

## **a. How do water and ions move in nanostructured charged polymeric membranes?**

### **Background**

The ability to design nanostructured membranes with concomitant high chemical specificity is an important goal of the membrane separations community. [?] The ability to control the nanoscale architecture of membranes will allow membranes to be designed precisely for the purpose of separating specific compounds. [?, ?] The membranes being studied in this project have pores sizes smaller than 1 nanometer, allowing size-selective filtration of small molecules and gases, as well as desalination of salt water based on charge repulsion and size-exclusion of hydrated salt ions.

The membrane system we propose to model in this project is composed of lyotropic liquid crystals (LLCs) which self-assemble into hexagonally packed, cylindrical pores with acidic groups facing toward the cylinder center, complexed to counterions. Subsequent cross-linking of vinyl groups on monomer tail ends forms a mechanically strong structure (Fig. 1). [?, ?, ?] The pores, at the level of resolution obtained by SAXS and TEM, appear to be straight and uniform in size (Fig. 1). This is very different from most commercially available membranes which have a pore size distribution with tortuous pathways that reduce selectivity and flux respectively. The ordered structure exhibited by LLC membranes makes them well suited for modeling using molecular dynamics simulations because the system has a defined structure that can be studied in detail with a reasonably sized unit cell.

Our initial systems were built by first parameterizing a single monomer using the General Amber Force Field (GAFF). Because the self-assembly process is long relative to times which we can simulate, monomers were rotated into layers of six monomers and stacked into cylinders to give a starting configuration close to where we expect the system to settle at equilibrium. While a single lyotropic liquid crystal consists of a small number of atoms (138), the entire unit cell consists of about 66K atoms. This number increases to approximately 100K when water molecules are added to the system.

Equilibration simulations have been performed which energy minimize and allow the system to stabilize over the course of 500 ns in vacuum. The final structure and trajectory are analyzed quantitatively by measuring pore size, distance between pores, distribution of sodium ions in the pore, simulating X-ray diffraction patterns and calculating ionic conductivity. These measurements are compared to values measured experimentally with SAXS and TEM imaging in order to validate the final structure. SAXS and TEM images were taken using dry membranes which justifies our initial study of the system in vacuum.

### **Proposed Experiments and Justification of Resources**

Vacuum simulations of this system with a unit cell consisting of 4 pores, each with twenty layers of six monomers, total 66K atoms. On Bridges, using MPI-enabled GROMACS 5.1.2, we obtain 99 ns/day using 224 cores, beyond which

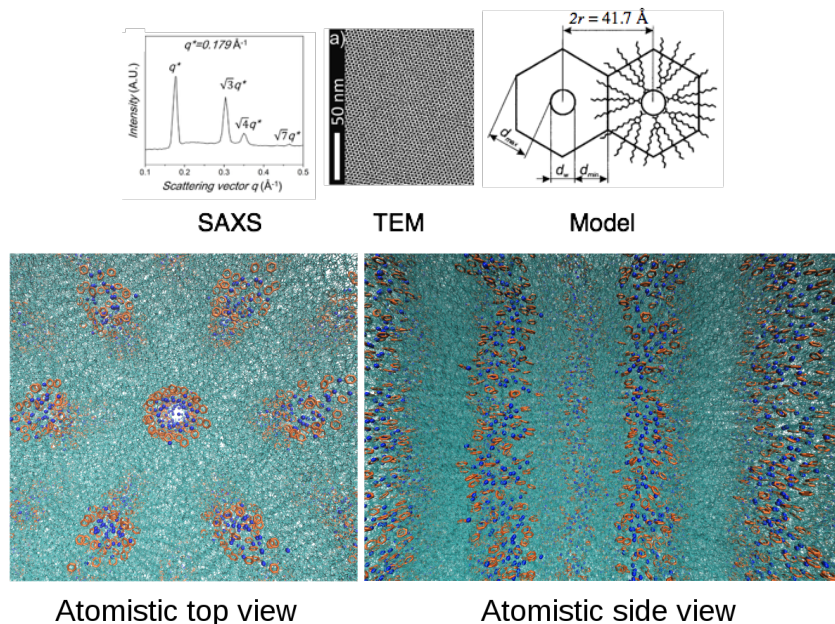


Figure 1: (above) SAXS and TEM experimental data, with hypothesized membrane structure. (Below, left) Top view of atomistic simulation, with hydrophilic head groups in orange, counterions in blue, and hydrophobic chains in blue. (Below, right) Side view of atomistic simulation showing counterions in dark blue and other atoms in light blue. These simulations demonstrate the heterogeneity in the modeled systems.

the scaling becomes mildly nonlinear (See Figure 4 in the “Code Performance and Scaling” document for more details). Simulations of the same system solvated in water results in a total of 100K atoms, which ran at 84.2 ns/day using 224 cores, at the limit of the near linear scaling regime. By creating a system with 80 layers, we reached 265K atoms, which scaled nearly linearly up to 29 ns/day on 224 cores. We will use these timings to estimate the time required for these experiments.

We repeated scaling studies with the same systems using GPU-enabled GRO-MACS 2016 with both the NVIDIA K80 and NVIDIA P100 GPUs available on Bridges. Using 4 K80 GPUs we obtain 58 ns/day, 56.4 ns/day and 10.1 ns/day for the 66K, 100K and 256K atom systems respectively. Using 4 P100 GPUs, we obtain 83.6 ns/day, 62.0 ns/day and 18.8 ns/day for each system respectively. Although in some cases, the system scales nearly linearly for greater than 4 GPUs, additional GPUs significantly extends queue wait times due to high demand. The cost will be the same, but starting simulations earlier will allow analysis to be performed during simulations and prevent valueless simulation

from preceding, saving computer time.

In order to study the relative stability of the two metastable states we have discovered, we will need to perform a computationally intensive free energy calculation using the Multistate Bennett Acceptance Ratio (MBAR) technique. MBAR estimates free energy differences with the currently lowest variance when compared to other estimators. In order to make a reasonable estimate, we will need to conduct simulations of all intermediate states which lead from one metastable state to another. Each configuration in the pathway mapping the two states must be sufficiently similar to adjacent configurations so that we achieve enough phase space overlap for a more precise calculation. We estimate that we will need at least 50 intermediate states, each run for 50 ns for a total of 2500 ns of simulation time. This will require  $(2500 \text{ ns} / (99.2 \text{ ns/day})) * 24 \text{ hrs/day} * 224 \text{ cores} = 135\text{K SU}$  using MPI on the RM partition,  $(2500 \text{ ns} / (58 \text{ ns/day})) * 24 \text{ hours} * 4 \text{ GPUs} * 1 \text{ SU} / \text{GPU-hour} = 4.1\text{K SU}$  on the K80 GPU nodes, and  $(2500 \text{ ns} / (83.6 \text{ ns/day})) * 24 \text{ hours} * 4 \text{ GPUs} * 2.5 \text{ SU} / \text{GPU-hour} = 7.2\text{K SU}$  on the P100 GPU nodes.

The resolution of simulated X-ray diffraction patterns is dependent on the size of the simulated unit cell. To create higher resolution patterns in the x, y or z directions, requires an increase in the respective dimension of the unit cell by adding more atoms. The fundamental reason for this limitation is the necessity of meeting the Bragg condition. When met, constructive X-ray interference occurs resulting in a signal which gives details about the position of atoms relative to each other. The lattice planes in the crystal, defined by the reciprocal space Miller indices h, k, and l, are separated by a distance, d. One can calculate all possible d values given all unit cell parameters. It is not trivial to see that an increase in box vector leads to a wider range of accessible hkl values and increases the spatial resolution. A simple way to estimate the simulated resolution in each direction is using the equations  $q_x = 2\pi/L_x$ ,  $q_y = 2\pi/L_y$ ,  $q_z = 2\pi/L_z$ . Our current resolution with a 66K atom system is  $0.078 \text{ \AA}^{-1}$  (Fig. 2). We propose a 4x increase in our z dimension resolution which will help us to distinguish reflections deemed a consequence of benzene ring pi-stacking (reflection occurs at 1.53 inverse angstroms) from simple alkane chain packing (reflection occurs between 1.4 and 1.57 inverse angstroms). We also hope to pick up finer details such as the sharp line that appears at .85 inverse angstroms experimentally but is absent in the simulated patterns. A system of this size is made of 265K atoms. Stacking equilibrated membrane layers directly on top of each will facilitate a fast equilibration of the large system. Equilibration simulations will be run for at least 50 ns followed by another 50 ns of simulation needed to collect enough information to simulate the XRD pattern. 100 ns of simulation time for a 265K atom system will require 19K SU on the RM nodes, 950 SU on K80 GPU nodes, and 1.3K SU on P100 GPU nodes.

Once equilibrated and cross-linked, we will examine the effect of membrane solvation on the pore structure and corresponding transport properties. We have learned that at least 1000 ns of total simulation is required to fully equilibrate the system with water.  $(1000 \text{ ns} / (86 \text{ ns/day})) * 24 \text{ hours} * 224 \text{ cores} = 62\text{K}$

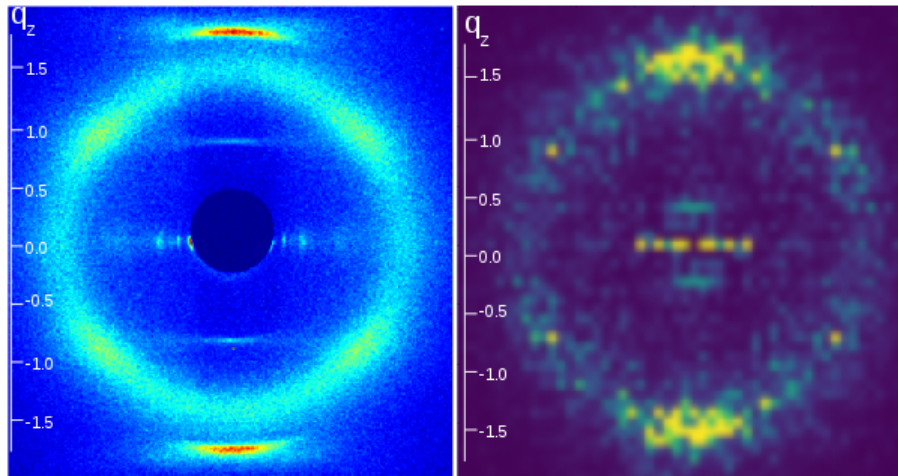


Figure 2: (left) Experimental Wide Angle X-ray Diffraction experiments contain reflections at  $q_z \approx 1.7\text{\AA}^{-1}$  indicating pi-stacking of benzene rings in the monomer head group. A weak, correlated reflection is present at  $q_z \approx 0.85\text{\AA}^{-1}$ . (right) Simulated X-ray diffraction shows a pi-stacking reflection at  $q_z \approx 1.5\text{\AA}^{-1}$ . The weak correlated reflection is not present but may be visible with higher resolution simulations

required SU, if we assume the initial membrane simulation boxes are sufficiently large.

We will carry out the same procedure with a new set of monomers similar in structure to the current monomer. Simple modifications can be made to the monomer structure as outlined in Figure 3. We will need to equilibrate each system using developed procedures which requires equilibration simulations of at least 500 ns. Of the possibilities outlined in Figure 3 we will try all variations of tail spacer and ionic headgroup and at least 3 counterions of different charge (+1, +2, +3). This results in a total of 24 systems that require 500 ns of equilibration time. These length simulations should also give us the required information to calculate physical properties such as ionic conductivity, and to simulate X-ray diffraction patterns. We will require xxx SU explore this space. To solvate all the test systems, an additional xxx SU will be necessary.

We will conduct transport studies on all systems in order to relate pore structure to macroscopic observables such as water flux and solute rejection. We will measure diffusion of various solutes across the membrane. Understanding the mechanisms of transport of each solute will provide us with sufficient understanding to suggest new modifications to monomer structure that should increase laboratory performance. We anticipate simulations of at least 50 ns will be necessary to get sufficient sampling for an informative mean square displacement (MSD) curve. The MSD curve will make it easier to identify potential transport mechanisms based on its shape.

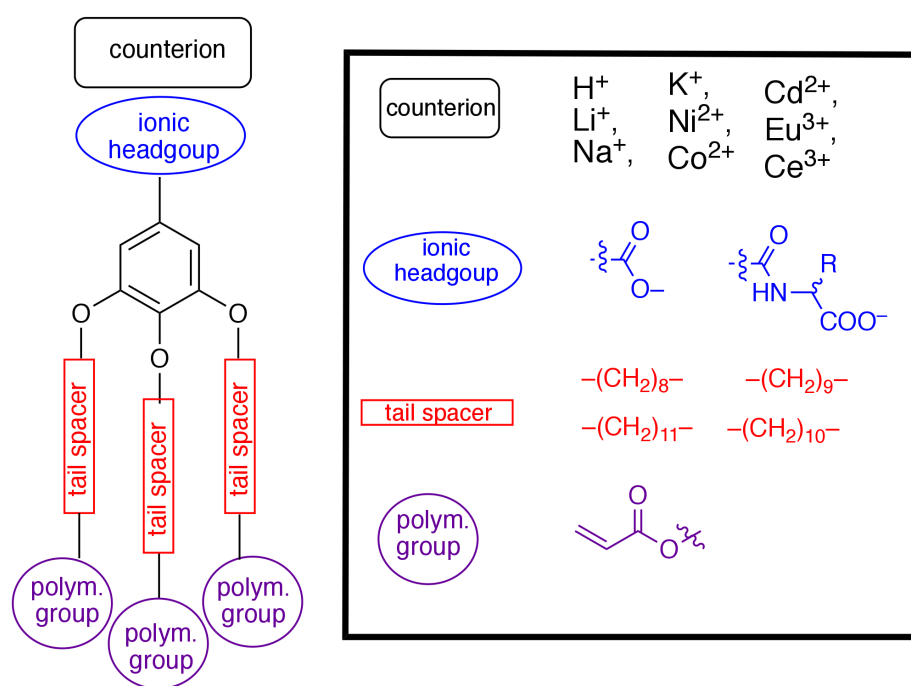


Figure 3: LLC chemistries we will examine in order to understand determinants of pore size, structure and solute transport

## Code Performance and Scaling

### I. GROMACS on larger systems

Typical simulations of membranes require between 50,000 and 300,000 atoms. We have tested the scaling of this range of jobs on Bridges. Bridges has the somewhat unusual configuration of 14 cores per processor (16 on P100 nodes), with 2 processors per node. GROMACS has three parallelization modes, via OpenMP threads, via MPI, and via GPU, and these 3 can be mixed, with multiple OpenMP threads per MPI rank with any number of compatible GPUs. OpenMP thread-based parallelization scales almost linearly for most systems of moderate size (tens of thousands of atoms) for jobs remaining on a single node. MPI scaling is almost as effective, but can additionally scale to many processors. For our GROMACS simulations, we use seven threads per MPI process, and 4 MPI processes per nodes. 14 threads per MPI process crashed for unknown reasons, and 28 threads per MPI process was slightly slower than 7 threads per MPI process. The smallest unit we tested for scaling is one node (28 cores) on the RM partition and 1 GPU on the GPU partitions (both P100 and K80 GPUs). We examine the scaling for three different size membrane systems, 66K atoms, 100K atoms, and 265K atoms.

We will run our simulations in regions of near linear scaling. Using the RM partition for all systems, the deviation from linearity occurs in the area of 224 cores (Fig. ??). Performance in the linear region gives approximately 0.44, 0.38 and 0.13 ns/day/core of simulation time for each system from smallest to largest. Once a system becomes moderately parallelized, GROMACS automatically chooses a number of separate ranks for running Particle Mesh Ewald (PME) calculations in parallel (i.e. long range molecular interactions) to the simulation. This greatly improves performance. One can manually change the number of PME ranks allocated in an attempt to further increase performance. At maximum, we have seen performance improvements of no greater than 5%, so we plan to run with the default runs.

Using the GPU partitions we tested scaling of all 3 systems using both the K80 and P100 nodes. Using the K80 nodes, deviation from linear scaling occurs in the area of 4 GPUs (1 node, 28 cores) for the 66K and 100K atom systems (Fig. ??). Linearity persists for the 100K atom system out to 12 GPUs (3 nodes, 84 cores). Linear region performance gives approximately 14.5, 14.1 and 2.6 ns/day/core of simulation time for each system from smallest to largest. Using the P100 nodes, deviation from linear scaling occurs in the area of 4 GPUs (2 nodes, 64 cores) for both the 66K and 100K atom systems, and at least 12 GPUs (6 nodes, 192 cores) for the 265K atoms system (Fig. ??). We likely will not use 6 nodes due to the demand for and relatively small number of P100 nodes, but the cost will be the same. Linear region performance gives approximately 20.9, 15.5 and 3.7 ns/day/GPU of simulation time for each system from smallest to largest.

All RM simulations were performed with GROMACS 5.1.2, compiled using

the GNU 4.8.5 compiler suite by staff at Pittsburgh Supercomputer Center, using the AVX2 accelerated instruction set. All GPU simulations were performed with GROMACS 2016, compiled in the same way. Simulations were run with a 2.0 fs time step using LINCS to constrain the bonds and PME electrostatics with a 1.2 nm cutoff.

## II. GROMACS replica exchange on crystal polymorph systems

We will perform temperature replica exchange simulations of the organic crystals of interest. We are interested in the free energies as a function of temperature, so replica exchange was not introduced solely to increase sampling; instead, replica exchange allows us to increase the sampling of the simulations that would already be performed. Crystal simulations are relatively small (1-2K atoms) and thus only scale well to 4 threads per simulation. Replica exchange functionality in GROMACS is done through MPI calls. Because the information exchange is relatively small, and only occurs roughly every 100 steps, it was not expected to not affect the overall scaling out to the 28-30 temperature points run with each simulation. However, it does appear to have a slight penalty on Bridges, especially at higher thread counts per replica. Thus, we will primarily carry out our simulations on with thread per replica.

## III. TINKER scaling

The TINKER molecular mechanics package will be used to perform simulations with a polarizable force field to allow reweighting to QM potentials. This code is not parallelizable, and requires for typical small crystal molecules 10 (for formamide) to 20 (benzene) CPU-hour per nanosecond.

## IV. Quantum ESPRESSO scaling

Another computational task that will be performed during the course of the time requested by this grant will be running Car-Parrinello molecular dynamics using the Quantum ESPRESSO package (coincidentally, also version 5.1.2). Timings were again obtained for these systems on the Bridges supercomputer, compiles using the GNU 4.8.5 compiler suite.

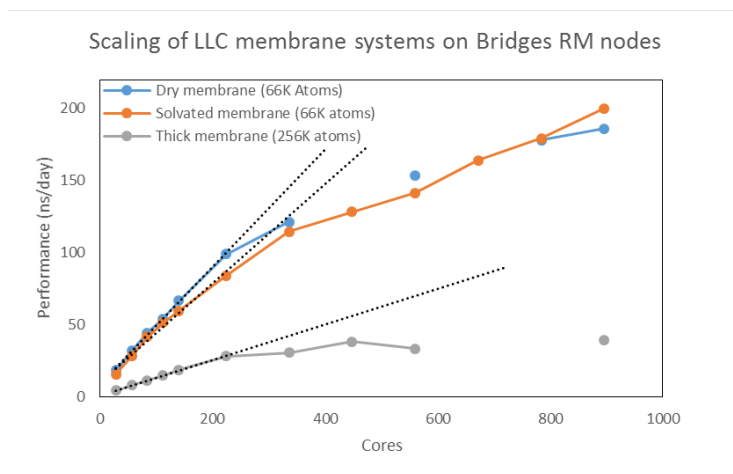
Tests (reported in ns/hour) with 1 unit cell of benzene (48 atoms, 312 electrons) and of formamide (24 atoms and 180 electrons) are shown in Fig. 6. Quantum ESPRESSO, as described online, is known to have relatively poor scaling to multiple processors, which we see here. The efficiency drops to approximately 90% when using two processors on one node (28 cores) and 75% when using 4 processors across two nodes (56 cores). In order to maximize efficiency and minimize total computational time, we propose to use 28 cores

on two processors on a single node for each of the Quantum ESPRESSO MD simulations. Since we are collecting stochastic information at thermal equilibrium, we do not need a high degree of parallelization, and can carry multiple simulations out independently. For these simulations, Car-Parrinello MD was performed with a time step of 5 Hartree a.u., an energy convergence threshold of 0.0001 a.u. and an electronic convergence threshold of  $10^{-8}$ . The kinetic energy cutoff was 40 Ry and the Van der Waals correction of DFT-D were used.

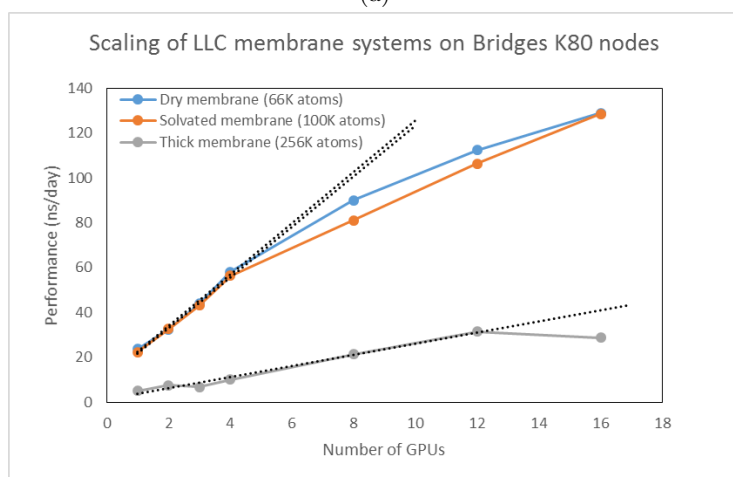
Quantum ESPRESSO will also be used to calculate single-point energies for reweighting, and to calculate quasi-harmonic approximations to compare to lattice energies and molecular dynamics simulations. These simulations will take approximately an orders of magnitude less time than the Car-Parrinello simulations.

Quantum ESPRESSO scaling performance is somewhat problematic, even though looking at online timings although it generally does not scale particularly well for small numbers of atoms. During the course of the year we will also investigate VASP to test its scaling. For the simulations in the current year, however, Quantum ESPRESSO should be sufficient.

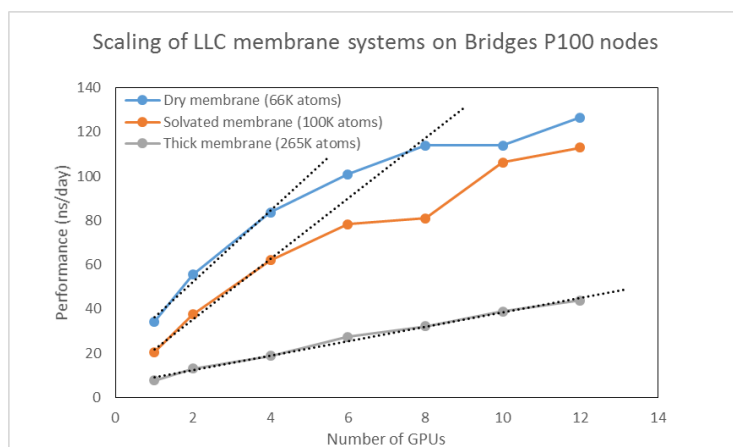




(a)



(b)



(c)

Figure 4: For systems of 66K to 265K atoms, GROMACS scales out linearly to (a) about 200 on RM nodes, (b) 4-12 GPUs on K80 nodes and (c) 4-12 GPUs on the P100 nodes

Figure 5: Small organic crystals scale moderately well up to four cores. Temperature replica exchange over 28 replicas incurs a small cost which is mostly independent of the number of threads used to run the simulation itself.

Figure 6: Scaling of Quantum ESPRESSO on Bridges. Quantum ESPRESSO, as compiled on Bridges, has problems scaling beyond 28 cores (one processors). However, since our calculations can for the most part be performed in parallel, scaling beyond one processor is not required.

## Summary of Scientific Discoveries

### Simulations of transport in nanostructured polymer membranes

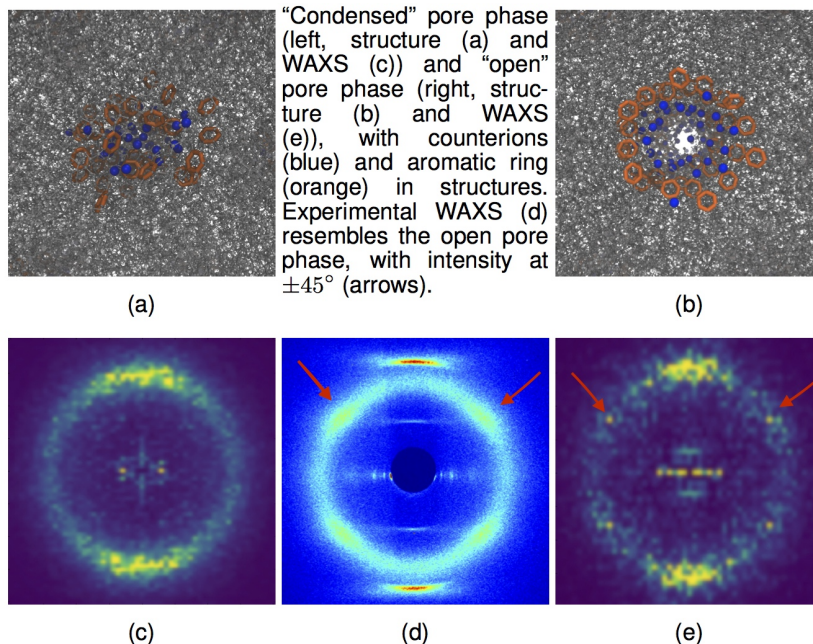
Our understanding of the microscopic structure of this type of LLC membrane has greatly increased with the aid of simulations run using Bridges.

Equilibration simulations of greater than 500 nanoseconds yield stable membrane configurations with the expected HII phase morphology. Various methods have been developed to characterize the equilibrated system. Generally, all equilibrium properties are compared to experimental measurements. We validated two methods for measuring ionic conductivity from atomistic simulations. Both methods require long simulations (at least 500 ns) in order to give accurate statistics. The distance between pores is an important structural parameter which we have measured using atomic coordinates and by simulating X-ray diffraction (XRD) experiments, a relatively undeveloped technique in MD for periodic systems such as ours. We have been able to generate structures which match experimental pore spacings within reason.

The X-ray diffraction simulations also give detailed information about membrane structure on the angstrom lengthscale. Our simulations have produced two dimensional X-ray diffraction patterns that contain all major features present in experimental studies. Producing a matching pattern was not trivial and resulted in the discovery of two metastable states. The two states are defined by the degree of local order inside the pore regions. The state which we had initially studied is characterized by a disordered pore region, however the X-ray diffraction pattern does not match experiment. We altered the starting configuration so that the benzene rings in the head group of each monomer were stacked in a parallel displaced configuration relative to each other. The resulting X-ray diffraction pattern of the new configuration, after equilibration, is a much closer match to experiment. Additionally, we were able to explain the spots that appear (red arrows in Fig. ??) contrary to how it was originally reported. The spots were assumed to be caused by the 40 degree tilt angle of the alkyl tails with respect to the plane of each stacked monomer layer, a common feature of liquid crystal systems. However, we were able to produce the same spots using configurations with an average tilt angle close to zero. Our most impactful finding remains as the discovery of two metastable states. This will be the subject of a publication which will be submitted in the coming months. In the future, we will conduct free energy calculations that will quantify the relative stability of the two states and help predict what experimental conditions might lead to each. We will also explore transport properties in each configuration to see which state provides optimal performance.

The next step for our system is to solvate it with water. In parallel to the preceding work, we have worked to develop methods which will be used to study the solvated system. While measuring ionic conductivity and running XRD simulations will work just the same, equilibration is non-trivial. We do not know exactly the equilibrium content of water or where the water is situated

Figure 7



in the membrane. While it is clear that most water should be in the hydrophilic pore region, simulations have shown that an appreciable amount of water can exist in the tail region near the slightly hydrophilic ester group. The best way to figure out how much water should be in the pore is to run very long equilibration simulations and allow the simulations to tell us. Our current approach is to create water baths at each face of the membrane and allow water to diffuse into the membrane. We have learned that we can equilibrate a membrane with water in 1000 nanoseconds. Studies of the hydrated, 'lyotropic', phase will be the subject of a future publication.