each simulation trajectory (see Section 2.2.6.1). There are two important factors to take into account when looking at differences between experimental and simulated XRD patterns. First, the simulated diffraction patterns have some noise, especially along the q_z axis at $q_r=0$, which is where several of the more interesting features are located. This is due to the angle averaging of the 3D structure factor around the q_z -axis, meaning there are fewer samples as q_r approaches 0. More importantly, our simulations do not appear to be long enough to sample truly independent configurations within their respective metastable basins (see Section 2.3.4), meaning we must take care in interpreting the results of the XRD. The amount of independent configurations required to converge along the q_z axis is analyzed in SI, Section A.16. The simulated patterns generated for all systems studied are shown and compared to experiment in Figure 2.6.

The simulated XRD patterns show moderate agreement with experiment. There are key qualitative and quantitative differences between the experimental and simulated locations and intensities of each major reflection. The integrated intensity and location of each simulated reflection relative to its experimental counterpart are shown in Table 2.2. Our approach for measuring the intensity of each reflection is described in Section 2.2.6.3.

In the next few subsections, we individually address each major reflection, and the similarities and differences between simulation and experiment. We start with relatively uncomplicated analyses of reflections that are very similar between experiment and simulation:

- The location of R-alkanes (Section 2.3.2.1)
- The location and intensity of R-pores (Section 2.3.2.2)

and then move onto more complicated explanations of those reflections whose characteristics do not match, or have not been fully explained by experiment:

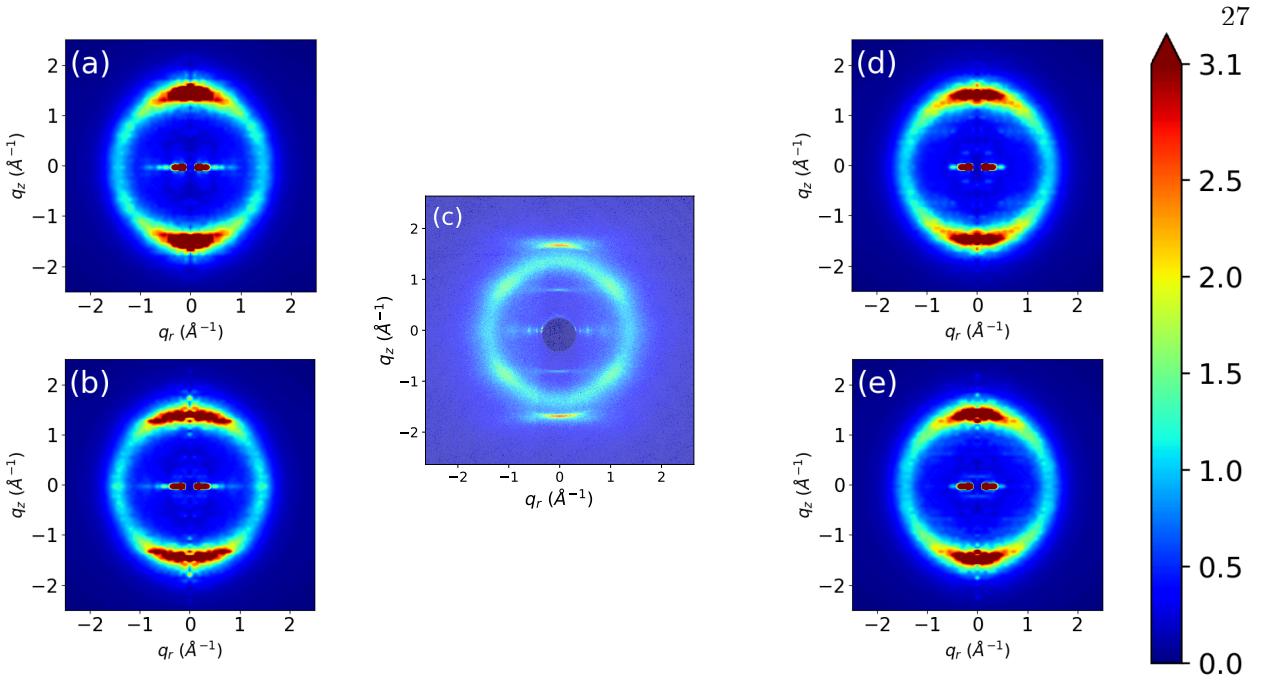


Figure 2.6: Simulated XRD patterns show some qualitative agreement with experiment. Shown is a comparison of the (a) Sandwiched, ordered basin (b) Sandwiched, disordered basin (d) Parallel displaced, ordered basin and (e) Parallel displaced, disordered basin configurations with (c) experimental WAXS. The major reflections of interest (See Figure 2.3a for their definitions) are present at varying degrees. In all cases, R-pores, R-alkanes and R- π are present to some degree. R-spots is also present, however it is generally partially covered by the broad R- π reflection. Quantitative comparisons of the relative intensities and locations of reflections of interest are presented in Table 2.2. In both XRD patterns generated from parallel displaced configurations, there is a faint line across $q_z \sim 0.7 \text{ \AA}^{-1}$, half the simulated value of R- π . Although it does not cross through $q_r = 0 \text{ \AA}^{-1}$, it is at the same q_z value where we expect R-double to be present. Still, R-double is not fully reproduced in any of the simulated patterns.

- The origin of R-spots (Section 2.3.2.3)
- The position, shape and intensity of R- π (Section 2.3.2.4)
- The origin of R-double (Section 2.3.2.9)

2.3.2.1 The Location of R-alkanes

We normalized the experimental and simulated diffraction patterns so that the average intensity within R-alkanes is equal to 1 because we believe that R-alkanes is the experimental feature

most likely to be reproduced by our simulations. We reached this conclusion because alkane parameters in GAFF and AMBER force fields generally reproduce data well. [26, 43]

R-alkanes appears close to its expected location. In all cases, the maximum intensity of R-alkanes along the q_r axis at $q_z=0$, which we use to estimate its location, appears at a $|\mathbf{q}|$ value that is, at most, 3.5% higher than experiment. The error in the simulated diffraction patterns is less than the error due binning in each frequency direction, so was not further investigated.

2.3.2.2 The Location and Intensity of R-pores

R-pores is about 10 times more intense than experiment in all simulated systems. Some or all of this discrepancy may be a consequence of our normalization of the experimental 2D-SAXS pattern, since the common peak between it and the 2D-WAXS patterns is partially obscured by the beamstop in the WAXS. If our simulations do indeed predict an intensity of R-pores that is too high, we hypothesize that this is primarily due to the relatively perfect infinite hexagonal array of pores in the simulated systems. In the real system, periodicity of the hexagonal array is disrupted by misalignment of the crystalline domains and the pore-to-pore distances fluctuate more over long time scales. The agreement in intensity in R-pores obtained here is sufficient for this study because this intensity is primarily controlled by longer-range organization that cannot be captured by simulation of only 4 pores and because we are only concerned with the structure of the individual pores themselves.

The location of the leading peak of R-pores is directly related to the average distance between pores. However, there is substantial uncertainty in the exact location relative to the bin size resulting from the Fourier space transformation. We can instead measure the average distance between columns with more precision in real space (See Section 2.2.6.2). We showed that we achieve experimentally consistent pore spacings in our simulations in Section 2.3.1.

2.3.2.3 The Origin of R-spots

The intensity of R-spots is close to experiment when generated from any of our simulated systems. We measured its intensity as outlined in Section 2.2.6.3.

In order to more clearly identify the structural reasons for R-spots, we equilibrated an ordered basin sandwiched configuration at 280 K to moderately increase the order and therefore intensity. Many force fields do not obtain the correct equilibrium structure at precisely the experimental temperature. Wang et al. highlighted the shortcoming of some of the most popular protein force fields in predicting the temperature dependence of protein structural ensembles [44]. It is therefore possible that our membrane system, simulated at 300 K, is under-structured in the tail region compared to the experimental structure at 300 K. We found that lowering the temperature indeed better resolved R-spots (Figure 2.7a), suggesting that our ‘effective’ temperature might indeed be too high, understabilizing structure.

Previous literature has attributed the R-spots reflection in this particular WAXS dataset as the result of tilted alkane chains [17]. This explanation is not unfounded as there are examples in literature of systems where tilted liquid crystals give rise to reflections that resemble R-spots [45, 46]. We measured the tilt angle of the alkane chains of the 280 K system by measuring the angle made by the vector extending from top to bottom of each tail with respect to the membrane plane. We found that it equilibrates to an average tilt angle of $-2 \pm 13^\circ$ (Figure 2.7b), far from the 37° tilt angle previously used to explain R-spots. Even when we placed position restraints on the monomers in a tilted initial configuration, monomers quickly reduced their tilt relative to the xy plane to an angle statistically indistinguishable from zero (see Section A.17 of the SI).

The evidence from simulations strongly suggest that the source of the R-spots reflection is packing of the tails in a hexagonal array. First, we demonstrate R-spots comes primarily from the tails, by removing all non-tail atoms from the trajectory and simulating the XRD pattern with the remaining atoms, which preserves R-spots (Figure 2.7c). Next, we plotted the center of masses of the first four tail atoms of all tails (Figure 2.7d). We measured the angle between each COM

and its nearest neighbor COMs with respect to the xy plane of the membrane. We see distinct peaks in the distribution of these angles located ca. -60° , 0° , and 60° , which is consistent with a hexagonally-packed configuration (Figure 2.7e). This ordering is primarily in the parts of the tail proximal to the heads, and dies off at the tail ends, where there is more space to fill causing them to pack nearly isotropically. See Section A.18 of the SI for a more detailed explanation of the calculation and packing distributions generated from different sections of the tails.

The peaks in the nearest neighbor angle distribution due to the hexagonal packing are consistent with the location of R-spots. The 2D-Fourier transform of a simple hexagonal array constructed based on the peak angles in Figure 2.7e shows reflections in the same locations as R-spots, in addition to vertical stacking reflections along the q_z axis that would intersect with R- π (Figure 2.7f).

2.3.2.4 The Position, Shape, and Intensity of R- π

The position, shape, and intensity of R- π , generated from simulations at 300 K have qualitative and quantitative differences from experiment. The reflections appear at lower q_z values, they are more intense, and the shape of its cross-sections, especially in the q_r direction are different relative to the experimental system. For comparison, see Figure 2.8 where we used cross-sections of the simulated XRD generated from the ordered parallel displaced configuration as an example. In this section, we explore the various contributions to these discrepancies, and use this information to speculate on what reasonable differences in the simulations could explain them.

2.3.2.5 Distance and Correlation Between Stacked Monomer Head Groups

R- π appears at a lower q_z value in our simulations versus experiment because monomers in the simulated system stack further apart than those in the experimental system (Table 2.1). We calculated z -direction pair distribution functions, $g(z)$, as described in Section 2.2.6.4. The resulting distributions are generally characterized by decaying oscillatory behavior where the average distance between peaks corresponds to the average distance between stacked monomer head groups (Figure A.24). We show that the system size in the z -direction does not significantly alter $g(z)$ in

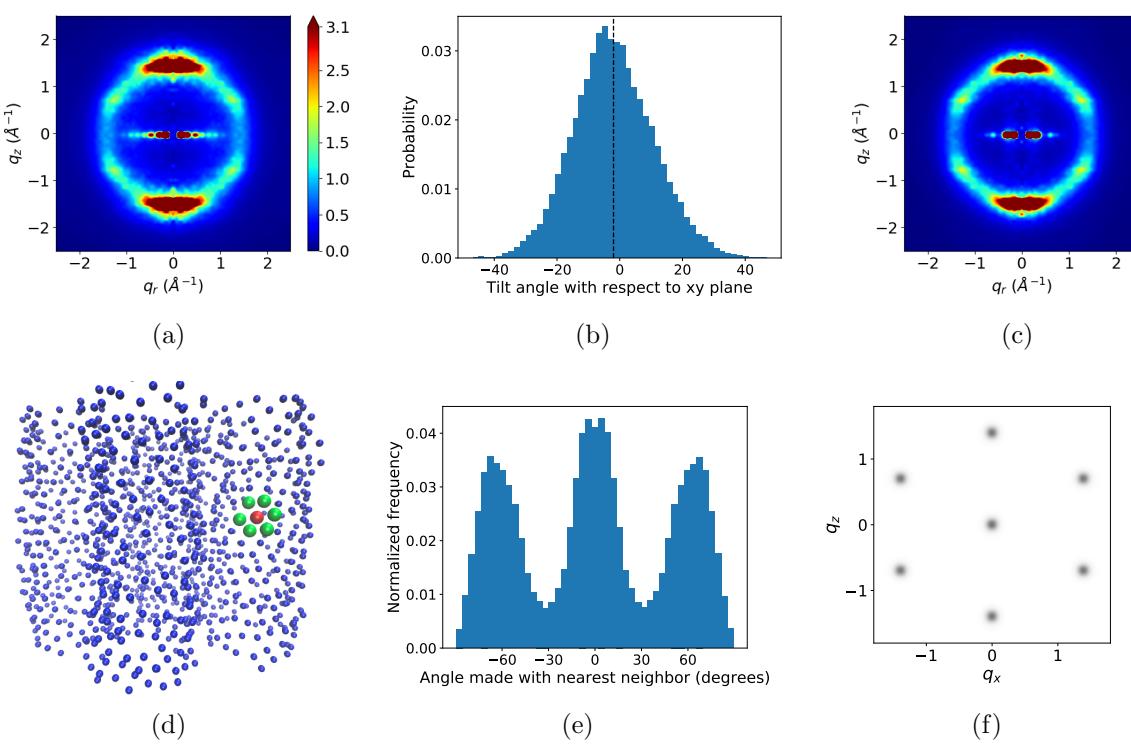


Figure 2.7: (a) R-spots increases in intensity when the temperature of the system is lowered to 280 K. (b) We measured the average angle made between each monomer alkane tail and the membrane plane. The average tilt angle (dashed line) is near -2° which is far from the 37° tilt angle previously used to explain R-spots. (c) To isolate the main cause of R-spots, we removed all atoms from the trajectory except for carbon atoms that constitute the tails. The simulated XRD pattern of the tails-only trajectory still shows R-spots. (d) Since the tails stay relatively flat, we plotted the center of mass of the first four carbon atoms of each tail originating from the head groups (for example, green colored centroids in the plot surround the red centroid in hexagonal fashion). Visually, the packing looks hexagonal. (e) We hypothesize that R-spots is the result of ordered tail packing. Defining the membrane plane to be 0° , we measured the angles between each COM and its nearest neighbor COMs for the equilibrated sandwiched configuration simulated at 280 K. Peaks appear in the distribution at -60° , 0° and 60° . (f) The Fourier transform of a hexagonally packed grid of points defined by the angles in (e) shows intensity at the same locations where we expect to see R-spots, as well as intensity along the q_z axis where R- π would appear.

Section A.19 of the SI. We calculated the equilibrated vertical stacking distance, d_{equil} , between monomers using Equation 2.3. The distance between stacked monomers is greater than experiment by 0.5–0.9 \AA across all cases, with disordered basins at the high end of that range. This behavior is not surprising since GAFF models atoms as point charges and does not appropriately model the

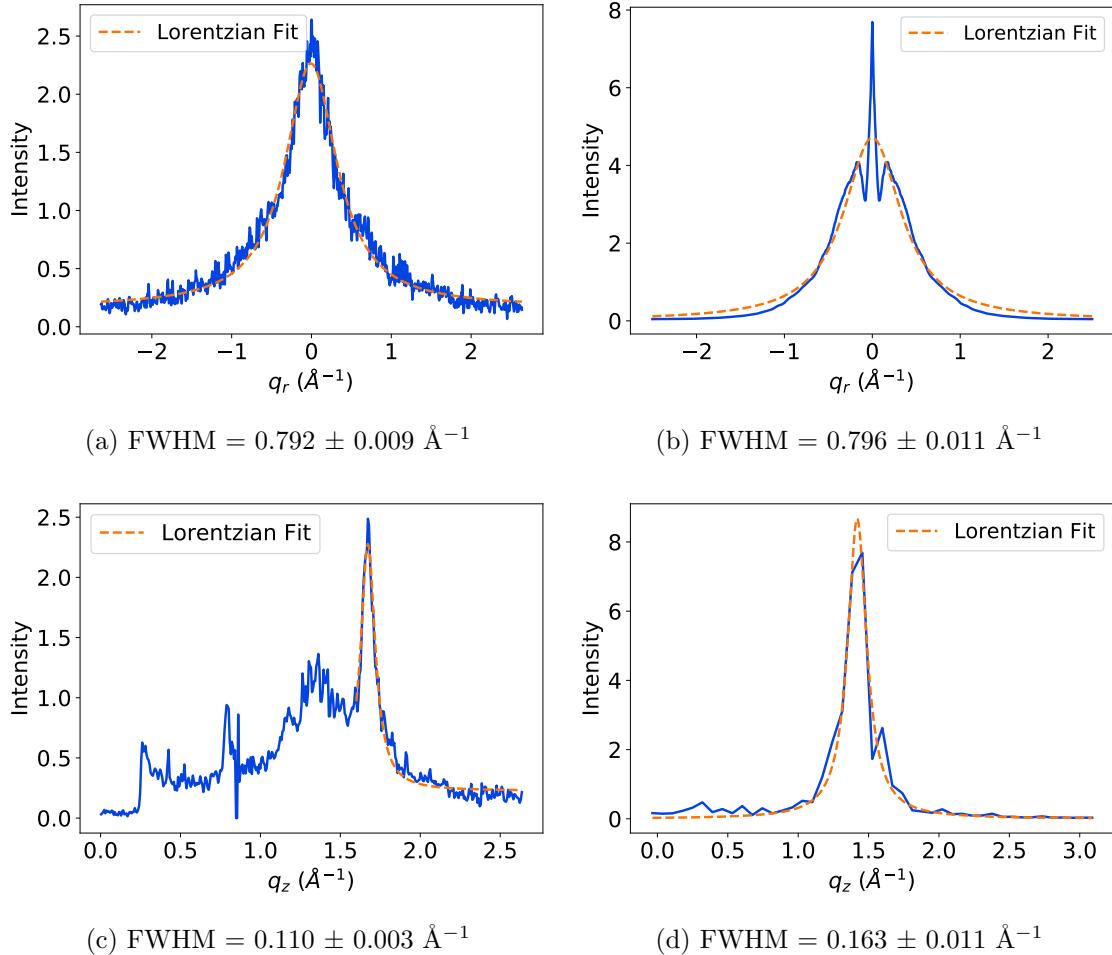


Figure 2.8: The maximum intensity of R- π generated from simulations of the ordered parallel displaced configuration is 3 times larger than experiment. The q_r cross-section of R- π ((a) and (b)) is qualitatively different between experiment (a) and simulation (b). We fit Lorentzian profiles to each peak and the FWHM (full width at half maximum) of the simulated pattern agrees with experiment within error. However the fit to the simulated data is affected by the three sharp peaks which appear near $q_r = 0$. The q_z cross-sections of R- π ((c) and (d)) are qualitatively similar. Each fits a Lorentzian profile well, however the FWHM of the simulated cross-section (d) is 48% larger than the experimental cross-section (c). Additionally, the peak of the simulated q_z cross-section is located at a lower q_z value than experiment.

aromatic $\pi - \pi$ interactions, which would make it more energetically favorable for the monomers to stack closer together [26].

The broadening of the q_z cross-section of R- π is related to z -directional correlation between scatters within monomer columns. The correlation length varies as the inverse of the full width at half maximum (FWHM) of the q_z cross-section of R- π [47]. Using this technique, we calculated the correlation length of the experimental system to be 9.0 Å. As scatterers become less correlated, we expect that R- π will broaden.

The correlation lengths of vertically stacked scatterers in our atomistic simulations are in reasonable agreement with experiment. Since it is not feasible for us to simulate membranes with much taller columns in order to obtain increased q_z resolution of our simulated XRD patterns, and because the simulated peak shape is convoluted by the extremely intense maximum of R- π , we measured correlation length by fitting Equation 2.2 to the peaks of $g(z)$. The correlation length of the parallel displaced, ordered basin system shows the closest agreement with experiment (Table 2.1). We could not extract a reliable correlation length from the disordered parallel displaced configuration since the peaks do not show clear patterns that can easily be fit to an exponential decay. The correlation length of both sandwiched systems are relatively low because the height of the first peak of $g(z)$ is so high.

Overall, we believe our simulations exhibit a realistic amount of distance correlation in the direction parallel to the pore axis between head groups within each column. The difference in the FWHM of the q_z cross-section of R- π between experiment and simulation is less than one bin size (Figures 2.8c and 2.8d). The more significant difference between the two cross-sections is their maximum intensity which we study in more detail below.

2.3.2.6 Shape and Intensity of the q_r Cross-section of R- π

In addition to the maximum intensity of R- π being too high, the shape of its q_r cross-section in the simulated XRD is qualitatively different from experiment (Figures 2.8a and 2.8b). We attempted to fit a Lorentzian profile to the q_r cross-section and although the FWHM agrees with

experiment within error, the fit is not optimal due to the three sharp peaks that appear near $q_r = 0$.

An exact quantitative comparison between the intensity and FWHM of the experimental and simulated cross-sections of R- π is not feasible for several reasons. Experimental peaks broaden due to effects that we can not easily simulate such as finite size crystalline domains and instrumental resolution [48]. In our simulated XRD patterns, the system size imposes limitations on the resolution, which makes it difficult to reliably fit peaks. This is especially problematic when comparing the FWHM of the q_z cross-section of R- π , where the experimental FWHM (0.11 \AA^{-1}) is similar to the bin size in the q_z direction (ca. 0.07 \AA^{-1}). Additionally, we model each atom with a Gaussian sphere of electron density, which is a simplification. However, we can explore reasons for which differences in the simulation structure could change the peak shapes and intensities to be in better agreement with experiment.

We studied the shape and intensity of R- π by setting up simplified systems where we represent monomer head groups as point scatterers, as described in Section 2.2.7. The amount of quenched and thermal disorder present in the atomistic system, which we applied to the simplified systems, are given in Table 2.3. The quenched disorder, the deviation of monomers from idealized symmetric configurations, is several times greater than thermal disorder, the deviations of the monomers from their average positions during the production simulations.

When we allow simplified monomer columns to move independently in the z -direction, the intensity of R- π drops and the q_r cross-section of the simulated diffraction patterns smooths out. We can demonstrate this by creating a simplified sandwiched configuration (as described above), with varying amounts of statistical correlation between columns (see Figure 2.9). In the low independent motion limit, the COM of columns are situated at the same z -coordinate. We allow increasing independence of columns by randomly shifting each column in the z -direction according to a uniform distribution bounded by $(0, f \times d)$ where f ranges between 0 and 1 and d is the vertical distance between scatterers. When $f = 1$, the columns move independently and the q_r cross-section smooths out completely (Figure 2.9c). Any amount of dependence results in some presence of sharp peaks,

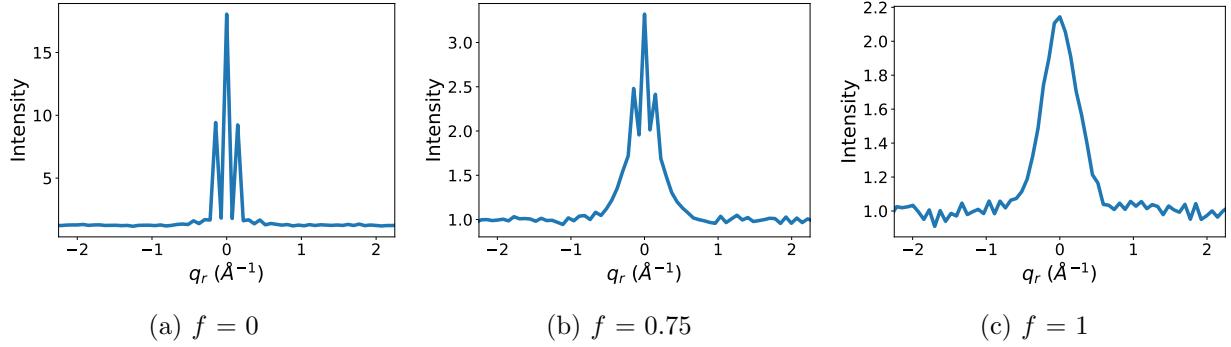


Figure 2.9: As we increase column independence using the f parameter, the intensity of R- π decreases and the q_r cross-section of the simulated diffraction pattern becomes more smooth. (a) When we place all columns at the same reference z coordinate, $f = 0$, the intensity of R- π is 18.3 and is characterized by sharp Bragg-like peaks. (b) When columns have a moderate amount of independence, $f = 0.75$, the intensity of R- π decreases 5-fold to 3.4. The edges of the peak are beginning to smooth out, but sharp peaks still exist. (c) When columns are completely independent, $f = 1$, the intensity of R- π decreases 9-fold to 2.1 and the cross-section is relatively smooth.

though significantly attenuated at f nearer 1. The intensity of R- π also decreases with increasing column independence. We see an 9-fold decrease in the intensity of R- π when $f = 1$ versus when $f = 0$ for our simplified model systems.

We also used our simplified systems in order to study the influence of increased disorder in each dimension. Increasing z -directional disorder reduces the intensity of R- π without changing the shape of its q_r cross-section. Increasing disorder on the xy plane, somewhat counterintuitively, reduces the FWHM of the q_r cross-section of R- π with an insignificant effect on its intensity. We explain these points in more depth in Section A.20.

2.3.2.7 Quenched Disorder in Ensembles of Configurations

Since the dynamics of this system are slow, consequently most of our analysis has been focused on a set of quenched configurations with some structural characteristics that do not vary significantly with time. We therefore attempted to learn more about their time-averaged structures by studying ensembles of shorter simulations started from more independent initial configurations. To do this, we created ensembles of 40 simulations in both the parallel displaced and sandwiched

configurations. We used the same equilibration scheme described in Section 2.2.5. However, we built the initial configurations with columns randomly displaced in the z -direction and looked at the results after 5 ns of unrestrained simulation, which should capture any quenched disorder built in from the initial configurations. We then simulated XRD patterns from the 40 trajectories, concatenated together. The q_r cross-sections of R- π generated from the two concatenated trajectories are shown in Figure 2.10.

Qualitatively, the q_r cross-sections of R- π generated from ensembles of simulations are nearly smooth, with only small spikes near $q_r=0$. These data suggest that the average intensity of R- π in the ensemble of sandwiched systems is about 3× higher than that in the ensemble of parallel displaced configurations. Compared to our long simulations, the intensity of R- π generated from ensembles of short simulations decreases by a factor of 3 and by a factor of 2 for the sandwiched and parallel displaced configurations respectively.

Overall, the intensity of R- π generated from the parallel displaced ensemble of short simulations shows the best agreement with experiment so far. For simplicity, we will limit the following discussion to the parallel displaced configuration. Data for the sandwiched configuration is presented in Section A.21.

These short independent simulations exhibit a symmetric distribution of quenched disorder providing evidence that the quenched disorder is randomized in each system. The only difference between each independent initial configuration is the random z -direction displacement that we apply to each column. The quenched disorder is then a result of the initial structural collapse that is directed by the randomized velocity vectors of the system’s atoms drawn from a Maxwell-Boltzmann distribution. We pooled the deviations of each head group COM from idealized positions from all independent simulations in order to create the distributions shown in Figure 2.11a (see Section 2.2.7). The distributions appear symmetric meaning that there is an equal chance for a head group to displace in the positive or negative z , r , or θ -direction.

We verified that this same symmetry exists for individual COMs. If we consider the COM of a given head group, we expect that the distribution of its deviations from its idealized position among

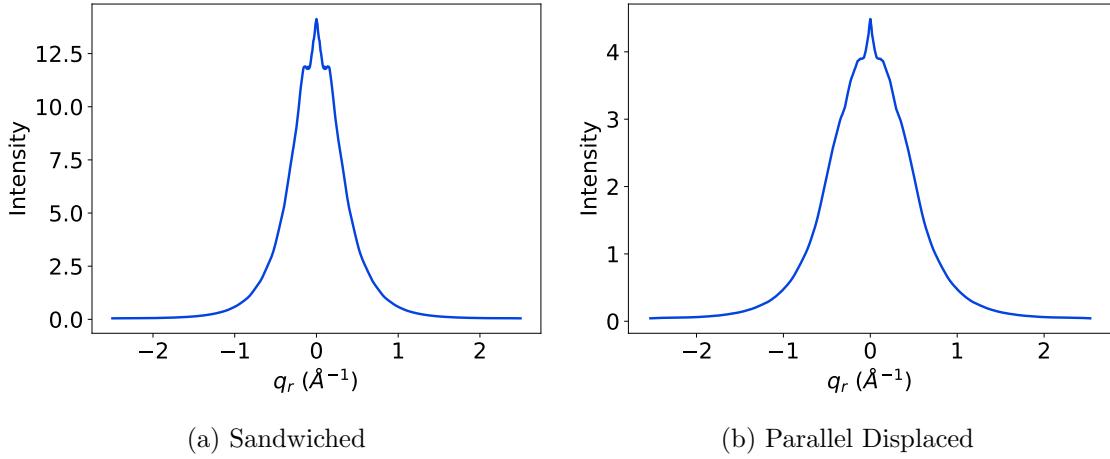
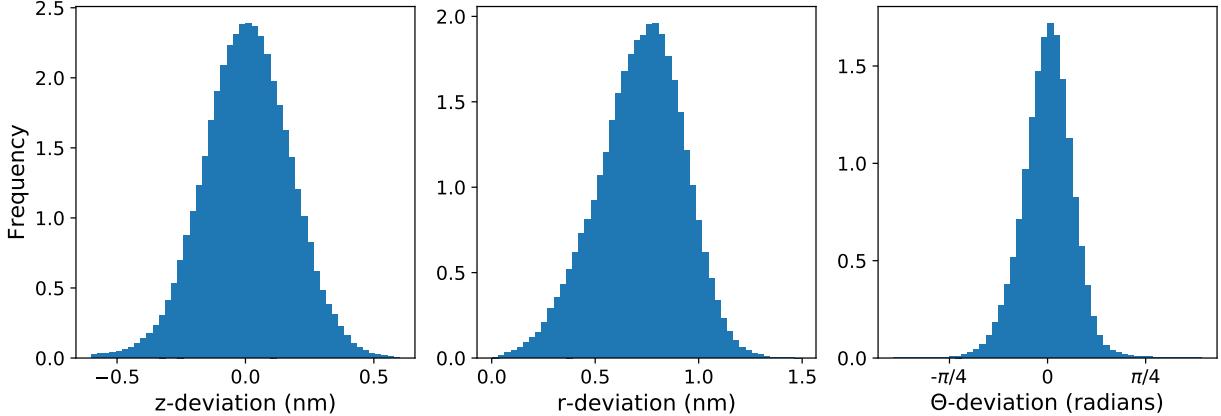


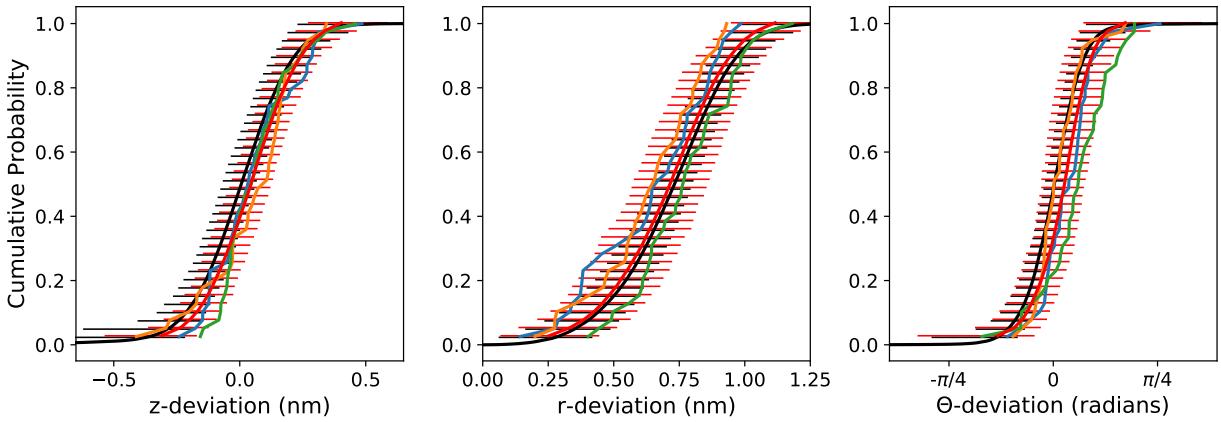
Figure 2.10: The q_r section of R- π is much smoother than Figure 2.8b when we generate simulated XRD patterns from an ensemble of 40 independent simulations, suggesting that much of the extra initial intensity may be due to the starting conditions. The maximum intensity of R- π is 3× higher in the sandwiched configuration (a) than in the parallel displaced configuration (b).

all the independent simulations should be symmetric about its idealized position. We can calculate the empirical cumulative distribution function (ECDF) of these distributions of deviations. The ECDFs of 3 randomly selected COMs are shown in Figure 2.11b to guide the eye (colored lines). The average of all COM ECDFs is represented by the red line with the 95% confidence interval represented by the red error bars. We also calculated the ECDF from the full distributions in Figure 2.11a (black line with error bars). We generated error bars that represent the 95% confidence intervals by performing 1000 trials where we randomly sampled 40 values (the same as the number of independent simulations) from the distributions in Figure 2.11a, then measured the deviation between the sampled ECDF and the full ECDF. These error bars are generally smaller than those generated from the 400 COM ECDFs, as there is a wider distribution of mean COM deviations than the mean deviations from distributions sampled from the full distribution. The similarity between entirely independently sampled ECDFs and the ECDFs from individual independent simulations supports the hypothesis that each quenched configuration is random and mostly independent.

The systems that we simulated for 400 ns and then used for most of our analysis are representative of the long term simulation behavior of any system chosen from the ensemble. The



(a) Distributions of deviations from ideal positions generated with data from all independent simulations



(b) Empirical cumulative distribution functions generated from (a) and from simulation ensembles.

Figure 2.11: (a) The distributions of the head group COM deviations from their idealized positions generated using pooled data from all frames of each independent configuration are symmetric, implying that there is equal probability for a head group to displace in the positive or negative z , r or θ -direction. (b) The COM position of a given head group is displaced randomly upon quenching. The ECDF generated from the pooled distributions (black) in (a) agree with the means of the ECDFs generated from each COM (red). The red error bars are larger than the black error bars since there is a wider distribution of mean COM deviations than the mean deviations of distributions sampled from the full distribution.

distributions of standard deviations generated from each of the ensemble trajectories (Figures 2.12a and 2.12b) are fairly tight with $(\sigma_{\sigma_z}, \sigma_{\sigma_r}, \sigma_{\sigma_\theta}) = (0.011 \text{ nm}, 0.007 \text{ nm}, 0.022 \text{ radians})$. We analyzed the first 5 ns of the 400 ns ordered parallel displaced configuration and observed the same amount of disorder (based on the σ values in Figure 2.12a) with indistinguishable mean values. Additionally,

we calculated the radial density of monomer components as a function of their distance from the pore centers (Figure 2.12c). All simulations have similar profiles. The distribution of sodium ions near the pore center is relatively noisy since the sodium ions have much more translational freedom than the head groups.

2.3.2.8 The Implications of the New-found Understanding of R- π

We can explain the discrepancies between R- π seen experimentally and in simulation using what we have shown in this section. The stacking distance between monomer head groups is too large, likely due to failures of the force field to appropriately model $\pi - \pi$ interactions. The intensity of the simulated R- π is too high, which may be a consequence of too much short-range ordering in the z -direction in addition to high degrees of correlation between columns. Finally, the q_r cross-section of R- π is not smooth, which is also likely a consequence of columns that are too highly correlated. If we limit ourselves to standard simulation techniques, it is necessary to average data generated from ensembles of independent simulation trajectories in order to begin converging on time-averaged descriptions of the system's molecular structure.

A secondary conclusion drawn from the evidence in this section is that the architecture of each column is likely in between our sandwiched and parallel displaced configurations. The parallel displaced configuration that we have simulated in this paper is an exaggerated manifestation of the parallel displaced $\pi - \pi$ stacking mode. Not only is the intensity of R- π calculated from the ensemble of parallel displaced simulations in close agreement with experiment, but the experimental WAXS pattern shows faint off-axis features at the same q_r value as R-double which are only present in simulated patterns generated from parallel displaced configurations. The monomers may prefer to be parallel displaced, but their displacement is likely only slightly shifted, on the xy plane, from the center of mass of their vertically adjacent neighbors (see Figure A.10 of the SI). In this way, columns can more easily act independently while maintaining some features of parallel displaced structure.

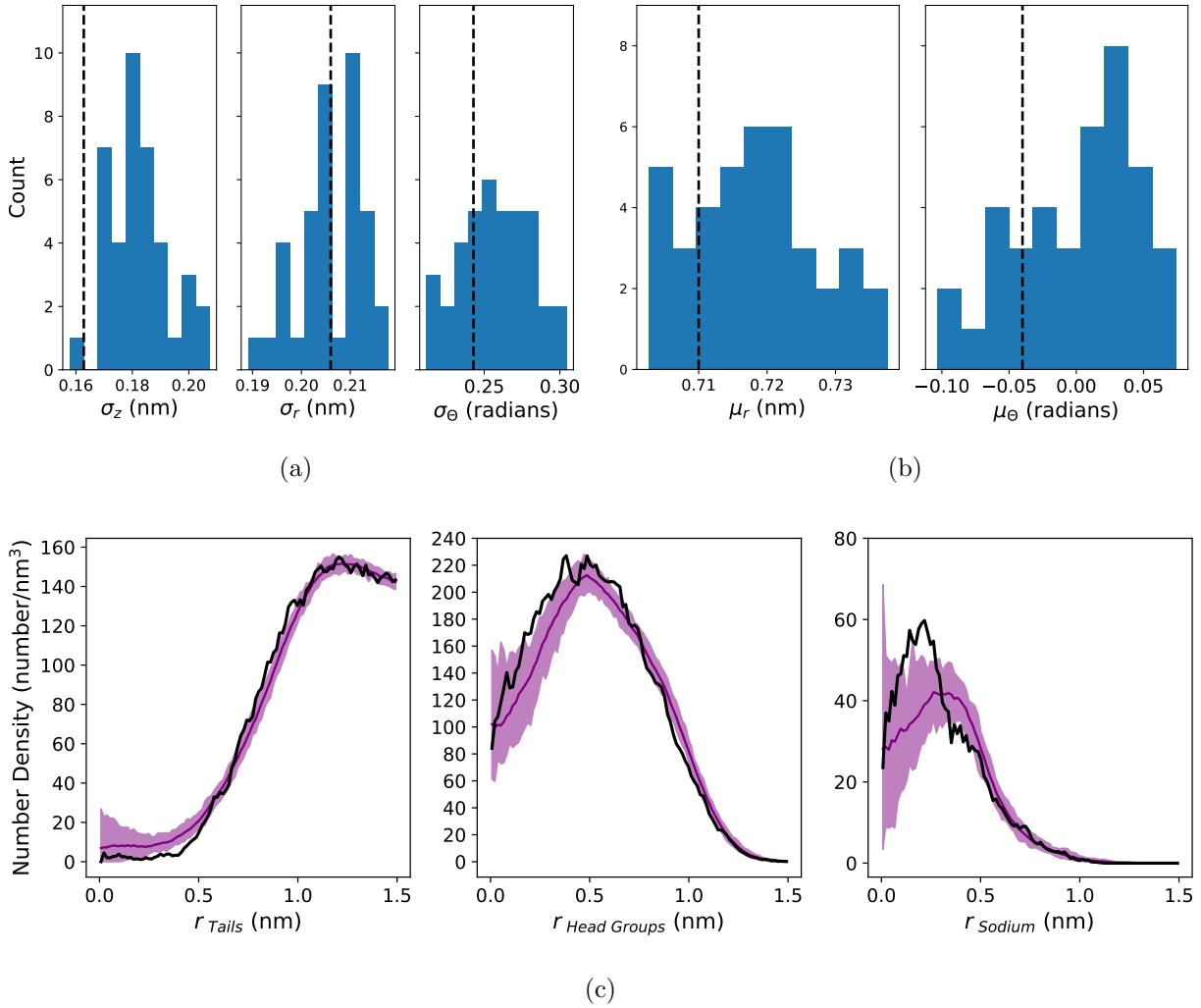


Figure 2.12: (a) The standard deviation of the distribution of quenched disorder from the first 5 ns of the main ordered parallel displaced system studied in this paper (black dashed line) is in agreement with the distribution of quenched disorder standard deviations calculated from the ensemble of simulations (histogram). (b) The mean values of r and θ from the first 5 ns of the main system trajectory (black dashed line) is in agreement with the distribution of mean values calculated from each simulation in the ensemble (histogram). The mean values of z are necessarily 0 so they are not plotted. (c) The radial densities of tail atoms, head group atoms and sodium atoms calculated from the first 5 ns of the main system simulation trajectory (black lines) and from the ensemble of trajectories (all other lines) look qualitatively similar.

2.3.2.9 Origin of R-double

R-double does not appear in any of the simulated diffraction patterns generated using the systems simulated up to this point. Here we hypothesize several configurational arrangements

which may lead to the appearance of R-double and show that we cannot achieve a long-term stable system that exhibits R-double without the inclusion of small amounts of water in the pores.

The appearance of R-double implies a vertical modulation in electron density every 7.4 Å. We are not able to achieve such modulation using our simple initial configurations. Although the position of monomers in parallel displaced configurations alternate every other layer, such a configuration will not produce R-double, but only off-axis reflections at the same q_z value [49]. There is not a unique solution that describes the origin of R-double. Extracting the exact relationship between a diffraction pattern and its real space configuration is well-known as the phase problem [50]. We have proposed configurations that result in the appearance of R-double below, and we can speculate which makes the most physical sense.

One way to produce R-double is if our initial configuration contains alternating parallel and anti-parallel carboxylate groups relative to the plane of the monomer's phenyl ring. It is difficult to physically justify this system. Systems built this way are only stable if position restraints are placed on all head group heavy atoms. Carboxylate groups quickly revert to the parallel position as restraints are released. There is an appreciable energy barrier that prevents rotation of carboxylate groups attached to phenyl rings since the group extends the system's π conjugation [51] (see Figure A.30 of the SI), which is significant even considering possible overestimates of the barrier height in the classical force field.

Another way to produce R-double is if we rotate monomers with respect to vertically adjacent monomers. In this configuration, the LLC monomers are rotated so that the vector created by the bond extending from the carboxylate carbon to the phenyl ring is oriented $\pm 15^\circ$ with respect to the vector extending from the head group COM to the pore center (Figure 2.13a). Every other monomer layer is rotated $+15^\circ$ and those in between are rotated -15° . This configuration allows monomer tails to sit between adjacent monomer tails which may be the most favorable way for them to pack. This configuration is only stable for a few nanoseconds and R-double quickly fades.

We can also produce R-double if the LLC monomers are not uniformly spaced in the z -direction, but instead are placed in pairs that stack less than 3.7 Å apart, with COMs that are

spaced 7.4 Å from neighboring pairs of monomers (Figure 2.13b). Simulations of unevenly spaced systems are only stable while position restraints are applied to heavy atoms of the phenyl rings. There is little evidence from quantum mechanical studies of stacked $\pi - \pi$ systems that such uneven stacking could be energetically stable. [52] As soon as we remove position restraints, our system immediately moves towards uniformly spaced monomers.

The addition of water to the system promotes the appearance of R-double, providing an answer to Question 3 posed in the introduction. We added water to the parallel displaced and sandwiched configurations in the ordered basin and equilibrated them according to the wet equilibration procedure. There is no experimental measurement of water concentration in these membranes so we tested a range of water concentrations from 1% to 5%. R-double appears transiently in the simulated XRD pattern of the parallel displaced configuration with 1 wt% water (Figure 2.13c). It is not initially present, but appears after 200 ns of simulation time. After 450 ns, it disappears again. Simulated XRD patterns of all other solvated systems tested are shown in Figure A.31, however R-double is not present.

R-double appears in the solvated system due to the structuring of the head groups. To demonstrate this, we removed the head groups from the trajectory used to produce Figure 2.13c in order to produce that shown in Figure 2.14a. R-double does not appear without the presence of the head groups. Water molecules must therefore play a role in the structuring of the head groups in these simulations since R-double does not appear in any dry simulations.

The addition of a small amount of water to the pores stabilizes non-uniform stacking of head groups much like that shown in Figure 2.13b. When two vertically stacked monomer head groups hydrogen bond with a shared water molecule, the monomers are drawn closer together (as illustrated in Figure 2.14b), which creates the asymmetry that allows R-double to appear. If a monomer head group shares a hydrogen-bonded water molecule with a head group above itself, it will be less likely to share a water molecule with a head group below it due to geometric constraints. The monomer head group below can just as easily share a water molecule with a head group below itself. There are a modest number of occurrences of this scenario, which we quantify in further

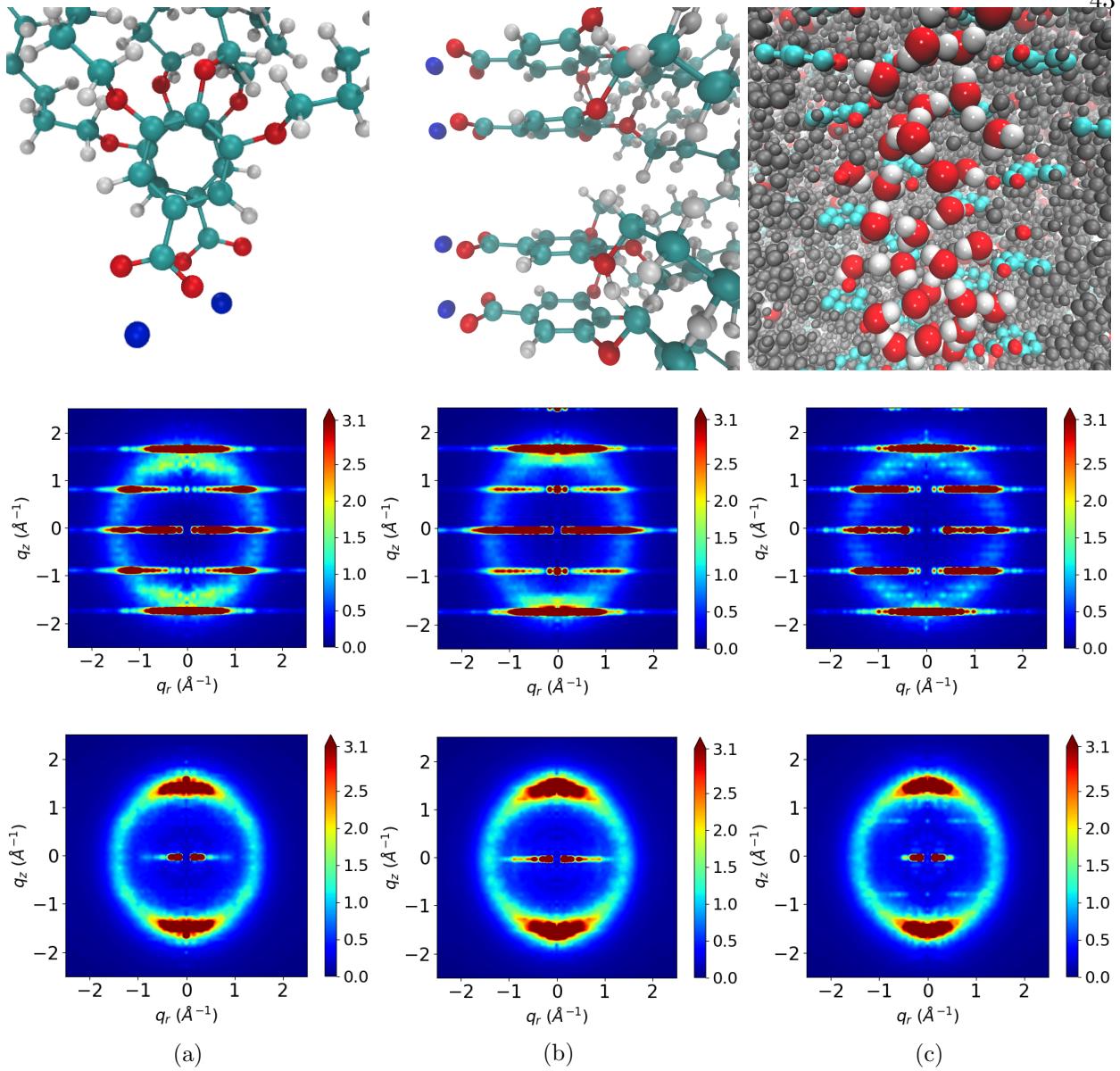


Figure 2.13: (a) When monomer head groups are rotated with respect to vertically adjacent monomers (top), R-double is visible while the heavy atoms of the head groups are held in place with position restraints (middle). R-double fades once the position restraints are released. (b) When monomers are non-uniformly spaced (top), R-double appears if all heavy atoms of the head groups are held in place with position restraints (middle). R-double quickly fades once the position restraints are released (bottom). (c) When we add 1 wt% water to the parallel displaced configuration in the ordered basin (top) R-double is not initially present during the restrained portion of equilibration (middle). After 200 ns of equilibration, R-double becomes visible and persists for another 200 ns (bottom).

detail in the SI, Section A.24.

Of all the configurations tested, it is most likely that R-double is induced by water molecules as described above since it is the only mechanism that can be observed without position restraints. The extent of the hydrogen-bonded network that forms is largely determined by the accuracy of the forcefield. It is possible that a more realistic force field would make the effect stronger or weaker. If this is truly the mechanism, it implies that the system studied by Feng et al.[17, 1] was not truly a thermotropic Col_h phase. Rather, they were very low water content H_{II} phases unintentionally created due to the neat monomers' hydroscopicity. Anecdotal evidence from other researchers suggests that membranes are less likely to assemble under completely dry conditions, supporting the idea that the “dry” structures do absorb some water. This detail may therefore be important for reproducing the results of Feng et al.

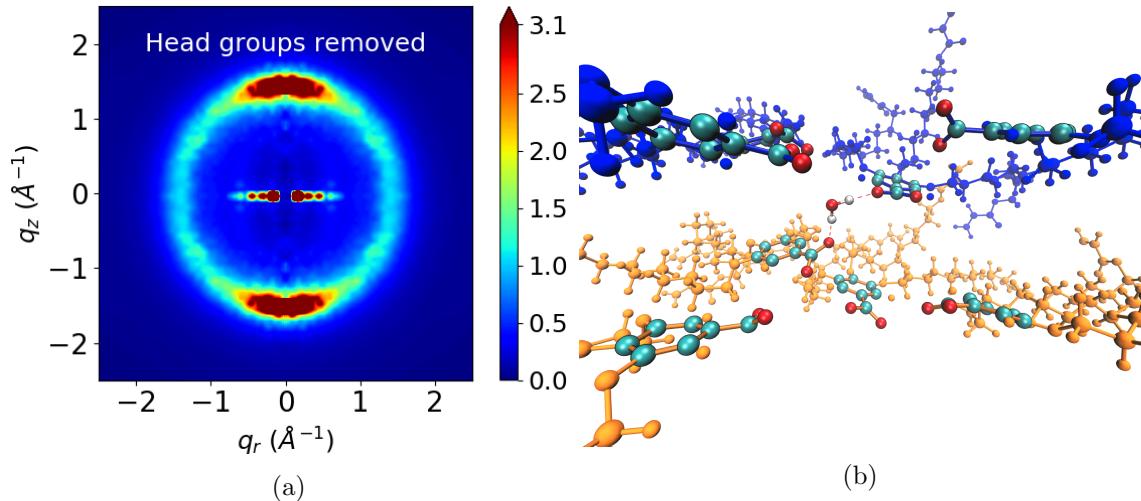


Figure 2.14: (a) The structure of the head groups is responsible for the appearance of R-double. When we remove head groups from the trajectory, the simulated diffraction pattern no longer shows R-double. (b) Monomer head groups above or below each other that hydrogen bond with a shared water molecule are drawn closer together in the z -direction. Blue monomers were stacked above orange monomers in the initial configuration.

2.3.3 Atomistic Structure of the Pore Columns

We are most interested in the structure and composition of the pores since we would like to study transport mechanisms within them. We have shown that the tails possess a certain degree

of order which is necessary in order to create the complex WAXS pattern shown experimentally, but they will not be involved in a separation process. We aim to further understand the pore architecture, in order to address Question 4 posed in the Introduction, and observe the differences, if any, between the different plausible equilibrated configurations studied so far. In general, the composition of each region, particularly within our definition of the pore region, is similar between all systems.

We plotted the number densities of heavy atoms in the head group (Figure 2.15a), carbon atoms in the tail region (Figure 2.15b) and all sodium ions (Figure 2.15c). For the head group region, we used heavy atoms making up the aromatic rings and carboxylate groups. For the tail region, we used carbon atoms of the monomer tails (See Figure A.13 of the SI for a diagram). We averaged the histograms over the equilibrated portion of the trajectory. There is a gradient in pore composition transitioning radially from the hydrophilic to the hydrophobic region, rather than an abrupt division. Based on size-exclusion experiments, the pore radius was estimated to be 0.6 nm [12]. However, the simulations do not confine sodium ions and head groups to just within this experimentally-defined pore region. For dry systems, 19% of sodium ions exist outside the pore region (except sandwiched, ordered basin, where 16% are outside the pore). Additionally, in all cases, about 3% of the plotted tail density is located within the pore region (except ordered sandwiched, where 1.5% are within the pore region). For the solvated system, the results are similar, however the head group density is shifted slightly radially outward, due to swelling of the pore by water.

The space in the pore region is filled with a mixture of sodium ions and head groups. The distributions appear somewhat different near $r=0$, but noise is higher since there is significantly less sampling as r approaches 0. Regardless, all systems, including the solvated system, have a significant number of head groups and sodium ions occupying the pore center. This observation highlights that the pore region is dense, not hollow, and may impede transport of solvent and solutes when compared to the previous idealized picture of a hollow tube conducive to transport.

Our observations suggest that the details of transport may be relatively independent of the

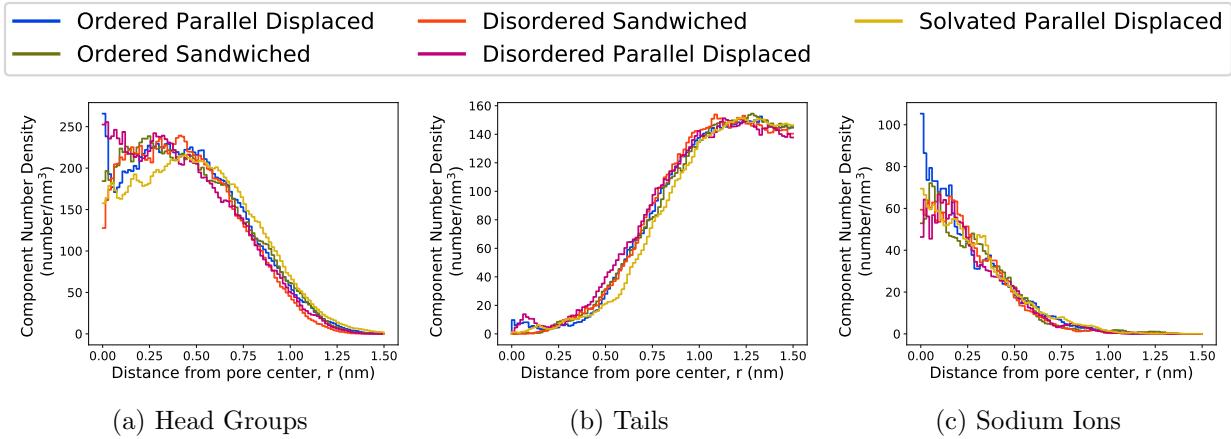


Figure 2.15: In all cases, the component radial distribution functions are similar. They exhibit a composition gradient transitioning from the hydrophilic to the hydrophobic regions. The biggest differences are at $r=0$ where noise is higher due to decreased sampling. The center of the pore is not hollow, but contains sodium ions and head groups, even when the system is solvated. This architecture may impede transport in the real system in a chemically-dependent manner. The solvated system has a lower density of head groups near the pore center which is likely due to the swelling that is necessary in order to fit water molecules in the pore region.

structural differences between different possible structures studied here. Despite the structural intricacies that give rise to differences between the various metastable basins, all pores are characterized by dense, primarily hydrophilic cores with a gradient towards a primarily hydrophobic tail region far from the pore center. This implies that we might see the same trends in transport properties from any of the systems studied so far. Since the ultimate goal of our work is to design new monomers using high throughput simulations, achieving that goal becomes tractable without the need to optimize every system with an experimental dataset.

2.3.4 Slow Dynamics

We observe slow dynamics in our system. Typical diffusion constants for columnar liquid crystals have been reported to be between 10^{-11} m²/s [53] and 10^{-14} m²/s [54]. In the case of LLCs, we would expect a diffusion coefficient at the low end of this range since hydrophilic regions of monomers can hydrogen bond, thus increasing the energetic barriers for motion. We measured

the diffusion constants of monomers in each of the main systems we studied (see Table A.4 of the SI) and find they are all on the order of 10^{-14} m²/s. However, since on this timescale, no monomers move a full ‘level’ above or below, this may be an overestimate. This slow speed is not entirely surprising considering how densely the monomers are packed as implied by diffraction experiments. The entangled tails likely restrict both vertical and lateral movement.

Consequently, there is not enough movement on the timescales we simulated for the system to consistently reach a definitive equilibrium structure. In all cases our monomers equilibrate to a stacking distance that is too large compared to experiment. As discussed in Section 2.3.2.4, while this may be in large part due to the force field’s inability to model aromatic interactions, it is also possible that the monomer tails do have enough time to pack as tightly as they could. More densely packed tails could allow the monomers to stack closer together.

We quantified the movement of the tails during our simulations by calculating the autocorrelation function of the dihedral angle formed around the bond between the head groups and the ether oxygens which attach the tails to the head group (see Figure A.35 of the SI). We exclude the dihedral from the middle tail since it is fundamentally different than the two symmetric outside tails. We limited these studies to the sandwiched configuration for simplicity.

The ether dihedrals become decorrelated on a reasonable timescale when we raise the system temperature. At 300 K (Figure A.35a), the autocorrelation function does not cross the *x*-axis until \sim 105 ns meaning that tails fully rotate on average about 4 times over the course of the 400 ns that we studied. Additionally, the correlation function plateau’s near a value of -0.2 which indicates that the tails are starting in an unfavorable configuration. We implemented distance restraints between the centers of mass of monomer head groups to preserve the hexagonal phase, then raised the temperature of the equilibrated 300 K ordered basin sandwiched system to 500 K. We witnessed decorrelation of the ether dihedrals after \sim 11 ns with a plateau at 0, (Figure A.35b) indicating a complete loss of memory. We annealed the resultant configuration back down to 300 K over 200 ns to see if the increased rotational freedom might allow the system to relax into to a more tightly packed configuration.

Decorrelating the ether dihedrals at high temperature, followed by thermal annealing does not improve packing in our model. In the ideal case, if all monomers stacked 3.7 Å apart, the *z*-dimension of our unit cell should be 7.4 nm. In the ordered basin, sandwiched system studied in this paper, the *z*-dimension of the unit cell equilibrated to 8.87 nm, which is roughly consistent with the stacking distance reported in Table 2.1. After 200 ns of annealing from 500 K to 300 K, the *z*-dimension of the unit cell was 9.22 nm. We repeated the annealing procedure over the course of 400 ns and the final *z*-dimension of the unit cell was 9.20 nm. Much longer annealing simulations may get the system to the correct density, but such simulations are beyond the scope of the current study.

2.3.5 Ionic Conductivity Measurements

We calculated the ionic conductivity of the parallel displaced configuration in the ordered basin with 1 wt% water since we believe its structure is the closest match to experiment. Our model estimates the ionic conductivity to be $(5.92 \pm 0.05) \times 10^{-5}$ S/m, about 5 times higher than the experimental value of $(1.3 \pm 0.1) \times 10^{-5}$ S/m. We verified the value calculated by the Nernst-Einstein relationship using the collective diffusion model[55] (see Section A.15 of the SI). The values calculated by the collective diffusion model agree with Nernst-Einstein values within error, however there is a much higher uncertainty that would require much longer simulations to lower.

The calculated value of ionic conductivity is 5 times higher than experiment likely because we simulated infinitely long, aligned pores and because our model over-predicts the diffusivity of sodium. The ionic conductivity measurement to which we are comparing was done with an 80 µm-thick film, nearly 10,000 times thicker than our simulated system. The thick film is likely imperfectly aligned and has defects leading to non-contiguous pores. It has been shown that there is a large dependence of ionic conductivity on the alignment of the pores. The ionic conductivity of an isotropically aligned film is ca. 85 times lower than that of the nearly aligned film to which we are comparing [17]. Additionally, sodium ions parameterized with AMBER parameters exhibit bulk water diffusion coefficients that are two times greater than experiment[56]. We cannot definitively

say how this ratio changes in the confined LLC pore environment, but it is likely that sodium diffuses faster than experiment in our system which contributes to a higher ionic conductivity.

2.3.6 Effect of Cross-linking

Experimentally, membranes are cross-linked for mechanical stability before use. We applied our cross-linking algorithm to equilibrated sandwiched and parallel displaced configurations in the ordered pore basin. We allowed the cross-linking algorithm to propagate until greater than 90% of the vinyl groups on the tails of the monomers were either involved in a cross-linking reaction or were terminated (see Section A.13 of the SI for further details of the cross-linking algorithm). We allowed the cross-linked configurations to simulate for 100 ns further in the NPT ensemble.

There are only minor changes to the physical characteristics of these systems when they are cross-linked (See Figure A.36). The ionic conductivity of the sandwiched configuration decreases while that of the parallel displaced configuration stays the same. The pore spacing decreases in both systems by 0.07 nm. The vertical monomer stacking distance increases in the sandwiched configuration and decreases in the parallel displaced configuration, however the values of cross-linked configurations fall within uncertainty of their un-cross-linked counterparts. The correlation length decreases in both systems, but is most pronounced in the parallel displaced system.

2.4 Conclusions

We have used detailed molecular modeling of the Col_h phase formed by Na-GA3C11 in order to study its nanoscopic structure. While there have been efforts to model formation of various liquid crystalline phases with molecular dynamics, to our knowledge there have been no studies which attempt to examine their structure with the same level of detail presented here.

We observed a number of metastable configurations, stable for hundreds of nanoseconds, which do not fit the experimental profile we have tried to match. We explored two classes of metastable basins which are dependent on the initial vertical stacking distance between monomers: the ordered and disordered pore basins. We expect that these metastable configurations will even-

tually rearrange and converge to a single equilibrium structure. We conducted extensive analysis in order to isolate structures which most closely resemble the true equilibrium structure.

We achieve maximum structural consistency with experiment, as determined by simulated 2D-XRD patterns, when we build our model in the ordered basin parallel displaced configuration with 5 monomer columns-per-pore and 1 wt% water. R-alkanes and R-pores appear where expected for the reasons originally predicted. We find that R-spots is likely due to ordered alkane chain packing. R- π appears at lower q_z values than experiment because monomer stack too far apart, and its intensity is far too high, likely because our initial configurations contain a high level of dependence between monomer columns. We can learn more about the time averaged structure of these systems by studying ensembles of independent simulations. Finally, we observed that our model can only produce R-double when we add small amounts of water to the system. This observation has possible implications for the reproducibility of the experimental results of Feng et al. [17, 1] since they specify that their membrane is synthesized dry.

Although exactly reproducing the experimental 2D-WAXS pattern required attention to minute structural details, we found that all systems showed similar radial distribution functions characterized by a gradual radial transition from a dense hydrophobic core to hydrophobic tails. This observation enables us to study new systems that have not been experimentally characterized with the expectation that any additional structural optimization will not greatly influence trends in transport property predictions. The compositional gradient itself raises questions about the nature of size-exclusion separations in systems without well-defined pore boundaries, which potentially could enable separations that vary with chemical identity as well as size.

Our system can reasonably estimate ionic conductivity. Our calculations are about 1 order of magnitude higher than experiment, however that is to be expected since we are simulating a perfectly straight and defect-free membrane.

Finally, we verified that our conclusions do not change when the system is cross-linked by the algorithm we implemented. The ionic conductivity drops by a factor of ~ 1.5 , in closer agreement with experiment, the pore spacing decreases and the membrane becomes thicker; however, all

changes are relatively minor, preserving most of the features well.

Future work, based on what has been learned in this study, may help further improve the structural agreement between experiment and simulation and test the sensitivity of our current conclusions to a more accurate structure. Polarizable force fields such as AMOEBA which explicitly include $\pi - \pi$ interactions between aromatic functional groups may be able to draw stacked monomers closer together [57, 58, 59]. We can study ensembles of long simulations in order to observe more averaged structure and to create more experimentally consistent XRD patterns. Extremely long simulations (on the order of μs) may be necessary to better quantify the time scales for large scale rearrangements, and the precise arrangements themselves.

With the structural understanding gained by these simulations, it will be possible to evaluate transport of various solutes within the system, and apply the knowledge gained from this study in order to suggest improvements to the existing system, as well as to evaluate new unsynthesized LLC systems.

Reflection	Normalized Reflection Intensity						
	Experiment	Parallel		Disordered		Disordered	
		Sandwiched	Displaced	Sandwiched	Parallel	Displaced	
R-alkanes	1.0	1.0	1.0	1.0	1.0	1.0	
R-pores	4.91	49.5	54.0	50.8	53.4		
R-spots	1.2	1.2	1.2	1.1	1.1		
R- π	2.8	44.0	7.7	8.4	10.1		
R-double	0.9	—	—	—	—		
Reflection Location ($ \mathbf{q} \text{ \AA}^{-1}$)							
Reflection	Experiment	Parallel		Disordered		Disordered	
		Sandwiched	Displaced	Sandwiched	Parallel	Displaced	
R-alkanes	1.39	1.44	1.44	1.42	1.43		
R-pores	0.176	0.170	0.170	0.173	0.172		
R-spots	1.39	1.44	1.44	1.42	1.43		
R- π	1.70	1.41	1.42	1.40	1.40		
R-double	0.85	—	—	—	—		

Table 2.2: The simulated XRD patterns of the systems tested, normalized so that the average intensity of R-alkanes in each pattern equals 1, show R-pores and R- π reflections that are significantly higher than experiment and R-spots reflections that are slightly lower than experiment. R-double does not appear in any simulated patterns, and thus has no measurable intensity. In terms of the locations of the reflections, R-pores, R-alkanes and R-spots appear at $|\mathbf{q}|$ values that are close to experiment, while R- π appears at a significantly lower $|\mathbf{q}|$ value than experiment.

System	Thermal Disorder			Quenched Disorder		
	σ_x (\text{\AA})	σ_y (rad)	σ_z (\text{\AA})	σ_z (\text{\AA})	σ_θ (rad)	σ_r (\text{\AA})
Ordered Sandwiched	0.31	0.33	0.34	1.48	0.43	2.30
Ordered Parallel Displaced	0.51	0.52	0.30	1.45	0.43	2.28
Disordered Sandwiched	0.41	0.57	0.32	1.58	0.43	2.93
Disordered Parallel Displaced	0.39	0.31	0.33	1.65	0.43	2.63

Table 2.3: Deviation of the positions of the center of mass of head groups from their average positions (thermal disorder) as well as their idealized positions (quenched disorder) to use as simulated disorder in our model system.

Chapter 3

Chemically Selective Transport in the H_{II} Phase

3.1 Introduction

Membranes capable of separating nm-sized solutes with high selectivity and permeability are highly desirable for a number of applications. For example, separation of salt from seawater or other briny sources can yield potable drinking water. [2] Removal of organic micropollutants, such as personal care products, pesticides and pharmaceuticals, from surface and groundwaters, can have large benefits for public health. [3] Finally, one can purify and recover dissolved species present in complex hydraulic fracturing flowback water waste streams which would help mitigate the effects of deep well injection and can be sold for profit. [5]

All of these separations are possible in part with current commercial membrane separation techniques, but all still suffer from serious drawbacks. Currently, reverse osmosis (RO) and nanofiltration (NF) dominate commercial membrane separations of small molecules. [60] RO membranes are typically dense polymer matrices that separate solutes based on differences in their solubility and permeability in the membrane material. [2] Although they can perform highly selective separations, high feed pressures and thus large amounts of energy are necessary in order to generate a useful permeate flux. [6] NF membranes have well-defined nm-sized pores which can give the same permeate flux as RO with lower applied pressure. [61] However, the pores are not uniform in size, limiting their selectivity. [4].

Under the right conditions, amphiphilic molecules can self-assemble into ordered nanostructures that may be capable of performing highly selective separations with permeabilities compara-

ble to NF. [62] The uniform-sized aqueous pores characteristic of these types of systems enforce a molecular-size cut-off, eliminating issues related to the polydispersity of NF pores. [12] In addition to size-exclusion, it may be possible to control the chemical environment within the nanopores by altering the identities of the hydrophilic moieties that form the pores, affording solute-specific separations. [20]

Amphiphilic lyotropic liquid crystals (LLCs) are known to form the type I bicontinuous cubic (Q_I) and inverted hexagonal (H_{II}) phases. [11, 5, 16, 15] When cross-linked, these materials can function as mechanically strong separations membranes. [12] The H_{II} phase has hexagonally packed, straight pores which is the ideal geometry for high throughput separations. However, the synthesis of H_{II} membranes with highly aligned hexagonal mesophases has been a challenge that has been met with limited success and has prevented the material's viability as a separations membrane. [17, 1] The Q_I phase consists of a three dimensional network of tortuous and interconnected pores. It's more facile synthesis has yielded it more success. The latest Q_I phase membranes have been shown to exhibit selectivities higher than commercial NF and RO membranes with permeabilities higher than commercial RO membranes. [5]

Although Q_I membranes have shown promising results, it is not clear how to controllably design them for specific separations beyond what has already been achieved. Dischinger et al. found that they do not separate purely on the basis of size. Empirical modeling revealed correlations between various solute physicochemical properties and observed solute rejection. Although there was some agreement with their empirical model, it does not offer a sufficiently detailed explanation of governing molecular interactions. [20]

Molecular dynamics (MD) simulations can give us mechanistic insights with atomistic resolution so that we can provide detailed descriptions of small solute transport within LLC membrane nanopores and intelligently design new membranes for solute-specific separations. In our previous work, we used MD simulations to determine the most likely structure of the hexagonal phase formed by the monomer Na-GA3C11, expanding upon the most recent experimental characterization. [63, 1] We developed techniques to simulate the hexagonal phase formed by neat monomer as

well as with varying amounts of water in the pores.

In this work, we present studies of the transport mechanisms exhibited by 20 uncharged polar solutes with varying size, chemical functionality and hydrophilic character within the pores of an H_{II} phase LLC membrane. The pores of the H_{II} phase have a similar topology to the Q_I phase but their geometry is simpler to model. In future studies, we will extend our techniques to the Q_I phase.

Figure 3.1 shows an atomistic rendering of the membrane in which we placed solutes. We have listed the solutes studied along with the abbreviations that we will use in the charts for the remainder of this paper in Table 3.1. Note that, in order limit our studies to uncharged species, we have only studied acetic acid in its protonated form. The pKa of gallic acid is 4.40 while the pKa of acetic acid is 4.75. [64, 65] Assuming that the pKa of gallic acid is similar to Na-GA3C11, it is possible for our sodium salted monomer to exist at the same time as protonated acetic acid. In any case, we are more interested in mechanisms of interaction with specific chemical functionality more than the behavior of specific molecules.

We begin our analysis by studying the structure of the membrane nanopores. Previously, we studied dry membrane systems and some systems with low water content. We showed that those systems have rather crowded pores with high concentrations of monomer head groups, sodium ions and water, where applicable. Here we build systems with appreciable amounts of water, 5 and 10% by weight, and analyze how the radial density of monomer components and water change compared to our previous results.

We then use the transport rates of water in and out of the pores in order to justify our study of solute transport restricted to the pore region. From a macroscopic perspective, it might make sense to hypothesize that water spends its time exclusively in the tube-like hydrophilic pore region. Our previous work showed that there is a gradual compositional transition from the hydrophilic to the hydrophobic region which means that solutes may not necessarily stay confined to the centers of the pores or even within the pore region. We show that, in our most viable system, water moves fastest within the pores suggesting that the area in and around the central pores should be the

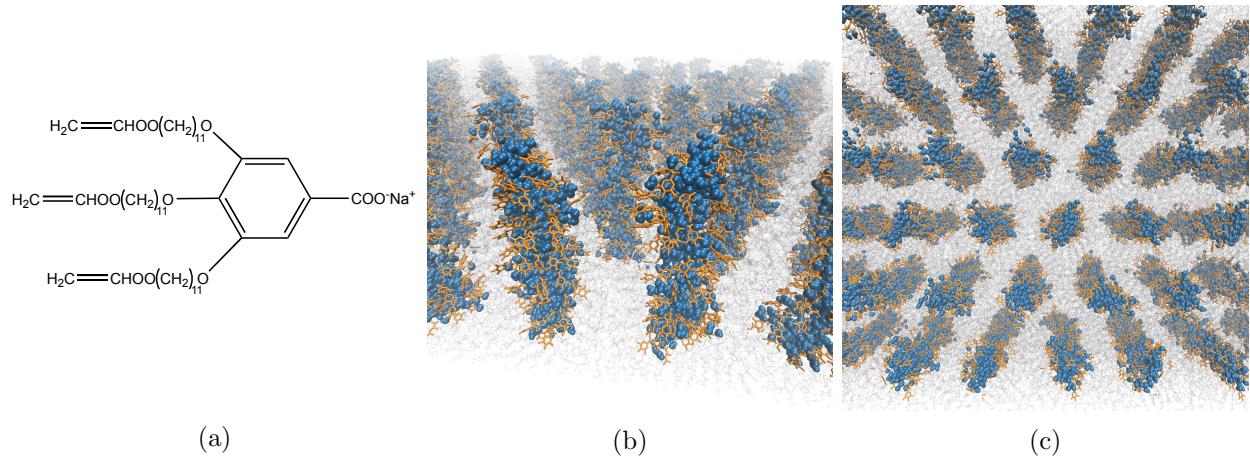


Figure 3.1: The H_{II} phase formed by LLC monomer Na-GA3C11 (a) consists of pores that are approximately straight (b) and packed hexagonally (c). In both (b) and (c), water molecules that occupy the pores are colored blue, the hydrophilic head groups of the LLC monomers are colored orange and all else is colored gray. The chemical environment within the nanopores is largely determined by the functionality of the head groups. Additionally, there is not a well-defined transition point from the aqueous pore region to the alkane-dense, hydrophobic tail region, generating further complexity.

Solute Name	Abbreviation	Structure	Solute Name	Abbreviation	Structure
methanol	MeOH		urea	URE	
ethanol	EtOH		acetamide	AcN	
propanol	PrOH		acetone	ACE	
butanol	BtOH		mercaptoethanol	ME	
ethylene glycol	EG		dimethyl sulfoxide	DMSO	
propylene glycol	PG		2,3-dimercapto-1-propanol	DMP	
glycerol	GLY		tetrahydrofuran	THF	
tetrose	TET		dimethylformamide	DMF	
ribose	RIB		propylene carbonate	PC	
acetic acid	AcOH		ethyl acetate	EAC	

Table 3.1: Names of solutes along with their molecular structures and the abbreviations which we use in this paper. Colors correspond to solute groupings and are used in many plots.

focus of our mechanistic studies.

In the remainder of the paper, we describe and quantify three trapping mechanisms which give rise to subdiffusive solute transport behavior. We describe the differences in solute motion, specifically their mean squared displacement (MSD), in conjunction with each solute's size and chemical functionality. We show how distinct interactions with various monomer components as well as membrane water content hinders the motion of solutes. We discuss how the interactions

between specific solutes and the membrane depend on the solute's chemical functionality.

There are also several questions beyond the scope of this study. First, we will not study the concentration dependence of the observed transport rates. Although the average MSD might change with concentration, we are focused on the underlying solute-membrane interactions that lead to the observed transport mechanisms which we conjecture will remain roughly the same regardless of concentration. Second, we will not study the chemical potential of solutes in the pores, which could give us a better understanding of equilibrium solute partitioning. However, this information is likely not needed to understand the principal mechanistic details in various membrane regions. This work instead represents a starting point for observing the types of interactions that occur between isolated solutes and the membrane.

3.2 Methods

Python scripts used to set up systems and conduct post-simulation trajectory analysis (as described below and in the Supporting Information) are available online at https://github.com/shirtsgroup/LLC_Membranes. The appropriate scripts to use for the subsequent calculations are summarized in Table A.5 of the Supporting Information. We ran all MD simulations and energy minimizations using GROMACS 2018.3. [31, 32, 33, 34]

3.2.1 Molecule Parameterization

We parameterized the interaction potential for the monomer and solutes using the General AMBER Force Field (GAFF) [26] with the Antechamber package [27] shipped with AmberTools16 [28]. We use GAFF as it has been used extensively with organic molecules. We assigned atomic charges using the am1bccsym method of `molcharge` included with QUACPAC from OpenEye Scientific Software.

We simulated water molecules using the TIP3P water model because it is the water model that GAFF and other AMBER family parameters have been optimized with. [66, 67]

3.2.2 System Setup

Stable H_{II} phases, assembled with monomer Na-GA3C11, can be formed using a broad range of water concentrations. In the literature, the system studied in this work is typically synthesized with close to 10 wt% water. [11, 19] However, Resel et al. noted that the system is likely fully hydrated with less than 7 wt% water. [68] Excess water fills space between hexagonal mesophases. We decided to test systems with two different water contents: 5 and 10 wt%.

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. We allowed a dry, previously equilibrated system to further equilibrate in coexistence with a 3 nm-thick (in the *z*-direction) layer of water. Although the water appears to equilibrate after 1000 ns of simulation, the density of water along the pore axis is close to zero at the membrane center, implying that there is a kinetic barrier to water filling the pores which we will be unable to overcome using conventional simulations. See Section A.29 of the Supporting Information for an expanded explanation and supporting figures.

We observed that water enters into the distal tail region of our system. We define the distal tail region to be ca. 1.5 nm from the pore center based on the minimum in the radial distribution function of water (see Section 3.2.11). Therefore we aimed to build our initial configurations with water in both regions, close to the expected equilibrium partition. The final amount of water present in the distal tail region may or may not be experimentally consistent but it is necessary for our results to be thermodynamically consistent; otherwise, water would slowly leak from the pore into the tail region over several hundred nanoseconds.

To determine the approximate ratio of water in the pores to tails, 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 of our initial configuration with water by running the GROMACS command `gmx solvate` on our initial configuration followed by the removal of any water molecules placed in the distal tail region. This method requires that we vary the monomer

starting radius since the vacant space dictates the pore's water content. We then placed a defined number of water molecules into the distal tail region one at a time with a short energy minimization between each insertion. 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 A.6. We found that systems started with more water in the pores than tails tended to equilibrate relatively quickly with an approximate ratio of 3 water molecules in the pores to 2 in the tails.

We removed the water reservoir and allowed the pore and tail water contents to equilibrate with 5 and 10 wt% total water with a 3:2 ratio of pore to tail water. We considered the water content equilibrated once the number of water molecules in each region plateaued. The 5 wt% system did not plateau until \sim 600 ns while the 10 wt % water system equilibrated within the first 100 ns of simulation (see Figure A.39a and Figure A.39b of the Supporting Information). The equilibrated 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. The water contents in each region does not change significantly in either case (see Figure A.40 of the Supporting Information). The ‘wet’ equilibration and cross-linking procedures follow the methods outlined in our previous work. [63]

The equilibrated unit cells have different geometries depending on water content. The average distance between pores is $4.15 +/ - 0.08$ nm and $4.41 +/ - 0.07$ nm in the 5 and 10 wt% water systems respectively. The length of the pores, accounting for tortuosity (see Section 3.2.11), is $9.03 +/ - 0.12$ nm and $8.07 +/ - 0.11$ nm in the 5 and 10 wt% water systems respectively. The approximate size of the pore region increases with water content (see Section 3.3.1), however we did not quantify the pore radius because its definition is ambiguous. As described in our previous work, there is not a hard partition between the pore and tail regions. [63] Rather, there is a gradual transition from the hydrophilic to hydrophobic region.

To study a given solute, we added 6 solute molecules, equally spaced in z , to the center of each pore of the equilibrated cross-linked configuration, for a total of 24 solute molecules. We chose 6 solutes per pore to provide a balance of a useful amount of data for generating statistics

and a low degree of interaction between solutes (see Section A.30 of the Supporting Information). At each insertion point we placed a randomly oriented solute molecule then ran a short energy minimization. We allowed the solutes to equilibrate at 300K for 5 ns using Berendsen pressure control then collected transport data over the course of 1 μ s MD simulations with pressure controlled by the Parrinello-Rahman barostat.

3.2.3 Mean Squared Displacement

We measured the time-averaged z -direction (parallel to the pores) mean squared displacement (MSD) of the centers of mass (COM) of each solute over the course of 1 μ s MD simulations using Equation 3.1:

$$\overline{z^2(\tau)} = \frac{1}{T - \tau} \int_0^{T-\tau} (z(t + \tau) - z(t))^2 dt \quad (3.1)$$

where τ is the time lag and T is the length of the trajectory [69]. The MSD, in standard treatments, grows according to Equation 3.2:

$$\langle z^2(t) \rangle = K_\alpha t^\alpha \quad (3.2)$$

where α is the anomalous exponent and K_α is the generalized diffusion coefficient. A value of $\alpha < 1$ indicates a subdiffusive process, while a value of $\alpha = 1$ and $\alpha > 1$ is characteristic of Brownian and superdiffusive motion respectively.

In practice, α corresponds to the growth of the *ensemble* MSD given by Equation 3.3 [69]:

$$\langle z^2(t) \rangle = \langle z(t) - z(0) \rangle \quad (3.3)$$

Since the ensemble MSD is calculated with respect to a reference position, it carries some dependence on its starting point. The time-averaged MSD averages over all possible time lags of a given length, effectively eliminating any initial configuration dependence and generating an increased number of observations. For ergodic systems, both types of MSDs will be equal. Since we have a small number of solutes with which to generate statistics and because we are not calculating values of α for this particular study, we will only use the time-averaged MSD.

We fixed the length of each simulated trajectory so that we could compare the total MSD between different solutes without the influence of the ageing phenomenon. In systems such as ours where solutes show hopping behavior between long periods of immobility, ageing is defined by the tendency of the average slope of an MSD curve to decrease as the length of trajectories are increased [70]. Since the maximum measured dwell time can be no longer than the total length of a simulated trajectory, longer dwell times are incorporated into the calculation as measurement time or trajectory length is increased, lowering the average MSD. Because the solute MSDs are non-linear and because of the ageing phenomenon, we did not attempt to calculate a diffusion constant as one might for a Brownian particle with a linear MSD. Instead, the reported MSD values represent the average MSD for a given solute after a 400 ns time lag. Our results are shown to be insensitive to our choice of time lag in Section A.31 of the Supporting Information.

The z -direction MSD of water in the 10 wt% system *is* linear to within statistical precision. Therefore, for the purpose of comparison, we calculated its diffusion constant by fitting a line to the linear region of the MSD curve. [71]. The diffusion constant is then equal to $m/2$ where m is the slope of the linear fit. The slope is divided by 2 (rather than 6) because we only measured particle displacement in one dimension.

3.2.4 Molecular Size Determination

In order to determine an effective radius for each solute, we divided in half the maximum pairwise distance between atoms of each solute over the course of a 2.5 ns simulation of solutes dissolved in a cubic box of water. Each box consisted of about 2100 water molecules and 6 solutes. Although there exist more involved methods for determining the hydrodynamic radius [72], we chose to use a simpler and more intuitive metric since we are only interested in observing trends in the solute MSD as a function of solute size.

3.2.5 The Stokes-Einstein Relationship

The Stokes-Einstein relationship expresses the diffusion coefficient of a hard spherical particle as a function that is inversely related to the particle's radius:

$$D = \frac{k_b T}{6\pi\eta f r} \quad (3.4)$$

where k_b and T are the Boltzmann constant and the system temperature respectively and η is the system's viscosity. Here we have also included the microfriction correction factor, f , introduced by Gierer and Wirtz [73, 74] for when solute size becomes on the order of solvent size since the solute can no longer be treated as a non-interacting hard sphere. f is defined in terms of the ratio of r_1 and r_2 , the radii of the solute and solvent molecules respectively:

$$f = \left(1.5 \frac{r_2}{r_1} + \frac{1}{1 + \frac{r_2}{r_1}} \right)^{-1} \quad (3.5)$$

For consistency with the presentation of our results, rather than calculate trends in D , we will calculate trends in solute MSDs. For Brownian particles, D is proportional to the MSD divided by time, t . Therefore, we analyze qualitative trends in solute MSDs according to:

$$MSD = \frac{k_b T t}{6\pi\eta f r} \quad (3.6)$$

We hold t constant in our simulations in order to make side-by-side comparisons.

Since we will be using Equation 3.6 for primarily qualitative observations, we made a number of simplifying assumptions so that it could easily be used for comparison between systems. First, we assume that all systems' viscosities are equal since we only make minor changes to the membrane's composition and accurately assessing the viscosity in a complex and inhomogeneous system such as ours is a challenge by itself. Second, we assume all particles are approximately spherical, neglecting the effect of solute shape on mobility. Chan et al. showed that the difference in diffusion constants of differently shaped particles, with constant molecular volume, did not differ by more than 25% with most deviating by less than 10%. [75]. Third, we assume that the ratio of solute and solvent radii as calculated using the methodology of Section 3.2.4, is equivalent to the the ratio of hydrodynamic

radii. For similarly shaped molecules, it has been shown that one can relate the hydrodynamic radius and radius of gyration (R_g) using a constant scaling factor. [76, 77, 78]. Assuming that our end-to-end distance reasonably approximates R_g for our small, relatively inflexible molecules, we believe this assumption is justified for qualitative demonstrations relevant to this study.

We fit Equation 3.6 so that it passed through the highest solute MSD. We assumed that the solute with the highest MSD exhibited the closest to Brownian behavior and therefore could set a rough boundary between subdiffusive, Brownian and superdiffusive MSDs. As a lower bound to our approximation, we also plotted the uncorrected version of Equation 3.4 ($f=1$), requiring it to converge to the same value as the corrected curve for large radii.

3.2.6 Hop Detection

In order to measure the length of hops, we first needed to detect when they occurred. We used an off-line change point detection algorithm, implemented in the python package **ruptures** [79], in order to determine at which points hops occurred in the time series of each solutes' COM. We used the 3 dimensional COM positions of each solute, rather than just the z -direction, in order to detect lateral hops. However, hop lengths are reported as their displacement in the z -direction. We reported the standard error in the average hop lengths by bootstrapping the empirical distribution of observed hop lengths.[80]

We determined the average frequency of hops by dividing the total number of hops by the total length of the simulation. We used the frequency of hops per 1000 ns as the λ parameter in a Poisson distribution. We calculated the 1σ confidence intervals of the probability distribution of hops per 1000 ns based on the corresponding Poisson cumulative distribution function and scaled them to a per-nanosecond basis.

We differentiated between the length of hops inside and outside of the pores. For 10 wt% water systems, we consider a solute to be in the pore if it is within 0.75 nm of a pore center. This radial cut-off maximizes the difference between average hop lengths in and out of the pore. See Section A.32 of the Supporting Information for further details on this optimization.

3.2.7 Time Spent in Pore Region

Using the cut-off defined above, we calculated the fraction of time that a solute spends within the pore region. Since this is a process with two possible outcomes, we calculated the standard error (SE) of our calculations based on the binomial theorem:

$$SE = \sqrt{\frac{p(1-p)}{n}} \quad (3.7)$$

where p is the probability that a solute is in the pore region and n is the sample size. Here n is the total number of transitions between each region.

3.2.8 Identification and Analysis of Hydrogen Bonds

Based on the geometric criteria proposed by Luzar and Chandler [81], we determined a hydrogen bond to exist 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°. Attempts to describe a hydrogen bond in the context of molecular simulations has yielded a number of definitions with no true consensus [82] especially since the geometry of hydrogen bonds has some dependence on the system being studied. The definition of Luzar and Chandler is easily visualized for trajectories using the `hbonds` representation of the Visual Molecular Dynamics (VMD) software package which allows us to directly check the validity of identified hydrogen bonds [83]. In Section A.33 of the Supporting Information, we show that our conclusions are insensitive to this definition within a reasonable range of distances and angles.

We determined the average percentage of solutes which actively participated in a hydrogen bond interaction with monomers each frame. Unless noted otherwise, we only counted unique solute–monomer hydrogen bond interactions, meaning solutes that hydrogen bond more than once simultaneously were classified as a single pairing event. We determined the standard error of this calculation by bootstrapping over each solute’s trajectory. For each bootstrap trial, we randomly chose 24 solutes with replacement and calculated the average active solute–monomer hydrogen bonds per frame.

3.2.9 Coordination Number

We quantified the coordination of solute constituent atoms with sodium ions. For each frame, we counted the number of coordinated molecules to a given solute atom based on a distance cut-off. Using four different methods, Rowley and Roux observed peaks in the radial distribution function for sodium coordinated with water at an O–Na distance of between 2.3 and 2.5 Å. [84] We used 2.5 Å as the distance cut-off in our calculations. We found that this approach is more useful than calculating the 3D spherical radial distribution function because it gives detailed frame-by-frame information rather than an average.

Using our procedure, we found that sodium ions in a solution of TIP3P water coordinate with an average of 3.6 water molecules. We created a 4 x 4 x 4 nm cubic box of water with the GROMACS tool, `gmx solvate`. We used `gmx genion` to replace water molecules with sodium and chloride ions in order to create a 0.1 M NaCl solution. We let the system simulate for 5 ns and reported the average number of coordinated water molecules per frame after discarding the first nanosecond of simulation.

We determined the average percentage of solutes actively coordinated to a sodium ion each frame. Our calculation procedure is analogous to that for solute–monomer hydrogen bond interactions outlined in the previous section.

3.2.10 Association Lifetimes

We quantified the length that two species stay associated via hydrogen bonding or coordination. For each unique pair, we measured the number of consecutive frames in which they stayed associated. We considered pairs that disassociated for a single time step and reformed on the next time step as a single continuous association event. We compiled the length of these events into a distribution of association lifetimes.

Hydrogen bond lifetimes appear to be distributed according to a power law or an exponential function (see Figure A.45 of the Supporting Information). A number of researchers provide evidence

that supports a power law distribution [85, 86]. However, these studies were done on extremely short timescales relative to ours, outputting positions every time step. Voloshin et al. studied hydrogen bonding on multiple timescales and observed exponential behavior on the longest timescales [87]. Due to memory limitations, we could not collect data frequently enough to provide a sufficient answer to this question so when comparing solutes, we use the 95th percentile of hydrogen bond dwell times. This places emphasis on solutes with long dwell times. We reported the standard error of this calculation by bootstrapping the distribution of dwell times.

The distribution of sodium association lifetimes appear similar to hydrogen bond life time distributions. Therefore, we reported association lifetimes in the same manner as hydrogen bond lifetimes.

3.2.11 Radial Distribution Functions

We measured the average radial distance of each solute of interest from the pore centers. We binned the radial distances and then normalized by the volume of the annulus defined by the bin edges. We generated 1σ confidence intervals by bootstrapping the RDFs of each individual solute 200 times. For each solute, we obtained an RDF averaged over the entire simulation trajectory. We randomly chose individual solute RDFs with replacement from the set of 24 and averaged each set for each bootstrap trial in order to generate a distribution from which to calculate uncertainties. This approach assumes that solute positions are uncorrelated to other solutes which we believe is valid due to the low degree of interaction between solutes described in Section 3.2.2.

Although the pores are often described as straight, they have a small degree of tortuosity which disrupts the RDF calculation. We tried to mitigate the effects of tortuosity by calculating the RDF with respect to splines that run through the pore centers. See Section A.35 of the Supporting Information for a graphical illustration. Each spline consists of 10 points, equally spaced in the z -direction, whose (x, y) coordinates are defined based on the COM of all head groups closest, in z , to the given point. When calculating RDFs, the radial distance from the pore center is based on the distance between the solute COM and the linearly interpolated (x, y) coordinates of the

pore center calculated based on the closest two spline points. Using the splines, we calculated the tortuosity 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.

In all RDF plots, we include the average RDF of the head groups as a reference. The head group RDF shown is the average of the head group RDFs for each solute system in the plot. We generated 1σ confidence intervals by bootstrapping the RDFs in the same way we did the solutes. We scaled the head group RDF so that it would be in frame next to solute RDFs.

3.3 Results and Discussion

3.3.1 Structure of Membrane Constituents

Before beginning our analysis of solute transport behavior, it is important to elucidate the topology of the membrane pores in solvated systems.

In contrast to our previous work with a dry version of this membrane, the region close to the pore center of the 5 and 10 wt% water systems is primarily filled with water and sodium ions. Figure 3.2 plots the RDF of each membrane constituent in the dry system as well as the 5 and 10 wt% water systems. In dry systems, the pore center is densely filled with sodium ions and head groups. In hydrated systems, water occupies, and is densest, at the center of each pore. The density of sodium ions is somewhat uniform in the pore center of the 5 wt% water system while it shows a maximum closer to the head groups in the 10 wt% water system. The peak density of sodium ions is not at the pore center in the 10 wt% water system likely because they are associated with the monomer's carboxylate head groups.

Pores in the 10 wt% water system are wider and less crowded by monomers than those in the 5 wt% water system. The peak head group density of 10 wt% water systems is about 0.2 nm further from the pore center than the 5 wt% water system. Based on these observations, we expect solute transport to be fastest in the 10 wt% water system.

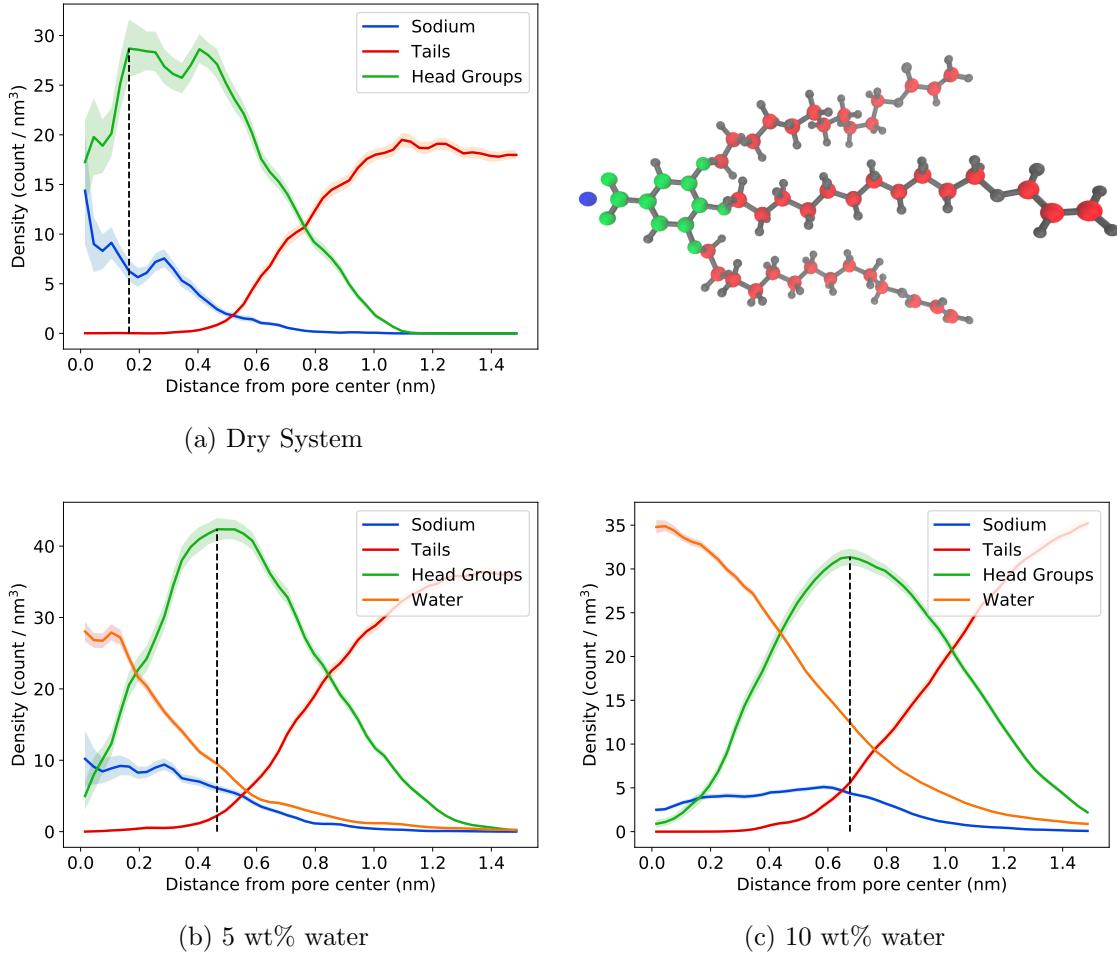


Figure 3.2: The radial densities of various monomer components paint a picture of the pore topology where the pore centers of hydrated systems are primarily, but not exclusively, composed of water and sodium ions. The monomer groups labeled in each plot correspond to the color-coded monomer pictured in the upper right corner of the figure. RDFs show the average number of atoms located at a given distance from the pore center normalized by the volume of the annular bin to which they belong. The vertical dashed black lines are positioned so that they intersect with the maximum head group density. (a) In the dry system, the density of head groups and sodium ions are highest 0.15 to 0.4 nm from the pore center. There is an appreciable density of head groups crowding the pore center. (b) In the 5 wt% system, monomer head groups retreat radially in order to make room for water molecules at the pore center. (c) Monomers in the 10 wt% system retreat an additional 0.2 nm to make room for more water.

There is an appreciable amount of water that partitions into the distal tail region of both systems. 28% and 31% of the total water is present in the distal tail regions of the 5 and 10 wt% water systems respectively. See Section A.29 of the Supporting Information for more details on the

equilibrations that led to this conclusion. The partition is due to a combination of the distal tail region's lower density (see Figure 3.3) as well as oxygen atoms at the ends of each monomer tail which can further stabilize water molecules.

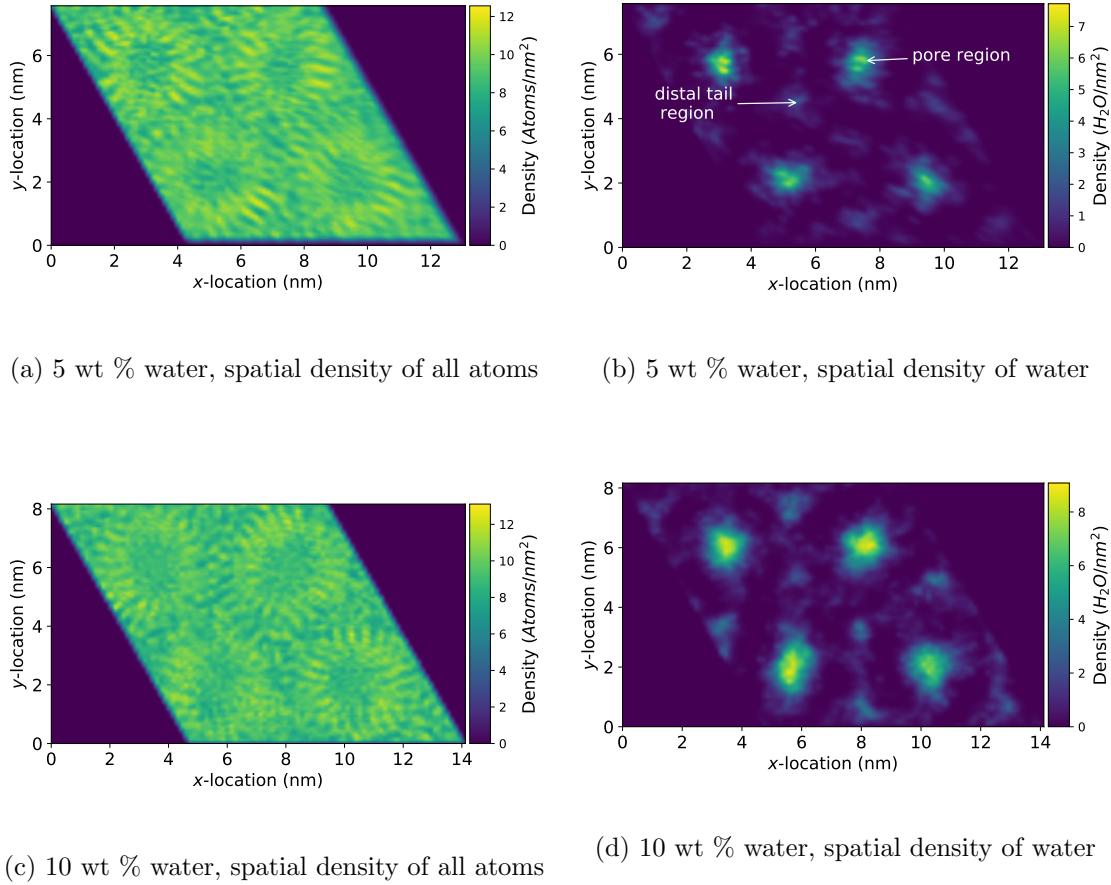


Figure 3.3: Looking down at the xy plane of the cross-linked LLC membranes, the lowest density regions of both membranes are in the pore center and in the distal tail region, furthest from the pore center (a & c). Water concentration is highest in regions of low monomer density (b & d).

We will restrict our studies of transport to solutes and water outside of the distal tail region because transport is fastest in the pores. We provide more rigorous justification for this restriction at the end of Section 3.3.2.1.

3.3.2 Mechanisms Governing Small Solute Transport

We observed transport of sodium, water and 20 other small polar solutes inside the membrane nanopores. First, we comment on transport of the membrane constituents, water and sodium, in a system absent of any additional solutes. We then present the general trends that we observe among all other solutes studied.

3.3.2.1 Water and Sodium Ions

Water and sodium's mobilities increase in larger and less crowded pores. In the 10 wt% water system, the MSD of water is about 51 times higher and the MSD of sodium is about 49 times higher compared to the 5 wt% water system. In the 10 wt% water system, water moves about 51 times faster than sodium and in the 5 wt% system, water moves about 49 times faster than sodium. Due to the linearity of water's MSD in the 10 wt% water system, we assumed long timescale Brownian behavior and measured its diffusivity to be $7.45 \times 10^{-7} \text{ cm}^2/\text{s}$ which is only 1.4% that of bulk TIP3P water. [88]

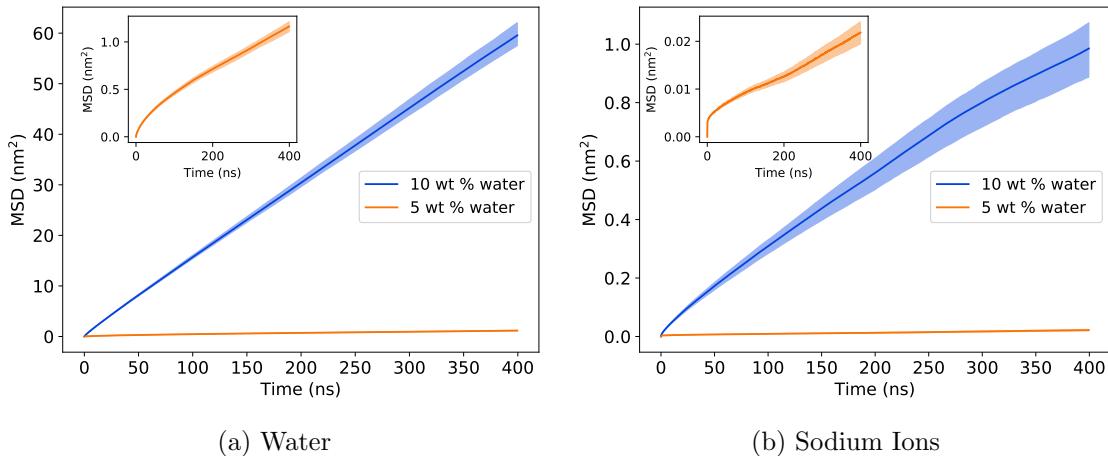


Figure 3.4: (a) The z -direction MSD of water in the 10 wt% water system is about 51 times higher than water in the 5 wt% water system. Water appears to exhibit Brownian motion in the 10 wt% system due to its linear MSD. (b) The MSD of sodium in the 10 wt% water system is about 49 times higher than sodium in the 5 wt% water system.

Sodium coordinates with both water and carboxylate moieties from the head groups in the

pores. Sodium coordinates with 3.4 water molecules in bulk solution (see Section 3.2.9), but on average in the pores, coordinates with 1.7 water molecules in the 10 wt% system and 1.2 water molecules in the 5 wt% system. The sodium ions additionally coordinate with 1 carboxylate group in the 10 wt% water system and 0.8 carboxylate groups in the 5 wt% water system. This behavior likely slows the motion of the sodium ions considerably compared to in bulk water.

In the 10 wt% water system, water molecules that spend the majority of their time in the distal tail region move significantly slower than those close to the pore center. In Figure 3.4a, we constructed the MSD curves based on water molecules that spent > 95% of their time outside the distal tail region. If we instead restrict our calculation to water molecules that spend > 95% of their time *in* the distal tail region, the MSD of water is about 8 times lower, decreasing from 59.6 to 8.2 nm².

In the 5 wt% water system, water molecules that spend the majority of their time in the distal tail region have MSDs slightly higher than those of water molecules close to the pore center. The MSD of distal tail water molecules is 1.7 nm² compared to 1.2 nm² outside the distal tails. This behavior is likely due to the much more crowded pore environment in the 5 wt% versus the 10 wt% water system. Monomer head groups are concentrated about 2 Å closer to the pore centers in the 5 wt% versus the 10 wt% water system (see Figure 3.2) which leaves less space for water molecules and restricts their movement among the relatively immobile monomer head groups. The consequently high energetic barrier to diffusion enhances the favorability of associative interactions between water and monomer components. The 95th percentile lifetime of hydrogen bonds between water molecules located near the pore center is 2.1 ns in the 5 wt% system compared to 0.5 ns in the 10 wt% system. However, since we only saved simulation output every 0.5 ns, it is likely that the true hydrogen bond lifetimes in the 10 wt% system are shorter than this value. Additionally, there are more frequent and longer periods of association between water and sodium ions. An average of 65% of water molecules in the pores of the 5 wt% water system associate with a sodium ion each frame compared to 44% in the 10 wt% water system. There are about half as many water molecules in the 5 wt% water system which allows a larger fraction of them to associate with sodium ions.

The lifetime of associations between sodium ions and water is 7.5 ns and 3.0 ns in 5 and 10 wt% water systems respectively.

Because the MSD of water in the pores of the 10 wt% system is significantly higher than that in the tails, we restrict our studies of transport in this paper to solutes in or near the pore center. Transport of solutes in the pores is likely to be the limiting factor that determines permeability and rejection. Although tail transport of water in the the 5 wt% water system is faster than pore transport, water transport overall is significantly slower than in the 10 wt% system. For practical applications, a membrane with water transport properties comparable to the 10 wt% system is far more feasible than the 5 wt% system.

3.3.2.2 Transport of Small Polar Solutes

We observe trends in transport properties that are dependent on the chemical environment within the nanopores rather than just solute size. Polar solutes are slowed by interactions between monomer functional groups and ions. A thorough understanding of these interactions will help us to create monomer design principles. We will begin our analysis by considering the collective trends observed across all systems and then focus on subsets of molecules with structural similarities.

As with water and sodium, the MSDs of the solutes studied in this work are significantly larger in the 10 wt% system than those in the 5 wt% water system (see Figures 3.5a and 3.5b). The fastest moving solute in both cases, methanol, has an MSD about 26 times larger in the 10 versus the 5 wt% water system. Clearly the equilibrium water content of a given LLC system will determine its viability for real separations.

The MSDs are not solely a function of solute size. We plotted each solute's radius against their MSD in Figures 3.5c and 3.5d. Of all the solutes, methanol is likely subject to the least hindrance by the membrane due its small size. Therefore we fit Equation 3.6 so that it intersects with methanol's average MSD and radius. The uncorrected version of Equation 3.6 ($f=1$) is plotted and converges to the same value as the corrected form for large radii (see Section 3.2.5). Although both curves are approximations, they illustrate that the majority of solutes in our study show

lower than expected MSDs. In most cases, the predicted MSDs even fall below the conservative uncorrected Stokes-Einstein estimate. It is clear that more complex mechanisms determine the MSDs of these solutes.

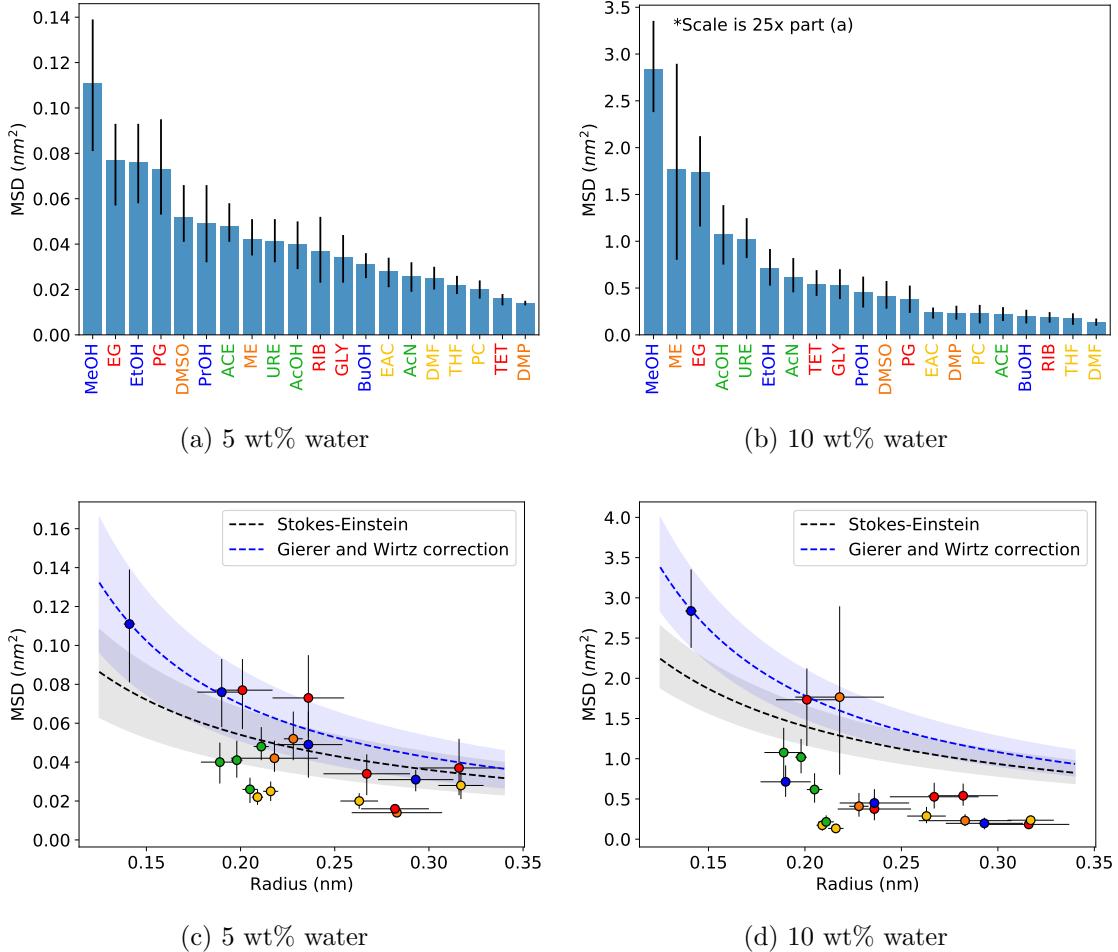


Figure 3.5: The MSDs of solutes in the 5 wt% water system (a) are significantly smaller than those of the solutes in the 10 wt% water system (b). The MSDs are not a monotonic function of molecular size (c and d). A significant number of solute MSDs fall below the theoretical lines predicted by the Stokes-Einstein equation and Gierer and Wirtz' corrected Stokes-Einstein equation.

On the timescales simulated in our study, solutes exhibit subdiffusive behavior. Figure 3.6a plots the z -coordinate versus time of 3 representative ethanol centers of mass in the 10 wt% water system. There are clear periods of entrapment separated by relatively large hops. The MSD calculated based on all ethanol molecules is plotted in Figure 3.6b and its shape is sublinear. The

long periods of entrapment lead in part to this sublinear, and thus subdiffusive, behavior.

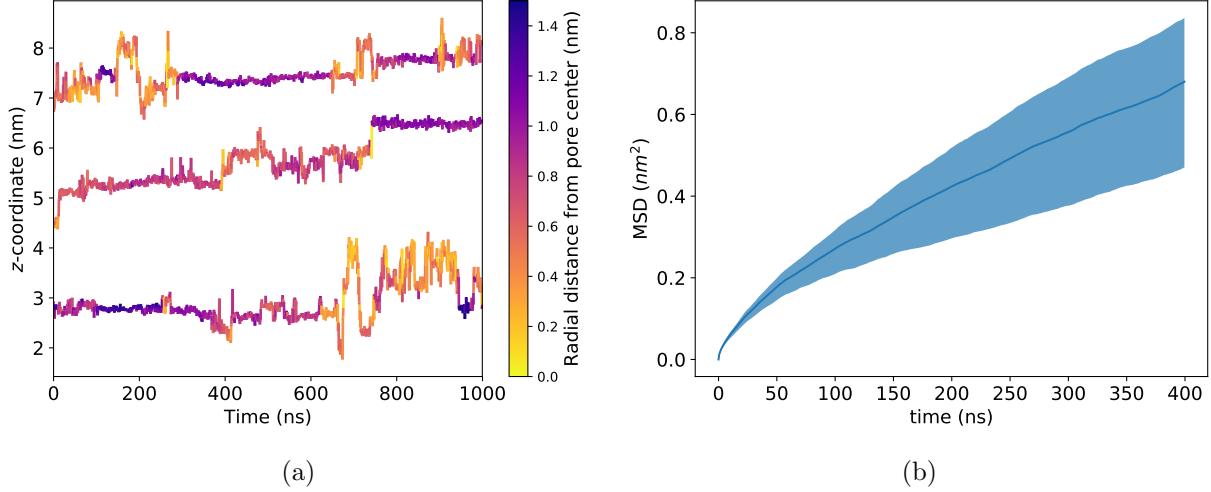


Figure 3.6: All solutes show subdiffusive transport behavior inside the membrane’s nanopores, similar to that exhibited by ethanol. (a) The z -coordinate trace of 3 representative ethanol COMs shows clear periods of entrapment separated by hops. In general, the longest dwell times occur when solutes are situated far from the pore center and more frequent hops occur when solutes are close to the pore center. (b) The time-averaged MSD of ethanol is sublinear which suggests transport is governed by an anomalous subdiffusion process.

The complex and inhomogeneous structure of the membrane leads to radially dependent transport mechanisms. In general, we observed that hops made in the pore region are about 59% larger than those made outside the pore region (see Figure 3.7a). There is a high resistance to movement in the alkane-dense head group and tail regions while, in the pore region, solutes can move relatively freely since it is primarily composed of water molecules. However, time spent in the pore region does not necessarily lead to more frequent hopping. The largest solutes in this study spend the most time in the pore region (see Figure 3.7b), but many hop with a below-average frequency (see Figure 3.7c). Therefore, trapping mechanisms controlled by membrane properties other than alkane density must be prevalent.

We observe a second trapping mechanism caused by preferential hydrogen bonding between hydrogen bond donor solutes and monomer head groups. To continue with our ethanol example, we observe that 64% of ethanol molecules donate hydrogen bonds and 3% of ethanol molecules

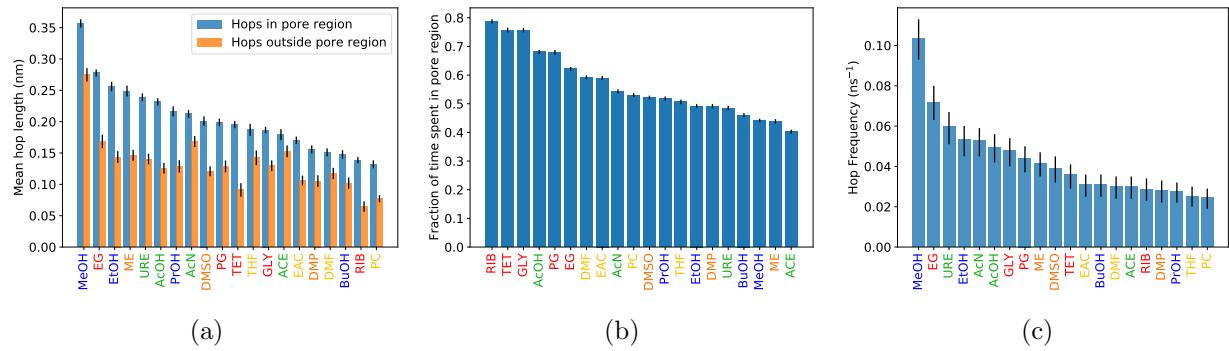


Figure 3.7: (a) Hops made in the pore region of the 10 wt% water system are, on average, 59% larger than those made outside the pore region. The trend in hop lengths is similar to the trend in MSDs shown in Figure 3.5b implying that solutes which make consistently larger hops have higher MSDs. The fraction of time spent by a solute in the pore region (b) does not necessarily lead to more frequent hopping (c). For example, ribose spends the largest fraction of time in the pore region, yet hops the fifth least frequently.

accept hydrogen bonds each frame. On average, 40% of hydrogen bonds donated by ethanol go to carboxylate head groups, 25% go to the ether linkages connecting the monomer tails to the head groups and the remaining 35% to water. There are about 46% less carboxylate oxygen atoms in the pore than there are water molecules yet more hydrogen bonds are donated to them. The stability of hydrogen bonds with carboxylate oxygen atoms is high because they have a negative charge with no neutralizing positive charges nearby. Additionally, on average, each sodium ion is coordinated to 1.7 water molecules meaning an appreciable fraction of the water molecules occupying the pore region are usually coordinated to a sodium ion which decreases their availability to accept hydrogen bonds from solutes.

The lifetime of hydrogen bonds between solutes and monomer head groups tends to be longer for solutes that hydrogen bond more frequently. In Figure 3.8 we see nearly the same ordering of the percentage of solutes hydrogen bonded to monomer head groups and the 95th percentile of hydrogen bond lifetimes. Solutes with multiple hydroxyl groups donate hydrogen bonds most frequently and tend to stay hydrogen bonded longer. Hydrogen bonds donated by nitrogen are far less common and tend to be short-lived as shown by acetamide and urea.

Finally, we observe slowing or immobilization of solutes that associate with sodium counterions. Much like water, the polarity of the solutes creates regions of high electron density, modeled using partial negative charges, which are stabilized through electrostatic interactions with sodium ions. In Figure 3.9, we've plotted the average percentage of solutes coordinated to a sodium ion each frame as well as the 95th percentile of sodium association lifetimes. The degree and length of coordination between solutes and sodium in the 5 wt% water system is higher in all cases than that in the 10 wt% water systems. The crowded pores of the 5 wt% water system force sodium ions in close proximity to solutes.

Carbonyl functional groups tend to associate with sodium the most. Nearly all of the most coordinated solutes contain a carbonyl group (except for DMSO which has an analogous sulfinyl group). There is a significant drop in sodium ion association for solutes that do not contain carbonyl groups or multiple hydroxyl groups to compensate. The corresponding association lifetimes follow a

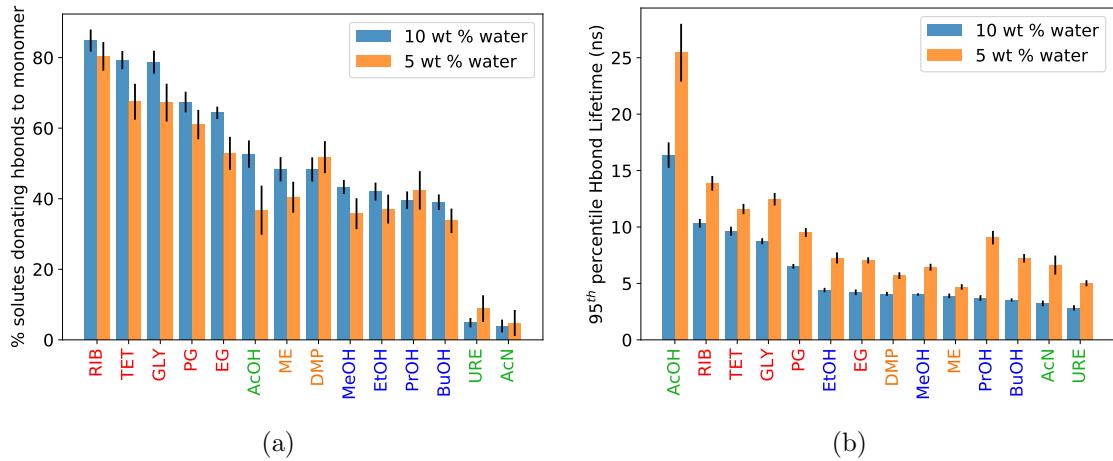


Figure 3.8: (a) Solutes capable of donating hydrogen bonds to monomer head groups do so to varying degrees. The reported percentages represent unique solute-monomer hydrogen bonds. Individual solutes that hydrogen bond with multiple head groups simultaneously are only counted once. (b) The lifetime of individual hydrogen bonds appears correlated to the percentage of solutes involved in hydrogen bond interactions. Hydrogen bond lifetimes tend to be longer for solutes that hydrogen bond frequently. Note that solutes incapable of donating hydrogen bonds are omitted from this figure.

similar trend, however the association lifetimes of highly coordinated solutes with multiple hydroxyl groups are generally lower since association between hydroxyl groups and sodium is apparently a weaker interaction. Carbonyl groups contain an exposed and highly electron-dense oxygen atom which interacts readily with sodium ions. Carbonyl groups with nitrogen substituents appear to interact with sodium more frequently than those with carbon substituents. Acetone associates with sodium significantly less than urea and acetamide.

Coordination of ions with oxygen has been observed in a variety of systems. Carvajal et al. noted that Na^+ coordinated with four THF molecules when they dissolved a sodium salt in the solvent. [89] Wu et al. observed coordination of Na^+ with the carbonyl group of a molecule used to design an organic electrode. [90] Finally, Shinoda et al. observed K^+ ions coordinated with the carbonyl and hydroxyl groups of carboxy-terminal groups in crystallized ATPase. [91]

Overall, the transport behavior exhibited by solutes in the 5 wt% water systems is similar to that shown by those in the 10 wt% system; however the timescales are much longer. We observe

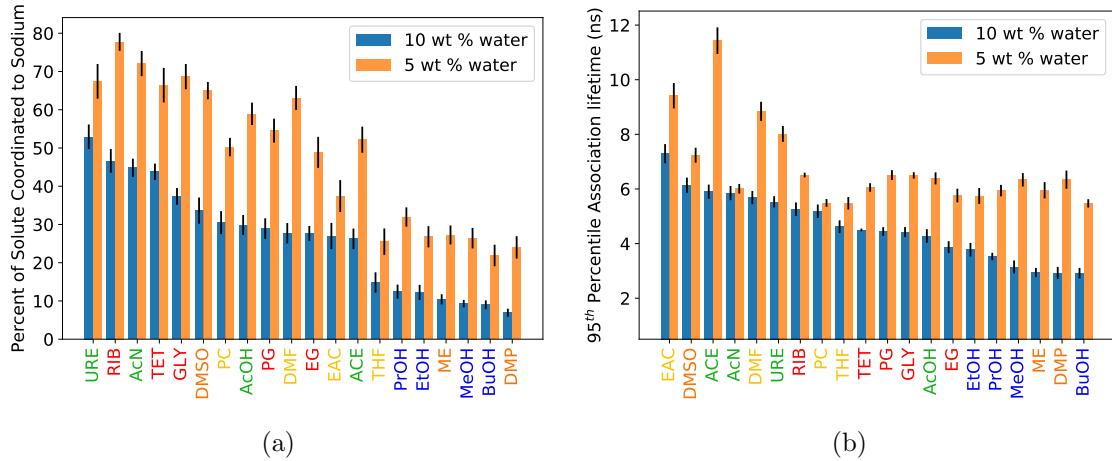


Figure 3.9: (a) Solutes, especially those with carbonyl groups, spend a significant fraction of time coordinated to sodium ions. (b) The length of time a solute-sodium pairs spends associated tends to be higher for pairs that associate more frequently.

subdiffusive behavior with intermittent hopping between periods of entrapment and evidence of the same three trapping mechanisms. The frequency and length of hops are both diminished in the 5 wt% system. Since there are only 24 solute molecules in each system, in order to obtain better time-averaged descriptions of solute transport mechanisms, we will focus the remainder of our analysis on transport in the 10 wt% water systems.

We will revisit our observations in the context of specific groups of molecules in the discussion that follows.

3.3.3 Transport of Simple Alcohols

The MSDs of methanol, ethanol, propanol and butanol descend in order of their size. Using methanol as a reference, the larger alcohols move slower than expected according to both the pure Stokes-Einstein relationship and the corrected relationship (see Figure 3.10a).

Methanol has the highest MSD of all solutes because it makes the most frequent and longest hops (see Figure 3.7). Somewhat counterintuitively, methanol, along with the other simple alcohols, spends a smaller fraction of its time in the pore region than most other solutes. However, even outside the pore region, hops made by methanol are larger than those made by almost any other

solute in the pore region. The small size of methanol allows it to move relatively unhindered.

The RDFs of longer chain alcohols show a sharp peak near the head groups (see Figure 3.10b). On average, the density of methanol in the pore center is only slightly less than its density near the head groups while all other alcohol molecules are most concentrated near the head groups.

All simple alcohols participate in a similar number of hydrogen bonding interactions with the monomer head groups, but with varying preference towards hydrogen bonds with the monomer carboxylate oxygen atoms over the ether oxygen atoms that connect the tails to the head groups (see Figure 3.10c). If all 5 hydrogen bonding acceptor sites on the monomer head groups were equal, we would expect the ratio of the number of hydrogen bonds between solutes and the two carboxylate oxygen atoms to the number of hydrogen bonds between solutes and the three ether groups to be 2/3. There is a clear preference towards hydrogen bonding with the carboxylate oxygen atoms for all simple alcohols. This is largely due to the high net charge of the carboxylate groups as well as the more highly crowded environment surrounding the ether oxygen atoms. Butanol shows the largest preference towards hydrogen bonds with carboxylate head groups. The radial distribution function of atoms located at opposite ends of butanol (Figure 3.10d) shows that, on average, oxygen atoms are situated closer to the pore centers than the most distal carbon atoms. This suggests that long chain alcohols will tend to orient themselves like the liquid crystal monomers, with hydrophilic components directed towards the pore centers.

3.3.4 Transport of Diols, Triols and Sugars

The order of the MSDs of the diols, triols and sugars are roughly consistent with their size; however, propylene glycol moves exceptionally slow (see Figure 3.11a). Ethylene glycol has the highest MSD followed by tetrose and glycerol, whose MSDs are similar, propylene glycol, the second smallest solute of this set, and finally ribose.

Transport is both facilitated and hindered by additional solute hydroxyl groups due to their influence on radial density and hydrogen bond frequency. Extra hydroxyl groups cause solutes to favor the water-rich pore region where there is the least hindrance to movement (see Figure 3.11b).

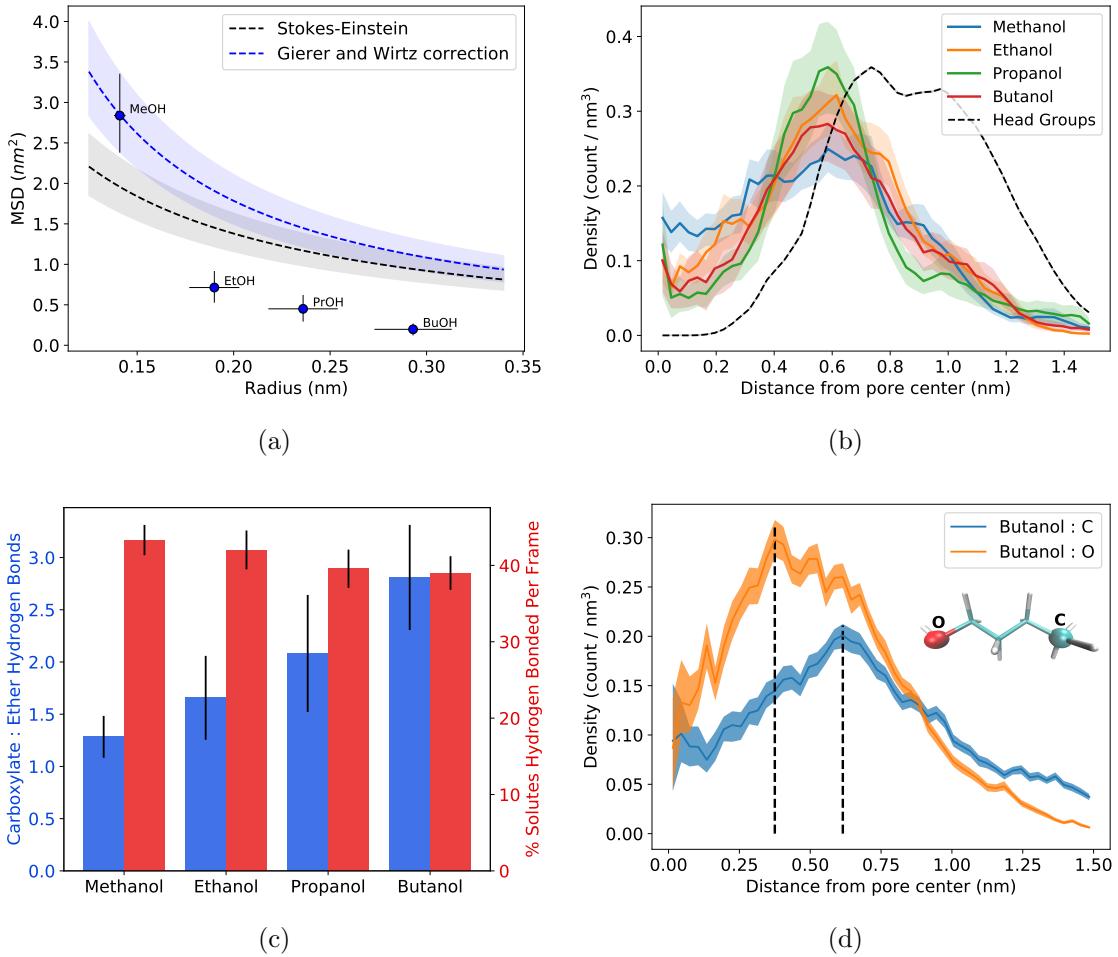


Figure 3.10: (a) The MSD of the simple alcohols decrease as a function of the solute size, however the MSDs of ethanol, propanol and butanol are considerably lower than expected based on the Stokes-Einstein equation. (b) The radial distribution functions of each simple alcohol shows a maximum close to the highest density of monomer head groups (normalized based on propanol's maximum density for easier visual comparison). Methanol spends the largest proportion of time, relative to the other alcohols, near the pore center, which may help explain its fast dynamics. (c) Despite relatively little difference in the total number of solutes actively participating in a hydrogen bond each frame, a given alcohol's preference towards hydrogen bonds with the carboxylate groups over ether linkages increases with increasing hydrophobic character. (d) The average location of butanol's oxygen atom is closer to the pore center than its most distal carbon atom, suggesting that the molecule is oriented with hydrophobic tails pointing away from the pore center.

Tetrose, ribose and glycerol are densest close to the pore center. They spend a greater fraction of their time in the pore region than any other solute (see Figure 3.7b). This is likely a consequence of both their hydrophilicity and large size which prevents them from partitioning into the head

group region. However, these extra hydroxyl groups facilitate a larger number of hydrogen bond interactions that work to hold solutes in place (see Figure 3.11c). It has been observed that hydrogen bonding in a system will generally reduce diffusivity. [92]

The number of hydrogen bonding interactions between solutes and head groups increases with the number of solute hydroxyl groups. These solutes frequently undergo simultaneous hydrogen bond interactions as shown in Figure 3.11c. For example, both hydroxyl groups of ethylene glycol can undergo hydrogen bonds with different hydrogen bond acceptors at the same time. In some cases, all 4 hydroxyl groups of ribose hydrogen bond to monomer head groups simultaneously. As a consequence, hydrogen bond lifetimes in these cases tend to be longer (see Figure 3.8b) since the solute positions are stabilized by multiple interactions. When one hydrogen bond is broken, the remaining unbroken hydrogen bonds keep the molecule in place and allow the previously broken bond to reform. Proximity to the pore center partially compensates for this effect in the cases of glycerol and tetrose, causing them to have relatively high MSDs for their size.

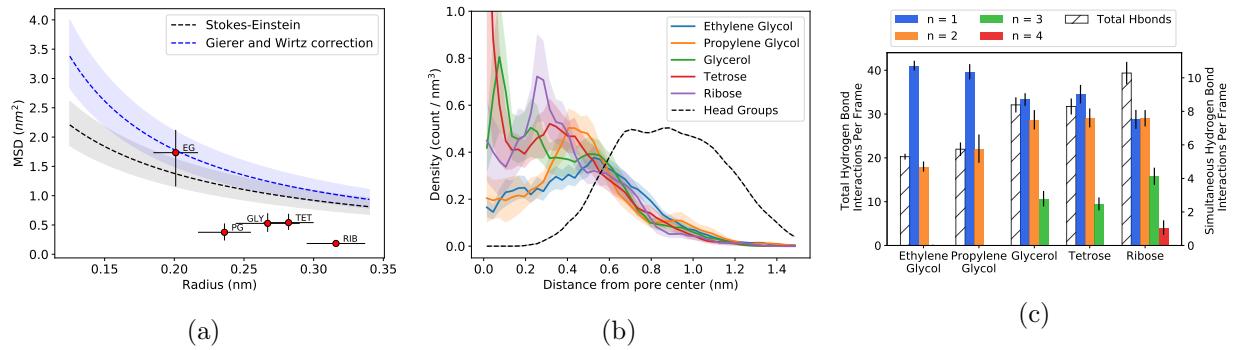


Figure 3.11: (a) The MSDs of diols, triols and sugars descend in order of their size, except for propylene glycol which moves exceptionally slow. The MSD of ethylene glycol is in close agreement with the theoretical lines, implying that the solute is subject to a similar amount of hindrance as methanol, the solute to which the theoretical lines were fit. (b) Glycerol, tetrose and ribose have high densities close to the pore center because they have a high number of hydrophilic groups and are relatively large. Ethylene glycol and propylene glycol are densest close to the head group region. (c) The number of hydrogen bond interactions between solutes and monomers increases as solutes gain additional hydroxyl groups. The number of hydrogen bonds made by a single solute in different locations simultaneously, n , also increases with the number of hydroxyl groups. In the most extreme case, all four hydroxyl groups of ribose ($n = 4$) are involved in a hydrogen bond interaction at the same time.

Of the two diols, ethylene glycol moves significantly faster than propylene glycol due to propylene glycol's affinity for the monomer head groups. Combined with an increase in size, the addition of a single methyl group to ethylene glycol increases propylene glycol's hydrophobic character and causes it to favor positions near monomer head groups (see Figure 3.11b). Both diols have comparable densities close to the pore center, however propylene glycol's density has a large peak near the monomer head groups relative to ethylene glycol. Propylene glycol can form more highly stabilized hydrogen bonds with carboxylate groups, explaining the slightly higher incidence of hydrogen bonds shown in Figure 3.11c. The 95th percentile of hydrogen bond lifetimes for propylene glycol with monomers is 9.5 ns compared to 7 ns for ethylene glycol. Somewhat counterintuitively, there is a relatively high density of ethylene glycol molecules beyond the head group region probably due to its relatively small size. This likely contributes to the somewhat large error bars on its MSD in Figure 3.5.

3.3.5 Transport of Ketones and Amides

The 4 ketone-like molecules tested show a range of transport behaviors. Urea, acetic acid, acetamide and acetone are all characterized by a carbonyl group with two attached heavy atoms. All are similar in size and are planar molecules due to the sp^2 hybridization of their carbonyl group. The fastest solutes of this grouping, acetic acid and urea, move about 3 times faster than the slowest, acetone.

The amides, urea and acetamide, hydrogen bond with head groups relatively infrequently, but regularly coordinate with sodium ions (see Figure 3.12). In an average frame, over 50% of acetic acid molecules participate in hydrogen bonds with monomer head groups while less than 10% of urea and acetamide molecules hydrogen bond with head groups. Urea and acetamide both have hydrogen bond donating nitrogen atoms, however nitrogen is a weaker hydrogen bond donor than oxygen due to its lower electronegativity. [93] Given their lower propensity to hydrogen bond, one might expect amides to partition out of the pore and/or to move through the pore quickly, perhaps faster than methanol. However, both RDFs contain sharp peaks situated between the pore

center and the head groups, but closer to the pore center than other solutes that hydrogen bond with carboxylate groups. Solutes that hydrogen bond frequently tend to show peaks in their RDFs near 0.5-0.6 nm from the pore center (see Figure 3.10b, for example) and those that coordinate with sodium ions more frequently tend to show peaks in their RDFs near 0.2-0.4 nm from the pore center. Both solutes spend about half of their time with their carbonyl oxygen atom coordinated to a sodium ion which restrains the solutes to within the pore region

Among the solutes in this set, only the carbonyl oxygen atoms coordinate with sodium ions. The nitrogen atoms do not coordinate at all despite a similar negative partial charge because the attached hydrogen atoms shield this interaction by making the NH_2 group approximately neutral.

Acetone has the lowest MSD of this set because it either coordinates with sodium or stays trapped near and behind the head groups. Acetone spends the smallest fraction of time in the pore region out of all solutes in this study (see Figure 3.7b). On average, acetone coordinates with sodium with the same frequency as acetic acid which is manifested as a peak in its RDF about 0.2 nm from the pore center. Acetic acid and the amides have other, unoccupied, hydrophilic groups while bound to sodium ions which increases their stability in the pore.

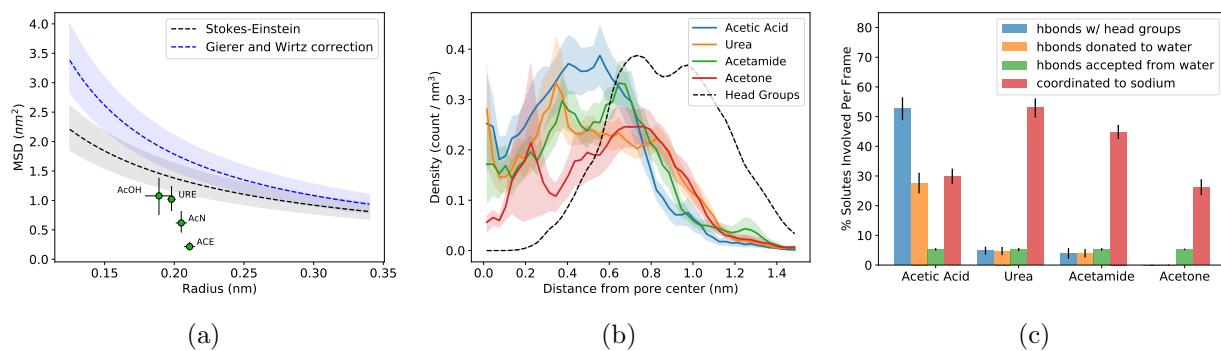


Figure 3.12: (a) The MSD of ketones and amides decreases and the deviation from Stokes-Einstein predicted behavior increases with molecular size. (b) The radial density near the pore center ($r = 0$) decreases with decreasing solute MSD. Peaks in the RDFs of urea, acetamide and acetone that appear 0.2 - 0.4 nm from the pore center are likely due to coordination with sodium ions. (c) The amides hydrogen bond with water far less than acetic acid, however they tend to coordinate with sodium ions more frequently. Acetone coordinates with sodium with same frequency as acetic acid.

3.3.6 Transport of Thiols

We also studied the transport properties of sulfur analogs of glycerol, ethylene glycol and acetone. We replaced all but one oxygen atom of ethylene glycol and glycerol with sulfur atoms to create dimercaptoethanol and 2,3-dimercapto-1-propanol. We replaced the carbonyl carbon of acetone with sulfur in order to create DMSO. Sulfur-containing compounds form weaker hydrogen bonds than nitrogen and oxygen-containing compounds due to their low electronegativity. [93] For this reason, thiols are less soluble in water than their hydroxyl group analogs.

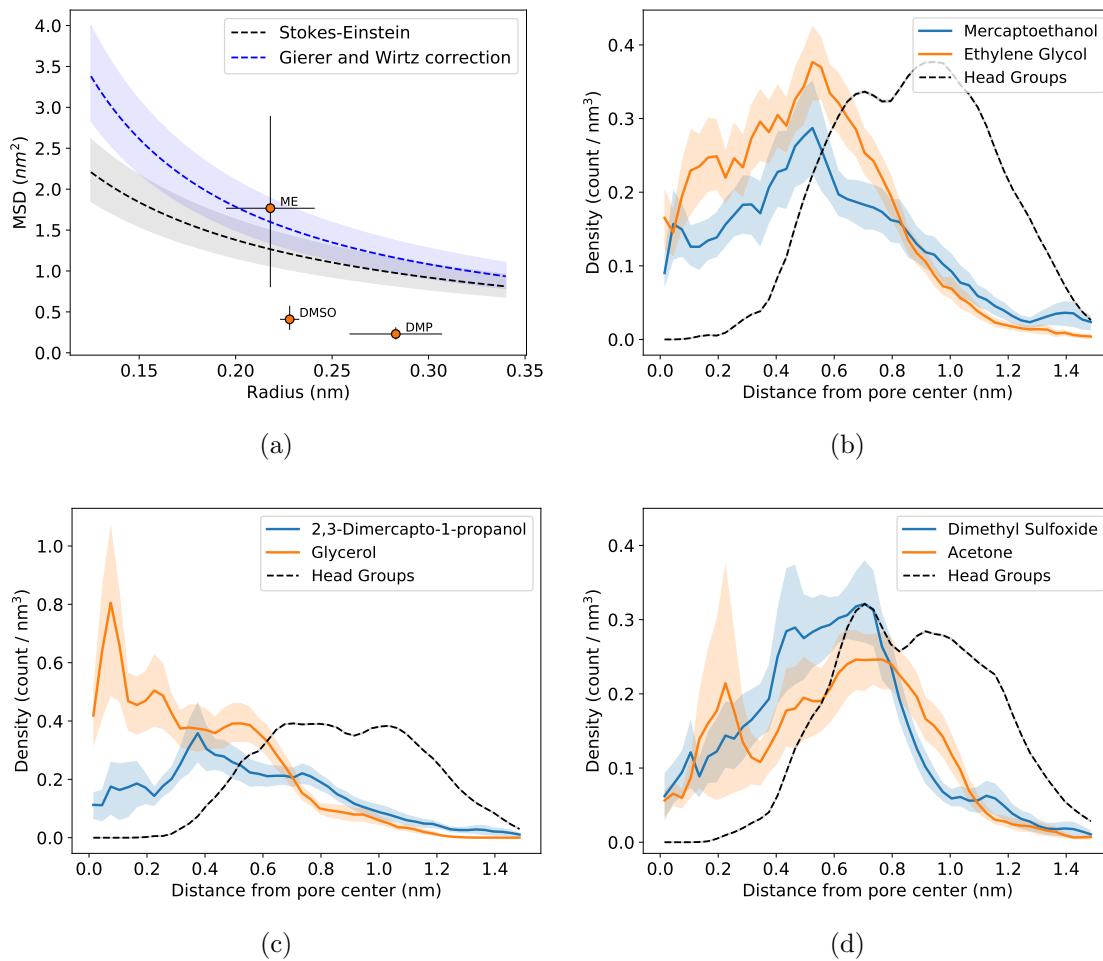


Figure 3.13: (a) The RDF of mercaptoethanol is similar to ethylene glycol except for its higher density in the tail region and consequently lower density in the pore region. (b) 2,3-dimercapto-1-propanol is densest near the head groups, unlike glycerol whose density is very high close to the pore center. (c) Overall, dimethylsulfoxide has a higher density than acetone within the pore region which may in part explain its marginally larger MSD.

Mercaptoethanol has a similar average MSD and RDF to ethylene glycol (see Figures 3.5b and 3.13b). There is a much larger uncertainty associated with mercaptoethanol's MSD. The range of behaviors shown by mercaptoethanol help explain the large variance of its MSD. Much can be accounted for by the higher density of mercaptoethanol molecules outside the pore region, where transport is inherently slower. Although both solutes hop with a similar magnitude inside and outside of the pore region, mercaptoethanol spends about 18% less time in the pore region (see Figure 3.7). This has a large impact on its hop frequency which is 42% lower than that of ethylene glycol. This implies that mercaptoethanol is relatively immobile outside of the pore, but moves quickly inside.

2,3-dimercapto-1-propanol exhibits slower transport than glycerol because more of it partitions into the tail region (see Figure 3.13c). 2,3-dimercapto-1-propanol spends 27% less time in the pore region than glycerol. Glycerol participates in about 2.5 times as many hydrogen bonds as 2,3-dimercapto-1-propanol (including all possible hydrogen bonding groups). Only 22% of the hydrogen bond interactions of 2,3-dimercapto-1-propanol involve sulfur. Because sulfur can only weakly hydrogen bond, 2,3-dimercapto-1-propanol is less soluble than glycerol in the water-filled pores and more readily partitions into the tail region.

DMSO has a comparable MSD to acetone even though it is a larger molecule (see Figure 3.5). DMSO spends 17% more time in the pore than acetone. On average, 35% of DMSO molecules are coordinated to a sodium ion each frame compared with 26% of acetone molecules. The pyramidal structure of DMSO may force it to spend more time closer to the pore center which increases its interaction with sodium ions. The tendency of DMSO to stay in the pore region counterbalances the sodium ion interactions to give it a higher MSD than acetone.

3.3.7 Solutes Unable to Donate Hydrogen Bonds

The slowest set of molecules we studied can accept hydrogen bonds, but cannot donate them. Among this set are the two slowest solutes in our study: THF and DMF. The MSDs of ethyl acetate, propylene carbonate and acetone are only marginally larger.

The solutes in this set have small hop lengths. 3 of the bottom 6 mean hop lengths are associated with ethyl acetate, propylene carbonate and dimethyl formamide (Figure 3.7a). Acetone and tetrahydrofuran perform slightly larger hops but with low hop frequencies (see Figure 3.7c).

The radial density of solutes near the pore center in this set is surprisingly high as shown in Figure 3.14b. Propylene carbonate and ethyl acetate are among the largest solutes in this study. Their size prevents them from easily entering the tail region and consequently leads to faster transport properties. However, this is not a hard rule. When a solute does overcome the barrier of entry beyond the pore region, it can become trapped. All solutes in this set show at least a small peak in the pore region which is caused by solutes that get trapped in the tail region for significant periods of time.

Carbonyl groups continue to show high degrees of association with sodium ions. Between 25 and 30% of propylene carbonate, dimethyl formamide and ethyl acetate molecules are coordinated with sodium ions for a given frame which is consistent with the coordination exhibited by acetone (see Figure 3.14c). The carbonyl group of the amides studied in the previous section associate with sodium nearly twice as frequently as compounds that don't contain nitrogen (see Figure 3.12c). Association between sodium and solutes in this set are also among the longest, only beaten by other solutes with carbonyl groups and ribose.

3.4 Conclusion

We have examined the transport characteristics of a series of small polar molecules in our model of the H_{II} phase formed by the liquid crystal monomer Na-GA3C11. We found that transport of solutes is governed by both water content as well as the interplay between the chemical functionality of the solute and the chemical functionality of the LLC monomer and ion making up the pore.

We have shown that the MSD of solutes, water and counter ions are highly dependent on LLC membrane water content. As more water is added to the system, the pores become more dilute in the monomer components. The MSDs of water and sodium ions are nearly 2 orders of

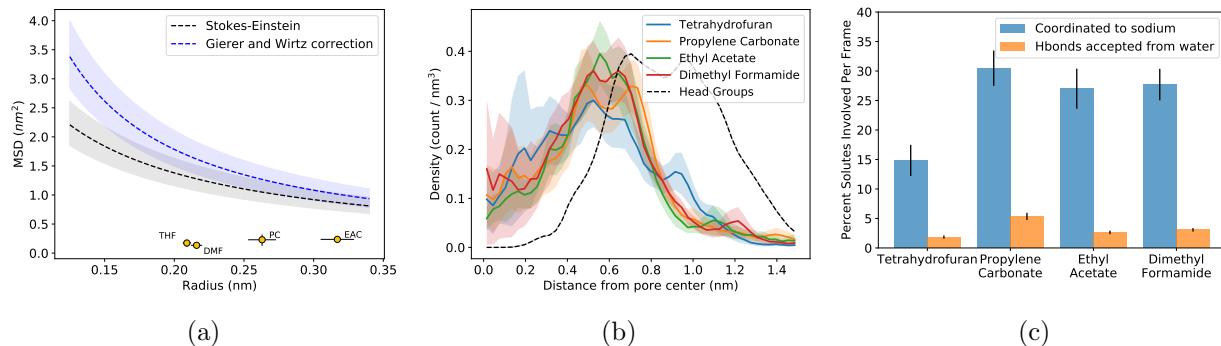


Figure 3.14: (a) The MSDs of solutes that can only receive hydrogen bonds are significantly lower than expected. (b) The solutes' radial density is surprisingly high in the pore region, but is balanced by an appreciable amount of solute trapped in the tails. (c) The low MSDs exhibited by each of these solutes is due to a combination of entrapment within the tail region and a high degree of coordination with sodium ions.

magnitude larger in the 10 wt% system than in the 5 wt% system. The amount of water in the pores thus deserves special attention when screening new monomers.

The transport of polar solutes in this system is significantly slower than that of water due to their interaction with the LLC monomers. We observed three primary mechanisms of solute entrapment within the pores.

- (1) Solutes can become entangled among monomer tails.
- (2) Solutes can donate hydrogen bonds to immobile monomers.
- (3) Solutes can associate with counterions.

The chemical functionality of solutes influences the degree to which they are affected by each trapping mechanism. Simple alcohols tend to partition into the tail region as their hydrophobic tail increases in length. Alcohols and sugars with multiple hydroxyl groups are very stable in the pore region, but multiple hydrogen bonds with monomer head groups prevent them from diffusing quickly. This effect is reduced when hydroxyl groups are replaced with thiol groups, but the solutes become less stable in the pore. Finally, solutes with carbonyl groups, which includes the ketones, amides and solutes that can only accept hydrogen bonds, associate with sodium ions, temporarily

slowing or immobilizing them.

Based on these trapping mechanisms, we may be able to suggest modifications that can be made to monomers that will modulate their effect on specific solute MSDs in order to make transport even more selective for particular solutes. Since solutes move slowly while entangled among the monomer tails, and since more than half of the solutes studied spend at least 40% of their time in the tails, one can try to design monomers that better control the partition of solutes between the pore and tail region. Solutes move into the tails due both to their size and hydrophilic character. One may be able to contain solutes, even small ones like methanol, by increasing the density of monomers close to the head groups through incorporation of cross-linkable groups in the vicinity. Alternatively, removal of the ether linkages between the head groups and the monomer tails may decrease the stability of any of the molecules we studied near the head groups and force them to spend more time within the pore.

One could instead focus on designing head groups with varying hydrogen bonding capabilities. Increasing the number of hydrogen bonding sites on the head groups may enhance the membrane's ability to trap solutes like sugars that have multiple hydroxyl groups. Conversely, one can decrease the number of hydrogen bonding sites, or replace them with weaker hydrogen bonding atoms like nitrogen, to trap less. In contrast to the system studied here, one can design the head groups so they can donate hydrogen bonds, which can enhance entrapment of solutes like acetone that can only accept hydrogen bonds.

Finally, one can attempt to control the degree to which solutes coordinate with counterions. Increasing the valence of the counterions may cause them to bind more strongly to solutes which may be a useful way to trap carbonyl-containing solutes like urea for longer periods of time. Of course, an increased valence would require proportionately less total counterions to balance the system charge which could result in the opposite effect. One could also design the head groups so that counterions are negatively charged which may eliminate their interaction with most solutes.

Our work demonstrates that there are important, atomistic features of nanostructured, self-assembled materials that molecular simulation can play a key role in elucidating. The complex

interplay between the inhomogeneous membrane structure, solute size and solute functionality leads to a variety of solute-membrane interactions, many of which are quite subtle. For example, association between polar groups and ions is seldom mentioned in the literature, but we could not give a complete mechanistic explanation of observed transport behavior without its consideration. We also observed that water and solutes partition into the tails, an area previously assumed to be devoid of solutes, which offers another route towards selectivity. The interactions we observed, along with those that may arise in other self-assembled nanoporous membranes, have the potential to produce an immense diversity of transport behavior, suggesting that a previously unexplored range of chemical selectivities may be possible. Atomistic molecular modeling is likely to be highly useful in uncovering important mechanisms that will inspire and accelerate novel design by experimentalists.

Bibliography

- [1] X. Feng, S. Nejati, M. G. Cowan, M. E. Tousley, B. R. Wiesnauer, R. D. Noble, M. Elimelech, D. L. Gin, and C. O. Osuji, “Thin Polymer Films with Continuous Vertically Aligned 1 nm Pores Fabricated by Soft Confinement,” *ACS Nano*, vol. 10, pp. 150–158, Jan. 2016.
- [2] C. Fritzmann, J. Löwenberg, T. Wintgens, and T. Melin, “State-of-the-Art of Reverse Osmosis Desalination,” *Desalination*, vol. 216, pp. 1–76, Oct. 2007.
- [3] R. P. Schwarzenbach, B. I. Escher, K. Fenner, T. B. Hofstetter, C. A. Johnson, U. v. Gunten, and B. Wehrli, “The Challenge of Micropollutants in Aquatic Systems,” *Science*, vol. 313, pp. 1072–1077, Aug. 2006.
- [4] J. R. Werber, C. O. Osuji, and M. Elimelech, “Materials for Next-Generation Desalination and Water Purification Membranes,” *Nat. Rev. Mater.*, vol. 1, p. 16018, May 2016.
- [5] S. M. Dischinger, J. Rosenblum, R. D. Noble, D. L. Gin, and K. G. Linden, “Application of a Lyotropic Liquid Crystal Nanofiltration Membrane for Hydraulic Fracturing Flowback Water: Selectivity and Implications for Treatment,” *J. Membr. Sci.*, vol. 543, pp. 319–327, Dec. 2017.
- [6] B. Van Der Bruggen, C. Vandecasteele, T. Van Gestel, W. Doyen, and R. Leysen, “A Review of Pressure-Driven Membrane Processes in Wastewater Treatment and Drinking Water Production,” *Environ. Prog.*, vol. 22, pp. 46–56, Apr. 2003.
- [7] T. Humplík, J. Lee, S. C. O’Hern, B. A. Fellman, M. A. Baig, S. F. Hassan, M. A. Atieh, F. Rahman, T. Laoui, R. Karnik, and E. N. Wang, “Nanostructured Materials for Water Desalination,” *Nanotechnology*, vol. 22, no. 29, p. 292001, 2011.
- [8] D. Cohen-Tanugi, L.-C. Lin, and J. Grossman, “Multilayer Nanoporous Graphene Membranes for Water Desalination,” *Nano Lett.*, vol. 16, pp. 1027–1033, Jan. 2016.
- [9] K. Hata, D. N. Futaba, K. Mizuno, T. Namai, M. Yumura, and S. Iijima, “Water-Assisted Highly Efficient Synthesis of Impurity-Free Single-Walled Carbon Nanotubes,” *Science*, vol. 306, pp. 1362–1364, Nov. 2004.
- [10] S. Maruyama, E. Einarsson, Y. Murakami, and T. Edamura, “Growth Process of Vertically Aligned Single-Walled Carbon Nanotubes,” *Chem. Phys. Lett.*, vol. 403, pp. 320–323, Feb. 2005.

- [11] R. C. Smith, W. M. Fischer, and D. L. Gin, “Ordered Poly(p-phenylenevinylene) Matrix Nanocomposites Via Lyotropic Liquid-Crystalline Monomers,” *J. Am. Chem. Soc.*, vol. 119, no. 17, pp. 4092–4093, 1997.
- [12] M. Zhou, T. J. Kidd, R. D. Noble, and D. L. Gin, “Supported Lyotropic Liquid-Crystal Polymer Membranes: Promising Materials for Molecular-Size-Selective Aqueous Nanofiltration,” *Adv. Mater.*, vol. 17, pp. 1850–1853, Aug. 2005.
- [13] F. G. Donnan, “Theory of Membrane Equilibria and Membrane Potentials in the Presence of Non-Dialysing Electrolytes. a Contribution to Physical-Chemical Physiology,” *J. Membr. Sci.*, vol. 100, pp. 45–55, Mar. 1995.
- [14] E. S. Hatakeyama, C. J. Gabriel, B. R. Wiesnauer, J. L. Lohr, M. Zhou, R. D. Noble, and D. L. Gin, “Water Filtration Performance of a Lyotropic Liquid Crystal Polymer Membrane with Uniform, Sub-1-Nm Pores,” *J. Membr. Sci.*, vol. 366, no. 1-2, pp. 62–72, 2011.
- [15] E. S. Hatakeyama, B. R. Wiesnauer, C. J. Gabriel, R. D. Noble, and D. L. Gin, “Nanoporous, Bicontinuous Cubic Lyotropic Liquid Crystal Networks via Polymerizable Gemini Ammonium Surfactants,” *Chem. Mater.*, vol. 22, pp. 4525–4527, Aug. 2010.
- [16] B. M. Carter, B. R. Wiesnauer, E. S. Hatakeyama, J. L. Barton, R. D. Noble, and D. L. Gin, “Glycerol-Based Bicontinuous Cubic Lyotropic Liquid Crystal Monomer System for the Fabrication of Thin-Film Membranes with Uniform Nanopores,” *Chem. Mater.*, vol. 24, pp. 4005–4007, Nov. 2012.
- [17] X. Feng, M. E. Tousley, M. G. Cowan, B. R. Wiesnauer, S. Nejati, Y. Choo, R. D. Noble, M. Elimelech, D. L. Gin, and C. O. Osuji, “Scalable Fabrication of Polymer Membranes with Vertically Aligned 1 nm Pores by Magnetic Field Directed Self-Assembly,” *ACS Nano*, vol. 8, pp. 11977–11986, Dec. 2014.
- [18] M. Matyka, A. Khalili, and Z. Koza, “Tortuosity-Porosity Relation in Porous Media Flow,” *Phys. Rev. E*, vol. 78, p. 026306, Aug. 2008.
- [19] M. Zhou, P. R. Nemade, X. Lu, X. Zeng, E. S. Hatakeyama, R. D. Noble, and D. L. Gin, “New Type of Membrane Material for Water Desalination Based on a Cross-Linked Bicontinuous Cubic Lyotropic Liquid Crystal Assembly,” *J. Am. Chem. Soc.*, vol. 129, pp. 9574–9575, Aug. 2007.
- [20] S. M. Dischinger, M. J. McGrath, K. R. Bourland, R. D. Noble, and D. L. Gin, “Effect of Post-Polymerization Anion-Exchange on the Rejection of Uncharged Aqueous Solutes in Nanoporous, Ionic, Lyotropic Liquid Crystal Polymer Membranes,” *J. Membr. Sci.*, vol. 529, pp. 72–79, May 2017.
- [21] R. Resel, G. Leising, P. Markart, M. Kriechbaum, R. Smith, and D. Gin, “Structural properties of polymerised lyotropic liquid crystals phases of 3,4,5-tris(ω -acryloxyalkoxy)benzoate salts,” *Macromol. Chem. Phys.*, vol. 201, pp. 1128–1133, July 2000.
- [22] T. J. McIntosh, S. A. Simon, and R. C. MacDonald, “The Organization of N-Alkanes in Lipid Bilayers,” *Biochim. Biophys. Acta, Biomembr.*, vol. 597, pp. 445–463, Apr. 1980.
- [23] A. S. Govind and N. V. Madhusudana, “A Simple Molecular Theory of Smectic-C Liquid Crystals,” *EPL*, vol. 55, p. 505, Aug. 2001.

- [24] X. Zhu, M. A. Scherbina, A. V. Bakirov, B. Gorzolnik, S. N. Chvalun, U. Beginn, and M. Möller, “Methacrylated Self-Organizing 2,3,4-Tris(alkoxy)benzenesulfonate: A New Concept Toward Ion-Selective Membranes,” *Chem. Mater.*, vol. 18, pp. 4667–4673, Sept. 2006.
- [25] P. Chaikin and T. Lubensky, *Principles of Condensed Matter Physics*. Cambridge, UK: Cambridge University Press, 1 ed., 1995.
- [26] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, “Development and Testing of a General Amber Force Field,” *J. Comput. Chem.*, vol. 25, pp. 1157–1174, July 2004.
- [27] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, “Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations,” *J. Mol. Graphics. Modell.*, vol. 25, pp. 247–260, Oct. 2006.
- [28] D. Case, R. Betz, W. Botello-Smith, D. Cerutti, T. Cheatham, III, T. Darden, R. Duke, T. Giese, H. Gohlke, A. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. Merz, G. Monard, H. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, D. Roe, A. Roitberg, C. Sagui, C. Simmerling, J. Swails, R. Walker, J. Wang, R. Wolf, X. Wu, L. Xiao, D. York, and P. Kollman, *AmberTools16*. University of California, San Francisco, Apr. 2016.
- [29] M. J. Cook and M. R. Wilson, “Development of an All-Atom Force Field for the Simulation of Liquid Crystal Molecules in Condensed Phases (LCFF),” *Mol. Cryst. Liq. Cryst.*, vol. 357, pp. 149–165, Mar. 2001.
- [30] N. J. Boyd and M. R. Wilson, “Optimization of the Gaff Force Field to Describe Liquid Crystal Molecules: The Path to a Dramatic Improvement in Transition Temperature Predictions,” *Phys. Chem. Chem. Phys.*, vol. 17, pp. 24851–24865, Sept. 2015.
- [31] H. Bekker, H. J. C. Berendsen, E. J. Dijkstra, S. Achterop, R. van Drunen, D. van der Spoel, A. Sijbers, H. Keegstra, B. Reitsma, and M. K. R. Renardus, *GROMACS: A Parallel Computer for Molecular Dynamics Simulations*. Physics Computing '92: Proceedings of the 4th International Conference, Praha, Czechoslovakia, Aug 24-28, 1992; World Scientific: Singapore, 1993.
- [32] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, “GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation,” *Comput. Phys. Commun.*, vol. 91, pp. 43–56, Sept. 1995.
- [33] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, “GROMACS: Fast, Flexible, and Free,” *J. Comput. Chem.*, vol. 26, pp. 1701–1718, Dec. 2005.
- [34] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, “GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation,” *J. Chem. Theory Comput.*, vol. 4, pp. 435–447, Mar. 2008.
- [35] J. Mondal, M. Mahanthappa, and A. Yethiraj, “Self-Assembly of Gemini Surfactants: A Computer Simulation Study,” *J. Phys. Chem. B*, vol. 117, pp. 4254–4262, Apr. 2013.

- [36] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, “Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations,” *J. Comput. Chem.*, vol. 30, pp. 2157–2164, Oct. 2009.
- [37] M. O. Sinnokrot, E. F. Valeev, and C. D. Sherrill, “Estimates of the Ab Initio Limit for π - π Interactions: The Benzene Dimer,” *J. Am. Chem. Soc.*, vol. 124, pp. 10887–10893, Sept. 2002.
- [38] J. D. Chodera, “A Simple Method for Automated Equilibration Detection in Molecular Simulations,” *J. Chem. Theory Comput.*, vol. 12, pp. 1799–1805, Jan. 2016.
- [39] T. Kuriabova, M. D. Betterton, and M. A. Glaser, “Linear Aggregation and Liquid-Crystalline Order: Comparison of Monte Carlo Simulation and Analytic Theory,” *J. Mater. Chem.*, vol. 20, pp. 10366–10383, Nov. 2010.
- [40] J. C. Slater, “Atomic Radii in Crystals,” *J. Chem. Phys.*, vol. 41, pp. 3199–3204, Nov. 1964.
- [41] T. E. Oliphant, “Python for Scientific Computing,” *Comput. Sci. Eng.*, vol. 9, pp. 10–20, May 2007.
- [42] A. Einstein, *Investigations on the Theory of the Brownian Movement*. Dover, New York, USA: Dover Publications, 1956.
- [43] C. Caleman, P. J. van Maaren, M. Hong, J. S. Hub, L. T. Costa, and D. van der Spoel, “Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant,” *J. Chem. Theory Comput.*, vol. 8, pp. 61–74, Jan. 2012.
- [44] L.-P. Wang, K. A. McKiernan, J. Gomes, K. A. Beauchamp, T. Head-Gordon, J. E. Rice, W. C. Swope, T. J. Martínez, and V. S. Pande, “Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15,” *J. Phys. Chem. B*, vol. 121, pp. 4023–4039, Apr. 2017.
- [45] V. Percec, M. R. Imam, M. Peterca, D. A. Wilson, R. Graf, H. W. Spiess, V. S. K. Balagurusamy, and P. A. Heiney, “Self-Assembly of Dendronized Triphenylenes into Helical Pyramidal Columns and Chiral Spheres,” *J. Am. Chem. Soc.*, vol. 131, pp. 7662–7677, June 2009.
- [46] R. I. Gearba, D. V. Anokhin, A. I. Bondar, W. Bras, M. Jahr, M. Lehmann, and D. A. Ivanov, “Homeotropic Alignment of Columnar Liquid Crystals in Open Films by Means of Surface Nanopatterning,” *Adv. Mater.*, vol. 19, pp. 815–820, Mar. 2007.
- [47] O. Young, A. R. Wildes, P. Manuel, B. Ouladdiaf, D. D. Khalyavin, G. Balakrishnan, and O. A. Petrenko, “Highly Frustrated Magnetism in SrHo2O4: Coexistence of Two Types of Short-Range Order,” *Phys. Rev. B*, vol. 88, p. 024411, July 2013.
- [48] G. S. Girolami, *X-ray Crystallography*. Herndon, VA, USA: University Science Books, 2016.
- [49] G. Harburn, C. A. Taylor, and T. R. Welberry, *Atlas of Optical Transforms*. Ithaca, New York, USA: Cornell University Press, 1975.
- [50] G. Taylor, “The Phase Problem,” *Acta Cryst D*, vol. 59, pp. 1881–1890, Nov. 2003.

- [51] F. Carey and R. Giuliano, *Organic Chemistry*. USA: McGraw-Hill Companies, Inc., 8 ed., 2011.
- [52] T. P. Tauer and C. D. Sherrill, “Beyond the Benzene Dimer: An Investigation of the Additivity of π - π Interactions,” *J. Phys. Chem. A*, vol. 109, pp. 10475–10478, Nov. 2005.
- [53] R. Y. Dong, D. Goldfarb, M. E. Moseley, Z. Luz, and H. Zimmermann, “Translational Diffusion in Discotic Mesophases Studied by the Nuclear Magnetic Resonance Pulsed Field Gradient Method,” *J. Phys. Chem.*, vol. 88, pp. 3148–3152, July 1984.
- [54] S. V. Dvinskikh, I. Furó, H. Zimmermann, and A. Maliniak, “Molecular Self-Diffusion in a Columnar Liquid Crystalline Phase Determined by Deuterium NMR,” *Phys. Rev. E*, vol. 65, p. 050702, May 2002.
- [55] Y. Liu and F. Zhu, “Collective Diffusion Model for Ion Conduction Through Microscopic Channels,” *Biophys. J.*, vol. 104, pp. 368–376, Jan. 2013.
- [56] C. D. Bruce, M. L. Berkowitz, L. Perera, and M. D. E. Forbes, “Molecular Dynamics Simulation of Sodium Dodecyl Sulfate Micelle in Water: Micellar Structural Characteristics and Counterion Distribution,” *J. Phys. Chem. B*, vol. 106, pp. 3788–3793, Apr. 2002.
- [57] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, “Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins,” *J. Chem. Theory Comput.*, vol. 9, pp. 4046–4063, Sept. 2013.
- [58] G. A. Kaminski, H. A. Stern, B. J. Berne, and R. A. Friesner, “Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from ab Initio Quantum Chemistry,” *J. Phys. Chem. A*, vol. 108, pp. 621–627, Jan. 2004.
- [59] P. E. M. Lopes, G. Lamoureux, B. Roux, and A. D. MacKerell, “Polarizable Empirical Force Field for Aromatic Compounds Based on the Classical Drude Oscillator,” *J. Phys. Chem. B*, vol. 111, pp. 2873–2885, Mar. 2007.
- [60] D. M. Warsinger, S. Chakraborty, E. W. Tow, M. H. Plumlee, C. Bellona, S. Loutatidou, L. Karimi, A. M. Mikelonis, A. Achilli, A. Ghassemi, L. P. Padhye, S. A. Snyder, S. Curcio, C. D. Vecitis, H. A. Arafat, and J. H. Lienhard, “A Review of Polymeric Membranes and Processes for Potable Water Reuse,” *Prog. Polym. Sci.*, vol. 81, pp. 209–237, June 2018.
- [61] N. Hilal, H. Al-Zoubi, N. A. Darwish, A. W. Mohamma, and M. Abu Arabi, “A Comprehensive Review of Nanofiltration Membranes:treatment, Pretreatment, Modelling, and Atomic Force Microscopy,” *Desalination*, vol. 170, pp. 281–308, Nov. 2004.
- [62] P. Alexandridis and B. Lindman, *Amphiphilic Block Copolymers: Self-Assembly and Applications*. Elsevier: Amsterdam, Oct. 2000.
- [63] B. J. Coscia, J. Yelk, M. A. Glaser, D. L. Gin, X. Feng, and M. R. Shirts, “Understanding the Nanoscale Structure of Inverted Hexagonal Phase Lyotropic Liquid Crystal Polymer Membranes,” *J. Phys. Chem. B*, vol. 123, pp. 289–309, Jan. 2019.
- [64] P. Chuysinuan, N. Chimnoi, S. Techasakul, and P. Supaphol, “Gallic Acid-Loaded Electrospun Poly(L-Lactic Acid) Fiber Mats and their Release Characteristic,” *Macromol. Chem. Phys.*, vol. 210, no. 10, pp. 814–822, 2009.

- [65] D. D. Perrin, B. Dempsey, and E. P. Serjeant, Pka Prediction for Organic Acids and Bases. Chapman and Hall: London, 1981.
- [66] P. Florová, P. Sklenovský, P. Banáš, and M. Otyepka, “Explicit Water Models Affect the Specific Solvation and Dynamics of Unfolded Peptides While the Conformational Behavior and Flexibility of Folded Peptides Remain Intact,” J. Chem. Theory Comput., vol. 6, pp. 3569–3579, Nov. 2010.
- [67] J. W. Ponder and D. A. Case, “Force Fields for Protein Simulations,” Adv. Protein Chem., vol. 66, pp. 27–85, Jan. 2003.
- [68] R. Resel, U. Theissl, C. Gadermaier, E. Zojer, M. Kriechbaum, H. Amenitsch, D. Gin, R. Smith, and G. Leising, “The H₂-Phase of the Lyotropic Liquid Crystal Sodium 3,4,5-Tris(omega-Acryloyloxyundecyloxy)benzoate,” Liq. Cryst., vol. 27, pp. 407–411, Mar. 2000.
- [69] Y. Meroz and I. M. Sokolov, “A Toolbox for Determining Subdiffusive Mechanisms,” Phys. Rep., vol. 573, pp. 1–29, Apr. 2015.
- [70] R. Metzler, J.-H. Jeon, A. G. Cherstvy, and E. Barkai, “Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking,” Phys. Chem. Chem. Phys., vol. 16, no. 44, pp. 24128–24164, 2014.
- [71] E. J. Maginn, R. A. Messerly, D. J. Carlson, D. R. Roe, and J. R. Elliott, “Best Practices for Computing Transport Properties 1. Self-Diffusivity and Viscosity from Equilibrium Molecular Dynamics [Article v1.0],” LiveCoMS, vol. 1, no. 1, pp. 6324–, 2018.
- [72] S. G. Schultz and A. K. Solomon, “Determination of the Effective Hydrodynamic Radii of Small Molecules by Viscometry,” J. Gen. Physiol., vol. 44, pp. 1189–1199, July 1961.
- [73] A. Gierer and K. Wirtz, “Molekulare Theorie der Mikroreibung - Molecular Theory of Microfriction,” Z. Naturforsch., vol. 8, pp. 532–538, 1953.
- [74] H. C. Chen and S. H. Chen, “Diffusion of Crown Ethers in Alcohols,” J. Phys. Chem., vol. 88, pp. 5118–5121, Oct. 1984.
- [75] T. C. Chan, H. T. Li, and K. Y. Li, “Effects of Shapes of Solute Molecules on Diffusion: A Study of Dependences on Solute Size, Solvent, and Temperature,” Dec. 2015.
- [76] H. Lee, R. M. Venable, A. D. MacKerell, and R. W. Pastor, “Molecular Dynamics Studies of Polyethylene Oxide and Polyethylene Glycol: Hydrodynamic Radius and Shape Anisotropy,” Biophys. J., vol. 95, pp. 1590–1599, Aug. 2008.
- [77] L. He and B. Niemeyer, “A Novel Correlation for Protein Diffusion Coefficients Based on Molecular Weight and Radius of Gyration,” Biotechnol. Progr., vol. 19, no. 2, pp. 544–548, 2003.
- [78] Z. Li, “Critical Particle Size Where the Stokes-Einstein Relation Breaks Down,” Phys. Rev. E, vol. 80, p. 061204, Dec. 2009.
- [79] C. Truong, L. Oudre, and N. Vayatis, “Ruptures: Change Point Detection in Python,” arXiv, vol. arXiv:1801.00826, Jan. 2018.

- [80] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. CRC Press: Boca Raton, May 1994.
- [81] A. Luzar and D. Chandler, “Effect of Environment on Hydrogen Bond Dynamics in Liquid Water,” *Phys. Rev. Lett.*, vol. 76, pp. 928–931, Feb. 1996.
- [82] D. Prada-Gracia, R. Shevchuk, and F. Rao, “The Quest for Self-Consistency in Hydrogen Bond Definitions,” *J. Chem. Phys.*, vol. 139, p. 084501, Aug. 2013.
- [83] W. Humphrey, A. Dalke, and K. Schulten, “VMD: Visual Molecular Dynamics,” *J. Mol. Graphics*, vol. 14, pp. 33–38, Feb. 1996.
- [84] C. N. Rowley and B. Roux, “The Solvation Structure of Na⁺ and K⁺ in Liquid Water Determined from High Level Ab Initio Molecular Dynamics Simulations,” *J. Chem. Theory Comput.*, vol. 8, pp. 3526–3535, Oct. 2012.
- [85] F. W. Starr, J. K. Nielsen, and H. E. Stanley, “Fast and Slow Dynamics of Hydrogen Bonds in Liquid Water,” *Phys. Rev. Lett.*, vol. 82, pp. 2294–2297, Mar. 1999.
- [86] H. F. M. C. Martiniano and N. Galamba, “Insights on Hydrogen-Bond Lifetimes in Liquid and Supercooled Water,” *J. Phys. Chem. B*, vol. 117, pp. 16188–16195, Dec. 2013.
- [87] V. P. Voloshin and Y. I. Naberukhin, “Hydrogen Bond Lifetime Distributions in Computer-Simulated Water,” *J. Struct. Chem.*, vol. 50, pp. 78–89, Feb. 2009.
- [88] M. W. Mahoney and W. L. Jorgensen, “Diffusion Constant of the Tip5p Model of Liquid Water,” *J. Chem. Phys.*, vol. 114, pp. 363–366, Dec. 2000.
- [89] C. Carvajal, K. J. Tölle, J. Smid, and M. Szwarc, “Studies of Solvation Phenomena of Ions and Ion Pairs in Dimethoxyethane and Tetrahydrofuran,” *J. Am. Chem. Soc.*, vol. 87, pp. 5548–5553, Dec. 1965.
- [90] X. Wu, S. Jin, Z. Zhang, L. Jiang, L. Mu, Y.-S. Hu, H. Li, X. Chen, M. Armand, L. Chen, and X. Huang, “Unraveling the Storage Mechanism in Organic Carbonyl Electrodes for Sodium-Ion Batteries,” *Sci. Adv.*, vol. 1, p. e1500330, Sept. 2015.
- [91] T. Shinoda, H. Ogawa, F. Cornelius, and C. Toyoshima, “Crystal Structure of the Sodium-Potassium Pump at 2.4 {\AA} Resolution,” *Nature*, vol. 459, pp. 446–450, May 2009.
- [92] G. Srinivas, S. Bhattacharyya, and B. Bagchi, “Computer Simulation and Mode Coupling Theory Study of the Effects of Specific Solute-Solvent Interactions on Diffusion: Crossover from a Sub-Slip to a Super-Stick Limit of Diffusion,” *J. Chem. Phys.*, vol. 110, pp. 4477–4482, Feb. 1999.
- [93] H. S. Biswal, “Hydrogen Bonds Involving Sulfur: New Insights from ab Initio Calculations and Gas Phase Laser Spectroscopy,” in *Noncovalent Forces* (S. Scheiner, ed.), Challenges and Advances in Computational Chemistry and Physics, pp. 15–45, Cham: Springer International Publishing, 2015.
- [94] ChemAxon, “MarvinSketch 17.13 2017 ChemAxon (<http://chemaxon.com>),” 2017.
- [95] N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, “Open Babel: An Open Chemical Toolbox,” *J. Cheminformatics*, vol. 3, p. 33, Oct. 2011.

- [96] “The Open Babel Package, version 2.4.1 <http://openbabel.org>.”
- [97] R. Walker and S. Tang, “Antechamber Tutorial (ambermd.org/tutorials/basic/tutorial4b/)”
- [98] A. W. Sousa da Silva and W. F. Vranken, “ACPYPE - AnteChamber PYthon Parser interface,” BMC Res Notes, vol. 5, p. 367, July 2012.
- [99] M. R. Shirts and J. D. Chodera, “Statistically Optimal Analysis of Samples from Multiple Equilibrium States,” J. Chem. Phys., vol. 129, p. 124105, Sept. 2008.
- [100] R. Young and P. Lovell, Introduction to Polymers. New York, USA: CRC Press, 3 ed., 2011.
- [101] A. R. Rakitin and G. R. Pack, “Necessity of Aromatic Carboxylate Anions to Be Planar to Induce Growth of Cationic Micelles,” Langmuir, vol. 21, pp. 837–840, Feb. 2005.
- [102] M. R. Nelson and R. F. Borkman, “Internal Rotation Barriers: Ab Initio Calculations on Substituted Ethyl Benzoates and Benzoic Acids as Models for Polyester Flexibility,” J. Mol. Struct. (THEOCHEM), vol. 432, pp. 247–255, June 1998.

Appendix A

Test

A.1 Setup and analysis scripts

All python and bash scripts used to set up systems and conduct post-simulation trajectory analysis, except for simulating X-ray diffraction (XRD) patterns, are available online at <https://github.com/bencoscia/l1csm>. Table A.1 provides more detail about specific scripts used for each type of analysis performed in the main text.

The python scripts used to simulated XRD patterns are publicly available online at <https://github.com/joeyelk/MD-Structure-Factor>. Given a GROMACS trajectory (.trr or .xtc) and a configuration file (.gro), simulate an XRD pattern using `main_gromacs.py`. Pass the flag `-manuscript_format` to generate patterns and cross-sections with the same format as those presented in the main text.

A.2 Further details regarding monomer parameterization

We parameterized monomers according to the following procedure:

- (1) *Create monomer structure file with connectivity:* We drew atomistic structures using MarvinSketch 17.13 [94] with all hydrogen atoms drawn out explicitly. We optimized the 3D geometry of the structure using the ‘Clean in 3D’ function of MarvinSketch. We saved the structure as a .mol file, then converted it to .pdb format using Open Babel 2.4.1 [95, 96].
- (2) *Assign GAFF atomtypes using antechamber:* Using the .pdb structure file as input, we

ran `antechamber` [27] using the AM1-BCC charge model. The net charge on the monomer is input as -1 since the sodium ion is kept as a separate residue. We use `LEaP` [28] and the output of `antechamber` to create Amber topology files. A detailed tutorial for using `LEaP` for parameterization can be accessed elsewhere [97].

- (3) *Create GROMACS topologies from Amber output:* The output of `LEaP` is a `.inpcrd` and a `.prmtop` file which are Amber topology files. Using `acpype.py` [98], we converted the `LEaP` output into GROMACS `.gro` and `.top` files.
- (4) *Perform a simulated annealing procedure on the monomer:* We created a cubic box around the monomer using the GROMACS command `gmx editconf`. The monomer was centered in the box with edges of the box spaced at least 3 nm from the monomer on all sides. We ran an energy minimization on the system with the steepest descent algorithm. Next we performed an NVT simulated annealing procedure. We linearly decreased the temperature of the system from 1000 K to 50 K over the course of 10 ns. We randomly chose a monomer configuration from the last 10% of the trajectory.
- (5) *Reassign charges with molcharge:* With the monomer configuration taken from the annealed trajectory, we reassigned charges using `molcharge` with the `am1bccsym` method in order to ensure charges are symmetric. This condition is not guaranteed with `antechamber`. The charges in the GROMACS topology file (`.top`) were replaced with the new charges calculated by `molcharge`.
- (6) *Anneal again to get final structure:* We repeated the same simulated annealing procedure using the monomer topology with `molcharge` charges. A random monomer configuration was pulled from the last 10% of the trajectory and was used to build all assemblies reported (Figure A.1).

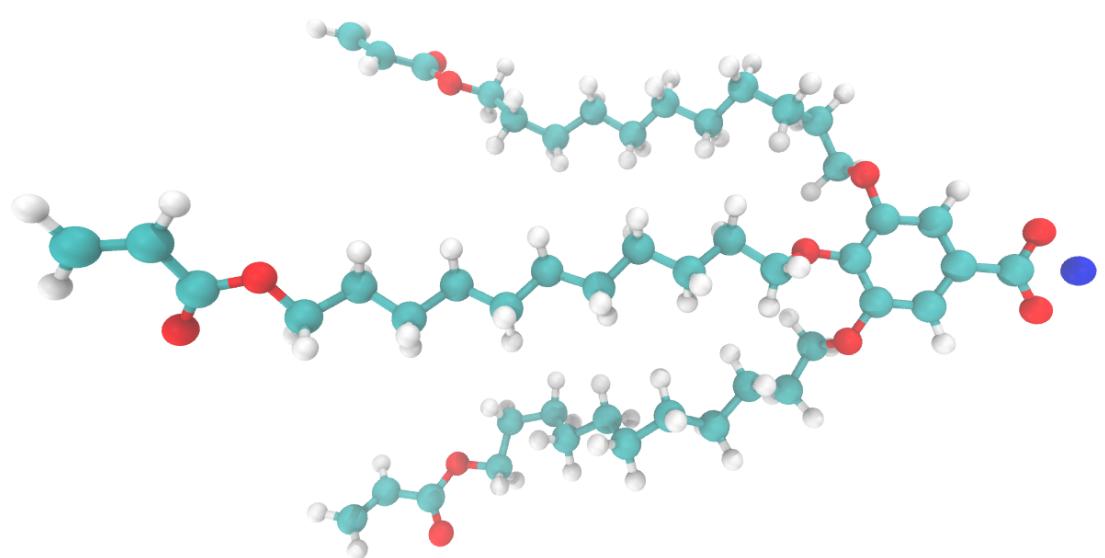


Figure A.1: Atomistic representation of the monomer Na-GA3C11. White atoms represent hydrogen, cyan atoms represent carbon, red atoms represent oxygen and the blue atom is sodium.

Script Name	Section	Description
/setup/param.sh	2.2.2	Automatically run the monomer parameterization scheme outlined in Section A.2 given an initial structure file in .pdb format.
/setup/build.py	2.2.3 & 2.2.4	Build an initial configuration in the sandwiched or parallel displaced configuration with the desired initial vertical monomer spacing, monomers-per-column, and columns-per-pore.
/setup/equil.sh	2.2.5	Wrapper script for running the GROMACS functions and python scripts needed in order to equilibrate an initial configuration.
/analysis/Structure_char.py	2.2.6.2	Calculate the distance between pores.
/analysis/correlation.py	2.2.6.4	Calculate the 3D correlation function from a trajectory and plot desired slices.
/analysis/regional_density.py	3.2.11	Calculate the density of a given monomer component as a function of its distance from the pore center.
/analysis/structure_factor.py	2.2.7	Generate customizable trajectories of point scatterers, then calculate their structure factor and plot slices of it.
/analysis/disorder.py	2.2.7	Measure the magnitude of quenched disorder.
/analysis/Ionic_Conductivity.py	2.2.8	Calculate the ionic conductivity using the Nernst-Einstein relationship and collective diffusion model.
/setup/xlink.py	2.2.9	Iteratively cross-link a system given an initial configuration.
/analysis/orientational_order.py	2.3.1	Calculate the nematic order parameter.
/analysis/tilt.py	2.3.2.3	Calculate the angle between the xy plane and the vector extending from the front to the back of each monomer tail.
/analysis/tail_packing.py	2.3.2.3	Calculate the distribution of angles between the center of mass of each monomer tail and its nearest neighbor monomer tails.
/analysis/hexagonal_packing.py	2.3.2.3	Calculate Fourier transform of 2D array of hexagonal scatterers as shown in Figure 2.7f of the main text.
/analysis/compare_disorder.py	2.3.2.4	Measure quenched disorder for each trajectory in an ensemble of simulations, then plot distributions of quenched disorder.
/analysis/hbonds_pairing.py	2.3.2.9	Identify all hydrogen bonds then count the number of monomer pairs that share hydrogen bonds.
/analysis/torsions.py	2.3.4	Calculate the autocorrelation function for a chosen dihedral.

Table A.1: The first column provides the names of the python scripts available in the `llcsim` GitHub repository that were used for system setup and post-simulation trajectory analysis. Paths preceding script names are relative to the `llcsim` root 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.

A.3 Attempted Self-assembly

We attempted self assembly of Na-GA3C11 monomers by using an isotropic configuration generated using Packmol [36] (Figure A.2a). The input file given to Packmol is shown below.

```

tolerance 2.0

output packed.pdb

filetype pdb

structure monomer.pdb

number 400

inside box 0. 0. 0. 84.3 84.3 84.3

end structure

```

We used the same number of monomers (400) that we used in all other simulations. Since Packmol does not have the capability to create monoclinic boxes, we used a cubic box with a volume of 660 nm³, 25% larger than those used to create ordered unit cells we studied. We tested systems with semi-isotropic and anisotropic pressure coupling since the shape of the unit cell will likely need to change in order to accommodate hexagonally packed pores.

We use the nematic order parameter as defined in Section A.14 to observe any progress towards system ordering. In both cases, we see that the nematic order parameter stays close to zero for the duration of the simulation time, indicating an isotropic arrangement of head groups (Figure A.2b). The values are slightly negative since there are only 400 monomers. Equation A.3 is bounded between -1 and 1, however, with a proper choice of nematic director, measured values will range between 0 and 1. With a large number of samples, an isotropic system would have an order parameter of 0.

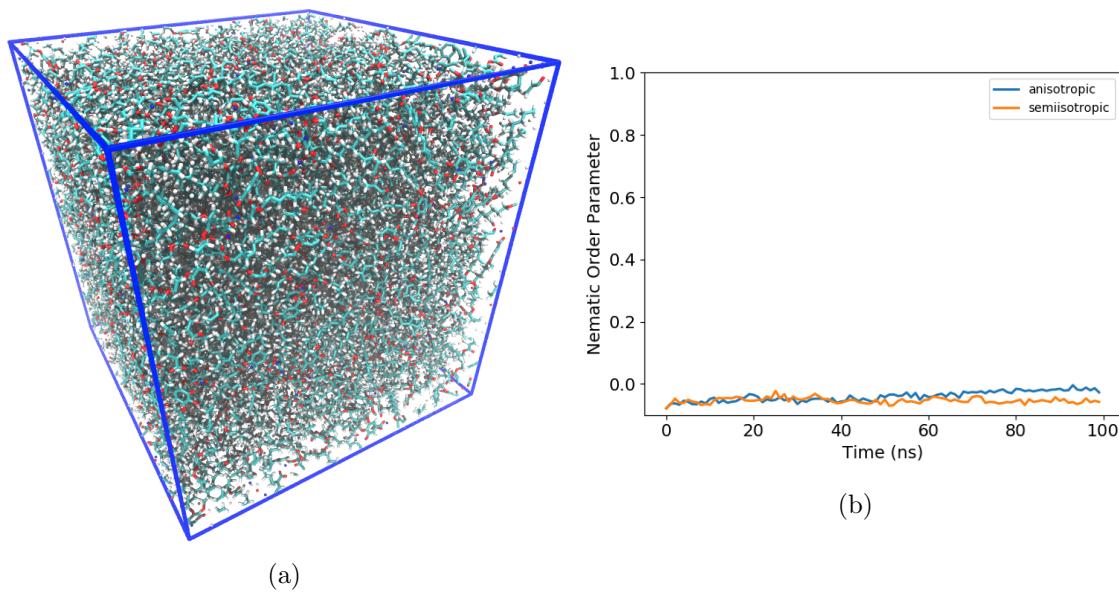


Figure A.2: (a) We created a box of isotropically packed monomers and allowed it to simulate for 100 ns using isotropic and anisotropic pressure coupling. (b) The nematic order parameter hovers close to zero for the duration of the simulation meaning the system maintains its isotropic alignment.

A.4 Monomer build procedure

Figure A.3 illustrates the procedure carried out by `build.py` in order to build initial configurations.

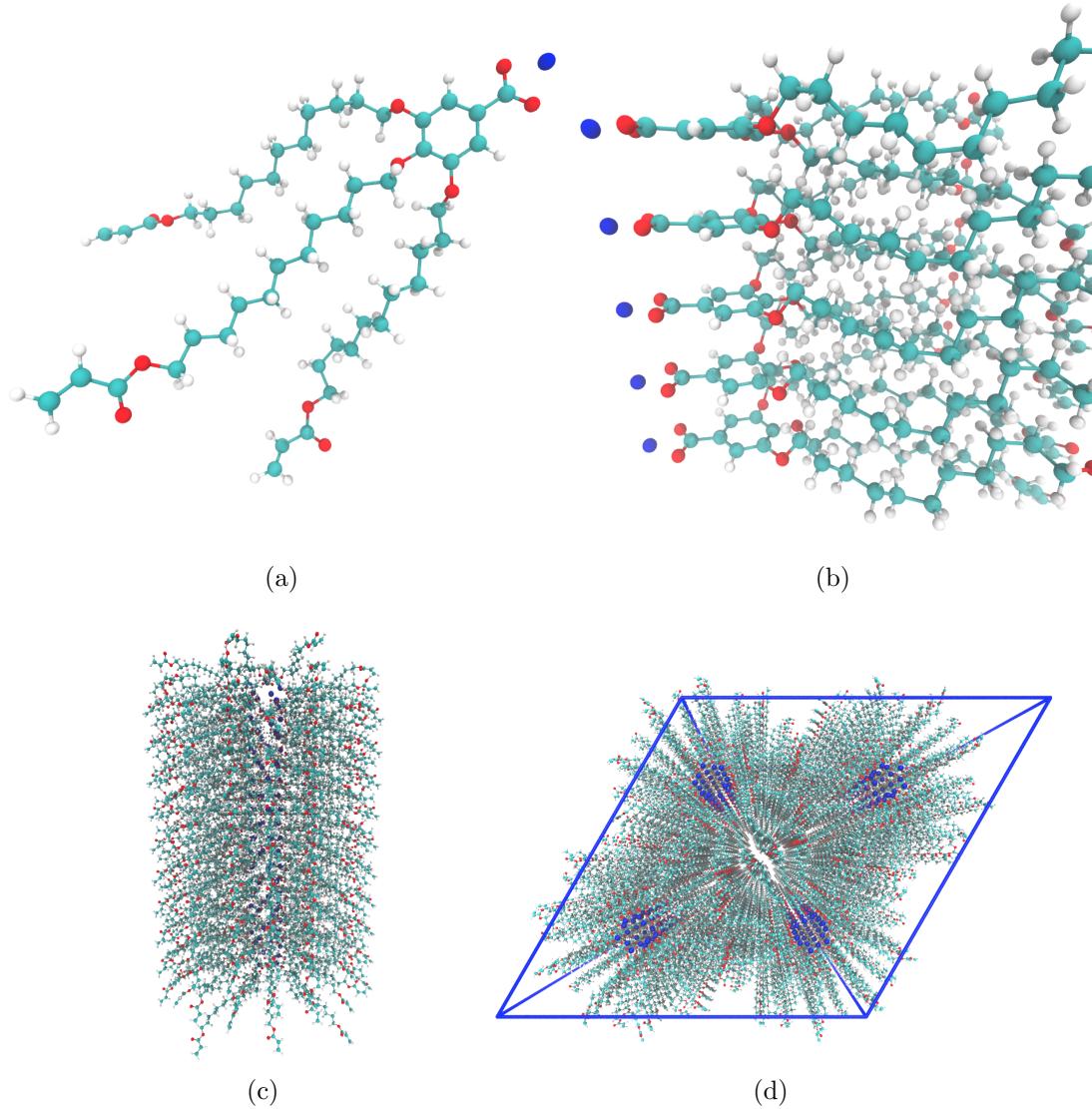


Figure A.3: (a) We parameterized a single monomer and annealed it to produce a low-energy configuration. (b) We assembled monomers into columns by stacking them on top of each other. (c) We duplicated each column and rotated them to create hydrophilic pore centers. We chose to stack twenty monomers into each column. (d) We duplicated the pores and placed them into a monoclinic unit cell with hexagonal symmetry.

A.5 Number of monomers per column

We chose to build all systems using 20 monomers per column, i.e. for a simulation with 5 columns around each pore, there are 100 monomers per pore, and 400 monomers per simulation. We made this choice by considering a number of factors that directly affect the quality of our results. Clearly, choosing a smaller number of layers will cause the system to suffer from larger finite size effects. Even in our 20 monomer-per-column system, axial correlations between stacked monomers persist throughout the entire correlation function (see Figure A.24). The correlation function of the system which we built with 40 monomers per column is very similar to that created with 20 monomers per column (it continues to oscillate for its full length (Figure A.4b)). If we continue adding monomers, the correlation function will eventually fully decay but it is not worth the computational expense. We do find that the correlation lengths calculated from 20 monomer per column simulations are in sufficient agreement with experiment.

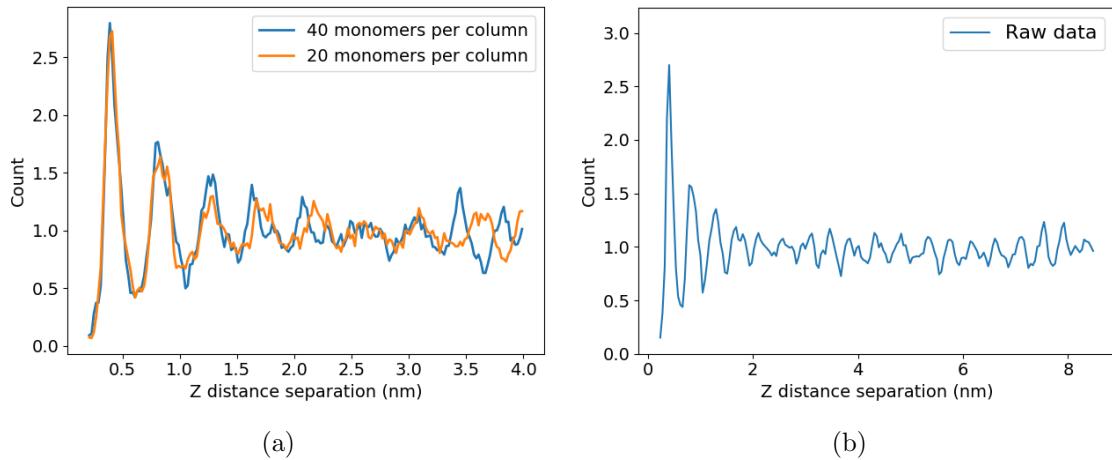


Figure A.4: (a) The correlation functions generated from ordered basin sandwiched configurations are nearly the same when the z -dimension of the system is doubled. (b) Oscillations persist throughout the length of the correlation function for the sandwiched system in the ordered basin built with 40 monomers per column. The full system is ca. 17 nm tall so the full correlation function covers half this length since periodicity forces the remaining length to be its mirror image.

We may be able to simulate systems using less than 20 monomers per column, but chose to use 20 since it gives fairly good resolution when simulating X-ray diffraction patterns. The size of

the Fourier space bin in each dimension is determined by $\frac{2\pi}{L}$ where L is the length of a given box vector. The z-box vector for 20 monomer-per-column systems is $\approx 85 \text{ \AA}$ which means we will see a reciprocal space resolution of 0.074 \AA^{-1} in the z direction. Further decreasing the resolution would increase the noise in the simulated patterns and make it more difficult to measure the intensity of reflections of interest.

A.6 Initial Configuration Dependence

We addressed any major dependence on initial configuration in the main text. There we showed that systems are stable when made with 4, 5, 6, 7, and 8 columns per pore. We also showed that systems are stable when monomer head groups are oriented in the parallel displaced and sandwiched configurations (see Figure A.9). Our model is most consistent with experiment when built with 5 columns per pore, with monomers stacked 3.7 \AA apart in the parallel displaced configuration.

There are three other parameters whose choice may influence the equilibrium structure: initial pore spacing, initial pore radius and initial stacking distance between monomers. Here we show the results of a sensitivity analysis performed on the three parameters. To reduce the size of the sensitivity analysis, we only tested systems built with 5 columns per pore with monomers stacked in the parallel displaced configuration. We equilibrated all systems according to the dry equilibration procedure.

A.6.1 Initial pore spacing

We tested five different initial pore spacings, defined as the distance between the central axis of each pore with all others (Figure A.5a). To reduce the number of variables, we held the pore radius constant, at 5 \AA and the distance between layers at 3.7 \AA since those were the values used in our optimal system in the main text. We prioritized ensuring that resulting configurations maintain the expected hexagonal symmetry. If initial pore spacing is too small, we observe repulsion between columns which disrupts equilibration. If the pore spacing is too large the pores squeeze together,

but in a distorted hexagonal array because the xy initial translation of the pores is somewhat erratic.

- (1) 39 Å : We tested a pore spacing of 39 Å in order to have a test system with an initial spacing below the experimental value. As soon as the restraining potential switches to 56 $\text{kJ mol}^{-1} \text{ nm}^{-2}$, the columns are able to repel resulting in a large jump in pore spacing (Figure A.5b).
- (2) 41 Å : We chose to test a pore spacing of 41 Å since it closely matches the experimental pore spacing. Again, we observe abrupt repulsion of columns once the restraining potential is switched to 56 $\text{kJ mol}^{-1} \text{ nm}^{-2}$ (Figure A.5c).
- (3) 45 Å : A pore spacing of 45 Å is about 10% larger than the experimental value. This systems exhibits the smallest response to our equilibration procedure (Figure A.5d) and we chose to use this value for all our simulations in the main text.
- (4) 50 Å : We tested a pore spacing of 50 Å, about 20% larger than the experimental value. Once the restraining potential reaches 56 $\text{kJ mol}^{-1} \text{ nm}^{-2}$, the pore spacing begins to decrease rapidly, following a linear trend (Figure A.5e).
- (5) 55 Å : We tested a pore spacing of 55 Å which is at a distance where monomers in each pore no longer intersect those of adjacent pores. Once the restraining potential reaches 56 $\text{kJ mol}^{-1} \text{ nm}^{-2}$, the pore spacing changes erratically until it begins to settle when the force constants are below 3 $\text{kJ mol}^{-1} \text{ nm}^{-2}$ (Figure A.5f). We recommend avoiding a system such as this where vacuum gaps between pore columns are introduced unnecessarily.

A.6.2 Initial pore radius

We tested 3 different pore radii, defined as shown in Figure A.6. For each system, we held the initial pore spacing constant at 45 Å and the initial stacking distance between monomers constant at 3.7 Å. Equilibrated values of pore radii are presented in Table A.2. The pore radius in each

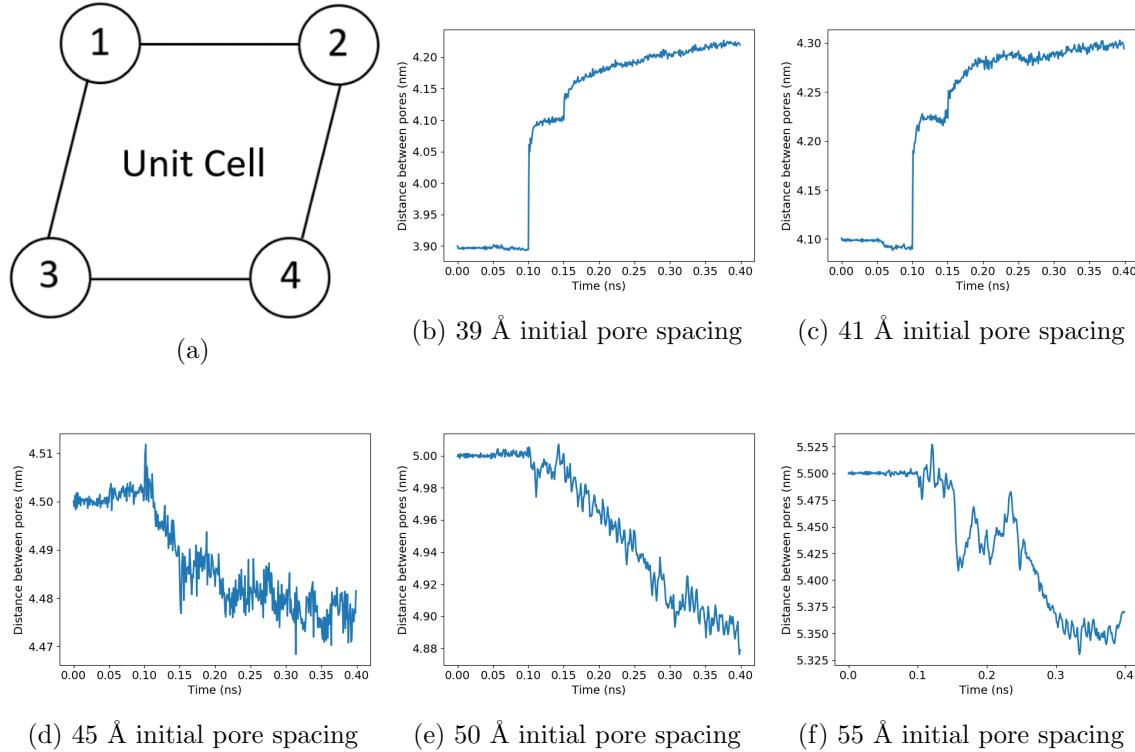


Figure A.5: (a) In a perfect hexagonal array, there are 5 distances that should be exactly equal. As pictured in this diagram the distance from 1-2, 1-3, 1-4, 2-4, and 3-4 are the same. All plots in (b)–(f) are the average of these distances and represent the pore spacing during the restrained portion of the dry equilibration procedure. Every 50 ps (0.05 ns) we reduce the force constant on the position restraints according to the sequence: 1000000, 3162, 56, 8, 3, 2, 1, 0 kJ mol⁻¹ nm⁻². When we chose an initial pore spacing below experimental value (b) or at the experimental value (c), there is an abrupt change in pore spacing when the position restraints are reduced to 56 kJ mol⁻¹ nm⁻². When we initially position pores 45 Å apart (d) the pore spacing remains relatively stable. When we initially place pores 50 Å apart (e), the pore spacing decreases nearly linearly once the restraints are reduced to 56 kJ mol⁻¹ nm⁻². When we initially place pores 55 Å apart (f), so that monomers do not intersect with adjacent pores, and position restraints are reduced to 56 kJ mol⁻¹ nm⁻², the pore spacing changes erratically before stabilizing when force constants are reduced below 3 kJ mol⁻¹ nm⁻².

simulation frame is calculated as the average distance of all carbonyl carbons (see Figure A.6) from their associated pore center. We generated statistics for the pore radii reported from the time series representing the average pore radius at each frame. We detected equilibration as described in Section 2.2.6.1 of the main text. We calculated the average and standard deviation of pore radii using only data points collected after equilibration was detected.

- (1) 2.5 Å : The smallest pore radius that we can achieve before energy minimization becomes problematic is 2.5 Å. After being run through the full dry equilibration procedure, the average pore radius is 0.40 ± 0.01 Å.
- (2) 5 Å : We tested a pore radius of 5 Å because it is slightly larger than the equilibrated pore radius of the system simulated with an initial pore radius of 2.5 Å. After being run through the full dry equilibration procedure, the average pore radius is 0.42 ± 0.01 Å, which agrees with the 2.5 Å configuration within uncertainty.
- (3) 8 Å : The largest pore radius that can achieved before energy minimization becomes problematic is 8 Å. One should use caution with such a structure because of the relatively large vacuum space that is created in the pore region of the initial configuration. After being run through the full dry equilibration procedure, we see a combination of cylindrical and slit-like pores (Figure A.7). Measuring the pore radius of this system does not have a concrete meaning since slits do not have a single radius, but its calculated value is still reported in Table A.2. We are wary of such a non-symmetric structure and choose not to use a pore radius of 8 Å for our starting configurations.

Initial Pore Radius	Equilibrated Pore Radius
2.5 Å	4.0 ± 0.1 Å
5 Å	4.2 ± 0.1 Å
8 Å	6.9 ± 0.1 Å

Table A.2: The average pore radii of systems built with an initial pore radius of 2.5 Å and 5 Å equilibrate to values that agree within error. If the pore radius is too large, slit pores may form.

A.6.3 Initial stacking distance between monomers

We tested 3 different initial monomer spacings, defined as the z -distance between the planes of aromatic rings. Systems built with layers stacked 3.7 Å and 5 Å apart are discussed extensively in the main text. We also tested a system built with layers stacked 10 Å apart.

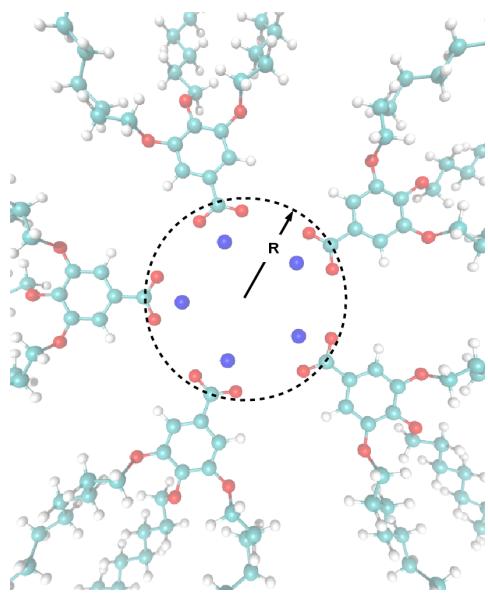


Figure A.6: When creating an initial configuration, we define the pore radius, R , based on the distance of the carbonyl carbon from the pore's central axis.

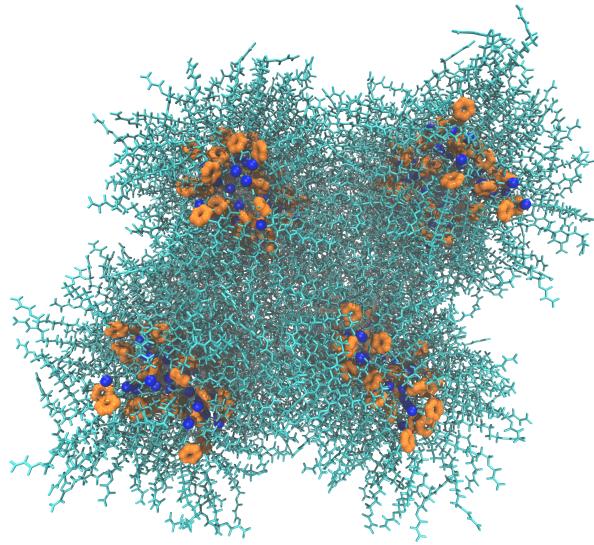


Figure A.7: A system that was built with an initial pore radius of 8 Å equilibrates to a structure that exhibits both cylindrical and slit-like pores. As pictured here, sodium ions are colored blue, carbon atoms in the aromatic ring of the head group are colored orange and all else is colored cyan.

Figure A.8 shows the structure of an assembly built with an initial layer spacing of 10 Å immediately after the restrained portion of the equilibration procedure. Since we used position restraints, the simulations were run in the NVT ensemble. When layer spacing is large, such as this situation, there is a significant amount of vacuum space which the monomer attempts to fill. Even if turning pressure control on allows the system to recover the geometry of the hexagonal phase, we would likely need much longer equilibration times, and it will almost certainly get trapped in a metastable configuration that bears no resemblance to the experimental profile.

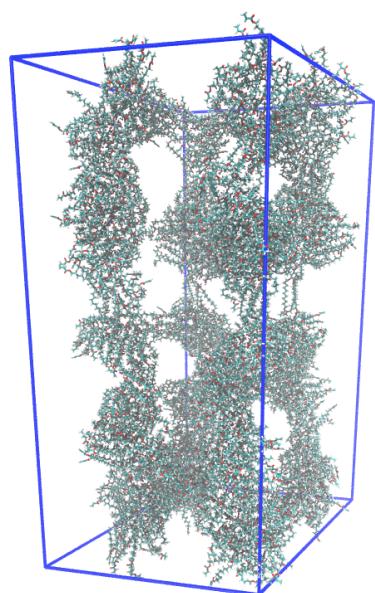


Figure A.8: When layers are initially stacked 10 Å apart and the system is equilibrated using the dry equilibration procedure, large vacuum gaps form as the monomers attempt to fill space.

A.7 Placement of monomer head groups

The monomer head groups consist of a substituted benzene ring which may participate in $\pi - \pi$ stacking interactions. There are three $\pi - \pi$ stacking modes known to occur in nature: sandwiched, parallel displaced and T-shaped (Figures A.9a–A.9c) [37]. We ruled out the T-shaped configuration because its vertical stacking distance is inconsistent with experimental spectroscopy. Additionally, steric interactions between the bulky alkane tails make this motif unfavorable. We chose to explore systems in both the sandwiched and parallel displaced configurations since their stacking distances are both consistent with experimental diffraction patterns.

Based on evidence from Section 2.3.2.4 of the main text, we believe it is most likely that the head groups stack somewhere between the sandwiched and parallel displaced modes which we studied (Figure A.10).

A.8 Equilibration Details

We developed equilibration schemes to create dry and wet configurations. Both schemes start with an initial configuration generated according to the previous guidelines. To create a dry configuration, we fix monomer head groups in the sandwiched or parallel-displaced configuration using position restraints with a force constant of 10^6 kJ mol $^{-1}$ nm $^{-2}$. We run a 50 ps simulation in the NVT ensemble which allows the monomer tails to settle without disrupting the ordering of the head groups. Doing so also mitigates system dependence on initial monomer configuration. Every 50 ps, we reduce the force constants by the square root of its previous value. Once the force constant is below 10 kJ mol $^{-1}$ nm $^{-2}$, we reduce the restraints in a sequence with values of 8, 3, 2, 1, and 0 kJ mol $^{-1}$ nm $^{-2}$ respectively. We allow the resulting unrestrained structure to equilibrate for 5 ns in the NPT ensemble with pressure controlled by the Berendsen barostat. Next, we run long NPT equilibration simulations for at least 400 ns using the Parrinello-Rahman barostat with a time constant of 10 ps.

In order to create a “wet” system, we solvated an initial configuration with water using gmx

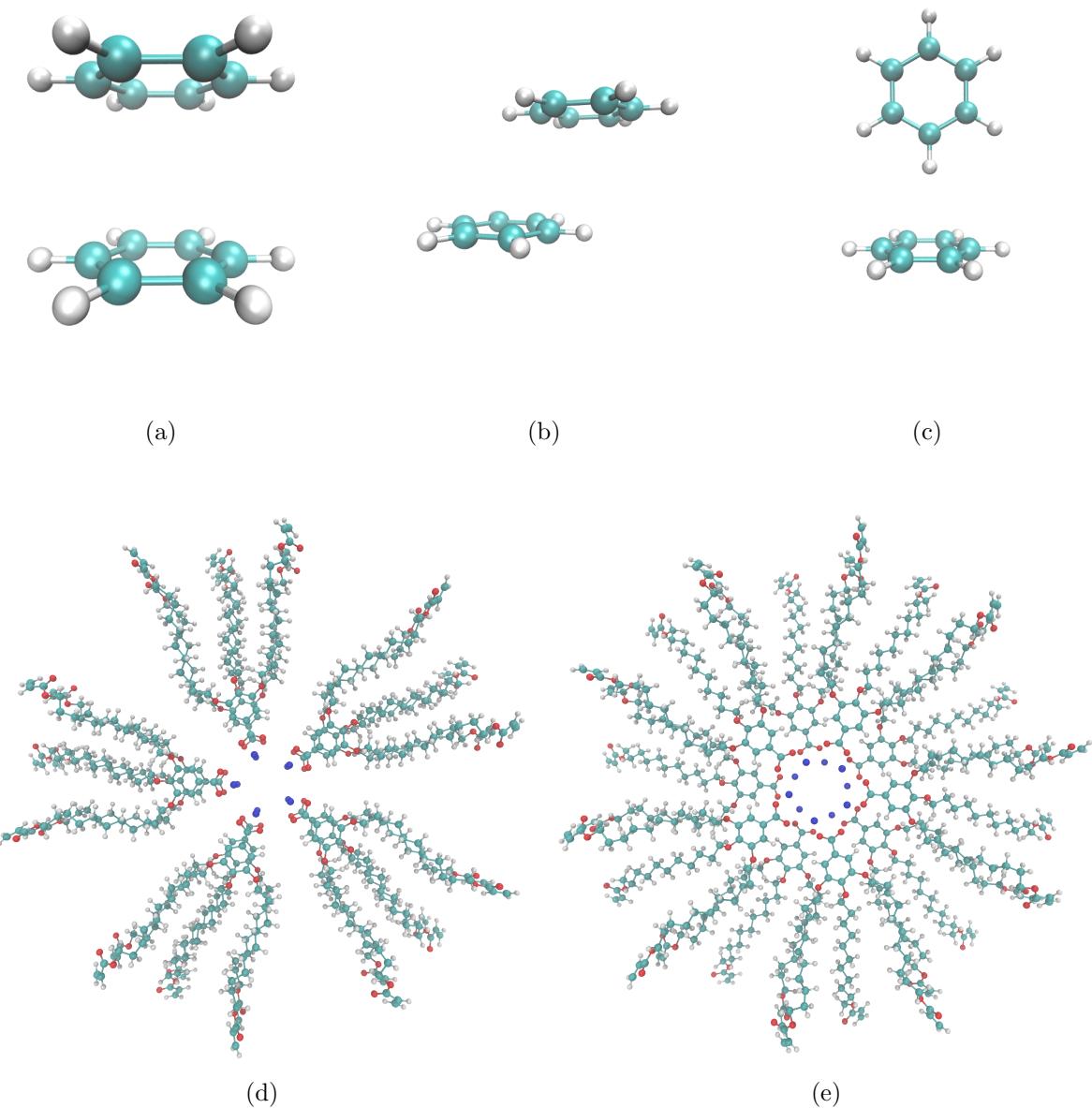


Figure A.9: (a) Sandwiched benzene dimers stack 3.8 Å apart. (b) Parallel-Displaced benzene dimers stack 3.4 Å vertically and 1.6 Å horizontally apart. (c) T-shaped benzene dimers stack 5.0 Å apart. (d) Monomers stacked in the sandwiched configuration (e) Monomers stacked in the parallel-displaced configuration.

solvate. We remove all water molecules placed outside the pore region. Then we randomly remove water molecules inside the pore region until the pores reach the desired concentration of water. The remainder of the equilibration follows the same procedure as the dry system.

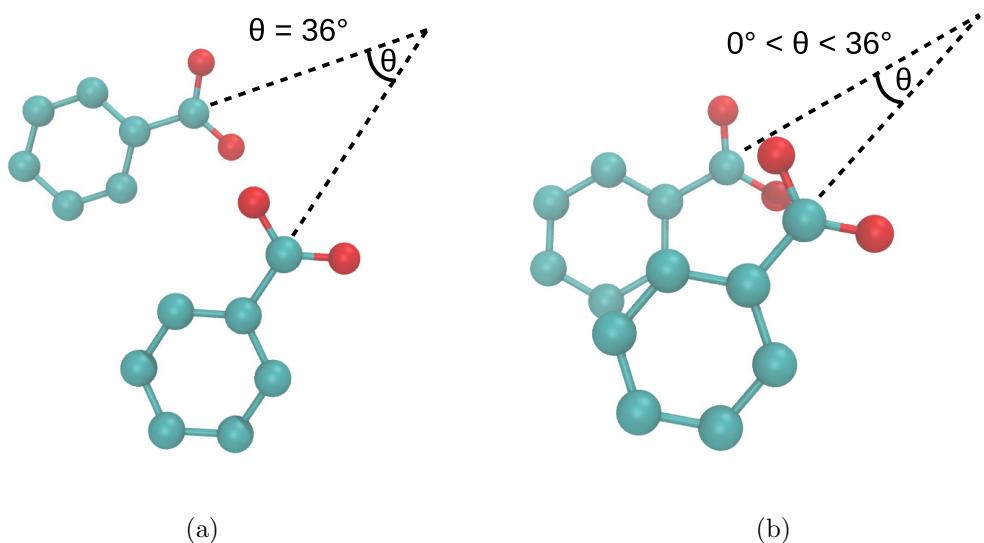


Figure A.10: Head groups are likely stacked in a configuration between sandwiched and parallel displaced. (a) In our parallel displaced configurations with 5 columns per pore, we placed head groups so the angle, θ , between vectors extending from vertically stacked head groups to the pore center equals 36° ($(360^\circ / n_{col}) / 2$, where n_{col} is the number of columns per pore). (b) In the true configuration, θ is likely distributed between 0° and 36° . When $\theta=0^\circ$, the system architecture is the same as a sandwiched configuration.

A.9 Calculation of pore-to-pore spacing statistics

We are interested in 5 pore-to-pore distances which should all be equal in a perfect hexagonal array (Figure A.5a). Only 4 of the 5 distances are independent. We can calculate a trajectory of spacing versus time for each of the 5 distances. We calculated the average pore-to-pore spacing and its uncertainty according to the following procedure:

- (1) We calculated the time when each of the pore-to-pore distances were equilibrated using `pymbar.timeseries.detectEquilibration()` [38, 99]. We began calculations after the largest of the five values.
- (2) We calculated how long it takes for the data in each of the 5 trajectories to become uncorrelated using `pymbar.timeseries.integratedAutocorrelationTime()` [38, 99].
- (3) We broke the full equilibrated trajectory into blocks of length τ , where τ is the maximum of the five autocorrelation times calculated. Each block contains five sub-trajectories of pore-to-pore spacings.
- (4) We generate statistics using the bootstrapping technique. For each bootstrap trial, we reconstruct an equilibrium trajectory by randomly sampling from the trajectory blocks.
- (5) The average pore-to-pore distance is the mean of all pore spacings among all bootstrap trials.
- (6) To calculate the uncertainty in pore-to-pore distance, we calculate the average pore spacing for each pore over all bootstrap trials. Using the 5 average pore-to-pore distances, we calculate the spread in a single pore-to-pore distance with:

$$s = \sqrt{\frac{1}{4} \sum_{i=1}^5 (x_i - \bar{x})^2} \quad (\text{A.1})$$

where \bar{x} is the average pore-to-pore distance. We divided the sum in Equation A.1 by 4 since there are only 4 independent pore spacings.

Alternatively, one can calculate the pore spacing as half of the x and y box vectors, however this technique does not fully capture the uncertainty in the individual pore-to-pore distances (Table A.3).

System	Pore Spacing (nm)	Pore Spacing (nm)
	Pore centers	$\frac{1}{2}(x \text{ box vector}) \text{ (nm)}$
Sandwiched, Ordered	4.18 ± 0.13	4.16 ± 0.01
Parallel Displaced, Ordered	4.17 ± 0.12	4.18 ± 0.01
Sandwiched, Disordered	4.05 ± 0.14	4.01 ± 0.01
Parallel Displaced, Disordered	4.03 ± 0.07	4.01 ± 0.01

Table A.3: The pore spacing measured in two ways agree within uncertainty. In the main text, we report values calculated by measuring the average distance between pore centers. Alternatively, one can estimate the pore spacing as half of the x and y box vectors. Both methods agree within uncertainty, but measuring the distance between pore centers does a better job of showing the spread of pore-to-pore distances.

A.10 2D Small Angle X-ray Scattering Data

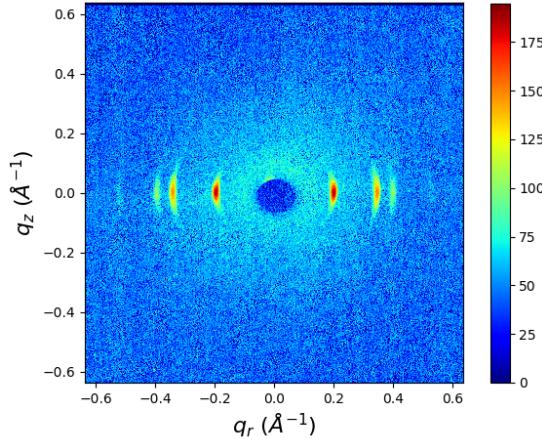


Figure A.11: Two dimensional small angle X-ray scattering experiments show reflections along the q_r axis at $q_z=0$. The spacing between the reflections is characteristic of a hexagonal phase and is directly related to the distance between pores. The reflections are arced because the pores are not perfectly aligned. A perfectly aligned system would show dots at each peak location. An isotropically aligned system would show concentric rings about the origin which pass through each peak location. The d_{100} and d_{110} peaks (the first two pairs of reflections from origin located at $|\mathbf{q}| \approx 0.18 \text{ \AA}^{-1}$ and $|\mathbf{q}| \approx 0.31 \text{ \AA}^{-1}$) are not visible in the 2D WAXS pattern. The edges of the d_{200} reflection (the third pair of reflections located at $|\mathbf{q}| \approx 0.36 \text{ \AA}^{-1}$) are slightly visible in the 2D WAXS pattern, so we used them to approximate the experimental intensity of R-pores.

A.11 Component radial density functions

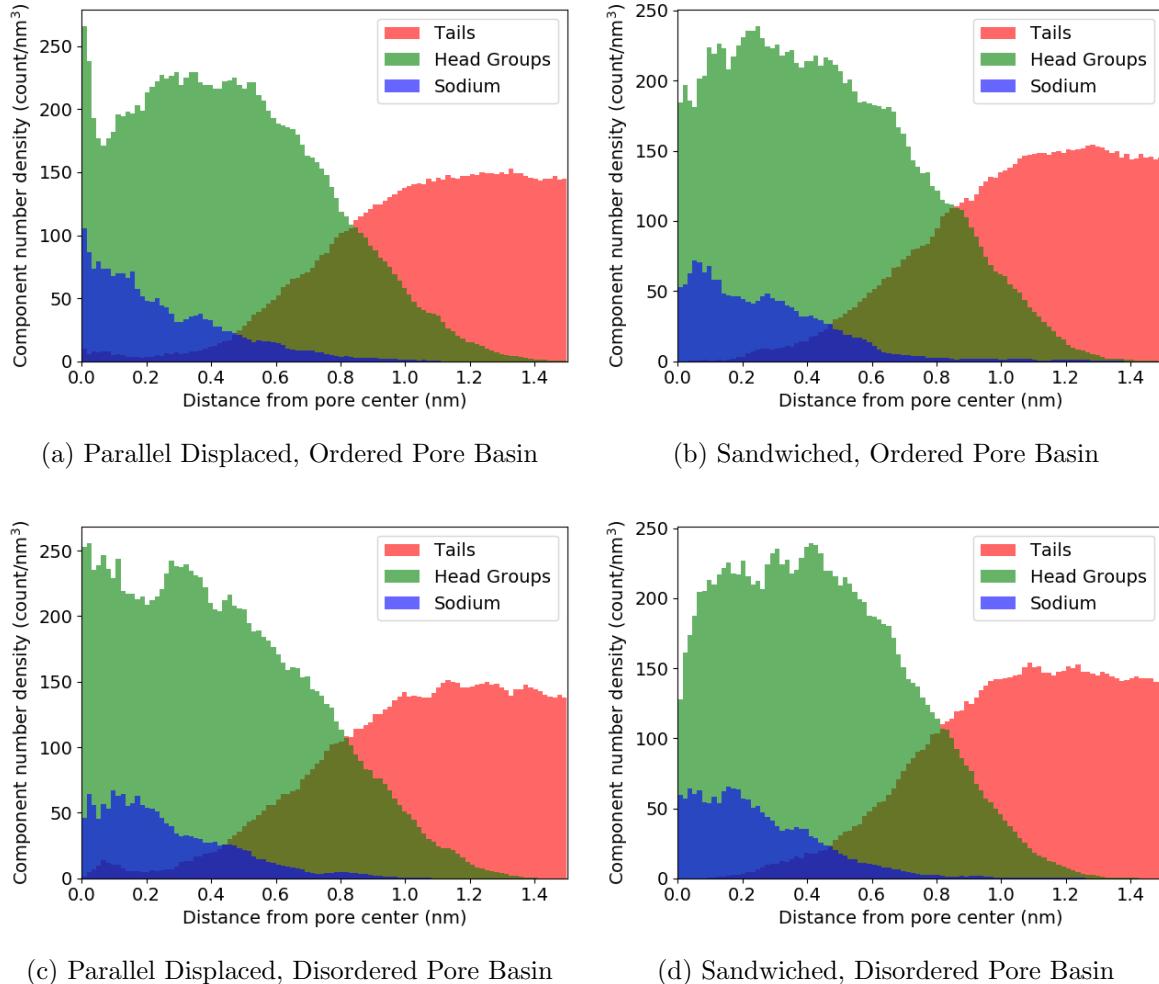


Figure A.12: In all cases, the component radial distribution functions exhibit a composition gradient transitioning from the hydrophilic to the hydrophobic regions. Systems with monomers initially stacked 3.7 \AA apart in the parallel displaced (a), and sandwiched (b) configurations both show the highest concentrations of ions and head groups away from the pore center. Systems with layers initially spaced 5 \AA apart (c) and (d), both show the highest concentrations of ions and head groups near the pore center, implying a more uniform, disordered pore.

A.12 Ionic Conductivity

In the main text, we calculated ionic conductivity using the Nernst-Einstein relationship. We also measured ionic conductivity using a second method, termed the collective diffusion model[55], for robustness (Figure A.15). The collective diffusion model measures the movement of the collective variable, Q , which is defined as the amount of charge transfer through the system and can be thought to represent the center of charge of the system. The conductance, γ , of the system can be calculated as:

$$\gamma = \frac{D_Q}{k_b T} \quad (\text{A.2})$$

D_Q is the diffusion coefficient of the collective variable Q and is calculated using the Einstein relation. We convert the resulting value to ionic conductivity by multiplying by channel length and dividing by the membrane cross sectional area. One can access a detailed derivation of the model elsewhere[55].

A.13 Cross-linking algorithm details

Crosslinking of the monomer Na-GA3C11 occurs via a UV initiated free radical polymerization (FRP) (Figure A.16). Head-to-tail addition takes place between terminal vinyl groups on each of the monomer tails. We only considered head-to-tail addition since it is the dominant propagation mode in the real system.

We based our cross-linking algorithm on the known reaction mechanism. FRPs require an initiator which bonds to the system, meaning new atoms are introduced into the system. For simplicity, we simulated the initiator as hydrogen and made it present in the simulation by including them as dummy atoms in all possible locations where an addition could occur. We carry out the cross-linking procedure iteratively. During each iteration, the algorithm selects eligible bonding carbon atoms based on a distance cut-off. The algorithm then updates the topology with new bonds and changes dummy hydrogen atoms to appropriate hydrogen atom types. We only considered head-to-tail addition due to its dominance in the real system [100]. We did not consider direction

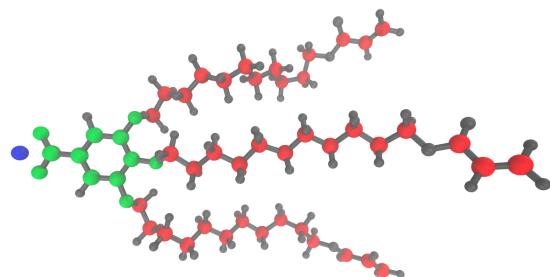


Figure A.13: The groups used for radial distribution calculations. Red atoms are in the tails group. Green atoms are in the head group region. The blue atom is sodium.

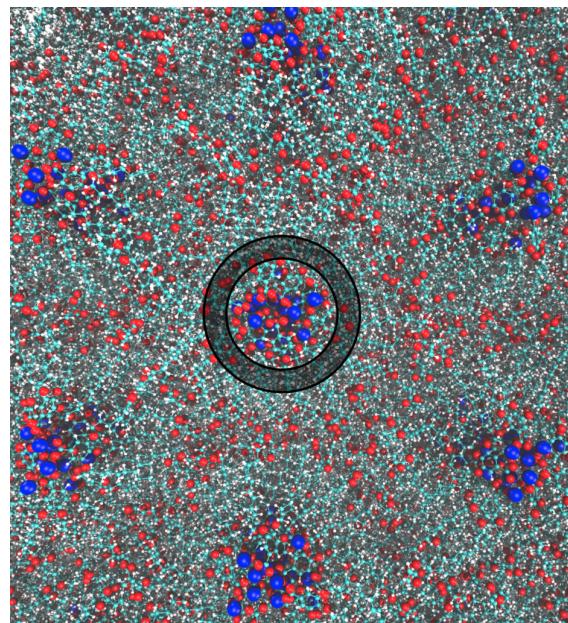


Figure A.14: Looking down onto the xy plane of the membrane, we binned the radial distance of all atoms in chosen groups from pore centers. The bins are defined by the annulus bounded by concentric circles centered at the pore centers, as shown. We normalize by dividing the count of atoms within each bin annulus by its volume.

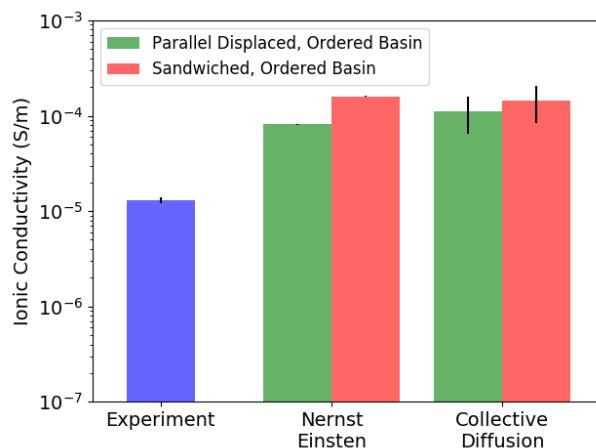


Figure A.15: The Collective Diffusion model and the Nernst-Einstein relation yield agreeing values of ionic conductivity for both types of ordered basin systems simulated. Both methods estimate the value of the ionic conductivity to be an order of magnitude higher than the experimental value. There is more noise in the collective diffusion model because there is inherently less data that can be used for its calculation. For that reason, we use only the Nernst-Einstein relationship in the main text.

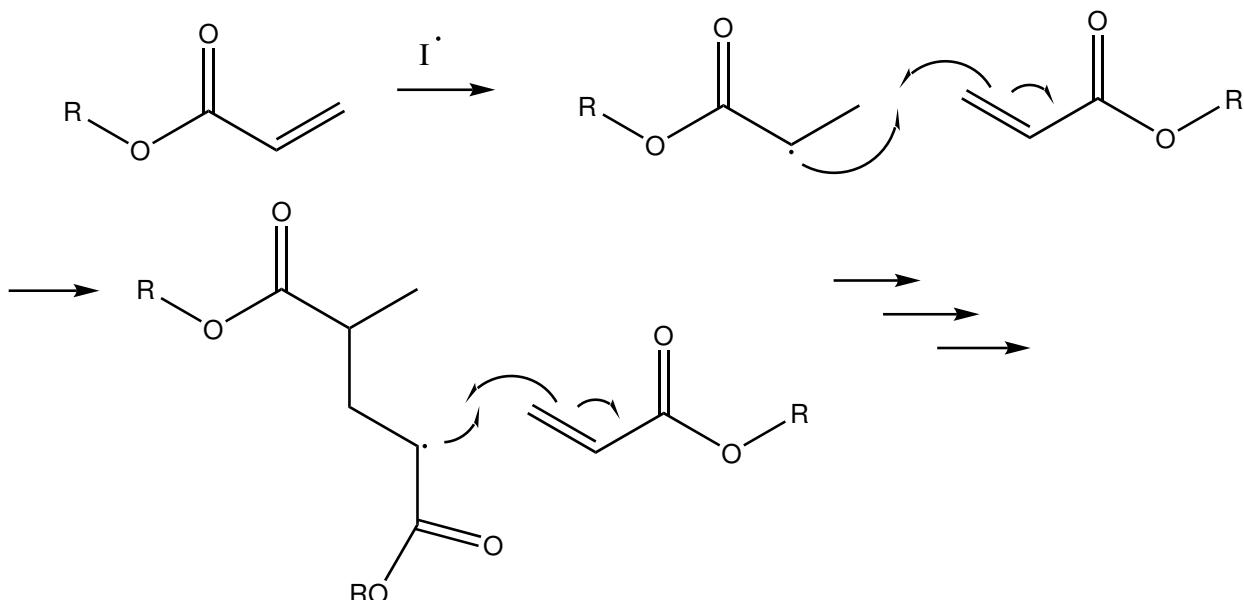


Figure A.16: Terminal vinyl groups present on each monomer tail react with free radical initiators to create monomers with terminal vinyl radicals. Vinyl radicals react with the vinyl groups of other monomers in order to propagate crosslinking.

of attack because the resultant mixture is racemic.

Our implementation requires long simulation times to achieve high cross-link densities. A

typical cross-linking procedure can take up to 24 hours. In order to collect equilibrated data, further NPT simulation is necessary. We typically run a cross-linked system for an additional 100 ns to allow the system to readjust. For those reasons we did not cross-link all systems tested.

A.14 The nematic order parameter

We calculated the nematic order parameter for our system in order to understand the degree of ordering among monomer head groups. Typically, the nematic order parameter is calculated for nematic liquid crystal systems which are characterized by unidirectional ordering of liquid crystal monomers. The preferred direction of monomers is defined by the unit director vector, \mathbf{n} . Assuming a single preferred direction of alignment, the nematic order parameter, S , is defined as: [25]

$$S = \frac{1}{2} \langle (3 \cos^2 \theta - 1) \rangle \quad (\text{A.3})$$

where θ is the angle between the molecular long axis and \mathbf{n} . In a perfectly ordered nematic liquid crystal system, the molecular axis of each monomer should be aligned with \mathbf{n} and give an order parameter of $S = 1$. We are interested in quantifying the degree of monomer head group alignment between systems. We use Equation A.3 to accomplish this by defining \mathbf{n} as the z -axis (or pore axis), and then measuring the angles, θ , between \mathbf{n} and the vectors perpendicular to the plane of the aromatic head groups (\mathbf{v} in Figure A.17).

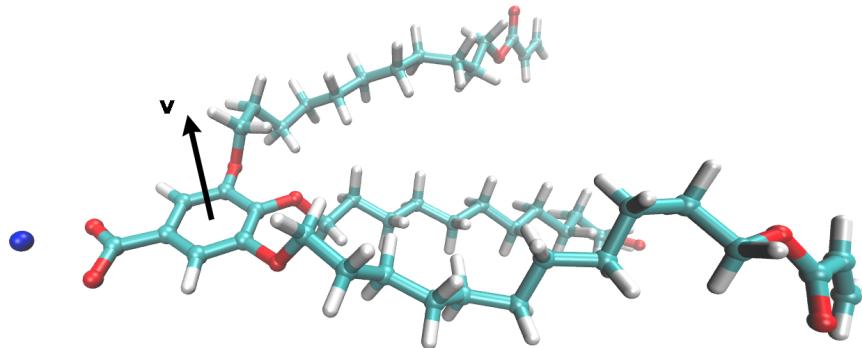


Figure A.17: We calculate the nematic order parameter by measuring the angle between \mathbf{v} and the z -axis.

The nematic order parameter is highest in sandwiched systems. Figure A.18 shows the distribution of angles between the nematic director and the vector perpendicular to the monomer head

groups. Systems in the ordered basin have higher nematic order parameters than their disordered basin counterparts. Both sandwiched systems have higher nematic order parameters than both parallel displaced systems. This may occur because monomers in the sandwiched configuration have their motion restricted by vertically adjacent monomers, while those in the parallel displaced configuration have more room to rotate.

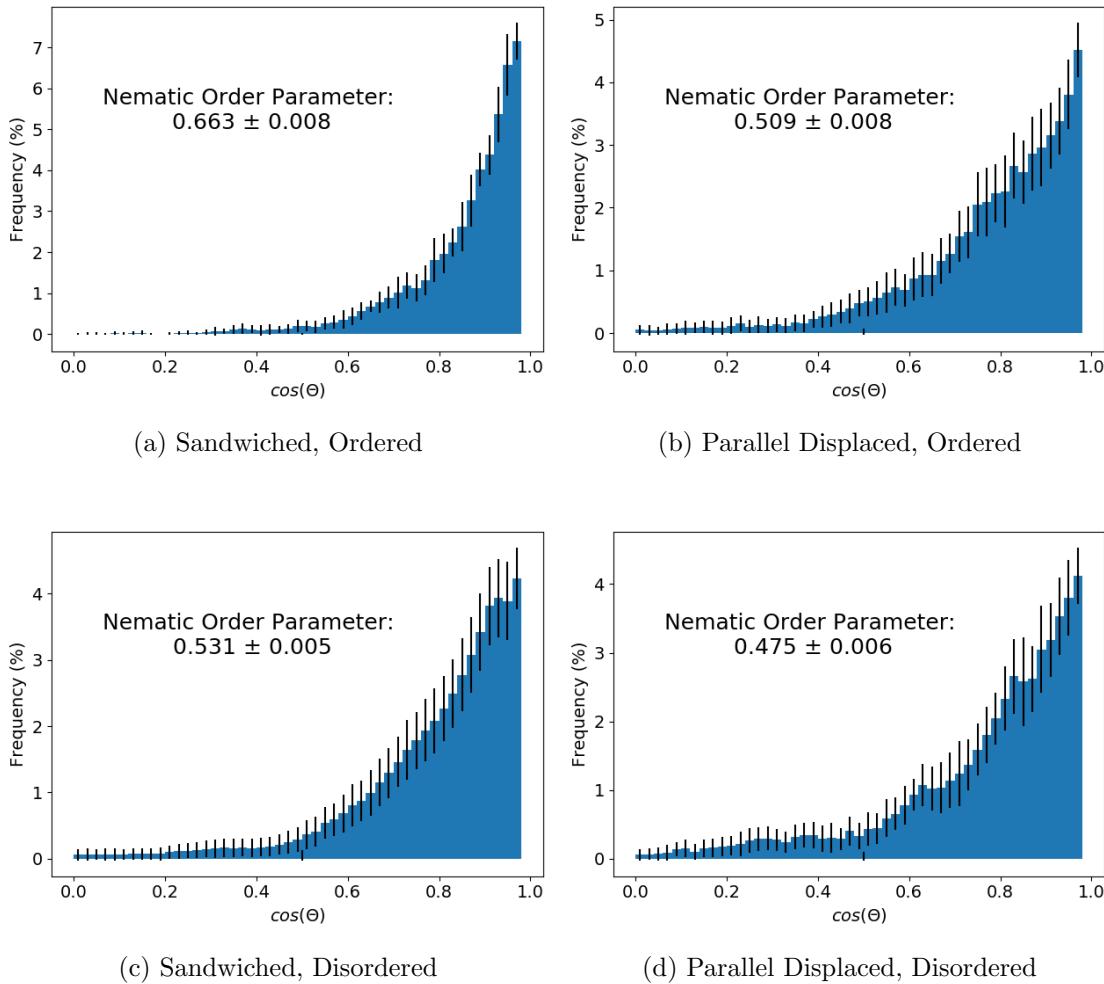


Figure A.18: The distribution of angles between nematic director vector (see Figure A.17) and the z -axis averaged over the equilibrated portion of each trajectory. The ordered sandwiched configuration has a higher nematic order parameter than the ordered parallel displaced configuration. The disordered sandwiched configuration has a higher nematic order parameter than the disordered parallel displaced configuration.

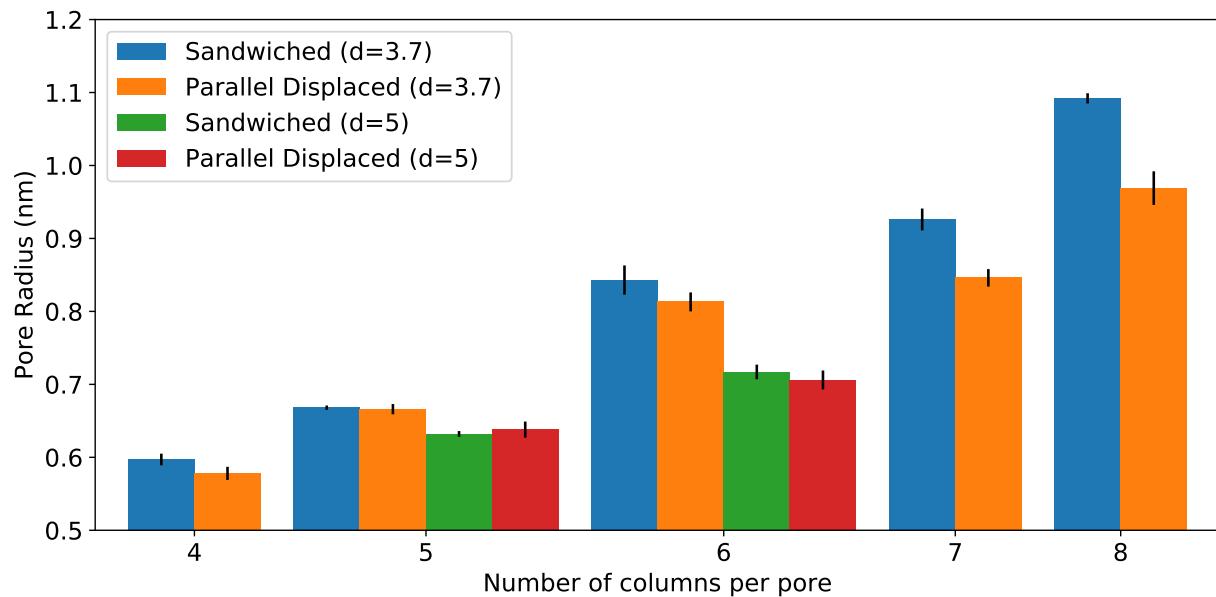
A.15 Pore radius as a function of monomers per column

Figure A.19: The pore radius increases as we increase the number of columns per pore.

A.16 Noise in simulated diffraction patterns

The magnitude of noise in the simulated diffraction patterns decreases as the number of independent configurations increases (Figure A.20). We simulated the structure factor of 100 randomly placed particles and varied the number of independent configurations (N). It is expected that the structure factor should resemble a two-dimensional Gaussian. We cannot sample enough independent configurations of our system to completely eliminate noise.

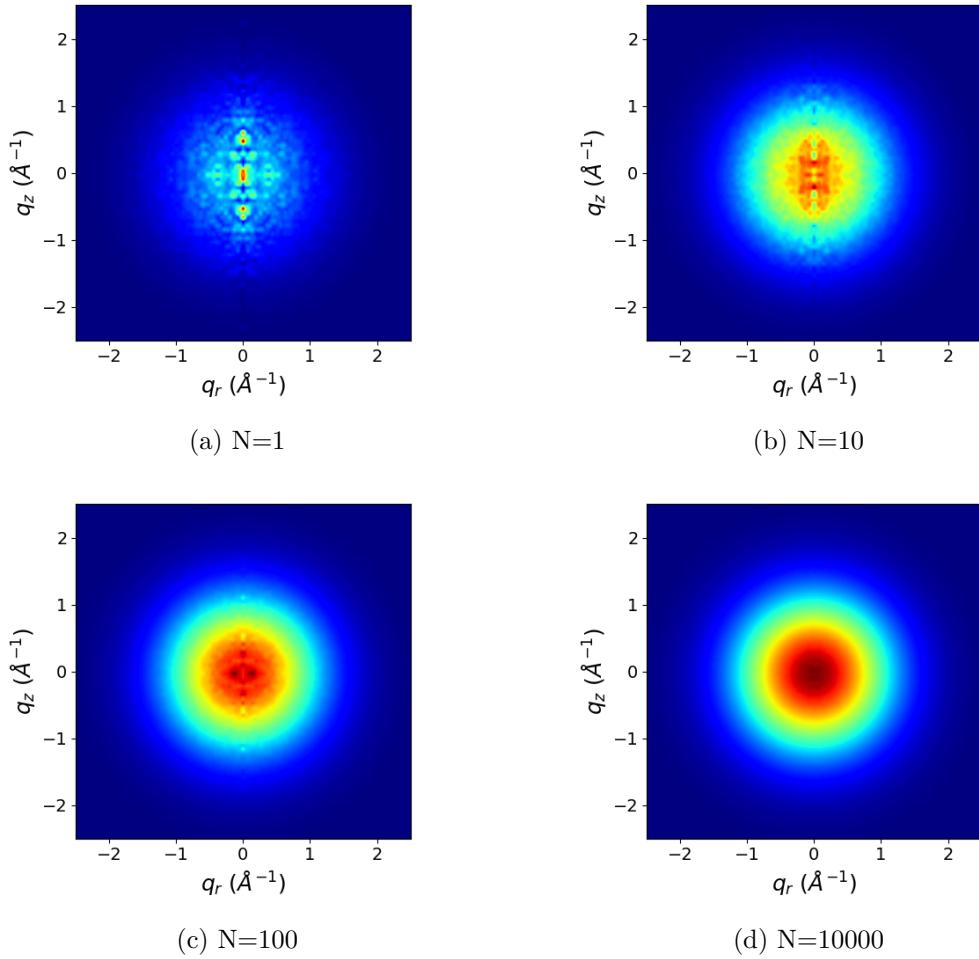


Figure A.20: The magnitude of noise, especially along $q_r=0$, decreases as the number of independent configurations (N) increases.

A.17 The effect of imposed tail tilt

Monomer tails prefer to lie flat rather than tilt with respect to the membrane plane. We created a monomer whose tails tilt ca. 30° with respect to the xy membrane plane. We added position restraints to the ends of the monomer tails with the same force constant as the head groups (see Section 2.2.5). As we slowly reduce the force constants, the tails begin to tend towards a negligible tilt angle (Figure A.21).

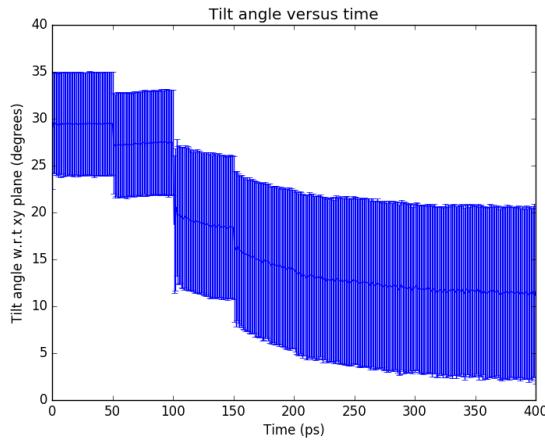


Figure A.21: As we reduce the force constants applied to the end of each tail of an initially tilted monomer configuration, the average tilt angle with respect to the membrane plane decreases until it is nearly negligible.

A.18 Tail organization

In order to understand the origin of R-spots, we looked for ordering among the tails. We divided the tail into three sections: tail-fronts, tail-middles and tail-ends. The sections are illustrated in Figure A.22.

We plotted the center of masses of each tail section and measured the angle between each center of mass and its nearest neighbor center of masses of the same tail section. For example, Figure A.23a shows a 3D plot of the center of masses of each tail front. We wanted to know the average location of nearest neighbors to each center of mass. The red bead in Figure A.23a is a randomly selected center of mass. The green beads are the nearest neighbor center of masses to the

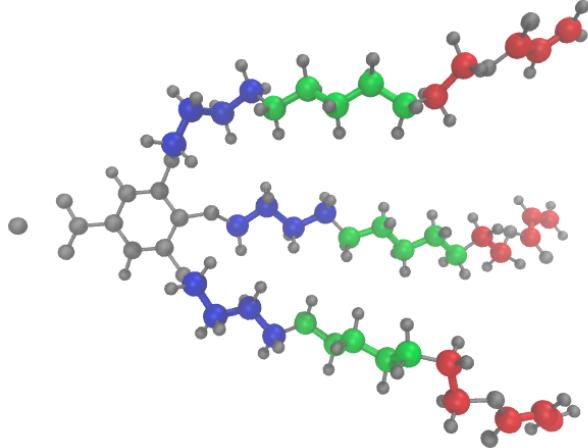


Figure A.22: We divided the tail into three regions: tail-fronts (blue), tail-middles (green) and tail-ends (red).

red bead. We measured the angle between the red bead and each green bead with respect to the xy plane. We repeated the same calculation for each center of mass at each simulation frame and were able to plot the average distribution shown in Figure A.23b. We repeated the calculation for the tail middles (Figure A.23c) and the tail ends (Figure A.23d). The amount of order decreases as we explore packing of tails farther from the head group.

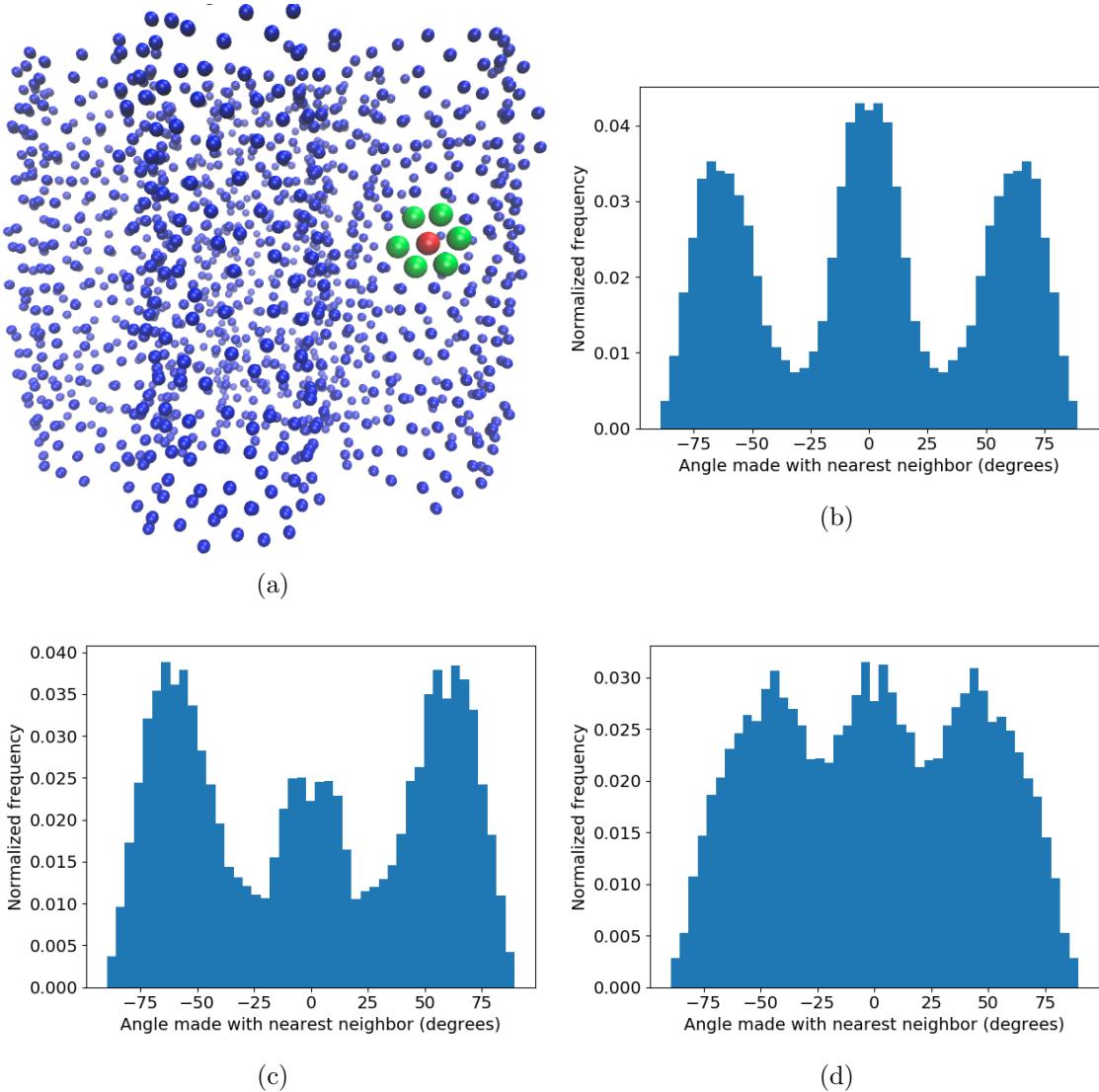


Figure A.23: Monomer tails pack together hexagonally. (a) The center of mass of each tail-front (See Figure A.22) is visualized as a blue sphere. The red sphere highlights an example of a tail-front center of mass with its nearest neighbors (green spheres) surrounding it in a hexagonal pattern. (b) We calculated the angle between each tail-front center of mass and its nearest neighbor tail-front center of masses and plotted the distribution. There are distinct peaks ca. -60° , 0° and 60° which is indicative of hexagonal packing. (c) We repeated the same procedure outlined in (a) and (b) with the tail-middles. There is still a high degree of order. (d) We repeated the procedure again with the tail-ends. There is far less order at the ends of the tails where there is more space to fill.

A.19 Correlation Functions

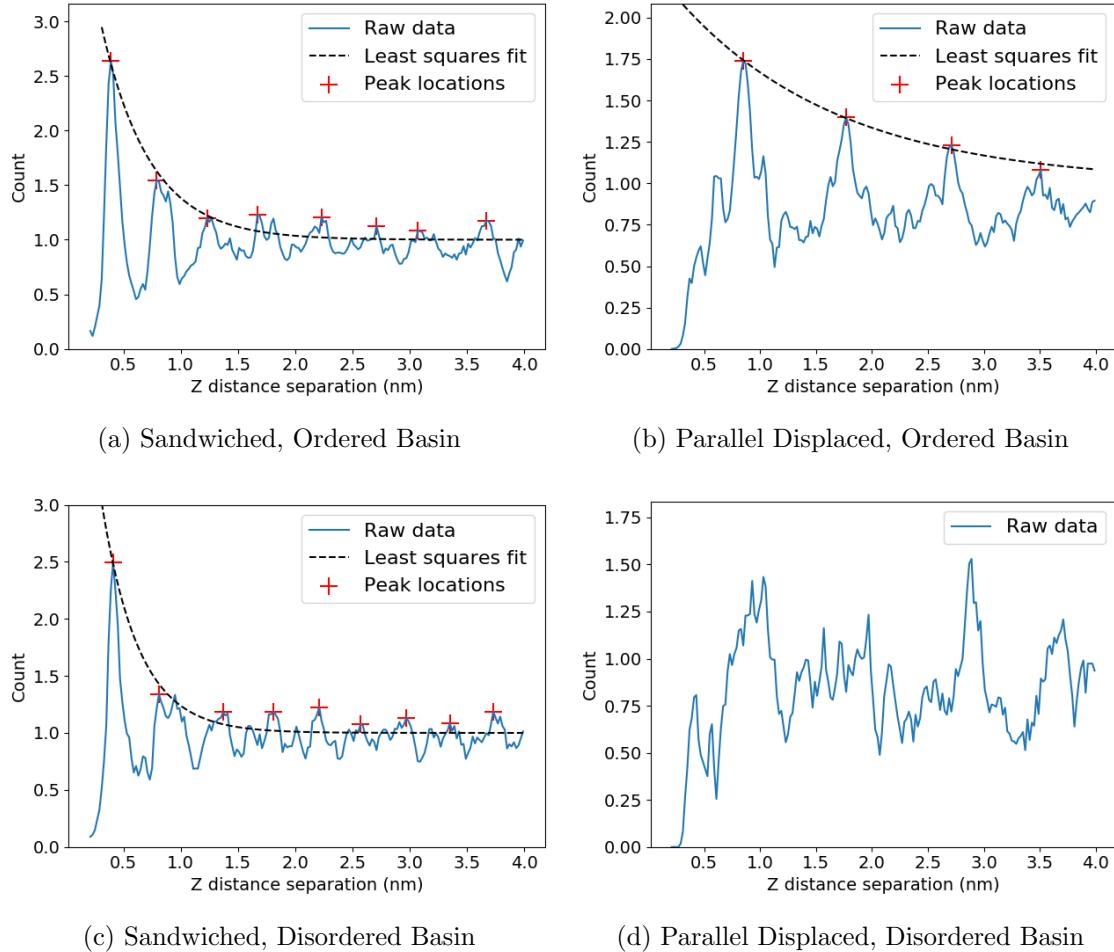


Figure A.24: 1D correlation functions of the center of masses of aromatic head groups, $g(z)$, show decaying oscillatory behavior and have consistent correlation lengths with experiment. We calculated the correlation length by fitting a decaying exponential function (Equation 2.2 of the main text) to the peaks of $g(z)$. The correlation length is longer for ordered basin systems (Table 2.1 of the main text). We did not attempt to calculate the correlation length for (d) because there are no clear peaks. We assume that this correlation length is less than the vertical distance between monomers.

The system size, in the z -direction, does not significantly alter $g(z)$. In Figure A.24, oscillations do not fully decay, so a taller system may be necessary to fully capture the correlation function. We therefore equilibrated a sandwiched system with twice as many monomers per column so that the system size doubled in the z -direction. The correlation length increases modestly

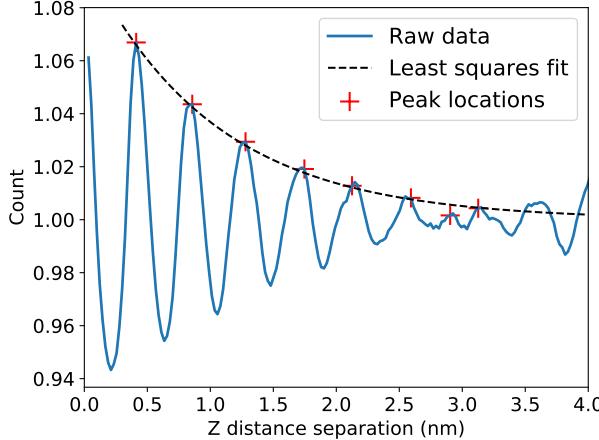


Figure A.25: When we average all z -slices of the 3D correlation function, the correlation length of the ordered sandwich configuration nearly triples compared to that shown in Figure A.24a.

from 4.5 ± 0.4 to 7.3 ± 1.2 Å. However, visually, the correlation functions are nearly identical and oscillatory behavior persists throughout $g(z)$ in both cases (See Figure A.4a).

It seems likely that the discrepancy between the shape and intensity of $R-\pi$ generated from our simulations versus experiment is primarily due to highly correlated columns in our highly symmetric initial configurations. If the columns were uncorrelated, $g(z)$ of the ordered sandwiched system would very closely resemble Figure A.24a when we average all z -slices of the 3D correlation function. Contributions to $g(z)$ from head groups in other columns within the same pore, and with head groups in different pores would cancel out. However, when we include all z -slices in $g(z)$ for our system, rather than restricting it to those z -slices within 2.1 Å of the center of the correlation function (see Section 2.2.6.4), $g(z)$ shows oscillatory behavior and its correlation length nearly triples to 12.4 ± 0.7 Å. (Figure A.25). The peaks of $g(z)$ are actually reinforced by head groups in neighboring columns which is not expected in a system where columns act independently.

A.20 Structure factors of simplified systems

In the Section 2.3.2.4 of the main text, we used simplified systems made up of point scatterers in order to understand the main factors that influence the cross-sections of $R-\pi$. We showed that both the shape and the intensity of $R-\pi$ are strongly influenced by the degree of correlation between

columns. Here we show the influence of varied amounts of noise on the xy plane and in the z -direction for select cases.

Increasing thermal noise in the z -direction will reduce the intensity of R- π , however the q_r profile remains unchanged. When we increased the thermal noise of the simplified parallel displaced system with fully correlated columns in the z -direction by 12% and 27%, we saw a 3 and 15-fold decrease in the maximum intensity of R- π respectively (Figure A.26a). Despite the decrease in intensity, the 3 cross-sections have the same shape (Figure A.26b), whose sharp Bragg-like peaks are not consistent with experiment.

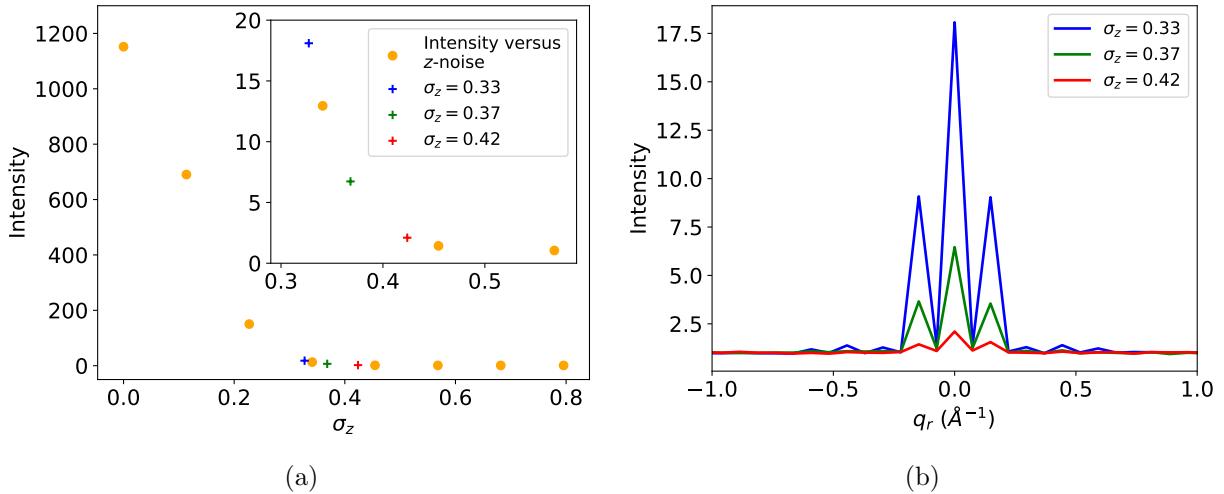


Figure A.26: We can decrease the maximum intensity of R- π by increasing thermal noise in the z -direction; however, the q_r cross-section profile does not change. (a) The intensity of R- π drops precipitously as we increase σ_z . We can decrease the intensity of R- π , relative to the background intensity, relative to systems with simulation levels of disorder (blue) by a factor of 3 when we increase σ_z by 12% (green) and by a factor of 15 when we increase σ_z by 27% (red). (b) Although the maximum intensity of R- π decreases with increasing σ_z , the peak shape remains the same.

Increasing thermal noise in the xy plane causes the FWHM of the q_r cross-section of R- π to, somewhat counterintuitively, decrease. Using a system with uncorrelated columns, we modified the disorder in both the r and θ directions by multiplying their values by a factor of 0.5 and 2 (Figure A.27). When we cut the noise in half, the FWHM increases by 88%. When we double the noise, the FWHM decreases by 51%. As we have fit the data in Figure 2.8b, the FWHM of R- π

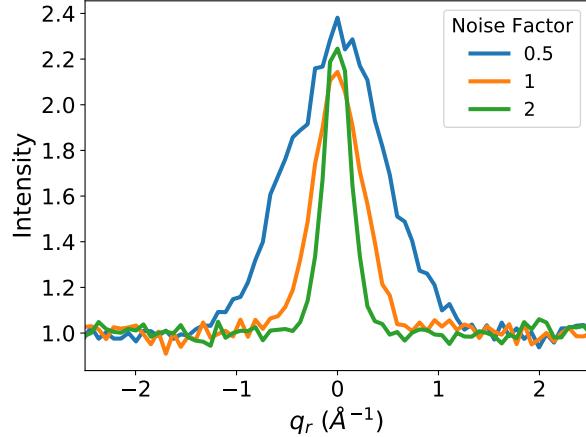
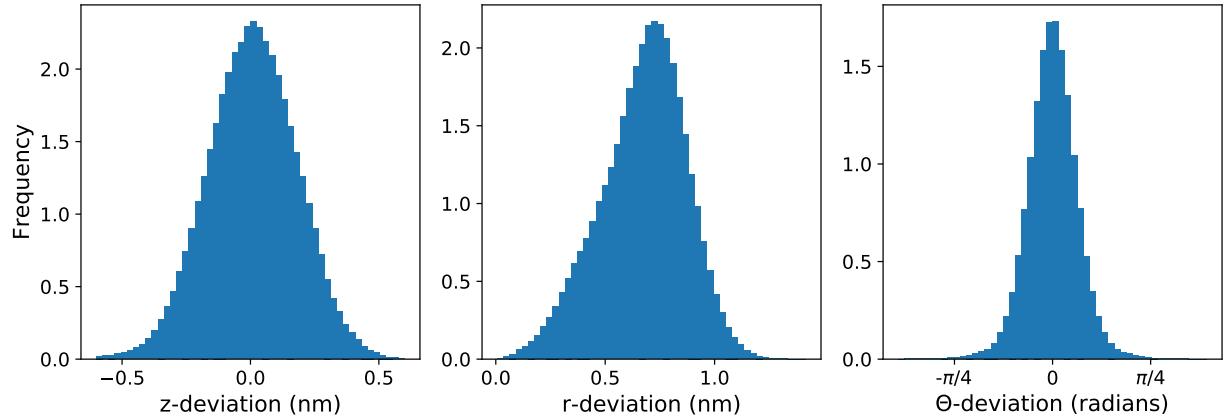


Figure A.27: Increasing the amount of disorder in the xy plane decreases the FWHM of the q_r cross-section of R- π . When we cut the amount of r and θ disorder in half (Noise factor = 0.5), the FWHM increases by 88%. When we double the disorder (Noise factor = 2), the FWHM decreases by 51%.

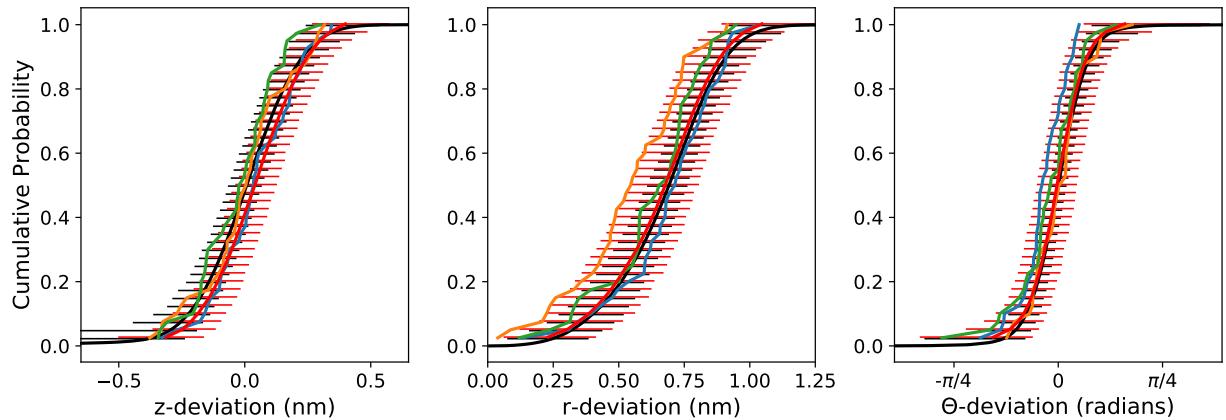
agrees well with experiment, meaning the quenched configuration contains about the right amount of disorder on the xy plane. However, the relatively low quality of the fit implies that the FWHM of a fixed Lorentzian fit to data from a single simulation might not be the best comparison to experiment and that some form of initial configuration dependence causes extra structure.

A.21 Ensemble of ordered sandwiched configurations

Here we show the results of the same analysis shown in Figures 2.10 and 2.11 of the main text for the ordered sandwiched ensemble of configurations.



(a) Distributions of deviations from ideal positions generated with data from all independent simulations



(b) Empirical cumulative distribution functions generated from (a) and from simulation ensembles.

Figure A.28: (a) The distributions of the head group COM deviations from their idealized positions generated using pooled data from all frames of each independent configuration are symmetric, implying that there is equal probability for a head group to displace in the positive or negative z , r or θ -direction. (b) The COM position of a given head group is displaced randomly upon quenching. The ECDF generated from the pooled distributions (black) in (a) agree with the means of the ECDFs generated from each COM (red). The red error bars are larger than the black error bars since there is a wider distribution of mean COM deviations than the mean deviations of distributions sampled from the full distribution.

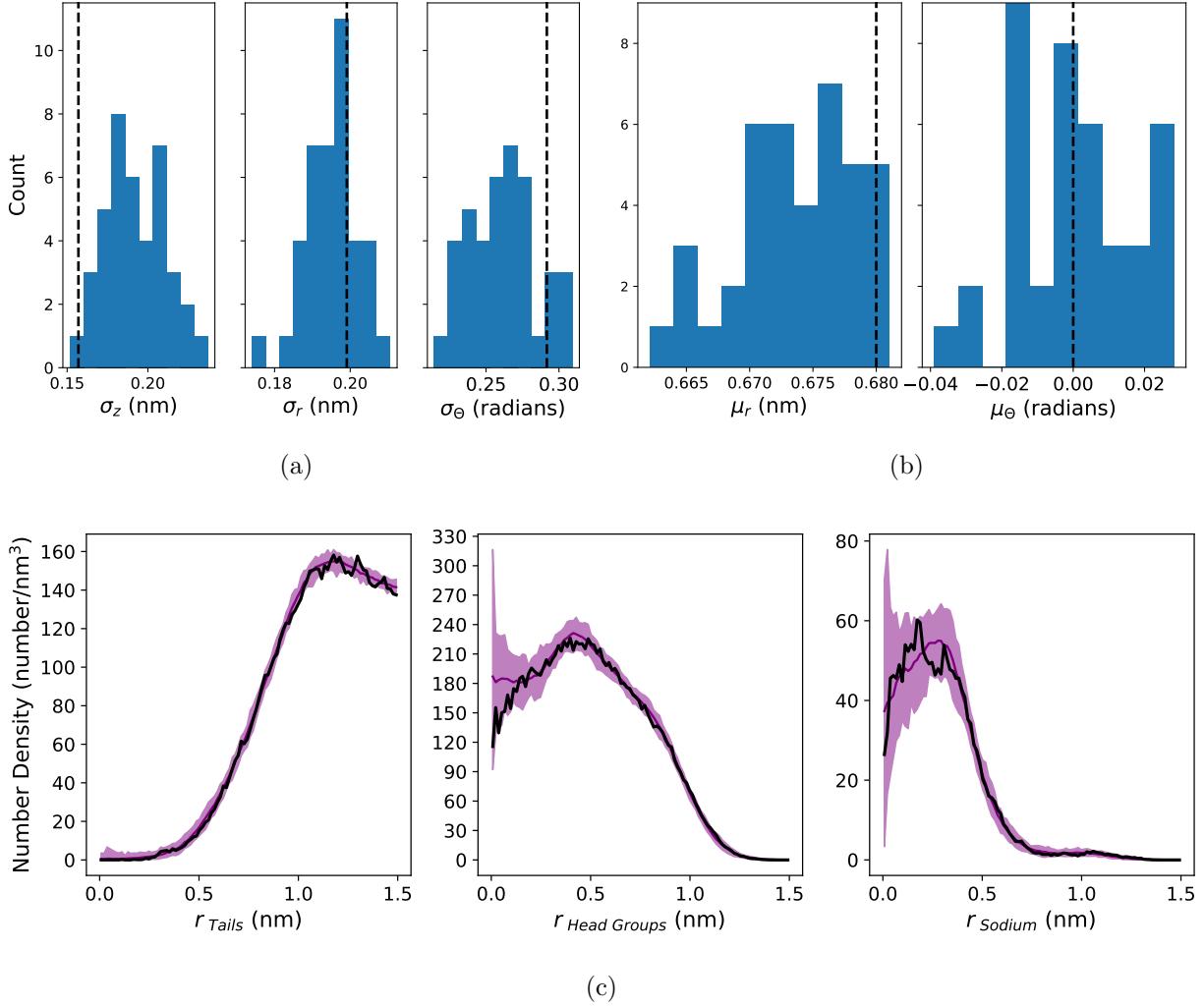


Figure A.29: (a) The standard deviation of the distribution of quenched disorder from the first 5 ns of the main ordered sandwiched system studied in this paper (black dashed line) is in agreement with the distribution of quenched disorder standard deviations calculated from the ensemble of simulations (histogram). (b) The mean values of r and θ from the first 5 ns of the main system trajectory (black dashed line) is in agreement with the distribution of mean values calculated from each simulation in the ensemble (histogram). The mean values of z are necessarily 0 so they are not plotted. (c) The radial densities of tail atoms, head group atoms and sodium atoms calculated from the first 5 ns of the main system simulation trajectory (black lines) and from the ensemble of trajectories (all other lines) look qualitatively similar.

A.22 Carboxylate Dihedrals

It is most energetically favorable for the carboxylate group attached to the monomer head groups to stay coplanar with the phenyl group to which it is attached (Figure A.30). Although GAFF may over-predict the penalty for deviating from its minimum energy configuration relative to those calculated by Rakitin and Pack [101] and Nelson and Borkman [102], they all show that there is an appreciable barrier to rotation about the carboxylate-phenyl bond. The potential energy reaches its maximum at rotation angles corresponding to a carboxylate group that is antiparallel to the plane of the phenyl ring to which it is attached (90° and 270°).

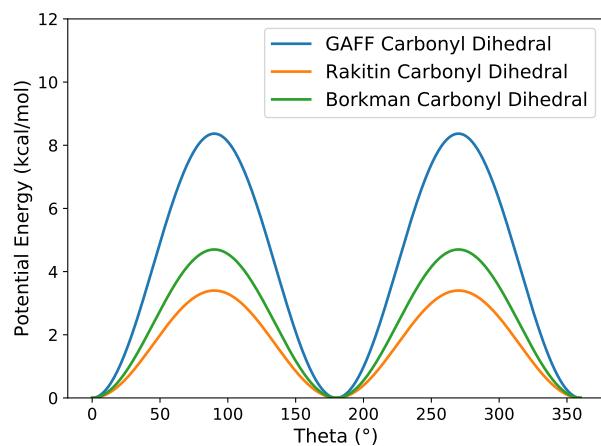


Figure A.30: A potential energy diagram for the dihedral angle about the carbon-carbon bond of a carboxylate group attached to an aromatic ring, using parameters from GAFF (blue), Rakitin (orange) and Borkman (green). It is energetically favorable for the carboxylate group to stay in plane with the aromatic ring. There is a large energy penalty as the dihedral becomes antiparallel to the plane of the aromatic ring.

A.23 All solvated systems

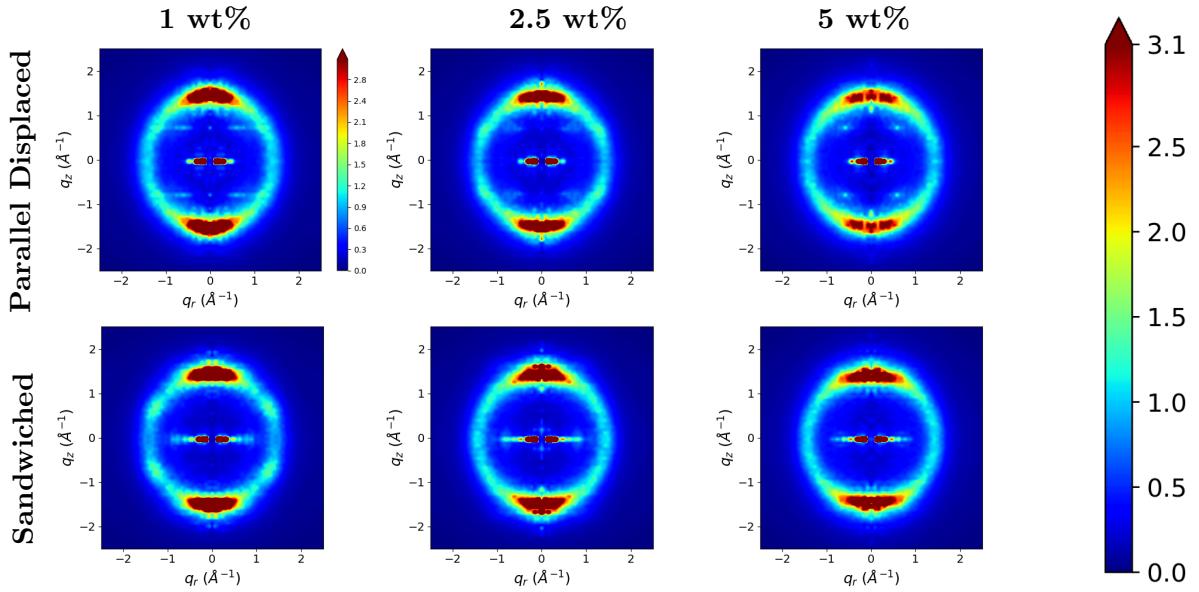


Figure A.31: Adding 1 wt% water to the ordered basin dry systems causes an increase in the intensity R-spots. R-double becomes visible in the parallel displaced configuration. Additional amounts of water increase discrepancies with experiment

A.24 The role of water in the appearance of R-double

The appearance of R-double is likely a consequence of structure induced when vertically stacked monomer pairs share hydrogen bonds with a common water molecule (See Section 2.3.2.9 of the main text). It is not a consequence of the structure of water itself because R-double still persists if we removed water from the trajectory which gives rise to the feature (Figure A.32). We quantified the occurrence of pairing interactions by binning the z -positions of the center of mass of the monomer head groups into 20 bins, the same number as the monomers per column. This works relatively well since there is a high degree of correlation between columns in our system. Peaks in the distributions represent the scenario where a monomer head group shares a water molecule with a head group vertically above it (See Figure A.33a). We calculated the discrete Fourier transform of the distributions in Figure A.33a which we used as a rough indicator of whether the head group arrangement will lead to R-double (Figure A.33b). We see the strongest indicator of asymmetric monomer stacking in Pore 1 where the peaks at 0.76 \AA^{-1} are clearly distinguishable, meaning there is periodicity at twice the π -stacking distance.

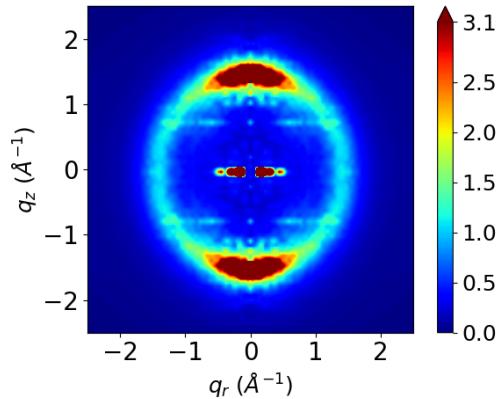


Figure A.32: When water is removed from the trajectory which produces R-double, R-double is still present since it is the result of uneven spacing between head groups.

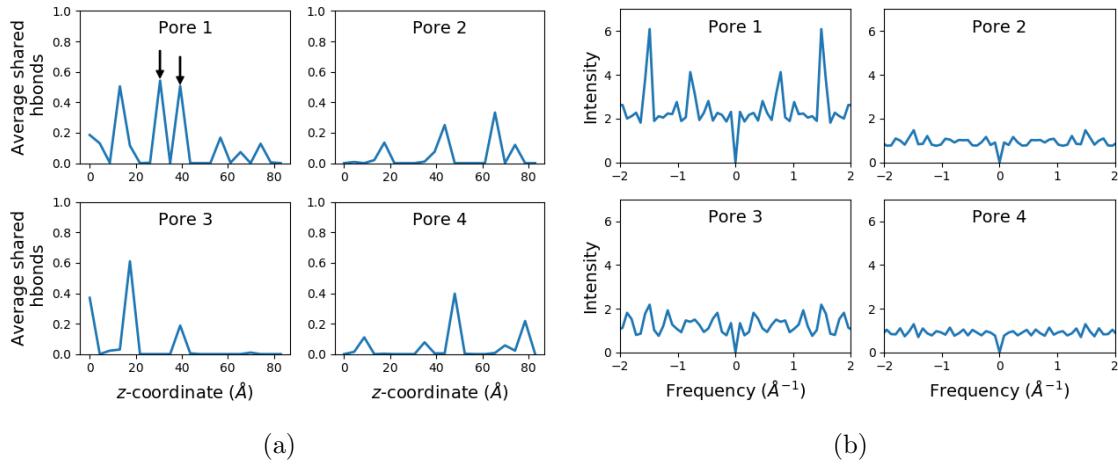


Figure A.33: (a) We group the z -coordinate of the center of mass of monomer head groups into 20 bins. Peaks in the distributions indicate a shared hydrogen bonded water molecule between a monomer head group and one in the bin vertically above it. The distance between the peaks roughly corresponds to the distance between pairs of monomers. The distance between peaks in pore 1 (indicated by arrows) is equal to two times the π -stacking distance. In this scenario, monomers are no longer evenly spaced in the z -direction since they are pulled closer together by hydrogen bonds. (b) We performed discrete Fourier transforms on each distribution in (a). In Pore 1, there is periodicity every 8.3 \AA , twice the π -stacking distance, as indicated by the intensity at 0.76 \AA^{-1} . In this case, the center of mass of two adjacent pairs is separated by twice the average monomer stacking distance, while the four monomers involved are unequally spaced. These conditions seem likely to give rise to R-double.

A.25 Diffusion of monomers

We calculated the diffusion coefficient of the monomer head groups by measuring the slope of the linear region of their mean squared displacement curves accumulated over the length of the equilibration simulations (Table A.4).

System	Diffusion Constant (m^2/s)
Sandwiched, Ordered	1.4×10^{-14}
Parallel Displaced, Ordered	2.0×10^{-14}
Sandwiched, Disordered	2.0×10^{-14}
Parallel Displaced, Disordered	3.5×10^{-14}
Parallel Displaced, 1 wt % water	7.1×10^{-14}

Table A.4: The diffusion constants for all systems are on the order of $10^{-14}m^2/s$.

A.26 Autocorrelation of monomer dihedrals

In Section 2.3.4 of the main text, we calculated the autocorrelation function of the outer tail either dihedrals pictured in Figure A.34. The plots of the autocorrelation functions versus time lag are shown in Figure A.35.

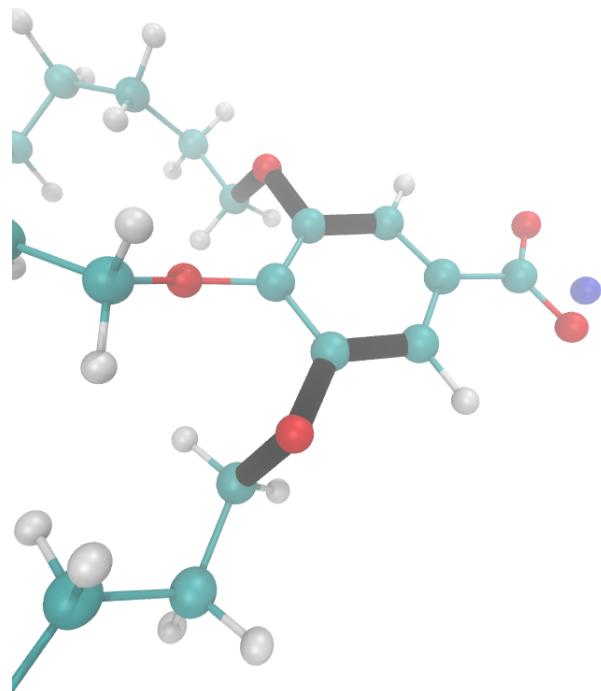


Figure A.34: We measured the autocorrelation functions of the dihedrals highlighted in black. We only included dihedrals associated with the outer tails in our calculation since the position of the center tail leads to fundamentally different interactions and dynamics.

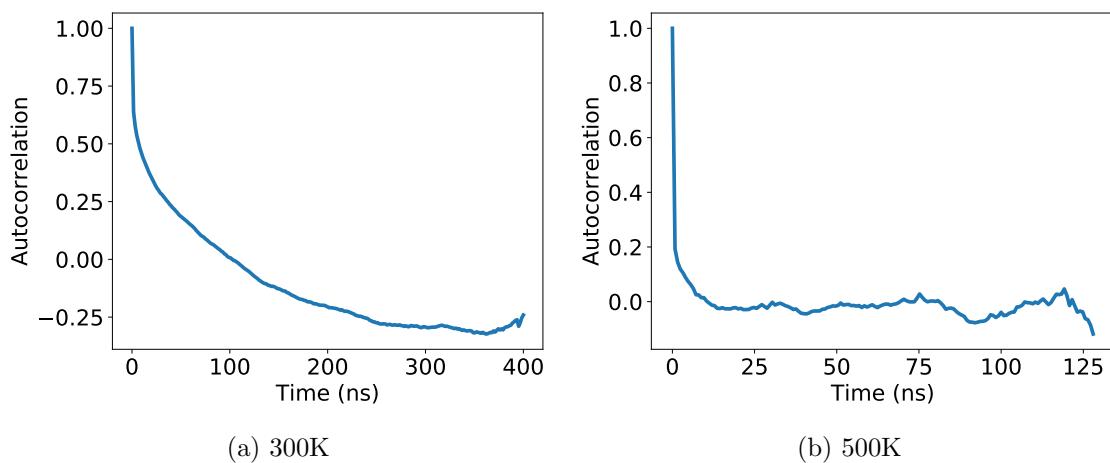


Figure A.35: The outer tail ether dihedrals become decorrelated on a reasonable timescale when we raise the system temperature. (a) The dihedrals do not become decorrelated for ca. 100 ns when the system is equilibrated at 300 K. (b) The ether dihedrals become decorrelated after 11 ns when the temperature of the system is increased to 500 K.

A.27 Influence of Cross-linking on physical properties

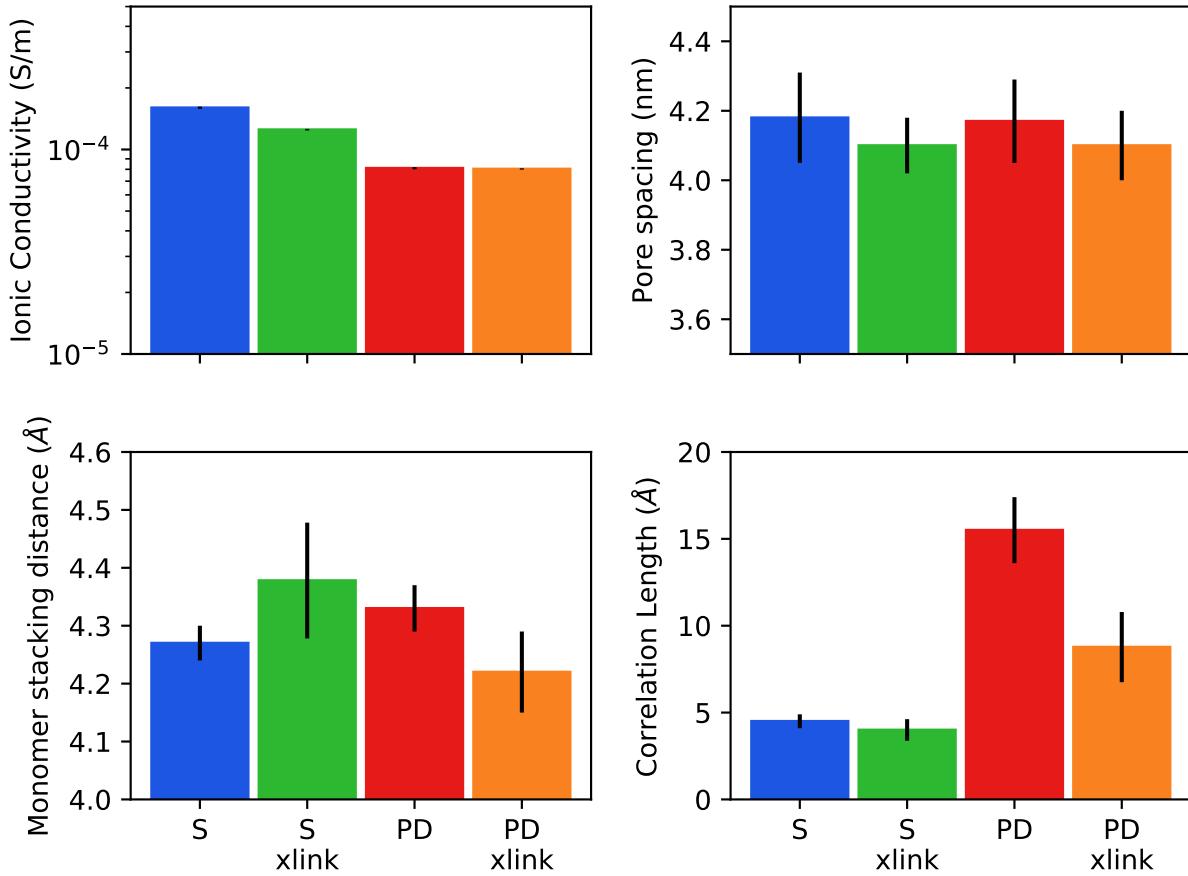


Figure A.36: There are minor differences between the physical properties of a crosslinked (xlink) system versus an uncrosslinked system. The sandwiched configuration (S) exhibits a smaller decrease in its ionic conductivity, while the parallel displaced (PD) configuration remains constant. The pore spacing of both systems decreases upon cross-linking. The vertical distance between stacked monomers increases in the sandwiched configuration and decreases in the parallel displaced configuration. The correlation length decreases in both configurations.

chapterSupporting Information to Chapter 3

A.28 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. Documenta-

tion for the `LLC_Membranes` repository is available at <https://llc-membranes.readthedocs.io/en/latest/>.

Table A.5 provides more detail about specific scripts used for each type of analysis performed in the main text.

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

Table A.5: 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.

A.29 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 A.37a). 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 A.37b). 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 A.37c). 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.

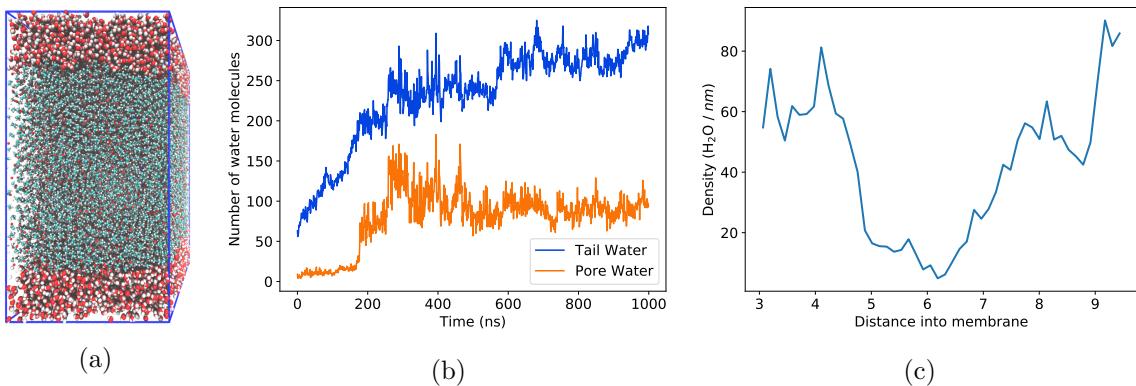


Figure A.37: (a) Using an equilibrated dry configuration, we inserted a 3 nm thick layer of water between periodic copies of the system in the z -direction. (b) Water slowly enters the membrane. Most water enters the tail region where the density of monomers is lowest. Water entering the pore plateaus 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.

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 A.6.

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 A.6: We chose a diverse combination of initial pore and tail water contents in order to study its effect on equilibrium water content.

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 A.38a and A.38b), 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 A.38c and A.38d) 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 can observe mechanisms as a function of water content in each region.

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 A.38. We considered the water content equilibrated once the water contents plateaued. The 5 wt% system

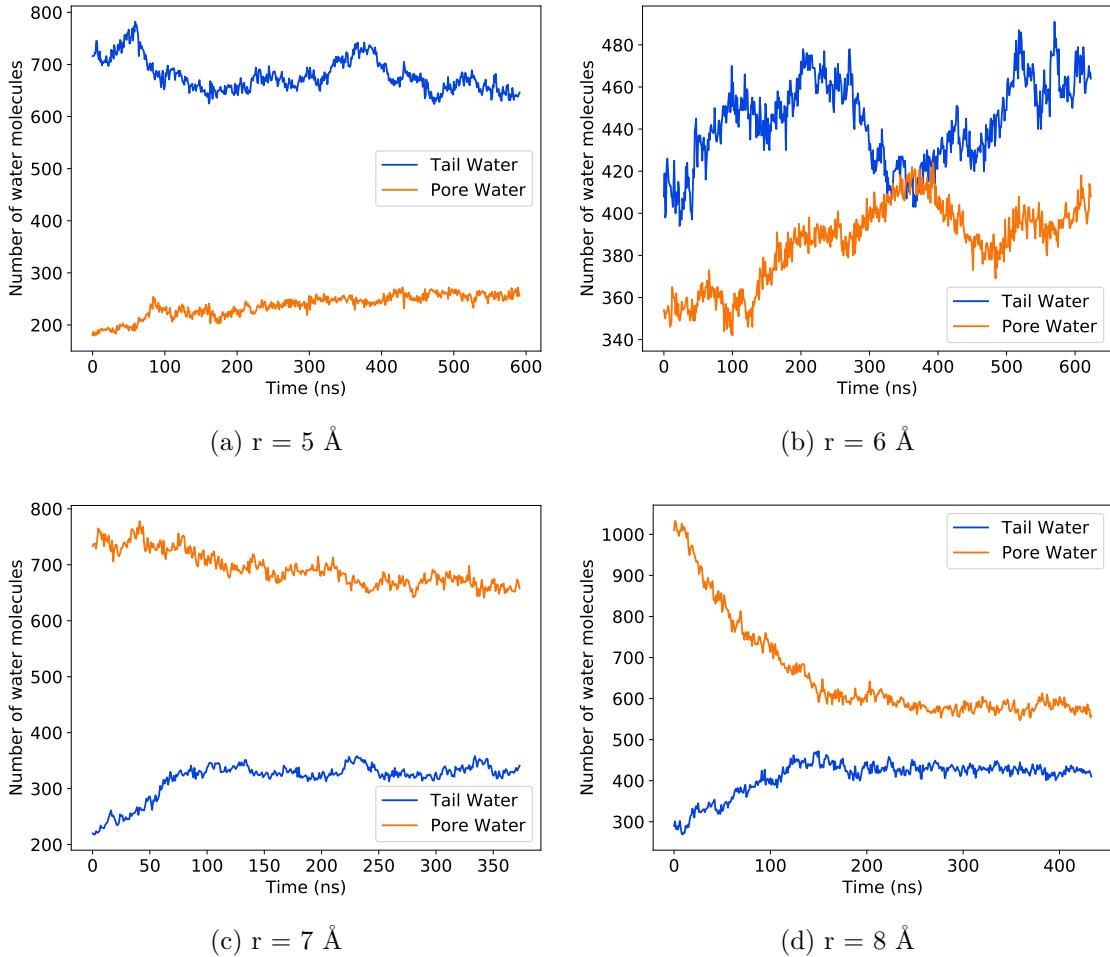


Figure A.38: There is likely more water in the pore region than in the tail region. When we create configurations with more water in the tails, equilibration is slow. When configurations start with more water in the pores, the water content in each region equilibrates quickly.

did not plateau until ~ 600 ns (Figure A.39a) while the 10 wt % water system equilibrated within the first 100 ns of simulation (Figure A.39b). The pores contain 72 % and 69 % of the total water in the 5 and 10 wt % systems respectively.

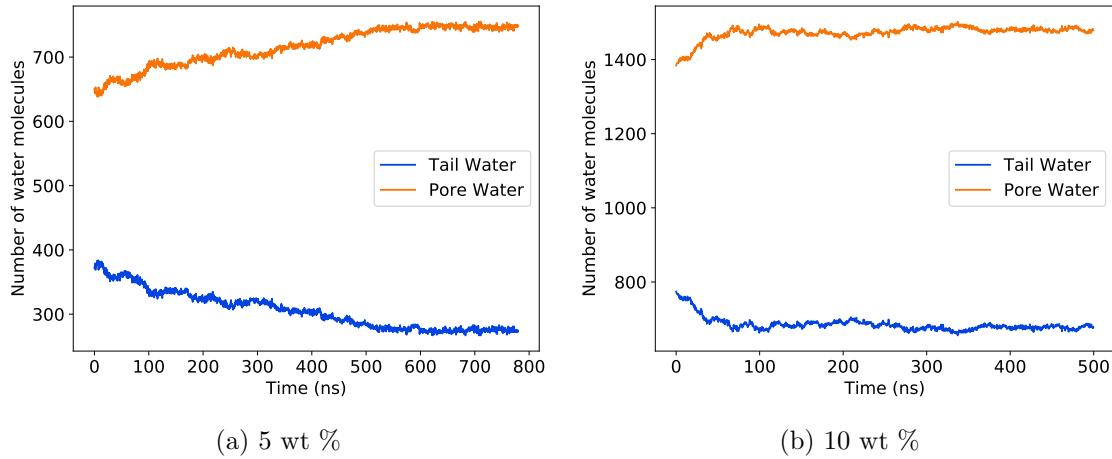


Figure A.39: We created solvated systems with one third of the total water initially placed in the tail region. (a) With 5 wt % total water, the water content equilibrates after 600 ns, with $\sim 72\%$ of the total water in the pores. (b) With 10 wt % total water, the water content equilibrates after 100 ns, with $\sim 69\%$ of the total water in the pores.

We cross-linked the equilibrated solvated systems, then allowed them to equilibrate further for 100 ns. The water contents in each region does not change significantly in either case.

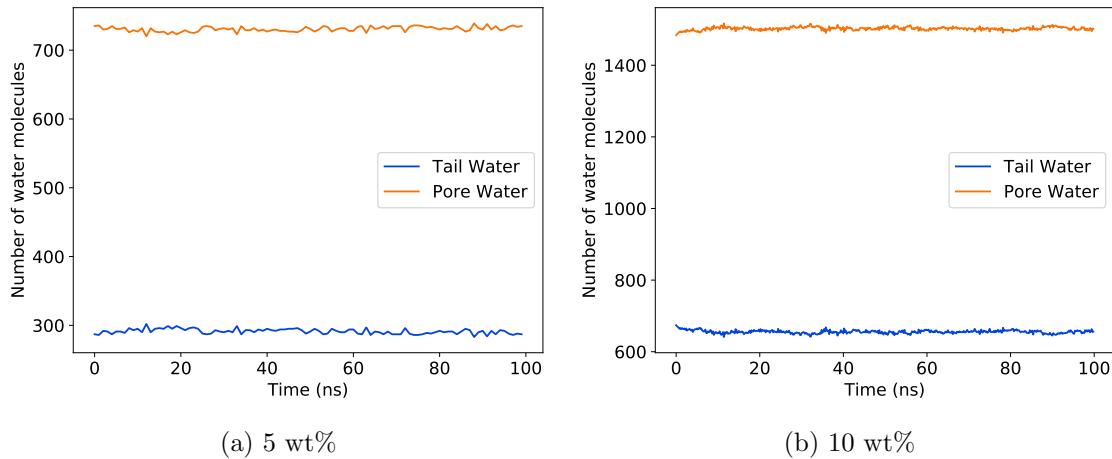


Figure A.40: The water content in the tail and pore region is not affected by cross-linking

A.30 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.

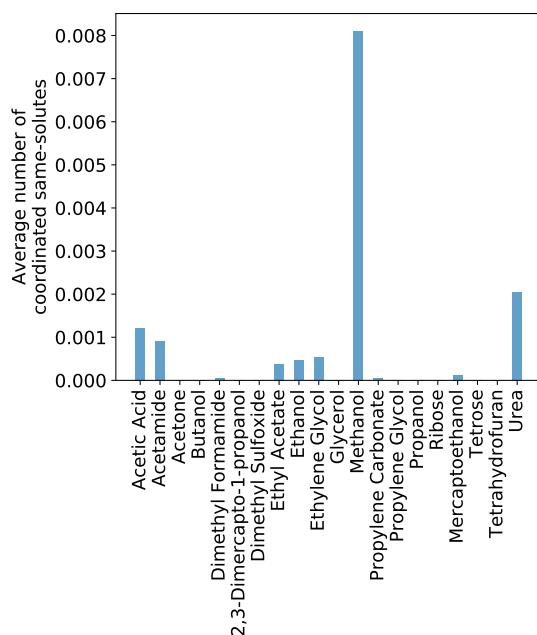


Figure A.41: Interactions between same-solutes are negligible when systems are built with 6 solutes per pore. Even methanol, which interacts with other same-solute molecules the most, does so less than 0.05 % of the time.

A.31 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 (see Figure A.42). Although there is minor re-ranking among solutes with similar MSDs, it does not affect any of the conclusions drawn in the main text.

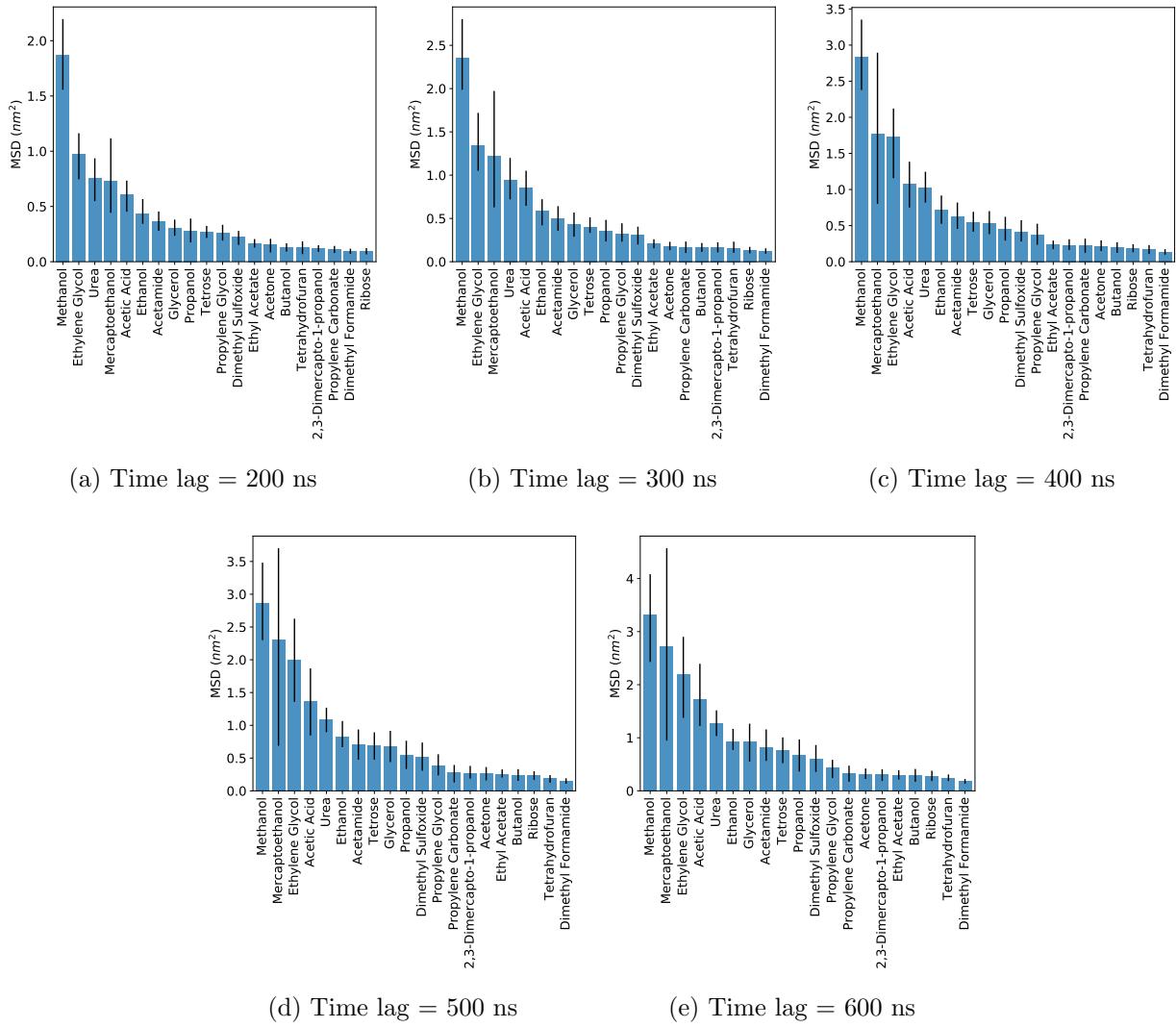


Figure A.42: The choice of time lag used to report solute MSDs does not influence the trends reported in the main text. We observe minor re-ranking of solutes with similar MSDs and note that longer time lags generally lead to large error bars on the MSDs, as expected.

A.32 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 A.43). 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.

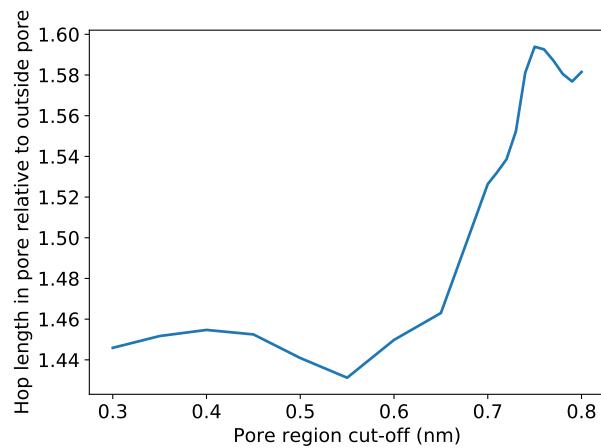


Figure A.43: Defining the cut-off 0.75 nm from the pore center maximizes the difference in hop lengths between solutes in and out of the pore. We varied the location of the cut-off defining the edge of the pore region in increments of 0.05 from 0.3 to 0.7. We varied the cut-off in increments of 0.01 from 0.7 to 0.8 in order to increase our resolution about the maximum.

A.33 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.

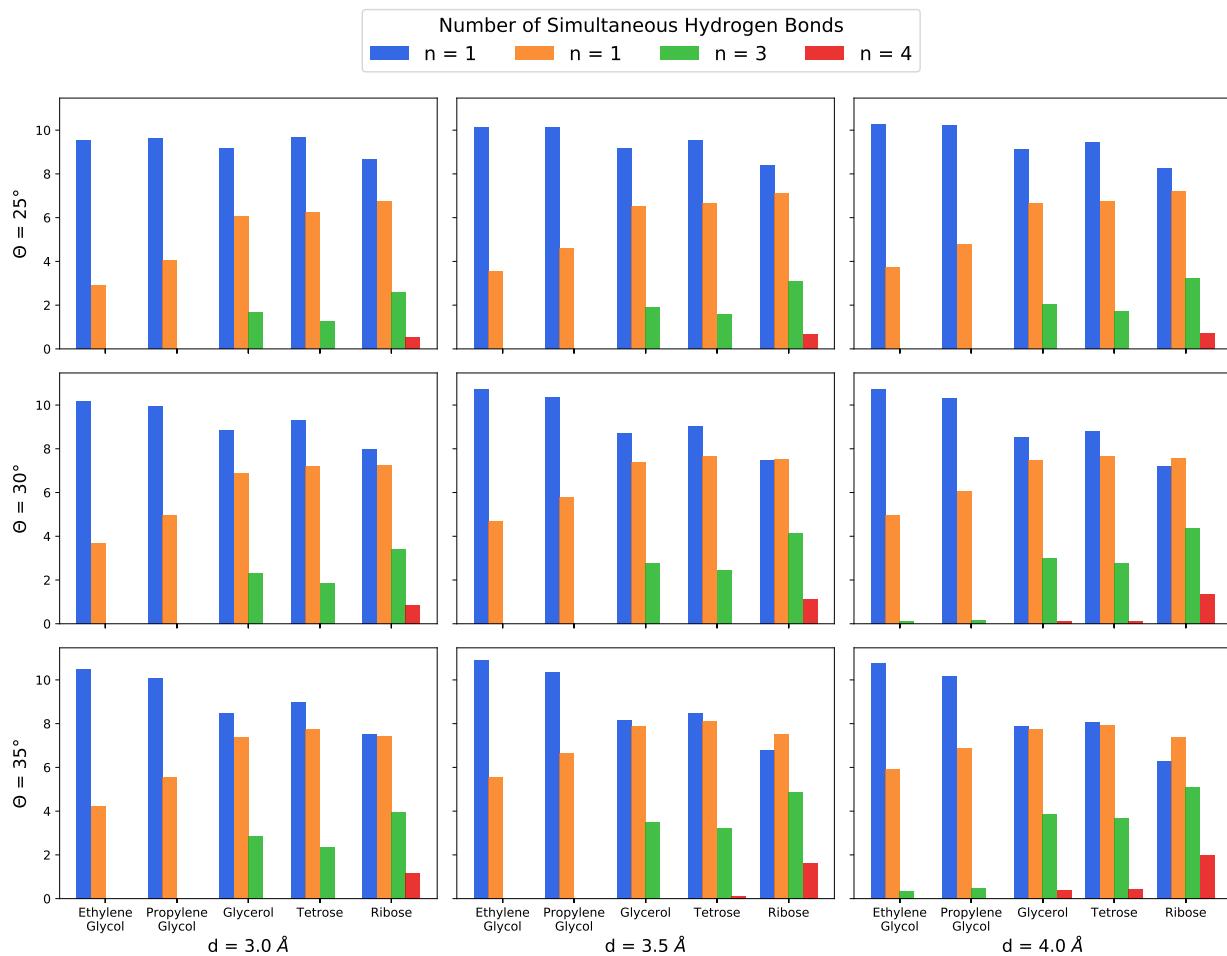


Figure A.44: We tested the sensitivity of our hydrogen bond detection to the chosen geometric criteria for hydrogen bonding. We varied the distance cut-off between 3.0 and 4.0 Å, and varied the angle cut-off between 25 and 35 °.

We tested the sensitivity of this criteria to ensure that our conclusions were not a strong function of the chosen criteria. In Figure A.44 we plotted the results of the same calculation performed to create Figure 3.11cc. 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 ($d = 4.0 \text{ \AA}$ and $\theta = 35^\circ$), glycerol, tetrose and ribose all show instances where they simultaneously hydrogen bond with 4 head group atoms. In the strictest case ($d = 3.0 \text{ \AA}$ and $\theta = 25^\circ$), 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 A.44:

- (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 glycol participates in more hydrogen bond interactions than ethylene glycol.

A.34 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 A.45.

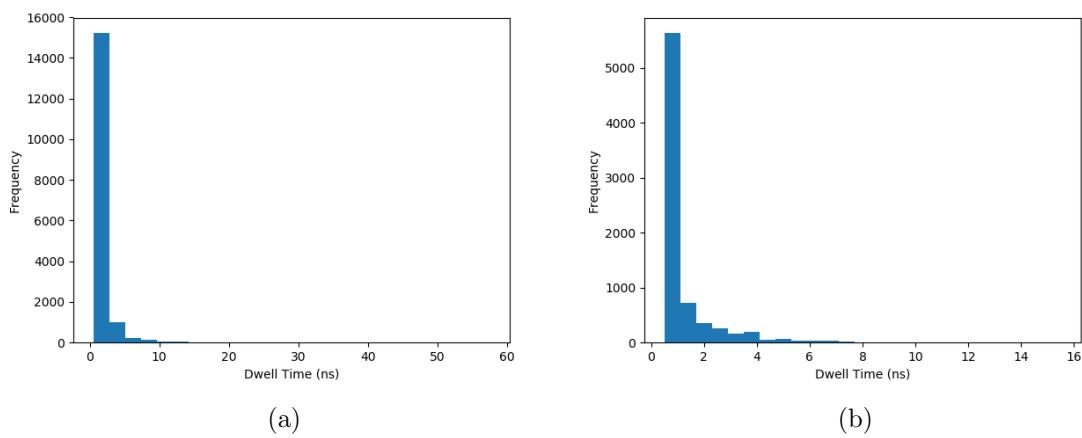


Figure A.45: The distribution of hydrogen bond lifetimes and association lifetimes of ethylene glycol both appear to be power law or exponentially distributed. We do not attempt to distinguish between the type of distribution and instead report the 95% confidence interval of association lifetimes.

A.35 Pore Splines

We captured the apparent z -dependence of the pore center locations by constructing splines running through each pore (See Figure A.46). 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.

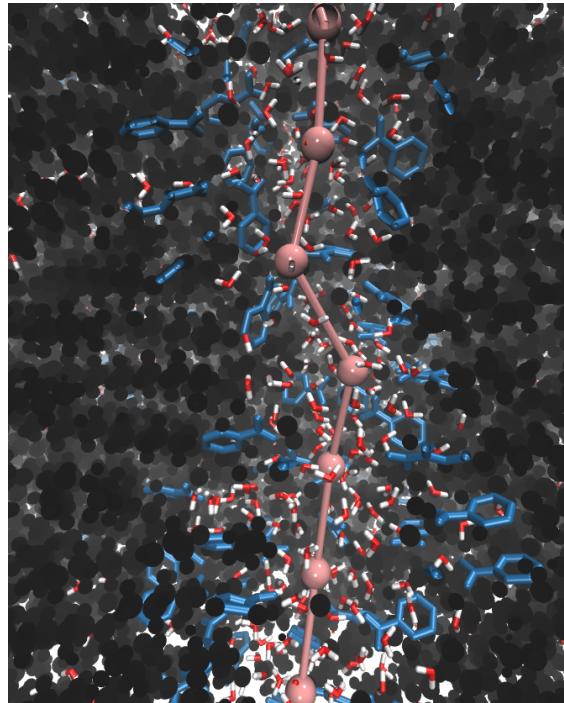


Figure A.46: We traced the center of each pore as a function of z using a spline (pink line). We constructed the spline in each pore using 10 points (pink spheres) whose positions we defined based on the center of mass of the head groups in closest proximity to the spline point in the z -direction.

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.