

Capturing subdiffusive solute dynamics and predicting long timescale selectivity in nanoscale pores with time series modeling

Benjamin J. Coscia Michael R. Shirts

January 6, 2020

Abstract

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 2 variations of the anomalous diffusion model. The first variation applies a single set of parameters to the solute displacements and the second applies 2 sets of parameters based on the solute's radial distance from the closest pore center. Next, we generalize Markov state models, treating the configurational state of the system as a Markov process, each of which has separate 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 our models reasonably reproduce the MSDs calculated from MD simulations. Finally, we demonstrate how one can use our models to estimate flux of a solute across a macroscopic-length pore and, based on those quantities, the membrane's selectivity towards each solute. Overall, this work helps to connect microscopic chemically-dependent solute motions and macroscopic membrane performance.

1 Introduction

Highly selective separations membranes are desirable in numerous applications. The ability to efficiently separate ions from saline water sources using membranes has been actively pursued for years in an effort to create potable water for people in water-scarce regions. [1] Even in relatively safe municipal water supplies, there is a need for membranes that can specifically separate potentially harmful organic micropollutants such as fertilizers, pesticides, pharmaceuticals and personal care products. [2] In the materials industry, there is a strong interest in creating breathable fabrics which selectively allow passage of water vapor. [3]

Lyotropic liquid crystals (LLCs) are a class of amphiphilic molecules that can be cross-linked into mechanically strong and highly selective nanoporous membranes. [4] They may provide a promising alternative to conventional membrane separation techniques by being selective based not only on solute size and charge, but on solute chemical functionality as well. [5] One can tune the functionality of LLC monomers in order to enhance or weaken specific solute-membrane interactions. [6] Two general types of LLC membranes are being actively developed. The inverted hexagonal, H_{II} , LLC phase has densely packed, uniform-sized pores on the order of 1 nm in size. A perfectly synthesized H_{II} phase has an ideal geometry for high throughput separations, however aligning the microscopic hexagonal mesophases on a large scale has shown limited success. [7, 8] Bicontinuous cubic, Q_I , phase LLCs share the uniform and nm-sized chemically complex pores of the H_{II} phase but its geometry consists of a tortuous network of 3D interconnected pores which do not require alignment. [9] While

its pore structure may decrease a Q_I membrane's permeability relative to the H_{II} phase, it is considerably easier to synthesize at scale.

Molecular modeling may make it possible to efficiently evaluate solute-specific separation membranes using the available chemical space of LLC monomers. To date, only a limited subset of LLC monomers have been studied experimentally in membrane applications. [9, 10, 11, 12, 13] Even within this small subset, they have demonstrated selectivities that could not be explained beyond speculation and vague empirical correlations. [5] Atomistic molecular modeling can provide the resolution necessary to identify molecular interactions that are key to separation mechanisms, allowing us to move beyond Edisonian design approaches.

In our previous work, we used molecular dynamics (MD) simulations to study the transport of 20 small polar molecules in an H_{II} phase LLC membrane. [14] We chose to study the H_{II} phase because it is a simpler geometry to model than the Q_I phase. In general, we observed subdiffusive transport behavior characterized by intermittent hops separated by periods of entrapment. We identified three mechanisms responsible for the solute trapping behavior: entanglement among monomer tails, hydrogen bonding with monomer head groups, and association with the monomer's sodium counter ions.

So far, our molecular models have provided valuable qualitative mechanistic insight with some quantitative support which already allows us to speculate about new LLC monomer designs. However, they would be of greater value to a larger set of researchers if we could provide quantitative predictions of macroscopic observables such as solute flux and selectivity. Due to the size of the types of systems we are studying (at least 62,000 atoms) it is prohibitively expensive and time-consuming to run simulations longer than those performed for this work ($\sim 5 \mu\text{s}$). Even with the relatively large size of our system, we only simulate 24 solute trajectories per unit cell in order to minimize solute-solute interactions. This results in relatively high uncertainties in observables such as mean squared displacement (MSD), preventing reliable long timescale predictions. We are in need of a way of studying the collective motion of a much larger set of solutes over much longer timescales.

Mathematical descriptions of transport in complex separations membranes are a powerful way to understand mechanisms and formulate design principles. [15, 16, 17] The complexity of a well-fit model generally parallels the complexity of the transport mechanism at hand, as well as the transport information the model conveys. In dense homogeneous membranes, the solution-diffusion model can extract diffusion and partition coefficients and has successfully predicted solute transport rates. [18] Analogously, pore-flow models yield predictions of diffusion coefficients and solute transport rates in nanoporous membranes. [19] Modern single particle tracking approaches have taken researchers beyond continuum modeling allowing them to characterize complex diffusive behavior. [20] At the molecular level, one can use molecular dynamics (MD) simulations to study both single particle dynamics and bulk transport properties with atomic-level insight. [14, 21] All of these approaches facilitate generation of hypotheses about the molecular origins of separations by attempting to give a more intuitive understanding of how solutes move as a function of their environment, in turn suggesting experiments that could be performed.

Using a bottom-up modeling approach, we can parameterize single particle behavior by extracting their trajectories from MD simulations and studying the properties that have the greatest influence on solute dynamics. With this information, we can learn how to construct an ensemble of characteristic single solute trajectories with these properties which would be useful for making long timescale predictions with computational ease and lower uncertainties. Fortunately, there is an abundance of information encoded in the complex solute trajectories. Into a model, we can incorporate fluctuations and time-dependent correlations in solute position as well as the length of time solutes are trapped. We can add further detail by integrating our time series analysis with our knowledge of the primary trapping mechanisms as well as the solute's changing chemical environment within the heterogeneous membrane.

In this work, we use the output of our MD simulations to construct two classes of mathematical models which aim to predict membrane performance while providing quantitative mechanistic insights. The functional form of these models is driven by mechanistic observa-

tions from our previous work and their inputs are parameterized using a substantial amount of data generated by long ($\geq 5 \mu\text{s}$) MD simulations.

Researchers have developed a rigorous theoretical foundation for describing the motion of particles exhibiting non-Brownian, or anomalous, transport behavior. [22, 23] The tools introduced by fractional calculus are instrumental to this theory. [24] They allow us to generalize the normally linear diffusion equation to fractional derivative orders, providing descriptions of a much more diverse set of behavior. [25]

Three well-known classes of behavior leading to anomalous subdiffusion are continuous time random walks (CTRW), fractional Brownian motion (FBM) and random walks on fractals (RWF).[26] These types of motion are frequently used alone or in combination to describe single particle trajectories. [27, 28] A CTRW is characterized by a distribution of hop lengths and dwell times, where trajectories are characterized by sequential independent random draws from each distribution.[29] FBM is common in crowded, viscoelastic environments where each jump comes from a Gaussian distribution but is anti-correlated to its previous step. [30, 31, 32] An RWF is imposed by a system's geometry. Systems with tortuous pathways and dead ends cause anti-correlated motion.[26, 33]

Our first modeling approach is based on the anomalous diffusion literature, applying the mathematical formalism describing subdiffusive transport to observed solute trajectories. Specifically, we treated our system as an FBM process subordinate to a CTRW, or subordinated FBM (sFBM) for short. We fit sFBM models in two ways: First, we parameterized solute motion using a single set of parameters fit to a hypothesized anomalous diffusion model. Second, we used a two-state approach by extracting two sets of model parameters calculated based on a solute's radial distance from the closest pore center. This allows us to include different dynamical behavior of solutes observed within the pore region versus within the monomer tails.

Where applicable, the sFBM model is an elegant way to characterize single particle trajectories and in addition to providing an easy and efficient way to simulate solute trajectories,

can help uncover the underlying solute-membrane interactions which result in observed transport mechanisms. It can help researchers formulate hypotheses to explain why a certain set of parameters is characteristic to a specific solute. Once one has knowledge of the primary mechanisms leading to anomalous diffusion behavior, one may gain additional insight by formulating a state-based model.

Markov state models (MSMs) are a popular class of models used to project long timescale system properties based on molecular simulation trajectories by identifying different dynamical modes and quantifying the rates of transitions between them. MSMs are frequently used to study systems with slow dynamics, such as protein folding. [34, 35] Researchers typically aim to come up with a low dimensional representation of the system based on features which preserve the process kinetics. This still often results in hundreds to thousands of distinct states. [36]

Our second modeling approach applies an extended MSM framework to a relatively small set of known states based on the three previously observed solute trapping mechanisms. Standard MSMs are typically applied to determine equilibrium populations of and the kinetics of transitions between various state configurations. [37] We extend the framework to include state-dependent fluctuations and correlations in solute position. The magnitude of solute fluctuations from its average trapped position and the degree of correlation with previous fluctuations is determined by the current state. To distinguish our approach from standard MSMs, we have named it the Markov state-dependent dynamical model (MSDDM).

We determine the degree of success of our modeling approaches in two ways. First, if we can closely reproduce solute MSDs measured from MD simulations with realizations of our models, then it is likely that the model sufficiently captures solute dynamics and can be used in a predictive capacity. Even if a model fails to reproduce the MD MSDs, there is value in uncovering the cause of the deviation. The second measure of success is by analysis of the parameters defining each model. The parameters themselves should provide clues that help explain solute behavior in a coherent way. A model that successfully reproduces MD MSDs,

but with nonsensical parameters has severely diminished value as it could be a consequence of luck.

The goal of this work is not to definitively determine which modeling approach is better but to evaluate their performance independently because they both have potential value dependent on a given research study's goals. The anomalous subdiffusion approach provides a systematic way to compare the dynamical behavior of different solutes based solely on the time series of their center of mass positions. The process of choosing the correct anomalous subdiffusion provides its own mechanistic insight since it requires a thorough analysis of solute time series behavior. One can simulate different solutes, compare the fit model parameters and relate them back to differences in solute size and chemical composition. The MSDDM characterizes explicitly defined trapping mechanisms, providing a clear picture of solute behavior while in each trapped state as well as the equilibrium occupation of each trapping state. It is possible to use the two models in tandem, the anomalous diffusion model to identify mechanisms, and the MSDDM to characterize mechanisms.

We evaluate the two modeling approaches by using them to characterize the dynamical behavior of the four fastest moving solutes studied in our previous work. Specifically, we study methanol, urea, ethylene glycol and acetic acid. In addition to moving quickly, allowing them to extensively explore membrane structural space, these solutes have a range of chemical functionality and experience each of the three trapping mechanisms to different extents.

Finally, we use both models in order to predict solute flux and selectivity in pores of macroscopic length, thus achieving a better understanding of macroscopic properties on the basis of microscopic dynamics. We show that anti-correlated solute hopping behavior severely reduces flux relative to uncorrelated behavior. We use the relative solute fluxes to estimate the membrane's selectivity between all pairs of solutes and demonstrate that, when solutes display different degrees of anti-correlated hopping behavior, selectivity becomes a function of pore length. With this improved understanding of macroscopic behavior, we can begin to think more critically about how to design membranes in order to selectively pass or

reject specific solutes.

2 Methods

We ran all MD simulations and energy minimizations using GROMACS 2018 [38, 39, 40, 41].

We performed all post-simulation trajectory analysis using Python scripts which are available online at

https://github.com/shirtsgroup/LLC_Membranes. The appropriate scripts to use for subsequent calculations are summarized in Table S1 of the Supporting Information.

2.1 Molecular Dynamics Simulations

We studied transport of solutes in the H_{II} phase using an atomistic molecular model of four pores in a monoclinic unit cell with 10% water by weight (see Figure 1). Approximately one third of the water molecules occupy the tail region with the rest near the pore center. We chose to focus our study on the 10 wt% water system because solutes move significantly faster than in the 5 wt% system studied previously. Appropriate stochastic modeling requires that solutes sample the accessible mechanisms with representative probability.

We chose to study a subset of the 4 fastest moving solutes from our previous work: methanol, acetic acid, urea and ethylene glycol. For each solute we created a separate system and to each system we added 6 solutes per pore for a total of 24 solutes. This number of solutes per pore provides a balance of a low degree of interaction between solutes, since we are primarily interested in solute-membrane interactions at present, as well as a sufficient amount of data from which to generate statistics on the time scales which we simulate. Further details on the setup and equilibration of these systems are described in our previous work.[14]

We extended the 1 μ s simulations of our previous work to 5 μ s in order to collect ample data. We simulated the system with a time step of 2 fs at a pressure of 1 bar and 300 K

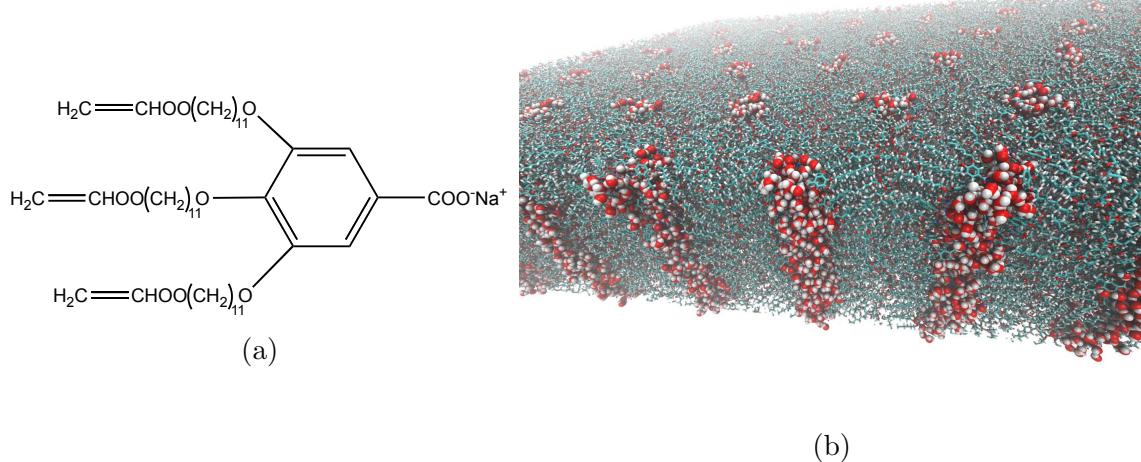


Figure 1: (a) The wedge-shaped liquid crystal monomer Na-GA3C11 will form the inverted hexagonal phase in the presence of water where the carboxylate head groups occupy the pore centers. (b) A cross-section of a periodically replicated atomistic unit cell used for simulations in this study reveals the membrane’s aqueous, hexagonally packed, straight and uniform sized pores. Water present in the tail region is omitted for clarity.

controlled by the Parrinello-Rahman barostat and the v-rescale thermostat respectively. We recorded frames every 0.5 ns.

We considered each system to be equilibrated when the solute partitioning between the pore and tail region reached apparent equilibrium and use only data after that point in our analysis of solute kinetics. We plot the solute partition versus time in Figure S1 of the Supporting Information in order to justify our choice of equilibration times for each solute.

2.2 The Anomalous Diffusion Model

Solutes in this system very clearly exhibit subdiffusive behavior, a type of anomalous diffusion. During an anomalous diffusion process, the mean squared displacement (MSD) does not grow linearly with time, but instead it follows a power law of the form:

$$\langle z^2(t) \rangle = K_\alpha t^\alpha \quad (1)$$

where α is the anomalous exponent and K_α is the generalized diffusion coefficient. We only calculate MSD with respect to the solutes’ center of mass z -coordinate, which is oriented

along the pore axis. A value of $\alpha < 1$ indicates a subdiffusive process, while values of $\alpha = 1$ and $\alpha > 1$ are characteristic of Brownian and superdiffusive motion respectively.

One can measure MSD in two ways. The time-averaged MSD measures all observed displacements over time lag τ :

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

where T is the length of the trajectory. The ensemble averaged MSD measures displacements with respect to a particle's initial position:

$$\langle z^2(t) \rangle = \langle z(t) - z(0) \rangle^2 \quad (3)$$

Fits to the *ensemble* averaged MSD will always reproduce the form of Equation 1.

The time averaged and ensemble averaged MSDs will give identical results unless a system displays non-ergodic behavior. For a pure CTRW, the power law distribution of trapping times leads to weak ergodicity breaking. In this case, the time-averaged MSD is linear while the ensemble averaged MSD has the form of Equation 1. [26] With power law trapping behavior, the time between hops diverges so there is no characteristic measurement time scale of solute motion. In fact, as measurement time increases, the average MSD of a CTRW tends to decrease, a phenomenon called aging, because trajectories with trapping times on the order of the measurement time get incorporated into the calculation. [42]

We chose to analyze just the time-averaged MSD because, compared to the ensemble average, it is a more statistically robust measure of the average distance a solute travels over time. Although theoretically possible, we do not use the ensemble averaged MSD to calculate α because we can extract the same parameter with higher precision using the dwell time distributions as described below.

2.2.1 Subordinated Fractional Brownian Motion

One can characterize the CTRW component of an sFBM process by the parameters which describe its dwell time and hop length distributions. We used the `ruptures` Python package in order to automatically identify mean shifts in solute trajectories, indicating hops.[43] We used the corresponding hop lengths and dwell times between hops to construct empirical distributions.

Dwell time distributions: For subdiffusive transport, the distribution of dwell times is expected to fit a power law distribution proportional to $t^{-1-\alpha}$. [26] Because we are limited to taking measurements at discrete intervals dictated by the output frequency of our simulation trajectories, we fit the empirical dwell times to a discrete power law distribution whose maximum likelihood α parameter we calculated by maximizing the likelihood function:

$$\mathcal{L}(\beta) = -n \ln \zeta(\beta, t_{min}) - \beta \sum_{i=1}^n \ln t_i \quad (4)$$

where $\beta = 1 + \alpha$, t_i are collected dwell time data points, n the total number of data points, and ζ is the Hurwitz zeta function where t_{min} is the smallest measured value of t . [44]

In practical applications, the heavy tail of power law distributions can result in arbitrarily long dwell times that are never observed in MD simulations. We cannot observe dwell times longer than the length of a single trajectory. In order to directly compare our anomalous diffusion model to finite-length MD trajectories we need to bound the dwell time distribution. We do this by adding an exponential cut-off to the power law so the dwell time distribution is now proportional to $t^{-1-\alpha}e^{-\lambda t}$. We determine MLEs of α and λ by maximizing the likelihood function: [44]

$$\mathcal{L}(\alpha, \lambda) = n(1 - \alpha) \ln \lambda - n \ln \Gamma(1 - \alpha, t_{min}\lambda) - \alpha \sum_{i=1}^n \ln t_i - \lambda \sum_{i=1}^n t_i \quad (5)$$

Correlated hop length distributions: The marginal distribution of hops by solutes under-

going an sFBM process is Gaussian, therefore we parameterize it by its standard deviation, σ . [22, 28, 45] The measured mean is always very close to zero so we assume that it is exactly zero in our time series simulations since we have no reason to expect drift in either direction based on pore symmetry.

sFBM implies that hops are correlated which we describe using the Hurst parameter, H . The autocovariance function of hop lengths has the analytical form: [30]

$$\gamma(k) = \frac{\sigma^2}{2} \left[|k - 1|^{2H} - 2|k|^{2H} + |k + 1|^{2H} \right] \quad (6)$$

where σ^2 is the variance of the underlying Gaussian distribution from which hops are drawn and k is the time lag, or number of increments between hops. The hop autocorrelation function is simply Equation 6 normalized by the variance. When $H < 0.5$, hops are negatively correlated, when $H = 0.5$ we recover Brownian motion and when $H > 0.5$, one observes positive correlation between hops. There are many methods for estimating the Hurst parameter for a time series. [46] It can be a difficult task because Equation 6 decays slowly to zero, especially when $H > 0.5$, meaning one needs to study large time lags with high frequency. Fortunately (from a mathematical perspective), all of our solutes show anti-correlated motion, so most of the information in Equation 6 is contained within the first few time lags (see Figure S2a of the Supporting Information). Any hope of fitting longer time lags to our data is lost in the noise since the analytical values are so close to zero. Therefore, we obtained H by performing a least squares fit of Equation 6 to the first ten 0.5 ns time lags of the empirically measured autocorrelation function.

2.2.2 Subordinated Fractional Lévy Motion

Since we also want to account for the possibility that distribution of hops is not Gaussian, we can model them with the more general class of Lévy stable distributions. For independent and identically distributed random variables, the generalized central limit theorem assures

convergence of the associated probability distribution function (PDF) to a Lévy stable PDF. [25] The characteristic equation which describes the Fourier transform of a Lévy stable PDF is:

$$p_{\alpha_h, \beta}(k; \mu, \sigma) = \exp \left[i\mu k - \sigma^{\alpha_h} |k|^{\alpha_h} \left(1 - i\beta \frac{k}{|k|} \omega(k, \alpha_h) \right) \right] \quad (7)$$

where

$$\omega(k, \alpha_h) = \begin{cases} \tan \frac{\pi \alpha_h}{2} & \text{if } \alpha_h \neq 1, 0 < \alpha_h < 2, \\ -\frac{2}{\pi} \ln |k| & \text{if } \alpha_h = 1 \end{cases}$$

α_h is the index of stability or Lévy index, β is the skewness parameter, μ is the shift parameter and σ is a scale parameter. The most familiar case, and one of three that can be expressed in terms of elementary functions, is the Gaussian PDF ($\alpha_h = 2, \beta = 0$). We assume symmetric distributions centered about 0 implying that β and μ are both 0.

Assuming hops are correlated, solute behavior may be described by subordinated fractional Lévy motion (sFLM). The Hurst parameter can again be used to describe hop correlations because they share the same autocorrelation structure. [47] The autocovariance function for FLM is:

$$\gamma(k) = \frac{C}{2} \left[|k - 1|^{2H} - 2|k|^{2H} + |k + 1|^{2H} \right], \quad C = \frac{E[L(1)^2]}{\Gamma(2H + 1) \sin(\pi H)} \quad (8)$$

where $E[L(1)^2]$ is the expected value of squared draws from the underlying Lévy distribution, effectively the distribution's variance. [48] In general, most Lévy stable distributions have an undefined variance due to their heavy tails. However, normalizing Equation 8 by the variance of a finite number of draws from a Lévy stable distribution results in the same autocorrelation structure as FBM. See Section S3 of the Supporting Information for numerical simulations illustrating this point.

Analogous to power law dwell times, the heavy tails of Lévy stable hop length distributions result in rare but arbitrarily long hops. These long and unrealistic hops result in

over-estimated simulated MSDs. We observe that the distribution of hops observed in our MD simulations are well approximated by Lévy stable distributions close to the mean, but they significantly under-sample the tails. We chose to truncate the Lévy stable distributions based on where the theoretical probability distribution function (PDF) starts to deviate from the empirically measured PDF (see Section S4.1 of the Supporting Information). [49]

Multiple Anomalous Diffusion Regimes

We observe different dynamical behavior when solutes move while inside the pore versus while in the tail region. This suggest two anomalous diffusion models of varying complexity. We created a simple, single mode model with a single set of parameters fit to solute trajectories. Our second, two mode model assigns a set of parameters to each of 2 modes based on the solute's radial location. Solutes in mode 1 are in the pore region, defined as less than 0.75 nm from any pore center. All else are in mode 2, the tail region. We determined this cut-off by maximizing the difference in dynamical behavior as described in our previous work. [14] Unfortunately, there were not enough sufficiently long sequences of hops in each mode to reliably calculate a Hurst parameter for each mode so we used the Hurst parameter from the single mode model for both modes of the two mode model.

For the two mode model, we defined a transition matrix describing the rate at which solutes moved between the tail and pore region. We assumed Markovian transitions between modes, meaning each transition had no memory of previously visited modes. We populated a 2×2 count matrix by incrementing the appropriate entry by 1 each time step. For example, if we observed the solute in the tails followed by a transition to the pores, we incremented the (2, 1) entry of the count matrix by 1. We generated the transition probability matrix from the count matrix by normalizing the entries in each row so that they summed to unity.

Simulating Anomalous Diffusion

We simulated models with all combinations of the types of dwell and hop length distributions described above, which are summarized in Table 1. All models include correlation between hops.

Dwell Distribution	Hop Length Distribution	Abbreviation
Power Law	Gaussian	sFBM
Power Law w/ Exponential Cut-off	Gaussian	sFBMcut
Power Law	Lévy Stable	sFLM
Power Law w/ Exponential Cut-off	Lévy Stable	sFLMcut

Table 1: We tested models with various modifications to the dwell and hop length distributions. We incorporate correlation structure into all models.

For each solute, we simulated 1,000 anomalous diffusion trajectories of length T_{sim} in order to directly compare our model’s predictions to MD simulations. T_{sim} varied between solutes due to differing solute equilibration times. We constructed trajectories by simulating sequences of dwell times and correlated hop lengths generated based on parameters randomly chosen from our bootstrapped parameter distributions. We propagated each trajectory until the total time equaled or exceeded $T_{sim} \mu\text{s}$, then truncated the last data point so that the total time exactly equaled $T_{sim} \mu\text{s}$ since valid comparisons are only possible between fixed length sFBM simulations. We reported the time-averaged MSD up to a 1000 ns time lag corresponding 95% confidence intervals.

When simulating 2 mode models, we determined the state sequence based on random draws weighted by the appropriate row of the probability transition matrix. We then drew hops and dwells based on the current state of the system. Since we calculated the transition probabilities from a finite data set, they have an associated uncertainty which we incorporated by re-sampling each row from a two dimensional Dirichlet distribution (which is also a beta distribution for the 2D case) with concentration parameters defined by the count matrix. [50]

We used the Python package `fbm` to generate exact simulations of FBM and our own

Python implementation of the algorithm by Stoev and Taqqu to simulate FLM. [51] Note that, to our knowledge, there are no known exact simulation algorithms for generating FLM trajectories. However, the algorithm we used sufficiently approximates draws from the marginal distribution and reasonably approximates the correlation structure on MD simulation timescales. We added an empirical correction to enhance the accuracy of the correlation structure (see Section S4.2 of the Supporting Information for validation of the approach).

2.3 The Markov State-Dependent Dynamical Model

A Markov state model (MSM) decomposes a time series into a set of discrete states with transitions between states defined by a transition probability matrix, T , which describes the conditional probability of jumping to a specific state given the previously observed state. [52, 53]

In this work, we define a total of 8 discrete states based on the 3 trapping mechanisms observed in our previous work. Therefore, there is no need to apply any algorithmic approaches to identify and decompose our system into discrete states. The states we have chosen include all combinations of trapping mechanisms in the pore and out of the pore (see Table 2). They assume that there are no significant kinetic effects resulting from solute conformational changes or pore size fluctuations. We use the same radial cut-off (0.75 nm) as in the AD model to differentiate the pore and tail region. We define a hydrogen bond to exist if the distance between donor, D, and acceptor, A, atoms is less than 3.5 Å and the angle formed by $D - H \cdots A$ is less than 30°. We define a sodium ion to be associated with an atom if they are within 2.5 Å of each other, as determined in our previous work [14]

We constructed the state transition probability matrix, T , based on observed solute trajectories. Using methods described in our previous work, we determined each solute's radial location and which, if any, trapping mechanisms affected it at each time step, then assigned the observation to a specific state according to the definitions in Table 2. [14]

1. In tails, not trapped	5. In pores, not trapped
2. In tails and hydrogen bonding	6. In pores and hydrogen bonding
3. In tails and associated with sodium	7. In pores and associated with sodium
4. In tails, hydrogen bonding and associated	8. In pores, hydrogen bonding and associated

Table 2: We defined 8 discrete states based on all combinations of previously observed solute trapping mechanisms.

Analogous to the mode transition matrix in Section 2.2, and based on the current and previous state observation, we incremented the appropriate entry of an $n \times n$ count matrix by 1, where n is the number of states.

We verified the Markovianity of state transitions by solutes in our MD simulations by analyzing the empirically measured transition matrix. An important condition of a Markov process is the satisfaction of detailed balance:

$$T_{i,j}P_i(t = \infty) = T_{j,i}P_j(t = \infty) \quad (9)$$

where P is the equilibrium distribution of states. In other words, the number of transitions from state i to j and from state j to i should be equal. A second test is to verify that T holds up to coarser time scale partitioning. As long as solutes sufficiently sample the states in Table 2, T should be invariant with time. [56] We verified these properties in Section S5 of the Supporting Information.

Adding to the standard MSM framework, we incorporated the dynamics of the solutes within each state as well as the dynamics of state transitions, which includes the overall configurational state of the solute and its environments. While MSMs are often used to estimate equilibrium populations of various states, adding state-dependent dynamics allows us to simulate solute trajectories. Hence why we refer to them as Markov State-Dependent Dynamical Models (MSDDMs).

We recorded the z -direction displacement at each time step in order to construct individual emission distributions for each state and transition between states. This results in 64 distinct emission distributions with some far more populated than others. We modeled

all of the emission distributions as symmetric Lévy stable distributions in order to maintain flexibility in parameterizing the distributions.

We use the Hurst parameter to describe this negative time series correlation. However, there is not sufficient data to accurately measure a Hurst parameter for each type of transition. We avoided this problem by combining all distributions associated with state transitions and treating all transitions as correlated emissions from a single Lévy stable distribution. This reduces the number of emission distributions from 64 to 9 (1 for each of the 8 states and 1 for transitions between states).

We simulated realizations of the MSDDM using the probability transition matrix and emission distributions. For each trajectory simulated, we chose an initial state randomly from a uniform distribution. We generated a full state sequence by randomly drawing subsequent states weighted by the rows of the probability transition matrix corresponding to the particle's current state. Again, because we are working with a finite data set, we incorporated transition probability uncertainties into the rows of the transition matrix by resampling them from a Dirichlet distribution. For each same-state subsequence of the full state sequence, we simulated an FLM process using the Hurst parameter of that state and the parameters of the corresponding emission distribution. Independently, we simulated the transition between each same-state sequence with an FLM process based on the Hurst parameter of transition sequences and the parameters of the single transition emission distribution. We used the same FLM simulation procedure described in Section 2.2.

2.4 Estimating Solute Flux

We determined the rate at which solutes cross macroscopic-length pores based on the Hill relation: [57]

$$J = \frac{1}{MFPT} \quad (10)$$

where J is the single particle solute flux and MFPT refers to the mean first passage time. To account for input concentration dependence of the flux, assuming that particles are uncorrelated, one can multiply Equation 10 by the total number of particles to get the total flux. In the context of our work, the MFPT describes the average length of time it takes a particle to move from the pore entrance to the pore exit.

We generated particle trajectories, parameterized with the above models, in order to construct a distribution of first passage times across a membrane pore of length L . For computational reasons dependent on the model used, we simulated 1,000-10,000 trajectory realizations, released at the pore entrance ($z = 0$). In the case of uncorrelated hops, one can continuously draw from the hop length distribution until $z \geq L$ (or $-L$ for the sake of computational efficiency). The length of time between the last time the particle crossed $z = 0$ and the end of the trajectory gives a single passage time. When particle hops are correlated, as they are in all cases of this work, we cannot continuously construct the particle trajectories. Rather, we must generate trajectories of length n and measure the length of the sub-trajectory which traverses from 0 to L without becoming negative.

We calculated the expected value of analytical fits to the passage time distributions in order to determine the MFPT for a given solute and pore length. One should not use the mean of the empirical passage time distribution because it is highly likely that the true MFPT will be underestimated unless 100% of a very large number of trajectories reach L . If a trajectory does not reach L within n steps, it is possible that a very long passage time has been excluded from the distribution.

To come up with an analytical equation describing the passage time distributions, one can frame the problem as a pulse of particles instantaneously released at the pore inlet ($z = 0$) which moves at a constant velocity, v , and spreads out as it approaches L . This spreading is parameterized by an effective diffusivity parameter, D . This approach gives results equivalent to if we had released each particle individually and then analyzed the positions of the ensemble of trajectories as a function of time. The analytical expression

describing the distribution of first passage times is: [58]

$$P(t) = -\frac{1}{\sqrt{\pi}} e^{-(L-vt)^2/(4Dt)} \left(-\frac{D(L-vt)}{4(Dt)^{3/2}} - \frac{v}{2\sqrt{Dt}} \right) \quad (11)$$

where the only free parameters for fitting are v and D . A derivation of Equation 11 is given in Section S6 of the Supporting Information. We calculated the expected value of Equation 11 in order to get the MFPT.

We used the ratio of solute fluxes in order to determine membrane selectivity, S_{ij} , towards solutes. Selectivity of solute i versus j is defined in terms of the ratio of solute permeabilities, P : [59]

$$S_{ij} = \frac{P_i}{P_j} \quad (12)$$

We can relate this to solute flux using Kedem and Katchalsky's equations for solvent volumetric flux, J_v , and solute flux, J_s : [60, 61]

$$J_v = L_p(\Delta P - \sigma\Delta\pi) \quad (13)$$

$$J_s = P_s\Delta C + (1-\sigma)CJ_v \quad (14)$$

where L_p is the pure water permeability, ΔP and $\Delta\pi$ are the trans-membrane hydraulic and osmotic pressure differences, σ is the reflection coefficient, P_s is the solute permeability, ΔC is the trans-membrane solute concentration difference and C is the mean solute concentration. Since our simulations do not include convective solute flux, we eliminate the second term of Equation 14 which allows us to derive a simple expression for selectivity in terms of solute flux:

$$S_{ij} = \frac{J_i/\Delta C_i}{J_j/\Delta C_j} \quad (15)$$

3 Results and Discussion

3.1 Anomalous Diffusion Modeling

3.1.1 Parameterizing Subordinated Fractional Brownian Motion

The evidence suggests that solutes in this system can be well-modeled by subordinated fractional Brownian motion. In Figure 2, we use urea to support this claim. Analogous figures for all other solutes are presented in the Supporting Information, Figures S8–S10. The solutes clearly perform jumps after periods of relative immobility (see Figure 2a for example). The near-Gaussian distribution of jump lengths and power law distribution of dwell times are both characteristic of CTRWs (Figures 2b and 2c). The apparent anti-correlation between hops suggests a fractional diffusion process is subordinate to the CTRW. We believe the subordinated process is FBM or FLM because their analytical correlation structures are close to those observed in our simulations (Figure 2d). Fractional motion is common in crowded viscoelastic environments where movement is highly influenced by the movement of surrounding components, such as monomer tails in our case. [62]

We modeled the distributions of hop lengths in two ways. First, we assumed the distribution to be Gaussian since it is possible to exactly simulate realizations of fractional Brownian motion. Second, we fit the distributions to Lévy stable distributions since it is more general than the Gaussian distribution. We plotted the MLE fits of both on top of urea’s hop length distribution in Figure 2b. The Lévy distribution does a better job capturing the somewhat heavy tails and high density near 0 of the hop length distribution. However, since there are no known exact simulation techniques for generating realizations of fractional Lévy motion, we must evaluate whether this more general fit is worthwhile.

We also modeled the distribution of dwell times in two ways. First, we assumed pure power law behavior since it is consistent with most theoretical descriptions of CTRWs. The data fits well to this model at short dwell times but the density of long dwell times is over-estimated. In our second approach, we truncate the power law distribution with an

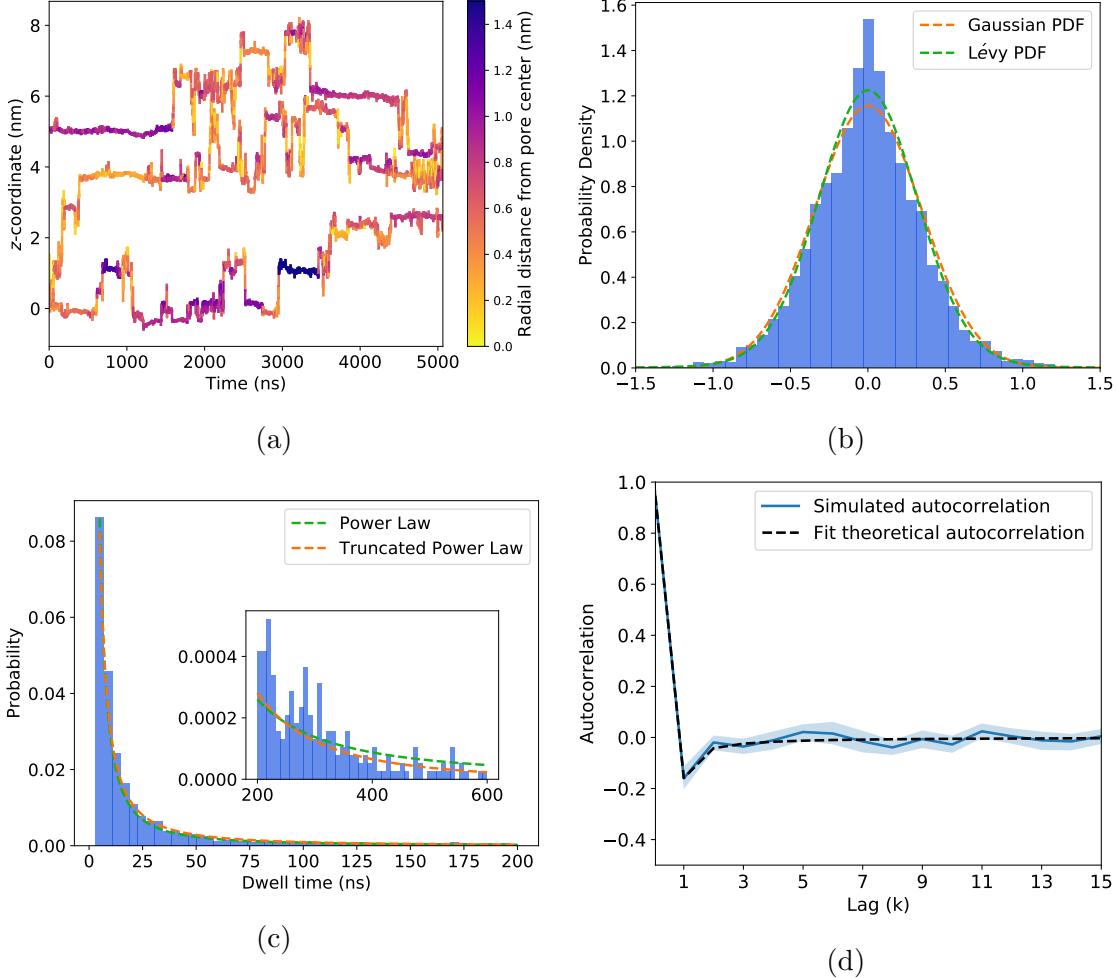


Figure 2: (a) Urea exhibits hops between periods of entrapment, characteristic of a CTRW. (b) The distribution of hop lengths fits well to a Gaussian as well as a more general Lévy stable distribution. The Lévy stable distribution does a better job of capturing the heavier tails and increased density near 0. (c) The distribution of dwell times is fit well by a power law but it may over-estimate the probability density at long dwell times. A power law truncated with an exponential cut-off may better describe the probability of long dwell times in our simulations. (d) The hops are negatively correlated to their previous hop. In combination, (a) – (d) support modeling solutes as subordinated fractional Brownian and Lévy motion. The behavior shown by urea in this figure is common to all solutes in this study (see Figures S8–S10 of the Supporting Information).

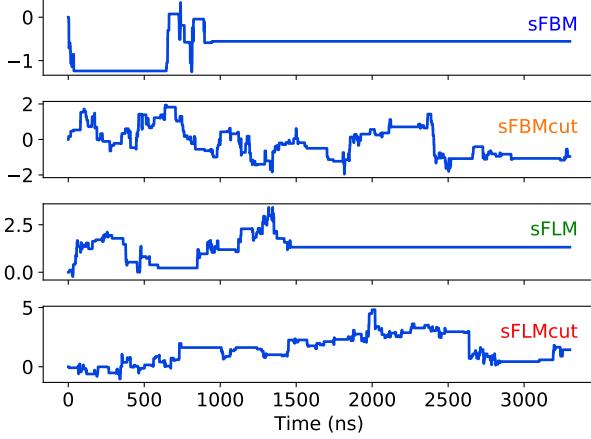


Figure 3: Simulations of 1 mode AD model trajectories for urea display qualitatively similar hopping and trapping behavior to that shown in Figure 2a. Dwell times appear exaggerated in the sFBM and sFLM models because the power law dwell time distributions are not truncated and have infinite variance.

exponential cut-off, lowering the probability of extremely long dwell times. We demonstrate this effect in the inset to Figure 2c, where the density of the truncated power law drops below that of the pure power law and tends towards 0 near a dwell time of 250 ns.

Realizations of the AD model based on the parameterization procedures just described yield qualitatively similar trajectories to those seen in our MD simulations. In Figure 3, we plot representative sample trajectories for each combination of the dwell time and hop distribution, labeled according to Table 1. The sFBMcut and sFLMcut in particular resemble the trajectories in Figure 2a. When we do not truncate the dwell time distribution, the trajectories tend to incorporate very long dwell times as shown by the sFBM and sFLM models.

3.1.2 Model validation on stationary trajectories

We evaluated the predictive capabilities of the AD modeling approach by training the model parameters on the first half of the equilibrated MD trajectory data and then comparing the MSD calculated from AD model realizations to the MSD calculated from the second half of the equilibrated MD trajectory data. This metric is only meaningful if the ensemble of

solute trajectories is stationary.

We observe that in many cases, the distribution of solutes within the membrane evolves slowly with time. We defined the perceived equilibration time point for each solute based on the time at which the number of solutes inside the pores and tails stabilized (Figure S1). Nonetheless, we observe evidence of non-stationary solute behavior after equilibration, on the μ s timescale. In Table 3, we compare the MSD of the solutes calculated using trajectory data from the first and second halves of the “equilibrated” simulation time. Ethylene glycol and methanol show considerable differences in their MSDs. Urea and acetic acid appear to demonstrate satisfactory stationarity. The shapes of their MSDs curves are also very similar (see Figure S11 of the Supporting Information).

Residue	MSD	
	First Half	Second Half
Urea	0.31 (0.23, 0.40)	0.35 (0.28, 0.43)
Ethylene Glycol	3.11 (2.06, 4.25)	1.03 (0.76, 1.29)
Methanol	2.24 (1.61, 2.92)	1.06 (0.77, 1.39)
Acetic Acid	1.58 (1.11, 2.03)	1.43 (1.10, 1.73)

Table 3: In order to be considered stationary, the MSD of the ensemble of solute trajectories should be the same independent of the portion of equilibrated trajectory that is analyzed. The MSD values in this table are averages taken after a 500 ns time lag, calculated independently from the first and second halves of the equilibrated solute trajectories. Urea and acetic acid both appear to satisfy the stationarity criteria, while ethylene glycol and methanol show significantly smaller MSDs when calculated from the second half of the equilibrated trajectories.

Using the approach described above, we validated our 1 and 2 mode models with urea and acetic acid, since their trajectories appear stationary. The MSDs resulting from 1000 realizations of the AD model are shown in Figure 4. We consider the model’s prediction to match well if the MSD lies within the 1σ confidence intervals of the MD MSDs. We also look for qualitative agreement in the shape of the curves.

The models are capable of reasonably predicting the MD MSD values of the second half of the solute trajectories based on parameters generated from the first half when the dwell time distributions are parameterized by a power law with an exponential cut-off. At long

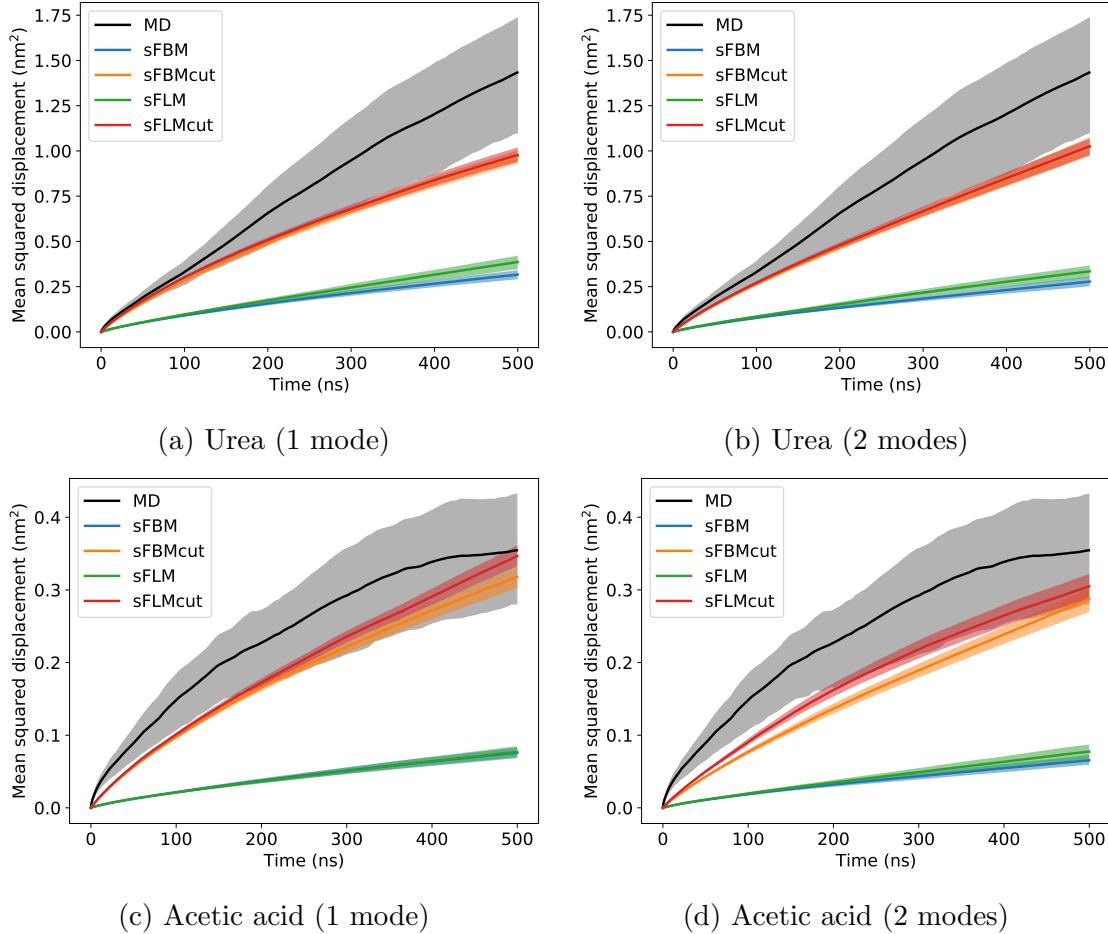


Figure 4: In most cases, when using power laws with exponential cut-offs (sFBMcut and sFLMcut), the MSD curves predicted by the AD model trained on the first half of the equilibrated data lie within the 1σ confidence intervals of the MD MSD curves generated from the second half of the equilibrated solute data. The models which use pure power laws systematically under-predict the MD MSD curves. Improper curvature of the MSD curves cause the magnitude of the urea's predicted MSDs to be under-predicted at long timescales and those of acetic acid to be under-predicted at short timescales.

timescales, the MSD of Urea is under-predicted for both the 1 and 2 mode models with the same true of acetic acid on short timescales. Without truncation of the power law distribution, the MD MSDs are underestimated in all cases because dwell times on the order of the MD simulation length are sampled and incorporated into the simulated anomalous diffusion trajectories. Considering the longest observed dwell time among all solutes was $1.3\ \mu\text{s}$ by ethylene glycol, we believe truncation is well-justified.

From a qualitative perspective, the models only have moderate success at predicting the shape of the MSD curves. Error in the MSD curvature can also help explain some of the error in the predicted MSD magnitudes. The curves predicted for Urea with both 1 and 2 modes appear to have too much curvature, which causes it to under-predict the MSD at long timescales, while those of acetic acid lack curvature, leading the AD model to under-predict the MSD at short timescales.

The under-estimate of Urea's MSD at long timescales is due to long timescale positional anti-correlation which may not be present in the molecularly detailed simulations of the system. The persistent curvature of the predicted MSD curves is a direct consequence of the Hurst parameter. Without anti-correlation, the process would be a pure CTRW for which one would expect the time averaged MSD curves to become linear. [26] It may be true that on the μs time scale, positional correlation is lost which would manifest as a transition from a sub-linear to linear MSD curve. A solution for more accurately modeling this behavior in the future may be to truncate the positional autocorrelation function. [63]

The under-estimate of the curvature of acetic acid's predicted MSD suggests that, in this case, the AD model over-estimates the Hurst parameter. This is not surprising because the Hurst parameter is challenging to quantify, especially with a relatively small amount of data (see Section 2.2.1 for further discussion of this challenge). A more accurate measurement of H would fix the shape of the MSD curve, but also lower the predicted MSD meaning we are either underestimating the width of the hop length distribution, favoring longer dwell times, or both.

This brief qualitative analysis suggests two shortcomings of the AD model. First, in a real system, positional anti-correlation may dissipate after a sufficiently long time lag, dependent on the solute being studied. Second, it is difficult to reliably parameterize the Hurst parameter which is important for accurately describing the curvature of the solute MSDs.

This analysis also suggests that working with only half of the data we collected ($\sim 2 \mu\text{s}$ post-equilibration) is not always sufficient for extracting reliable parameter estimates. In most cases, the magnitude of the MSD predictions after a 500 ns time lag are within or close to within error of the MD MSDs (for sFBMcut and sFLMcut), but still appear to systematically under-predict the mean. We may be operating on the border of the minimum amount of data required to accurately parameterize the AD model. Longer simulations and more independent trajectories may be necessary. Therefore, in the next section we will work with parameters fit to the full equilibrated portion of the solute trajectories, doubling the data.

3.1.3 Model predictions using all equilibrated data

We obtain reasonable predictions of the MD simulated MSDs when we parameterize the AD model with all available data after the perceived equilibration time. Although we concluded that methanol and ethylene glycol trajectories are not yet stationary, we include them in all analysis going forward because it is still instructive. The MSD curves generated from 1 and 2 mode models are overlayed with the MD simulated MSDs for comparison in Figures 5 and 6 respectively. The associated parameters are presented in Figures 7 and 8.

After a 1000 ns time lag, the 1 and 2 mode AD models do a fairly good job of predicting the magnitude of the MD MSD curves when we use truncated power laws to describe the dwell time distributions. In most cases, the sFBMcut and sFLMcut models give very similar results probably because their α_h parameter values are relatively close to 2, meaning the hop length distributions are nearly Gaussian even when fit to a more general Lévy distribution.

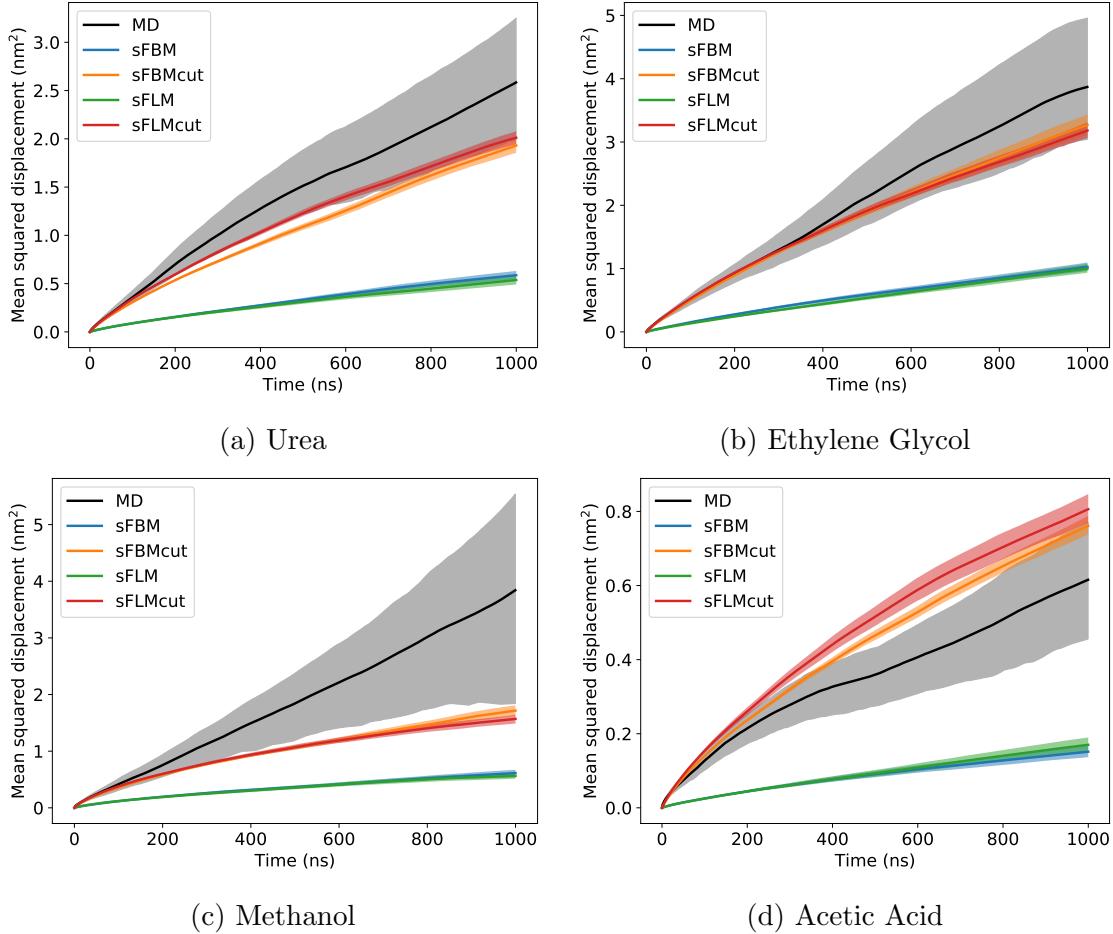


Figure 5: In most cases, MSDs generated from realizations of our 1 mode anomalous lie within or near the 1σ confidence intervals of MD-generated data when using an exponential cutoff on the dwell time distribution (sFBMcut and SFLMcut). When we do not apply an exponential cut-off (sFBM and SFLM), MSDs are underpredicted. Drawing hops from a truncated Lévy stable distribution yields MSDs similar to when hops are drawn from Gaussian distributions (sFLM and sFLMcut). In most cases, the MSDs are under-predicted at long timescales because the AD models show pronounced curvature which the MD simulations lack.

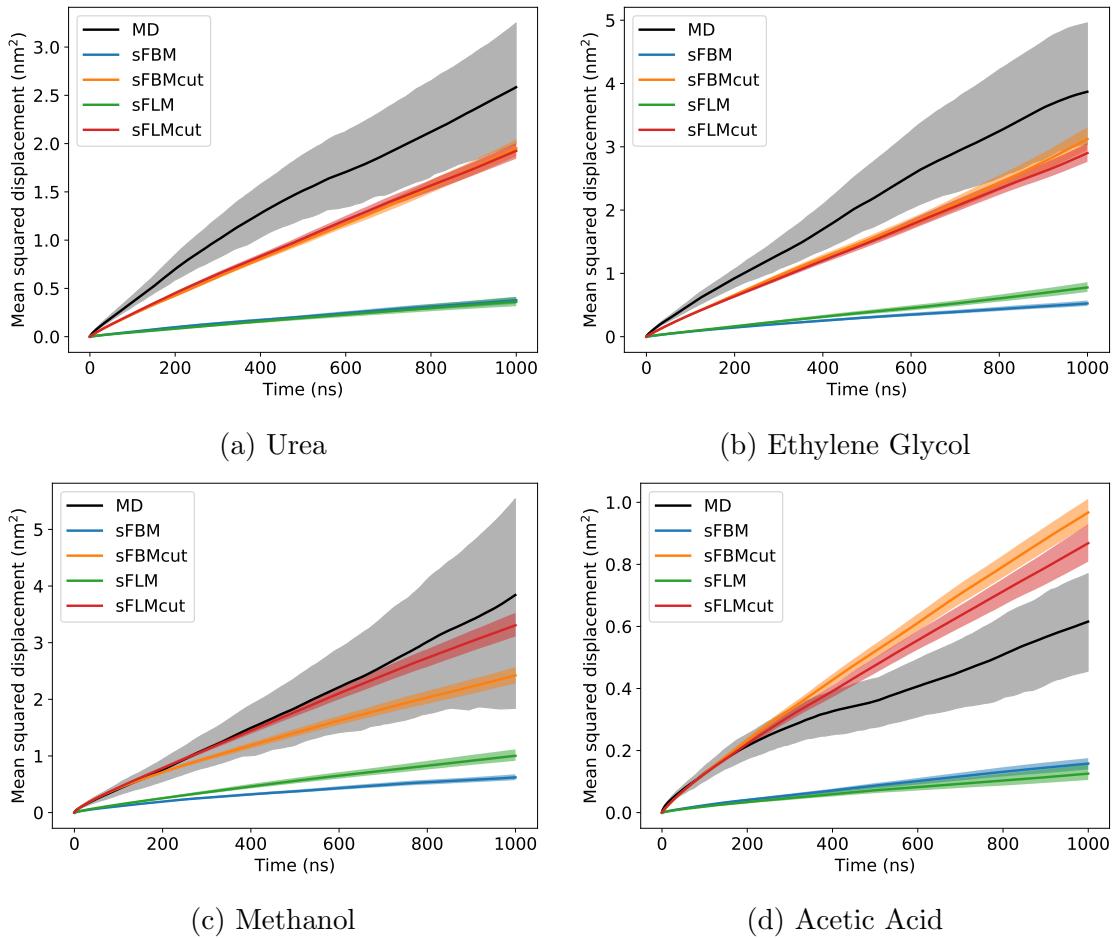


Figure 6: MSDs generated from realizations of our 2-mode anomalous diffusion model predict the magnitude of the MSDs to a similar degree of accuracy as the 1 mode model. Relative to the predictions of the 1 mode model the MSDs lack curvature because the hop correlation structure is broken every time a transition between tails occurs.

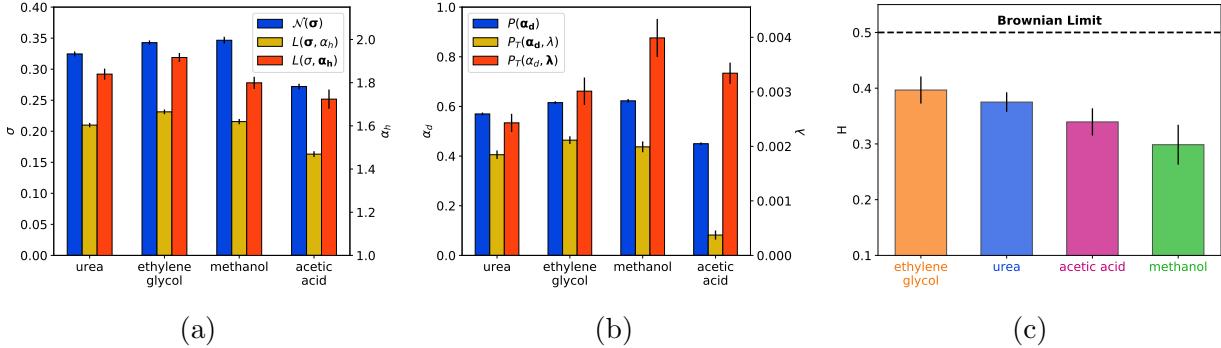


Figure 7: The parameters of the 1 mode model reveal differences in solute dynamics. (a) We parameterized Gaussian, $\mathcal{N}(\sigma)$, and Lévy stable, $L(\sigma, \alpha_h)$, distributions to describe solute hop lengths. We assume the mean (μ) to be zero for these distributions and there to be no skewness ($\beta = 0$) in the Lévy stable distributions. High values of σ and lower values of α_h result in larger hops. (b) We parameterized a pure power law, $P(\alpha)$, and a truncated power law, $P_T(\alpha, \lambda)$, distribution to describe solute dwell times. Lower values of α lead to heavier power law tails and higher values of λ truncate the distribution at lower dwell times. (c) Finally, we parameterized the hop autocorrelation function, $\gamma(H)$, to describe the degree of correlation between hops. Higher values of H display closer to Brownian behavior.

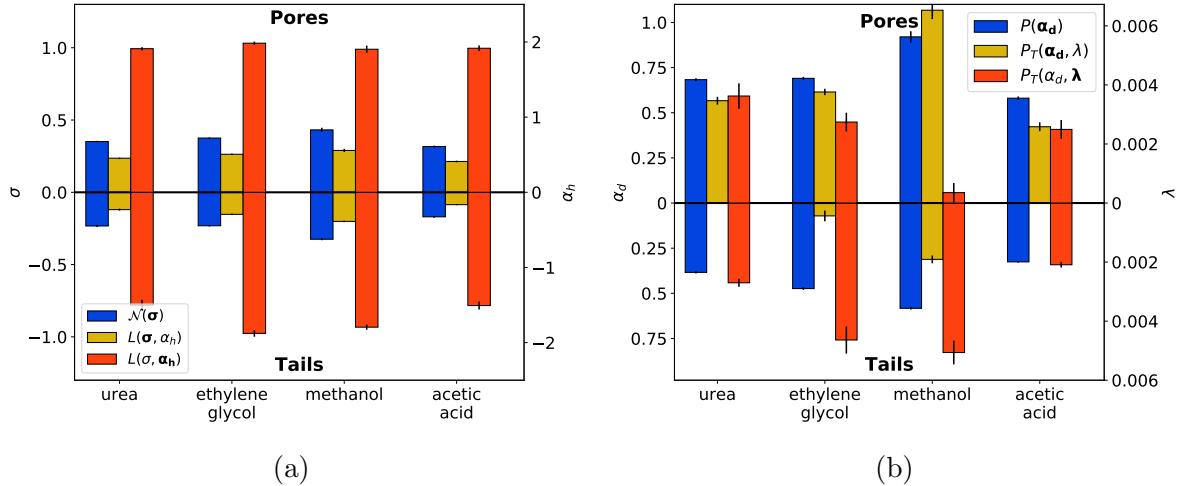


Figure 8: The two mode model parameterizes solute behavior in the pore and tails separately. We consider solutes to be within the pore region if they are 0.75 nm from a given pore center, otherwise they are in the tails. (a) Generally, movement is much more restricted in the tail region shown by lower σ values (smaller hops) for the normal and Lévy stable distributions. Values of α are significantly lower for urea and acetic acid meaning there is a larger probability that they will take large hops. (b) Dwell times are longer in the tails. Lower values of α correspond to power laws with heavier tails and thus higher probabilities of long dwell times. There is no easily discernable trend in λ of the truncated power law distribution. Note that we used the same Hurst parameter for both modes (shown in Figure 7c) due to a low number of sufficiently long sequences of hops in each mode.

The deviation between the 1 mode MSD predictions and MD are primarily due to differences in their curvature at long time lags. The shape and magnitude of the predicted curves appears accurate relative to MD at short time lags. However, they to undershoot the MD MSD mean as the time lag increases. As discussed in the previous section, long time positional anti-correlation may not exist in the MD system which results in a shift from sub-linear to linear MSD behavior. Acetic acid exemplifies this point. At first glance, acetic acid's predicted MSD (sFLMcut and sFBMcut) appears to match the curvature of MD quite well, but closer examination reveals that the MD MSD curve actually shifts from a sub-linear to a linear regime around 500 ns.

The 2 mode models shows considerably less curvature at large time lags than the 1 mode model because the implementation of the 2 mode model prevents long timescale correlation. The width of the hop distribution changes every time a switch between pores and tails occurs. We are unaware of an appropriate technique which can correlate hops that come from different hop distributions, therefore every time a mode switch occurs, the correlation structure is broken. Solutes which switch between modes the least show the greatest curvature. Methanol spends $> 90\%$ of its time in the tails, so mode transitions are relatively rare and the predicted MSD has significant curvature. This artificially solves the problem of long timescale correlation, however we do not recommend this as it has no theoretical basis.

The model parameters for the 1 and 2 mode models tell stories about each solute's behavior that help explain the difference between the MSDs of different solutes. Higher values of σ and lower values of α_h indicate larger average hop length magnitudes by increasing the hop length distribution's width and tail density respectively. Higher values of α indicate a lower probability of long dwell times. Higher values of λ truncate the power law distribution earlier preventing extremely long dwell times. Values of H closer to the Brownian limit of 0.5 indicate a lower degree of negative correlation between hops. All of these model parameters contribute to an overall increase in the predicted MSD and can be correlated to physical behavior that explains differences in the MD MSDs.

Turning first to the parameters of the 1 mode model, we can begin to break down the trends in solute MSDs. The parameters belonging to ethylene glycol and methanol are relatively similar which is consistent with their similar MD MSDs. Methanol tends to stay trapped for less time and takes larger hops but the most substantial difference is with respect to their Hurst parameters. In fact, methanol has the lowest H of all the solutes studied. It is possible that we have underestimated H since it appears that our model under-predicts the methanol MSD curve relative to other solutes. However, it remains plausible that methanol does have a low H value because it spends the majority of its time outside the pore region where collisions with tails are frequent. Urea has the third highest MSD which is primarily a consequence of more frequent and longer dwell times, indicated by lower values of α and λ . Urea's hop lengths (σ) and correlation (H) are comparable to ethylene glycol and methanol. Acetic acid has the smallest MSD among the solutes studied due to longer periods of entrapment and shorter hops. Its trapping behavior is parameterized by an α value significantly lower than other solutes, but an intermediate λ value, suggesting it experiences many medium length periods of entrapment. Its hops are smaller but are slightly compensated by a heavier tailed distribution (lower α_h) than the other solutes.

We can use the 2 mode model to gain an even deeper understanding of solute behavior in the pore versus in the tails. It is clear that solutes are significantly slowed while they are in the tail region where long dwell times are more probable (smaller α) and hops are smaller (smaller σ). Each solute spends a different amount of time in the tails (see Figure 9). Urea and acetic acid spend slightly more than half of their time in the tails (56% and 62% of their time respectively) while ethylene glycol spends about 44% of its time in the tails. Urea and acetic acid's compact and flat structure allows it to more easily partition into the tails while ethylene glycol prefers the pore region due to its two hydrophilic hydroxyl groups. Meanwhile methanol spends 91% of its time in the tails likely due to its small size. The value of α_h for urea and acetic acid in the tails is 1.50, meaning its hop distribution is heavy tailed relative to ethylene glycol and methanol whose α_h values are more consistent with a

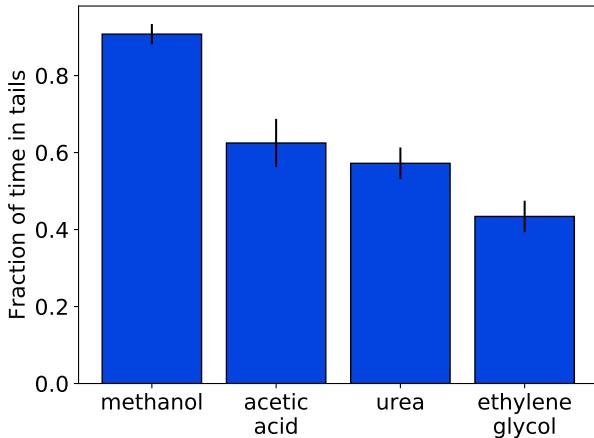


Figure 9: Solutes spend various amounts of time in the tail and pore region dependent on their size, shape and chemical functionality. Methanol’s small size favors occupation of the much larger accessible volume in the tails. Urea and acetic acid are fairly stable in both regions since they are small and hydrophilic. Ethylene glycol has a slight preference for the pores likely because it is a larger molecule with two hydrophilic hydroxyl groups.

Gaussian distribution ($\alpha=2$). Acetic acid and urea are structurally similar molecules, both planar with two heavy atoms attached to a carbonyl group. Their small size and rigid shape may allow them to occasionally slip through gaps in the tails. Meanwhile, methanol is small enough that it does not need to make larger jumps to escape traps.

Overall, the AD model does a reasonable job of predicting solute MSDs and its parameters can help us further understand solute dynamics. Hops appear to be well modeled as anti-correlated draws from Gaussian and Lévy stable distributions. The data strongly suggests that one must truncate the power law dwell time distributions in order to obtain accurate MSD estimates. We can further understand solute dynamics by adding radially dependent parameter distributions as in the 2 mode model. A significant amount of solute trajectory data is necessary in order to achieve good parameter estimates. The exact amount of data required is dependent on the solute studied.

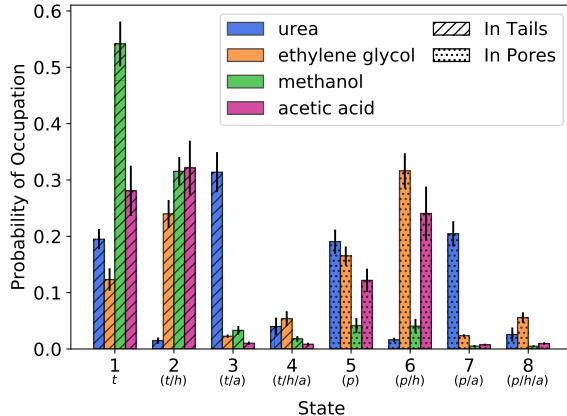


Figure 10: Solutes spend varying amounts of time under the influence of each trapping mechanism. To aid the reader, we labeled each state with an abbreviation which identifies the combination of conditions to which each solute is subject in each state: t - tails, p - pores, h - hydrogen bonded, a - associated with sodium. Solutes tend to favor the same types of interactions (e.g. hydrogen bonding and/or associating with sodium ions) independent of whether they are in the pores or the tails.

3.2 The Markov State-Dependent Dynamical Model

The AD model is useful if one does not know exact transport mechanisms in a system since it only requires time series data. However, since we have already studied transport mechanisms in detail in our previous work, we can attempt to model transport as transitions between known discrete states, defined in Table 2, with state-dependent positional fluctuations, which we refer to as the Markov state-dependent dynamical model (MSDDM).

Solute size and chemical functionality influence which states are visited most frequently. In Figure 10, we plotted the probabilities of occupying a given state at any time. Solutes tend to favor the same types of interactions independent of which region they are in.

Urea spends the largest fraction of its time trapped via association with sodium ions. It does so 31% of the total time while in the tails (state 3) and 21% of the total time while in the pores (state 7). Note that sodium does not drift significantly into the tails but sits close to the pore/tail region boundary. The electron-dense and unshielded oxygen atom of urea's carbonyl group is prone to associate with positively charged sodium ions. The nitrogen atoms of urea are only weak hydrogen bond donors.

Ethylene glycol spends the largest fraction of its time trapped in a hydrogen bonded state. It does so 24% of the total time while in the tails (state 2) and 32% of the total time while in the pores (state 6). The two hydroxyl groups of ethylene glycol readily donate their hydrogen atoms to the carboxylate head groups and the ether linkages between the head groups and monomer tails.

Methanol spends most of its time unbound in the tail region (state 1) and spends a significant portion of time hydrogen bonded while in the tail regions. Tail region hydrogen bonds are donated from methanol to the ether linkages between the monomer head groups and tails.

Finally, acetic acid spends the majority of its time hydrogen bonding both in and out of the pore (states 2 and 6). Although it has an unshielded carbonyl group in its structure, association with sodium ions in this environment is apparently a much weaker interaction than hydrogen bonding.

To create an MSDDM for each solute, we determined the state sequence associated with each solute trajectory and then generated emission distributions of fluctuations within each state as well as transitions between states. In theory, one could parameterize separate transition distributions for those which occur in the tails versus in the pores, however this would lead to a broken correlation structure similar to that seen in the 2 mode AD model.

We observe correlated emissions drawn from Lévy stable distributions. The deviation of the emission distributions from Gaussian behavior is far more pronounced than that seen in the hop length distributions of the previous section, therefore we did not consider the Gaussian case (see Figure 11a). The correlation structure between hops resembles that of FLM (Figure 11b). The parameters of the Lévy stable distributions along with their Hurst parameters are visualized in Figure 12 (and tabulated in the Supporting Information, Table S4).

Most of a solute's MSD is a consequence of transitions between the 8 states in Table 2. Motion while trapped in a state is highly anti-correlated as indicated by their consistently low

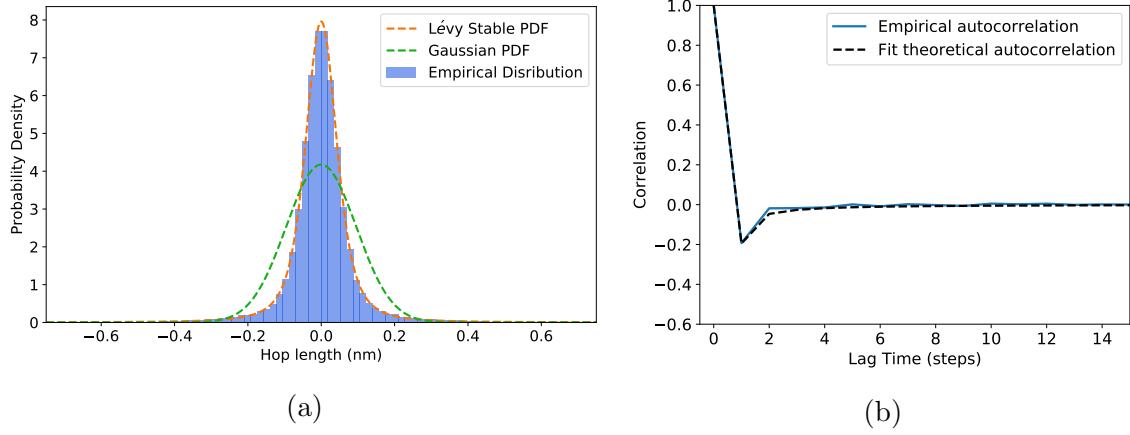


Figure 11: (a) Emission distributions are non-Gaussian and heavy-tailed. Shown here is the emission distribution for transitions between states. The maximum likelihood Gaussian fit severely underestimates the empirical density of hops near and far from zero while overestimating the density of hops at intermediate values. (b) Jumps drawn from the transition distribution are negatively correlated to each other. The normalized version of Equation 8 fits well to the data suggesting FLM is an appropriate way to model jumps.

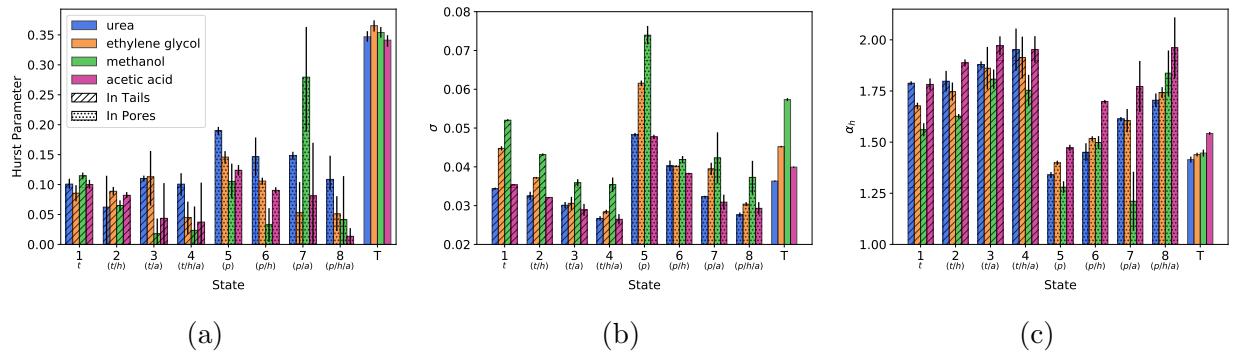


Figure 12: The parameters of the MSDDM are strong functions of trapping mechanisms. We observe different parameters but with similar trends between the tail and pore region. The states are defined in Table 2. See Figure 10 for a description of the abbreviation under each state number. The legend in (a) applies to all subplots. (a) Motion is highly anti-correlated in trapped states. As the number of simultaneously influencing trapping mechanisms increases, the Hurst parameter decreases. The Hurst parameter is highest during transitions between states (state T). (b) As more trapping mechanisms simultaneously influence solutes, the width of the hop length distribution decreases. The largest hops occur when solutes are unbound in the pores. (c) The weight of the hop length distribution’s tails, parameterized by α_h , increases as more trapping mechanisms influence solutes simultaneously. The transition hop distributions have among the heaviest tails.

Hurst parameters. Perfectly anti-correlated motion ($H=0$) results in no contribution to the solute's MSD. There is a weak negative trend in the Hurst parameter values as the number of simultaneously influencing trapping mechanisms increases (Figure 12a). Surprisingly, state 5 H values are low despite the solutes being subject to no trapping mechanisms. Highly anti-correlated behavior in the pores in the absence of any trapping mechanism is likely caused by collisions with monomer head groups within the pore region. The Hurst parameters for transitional (T) emissions are up to 18 times higher than emissions from trapped states. The value of α_h for transition emissions is also relatively low giving higher probabilities to larger hops.

As solutes are influenced by more trapping interactions simultaneously (e.g. hydrogen bonding *and* association with sodium versus just hydrogen bonding), the width of the hop length distribution, σ , decreases while its Lévy index, α_h , increases. Treating states in the tail and pore regions independently, σ is largest and α_h is smallest when solutes are not hydrogen bonding or associating with sodium (states 1 and 5). Solutes are free to move and take occasionally large hops. The smallest σ and highest α_h values are measured when solutes are hydrogen bonding and associating with sodium at the same time (states 4 and 8). Motion is restricted by multiple tethers which maintains a relatively narrow distribution of hop lengths.

Qualitatively, realizations of the MSDDM are not consistent with MD solute trajectories. Using the parameters shown in Figure 12 we generated realizations of the MSDDM and plotted a typical realization for each solute in Figure 13. Except in rare cases, the solutes do not make large hops. There are two reasons for this behavior. First, the width of hop length distributions are much smaller than those of the AD model. Closer examination of urea's MD trajectories shown in Figure 2a reveal that hops tend to be an accumulation of a series of hops in the same direction. All of the hops in the MSDDM are negatively correlated which prevents this from happening. The second reason is a consequence of using a single hop length distribution for transitions. Many transitions occur between two trapped

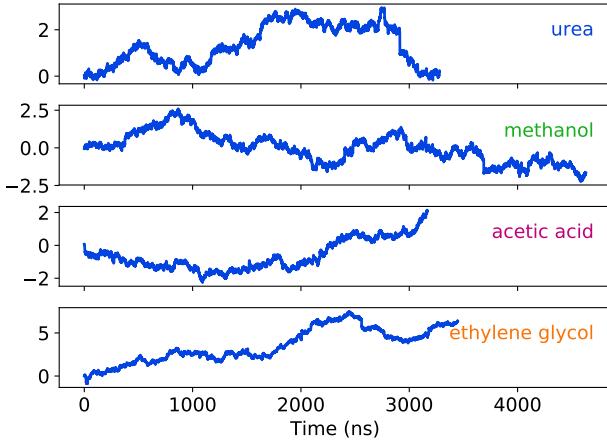


Figure 13: Realizations of the MSDDM for each solute do not reproduce the hopping and trapping behavior observed in our MD simulations. Except for a few large hops, the trajectories are qualitatively similar to what one might expect for Brownian motion. Note that the length of the realizations are equal to the length of the equilibrated portion of the MD trajectories for each solute.

states where the transitional hops are actually very small. Our model ignores this physical restriction which can cause the solute to drift rather than stay trapped.

Despite qualitative differences in trajectory realizations, the dynamics of urea, ethylene glycol and, to a lesser extent, methanol appear to be well-captured by the MSDDM. We simulated 1000 MSDDM trajectory realizations for each solute, as described in Section 2.3, then calculated their MSDs (see Figure 14). In most cases, the MSDDM predicts the magnitude of the MD MSDs within their 1σ confidence intervals. One could argue that the curves of ethylene glycol and urea are statistically indistinguishable. However, it's possible that they may diverge on longer timescales due to differences in their curvature. Similarly, methanol's MD MSD is nearly linear while the MSDDM prediction is curving downward implying that it may underestimate the MSD on longer timescales.

The predicted MSD of acetic acid is severely over-estimated most likely due to poor parameterization. As stated earlier, most of a solute's motion is due to displacements during state transitions. Acetic acid has a transitional Hurst parameter close to urea's and a transitional σ value that is higher than urea's. This suggests that the heavier tails of urea's

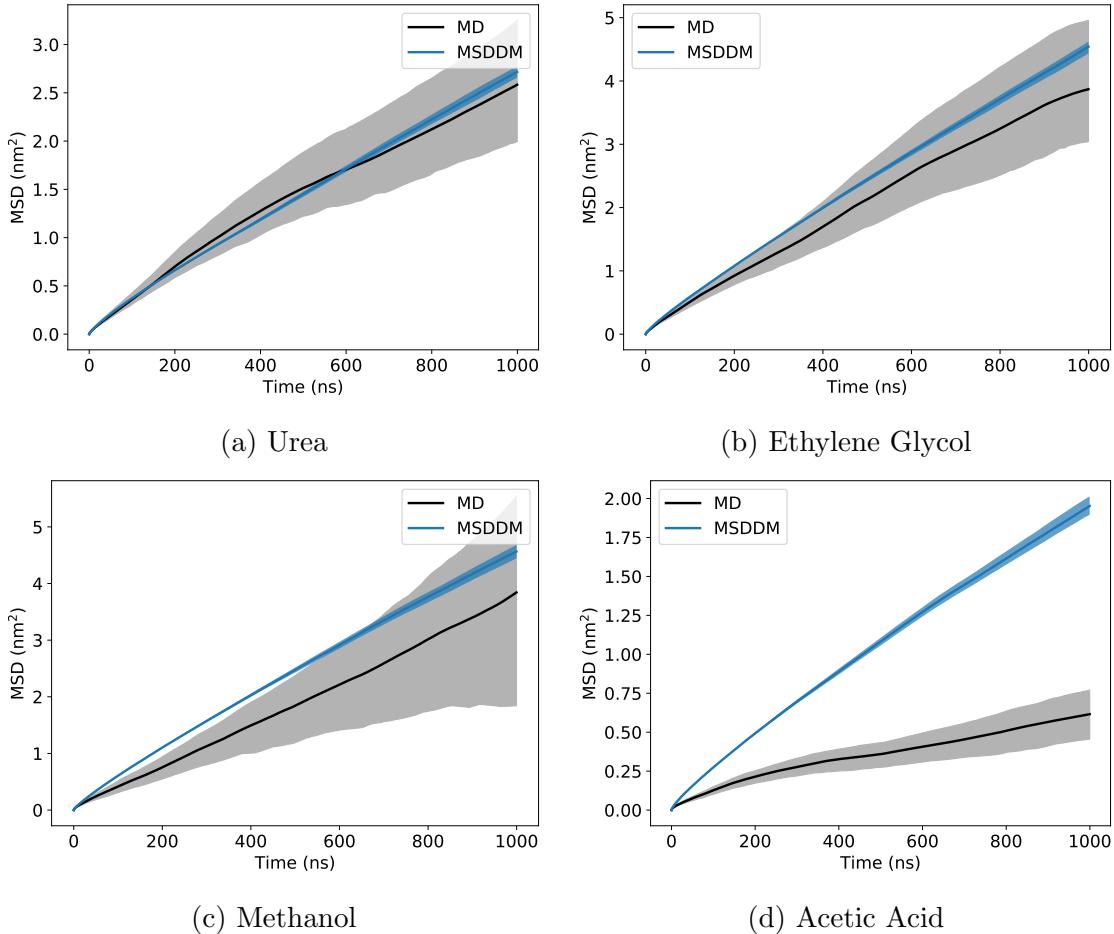


Figure 14: In most cases, the magnitude of the MSD curves predicted by the MSDDM agree well with those generated from MD simulations. The predicted MSD curves of urea and ethylene glycol lie within the 1σ confidence intervals of MD for all time lags. Methanol over-predicts the MSD at small time lags and acetic acid grossly over-predicts the MSD at all time lags. Like the anomalous diffusion model, the MSDDM doesn't fully capture the curvature of the MD MSD curves.

transitional hop distribution (greater α_h value) are in large part responsible for urea's higher predicted MSD. Although, note that acetic acid's predicted MSD actually lies just within the lower bound of Urea's MD MSD 1σ confidence interval. There is also the strong possibility that the over-estimate is a consequence of lumping together all of acetic acid's transitional hops into a single correlated distribution.

The MSDDM suffers similar shortcomings to the AD model. First, curvature on long timescales due to positional anti-correlation may result in underestimates of the MSD at large time lags. As suggested earlier, truncating the positional autocorrelation function may help solve this issue. Second, parameterization can be challenging. In addition to possibly poor estimates of the Hurst parameter, we made the simplifying assumption that we could lump all transitional hops into a single distribution. While this appears to work in most cases, it is the source of qualitative mismatches between MSDDM realizations and MD simulation trajectories and it may be the source of the over-estimate of acetic acid's MSD.

The MSDDM offers mechanistic insight and, in 3 out of 4 cases, gives reasonable predictions of solute MSDs. To successfully apply the model, one must know representative discrete states beforehand or apply some technique to identify them. At the very least, we can used the MSDDM to obtain a detailed quantitative picture of dynamics within each state in terms of the magnitude of their fluctuations and correlation with previous motion. If we have confidence in their predictive ability, then we can project solute trajectories out to macroscopic timescales and estimate macroscopic observables such as solute flux and selectivity.

3.3 Solute Flux

We used the 1-mode sFBMcut (see Table 1) anomalous diffusion model and the MSDDM in order to demonstrate how one can use their realizations in order to calculate the flux (see section 2.4) of solutes given model parameters extracted from MD simulations. The 1-mode sFBMcut model generates predictions similar to the 1-mode sFLMcut model at a lower computational cost and is far more accurate than sFBM and sFLM. We showed that

the 2-mode AD model has a broken correlation structure, so we do not consider it here.

It is computationally infeasible to simulate trajectories long enough that they traverse the length of a macroscopic pore. To date, the thinnest H_{II} LLC membrane synthesized with the monomer in this work was 7 μm thick. Using 24 cores to simulate trajectory realizations in parallel, it takes on the order of 1 day to simulate 10000 sFBMcut realizations of solutes traversing a 50 nm pore. The computational requirements of the MSDDM are about 10 times longer. For both models, the RAM requirements and performance scales greater than linearly and thus would take an infeasible amount of memory and time to simulate transport through a pore over 100 times longer.

We used simulated trajectories which traverse computationally-reasonable length pores in order to construct an empirical model which one can use to estimate particle flux for arbitrary length pores. We fit Equation 11 to the empirical distribution of first passage times in Figure 15a and used the expected value of the analytical equation to calculate flux from Equation 10. As shown in Figures 15b and 15c, the flux appears to scale according to a power law of the form:

$$J(L) = cL^{-\beta} \quad (16)$$

We determined that it is only necessary to simulate 1000 realizations of each model in order to parameterize Equation 16. We used 10000 trajectory realizations in order to generate each point in Figure 15b, however, in Figure S14 of the Supporting information, we show that one can still show statistically significant differences in solute flux, with minimal change to the mean flux values, using as few as 100 trajectory realizations. For better precision, we recommend simulating at least 1000 realizations. Since the computational cost of the MSDDM is greater than the AD model, we generated each point of the flux curves in Figure 15c using only 1000 realizations at each pore length.

The scaling of solute flux with pore length is primarily influenced by anti-correlation between solute hops. In Figure 16a, we show that β , as derived from the sFBMcut model, is inversely related to the Hurst parameter. This makes intuitive sense since higher degrees

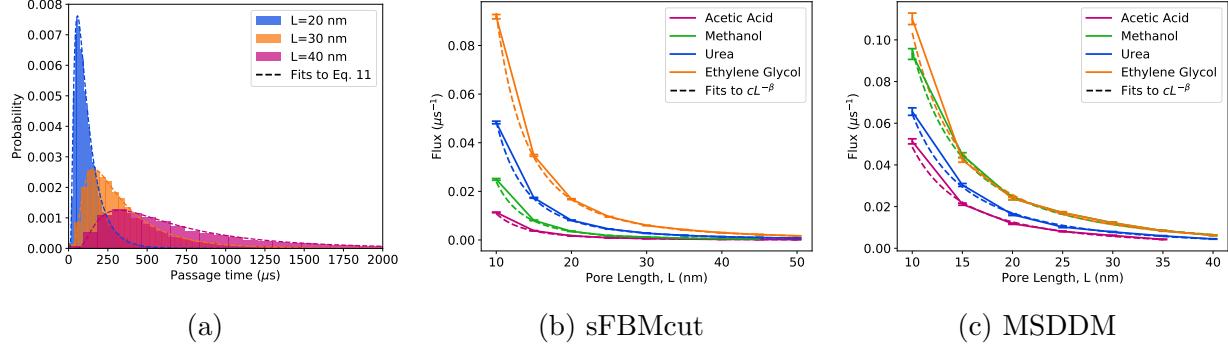


Figure 15: (a) The distributions of first passage times generated from the sFBMcut model fit well to Equation 11. We show similar fits for the remaining solutes as well as the MSDDM in Figures S12 and S13 of the Supporting Information. (b) The single particle flux measured by the sFBMcut AD model and (c) the MSDDM decays with increasing pore length. The rankings of solute fluxes are consistent with the MSDs predicted by each model. The ranking of the MSDDM is most consistent with MD MSDs. The MSDDM predicts higher fluxes than the sFBMcut model. We fit the single particle solute flux versus pore length, L , to a power law function of the form $cL^{-\beta}$ (dashed lines in (b) and (c)).

of anti-correlation should slow the rate at which solutes cross the membrane pore. β , as derived from the MSDDM, appears approximately constant but this is consistent with nearly invariant transitional Hurst parameters (see Figure 12a). However, Figure 16a does raise the question as to why β is so much lower than the sFBMcut model.

To understand the behavior of the MSDDM, we studied the β parameter derived from variations on each model, visualized in Figure 16b. First, we removed anti-correlation between hops in the sFBMcut trajectory (set $H=0.5$). The β parameter drops to values similar to the MSDDM. When we remove anti-correlation *and* dwell times between hops (set dwell times equal to 1), effectively simulating Brownian motion, we observe similar β parameters. These two variations imply that, in the AD models, the β parameter is only a function of H and not dependent on dwell times (α, λ) or hop lengths (σ). The MSDDM is fundamentally similar to the AD model in that one makes correlated draws from a hop distribution. Therefore, we should expect the β value of the MSDDM to be larger than that of Brownian motion.

The value of β for the MSDDM is under-estimated due to assumptions of the model itself

as well as inaccuracies in the correlation structure of very long FLM trajectories. Turning first to the model itself, we have designed it to ensure that the magnitude of the hops in the series of transitions between states are anti-correlated from start to finish. This assumes that the transition correlation structure is unaffected by the sub-trajectories between each state transition. Each time a state transition occurs, one must initialize a new time series sub-trajectory with its own correlation structure. Since we add the transitional hop lengths to each end of trapped state sub-trajectories, the transitional hop lengths are shifted with respect to one another, decreasing correlation between them. We tested this reasoning by modifying the MSDDM to completely immobilize particles except when they transition between states. Surprisingly, β is still close to the Brownian value. Further experimentation reveals that this is actually a consequence of the FLM simulation procedure. Simulating FLM requires Riemann-sum approximations of the stochastic integrals defining the process. To generate the curves in Figure 15c, we needed to correlate 25–1000 times more hops than for the MSD predictions in Figure 14. In short, it is computationally infeasible to use enough terms to accurately incorporate long timescale correlations into our long MSDDM realizations. Thus at long timescales, we lose correlation between transitional jumps. We confirmed this hypothesis by using fractional Brownian motion, for which we have an exact simulation method, in place of FLM in the MSDDM algorithm. When we use FBM, β increases well above the Brownian value. β increases even further if we immobilize the trapped states. Thus the low value of β of the MSDDM is a consequence of model assumptions and inexact simulation of FLM.

The values of c are directly related to the magnitude of solute flux and, where applicable, are functions of σ , α , λ and α_h . In Figure 17a, we plot the c values measured for each solute with each model from lowest to highest. In general, c values generated by the MSDDM model are smaller than those generated by the sFBMcut AD model because the AD model has to compensate for high β values. Comparison of Figure 17a with Figures 15b and 15c reveals that the ranking of the c parameters is consistent with the ranking of solute

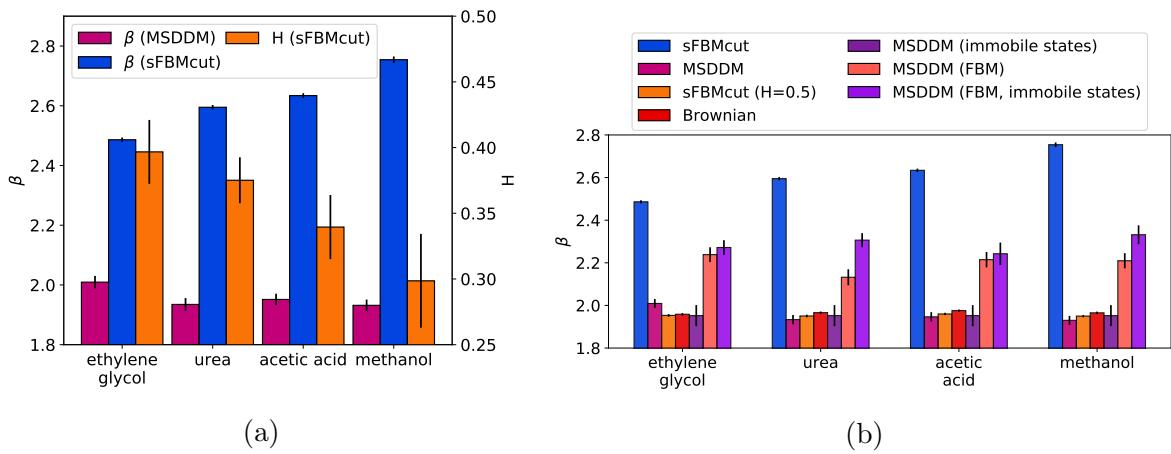


Figure 16: (a) The β values of the sFBMcut model appear to be inversely proportional to the Hurst parameters. The β values of the MSDDM are much lower with much less variation than the sFBMcut model. (b) The β parameters of the MSDDM are low because sub-trajectories between state transitions decorrelate transitional hops and because our FLM simulation procedure does not accurately correlate hops after long time lags. The high β parameters of the sFBMcut model are a consequence of anti-correlation between hops. Removing hop correlations causes β to drop down close to MSDDM values (sFBMcut ($H=0.5$)). Removing dwell times in addition to hop correlation yields a similar value of β (Brownian). Immobilizing particles while in a trapped state also yields a similar value of β (MSDDM (immobile states)). Replacing FLM with FBM in the MSDDM raises β well above the Brownian value (MSDDM (FBM)). Replacing FLM with FBM and immobilizing particles in trapped states further raises the value of β .

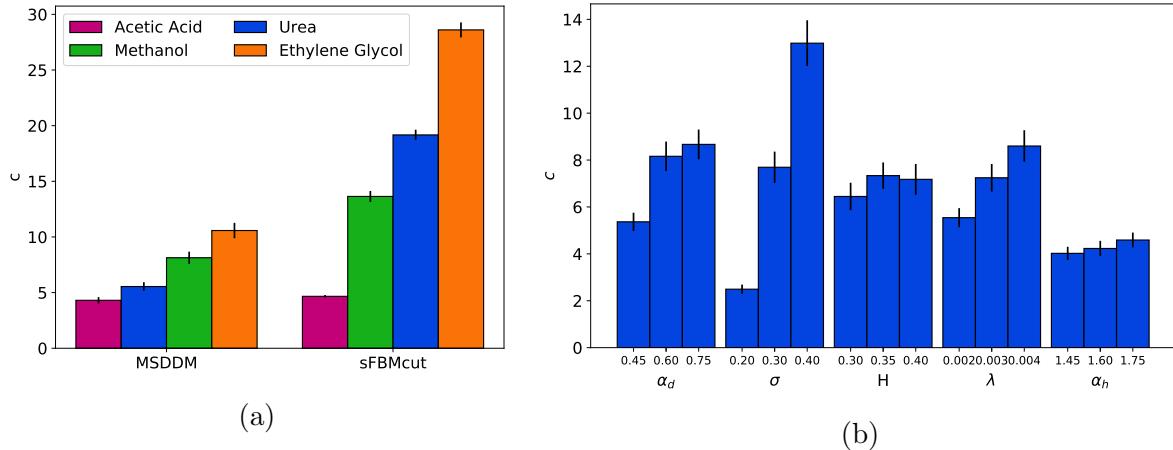


Figure 17: (a) The c parameter increases with solute flux (compare ranking with Figures 15b and 15c). The c parameters of the MSDDM are smaller than the sFBMcut AD model because the AD model must compensate for its strong length dependence (higher β). (b) Changes to parameters which increase the rate of solute displacement result in larger values of c . To test the dependence of c on α , σ , H and λ , we chose a single set of parameters, representative of solutes parameterized by the sFBMcut AD model, and generated realizations of the model by varying each parameter independently about the same base parameter set. Therefore, the middle bars for these four 4 parameters are generated from the same set of parameters. Decreased dwell times (increased α), increased hop lengths (increased σ), and a lower cut-off to the dwell time distribution (increased λ) lead to increases in c . c responds linearly to increases in σ and λ and non-linearly to α . The data suggests that the c parameters do not depend on hop anti-correlation (H). To test α_h , we used the set of parameters defining the MSDDM of Urea (since sFBMcut does not use a Lévy distribution) and varied the α_h parameter of the transitional hops. As α_h increases, c increases slightly since the density of the Lévy stable PDF is shifted away from rare hops and broadened at its center. We expect similar dependence of c on σ and H of the MSDDM.

flux. Therefore, it is reasonable to hypothesize that any parameter which acts to modify the rate at which solutes move through the membrane pores will affect c for various physical reasons. In Figure 17b we demonstrate that decreased dwell times (increased α), increased hop lengths (increased σ), cutting off the dwell time distribution at shorter times (increased λ) and decreased tail weights of the hop length distribution (increased α_h) independently lead to an increase in c . Figure 17b also suggests that c is not dependent on hop anti-correlation (H).

The power law decay of the flux with pore length implies the following relationship for

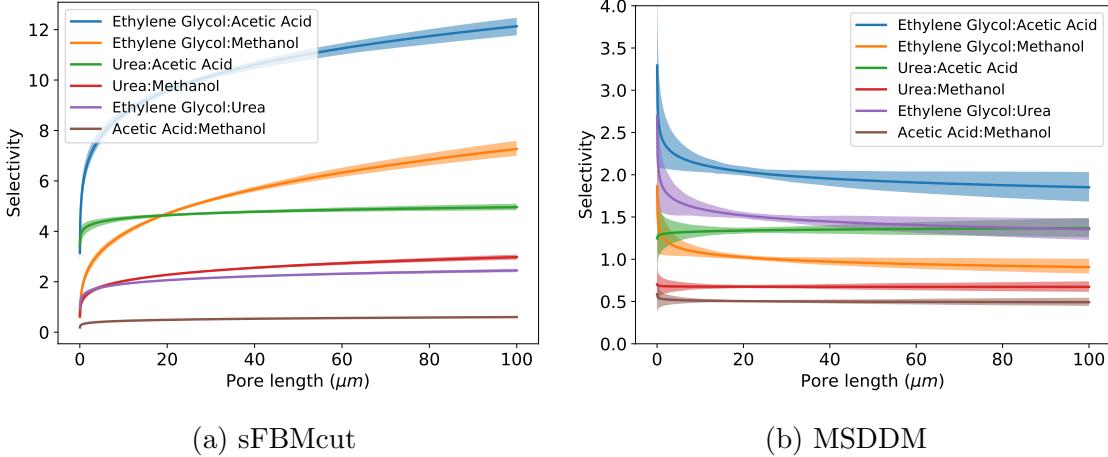


Figure 18: The selectivity of 2 species changes monotonically with pore length. The strength of dependence on pore length depends on the difference between β values. The selectivities predicted by the sFBMcut AD model (a) are much larger in magnitude than those predicted by the MSDDM (b) because predictions of the AD model result in a much larger variation in flux (compare Figures 15b and 15c).

selectivity via substitution of Equation 16 into Equation 15:

$$S_{ij}(L) = \left(\frac{c_i}{c_j} \right) L^{(\beta_i - \beta_j)} \quad (17)$$

In Figure 18, we plot Equation 17 for pore lengths ranging from those studied in Figure 15 to macroscopic-length pores. Greater differences in β values leads to selectivities that are stronger functions of pore length. For $(c_i/c_j) = 1$, LLC membranes will be more selective towards passage of solutes with lower β values versus those with high β values. When β values are equal, as is the case for uncorrelated motion, selectivity only depends on the pre-factors, c . In most cases, the length dependence of selectivity plateaus quickly which implies that a coating of LLC membrane with a thickness on the order of those already made experimentally should be sufficiently optimal.

The selectivities predicted by the sFBMcut model are both higher in magnitude and ordered differently than the those predicted by the MSDDM. The magnitudes of the sFBM-cut predictions are higher because the predicted solute fluxes vary more widely than those predicted by the MSDDM (compare Figures 15b and 15c). The ordering is different mostly

because the ordering of the flux predictions is different.

We cannot say with absolute certainty which selectivity curves are more trustworthy. The sFBMcut model uses exact simulation techniques and yields qualitatively accurate solute trajectories, but it may overestimate long timescale correlation present in the real system. For this reason, we might expect the β values of the true flux curves to be closer to the Brownian value as exhibited by the MSDDM. But the MSDDM yields qualitatively inaccurate solute trajectories and we do not have a good understanding of how to precisely control the point at which solute hops become decorrelated.

However, we may gain the most by using the two models in tandem. In both cases, the data suggests that this particular LLC membrane might be most useful for selectively separating ethylene glycol from acetic acid. Ethylene glycol exhibits the highest flux while acetic acid has the lowest. Ethylene glycol has the weakest dependence on pore length (lowest β) according to the sFBMcut model, while acetic acid has the second highest dependence. This analysis certainly merits further experimental exploration.

4 Conclusions

We have tested two different mathematical frameworks for describing solute motion by applying them to an H_{II} phase LLC membrane. The values obtained for the parameters when fitting the models to the time series data offer important mechanistic insight on the molecular details of transport. Subordinated fractional Brownian and Lévy motion have a strong theoretical foundation in the anomalous diffusion literature. Our single mode model quantifies and allows comparison of the hopping and trapping behavior among solutes. A two mode model that describes dynamics based on whether a solute is in or out of the pore region allows us to break down individual solute motion into the two distinct regimes and we showed that solute motion is clearly restricted while in the tail region. Our Markov state-dependent dynamical model uses explicitly defined trapping mechanisms and gives a

nice description of transitions between these observed states, the equilibrium distribution of solutes among states as well as the type of stochastic behavior shown in each state.

All of the models we tested have moderate success reproducing the MSDs of solute trajectories generated by MD simulations. In almost all cases, large portions of the MSDs predicted by our models fall close to or within the 1σ confidence intervals of MD. Aside from inaccuracies in parameter estimates, the most significant shortcoming of all of our models is their ability to accurately portray the curvature of the MSD curves. The MD MSD curves tend to become linear on long timescales, but the MSD curves generated by our models have persistent curvature that causes them to under-predict the MSD on very long timescales. This is a consequence of the correlation structure associated with fractional motion. In the future, we could address this by implementing a truncation to the positional autocorrelation function.

We demonstrated how one could used our models in order to to determine macroscopic flux and selectivity. We showed that, when using the AD model, solute flux decreases with pore length at a rate faster than pure Brownian motion due to anti-correlation between hops. However, using the MSDDM, we observed a decay in flux similar to Brownian motion. This is primarily due to the inexact method that we use to correlate hops of the fractional Lévy motion process. When we replace FLM with FBM in the MSDDM algorithm, we observe a flux decay more consistent with that predicted by the AD model. Finally, we used the ratio of solute fluxes in order to calculate selectivity. Due differences in hop anti-correlation, we observe a length dependent selectivity. However, this length dependence mostly plateaus at length scales that can be synthesized experimentally. The data suggests that this particular membrane may be useful for separating ethylene glycol and acetic acid.

Of course, our approach to modeling LLC membrane selectivity neglects a number of factors that influence real-world aqueous membrane separations. The most glaring is that we neglected convective flux which should work to increase overall solute flux and lower selectivity (see Equation 14). Defects in the pore structure will have a strong influence

on solute flux which may exacerbate the length dependence of flux. Even in a perfectly structured membrane, the barrier to pore entry from the bulk solution will have its own influence on experimentally observed selectivities. This energetic barrier can be made even worse by foulants deposited on the membrane's surface. Therefore, our predictions should be primarily used to identify the most promising types of selective separations for an H_{II} LLC membrane made from a specific monomer.

Our mathematical models help us think about how to design LLC monomers for solute-specific separations by forcing us to think in terms of controlling solute trapping. The anomalous diffusion model is quite flexible because it does not require knowledge of specific trapping mechanisms. Screening a set of solutes and applying the anomalous diffusion model can help uncover trapping mechanisms by forcing the scientist to identify features common to solutes with long or short hop lengths and dwell times. The MSDDM is a powerful way to characterize explicit trapping mechanisms. It clearly partitions a trajectory into discrete mechanisms and quantifies their relative dominance in the real system. For example, it is clear that sodium ion association is primarily responsible for trapping urea while hydrogen bonding dominates ethylene glycol. If one re-designs an LLC monomer to eliminate one source of trapping, then high selectivities are possible. Overall, in tandem or separately, our models facilitate new ways of approaching complex separation problems.

Supporting Information

Detailed explanations and expansions upon the results and procedures mentioned in the main text are described in the Supporting Information. This information is available free of charge via the Internet at <http://pubs.acs.org>.

Acknowledgments

This work was supported in part by 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. 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.

We would also like to thank Richard Noble for helping us make the connection between the empirical mean first passage time distribution and the analytical equation to which we fit the distributions (Equation 11).

References

- [1] 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.
- [2] M. O. Barbosa, N. F. F. Moreira, A. R. Ribeiro, M. F. R. Pereira, and A. M. T. Silva, “Occurrence and Removal of Organic Micropollutants: An Overview of the Watch List of Eu Decision 2015/495,” *Water Research*, vol. 94, pp. 257–279, May 2016.
- [3] J. E. Mondloch, M. J. Katz, W. C. Isley Iii, P. Ghosh, P. Liao, W. Bury, G. W. Wagner, M. G. Hall, J. B. DeCoste, G. W. Peterson, R. Q. Snurr, C. J. Cramer, J. T. Hupp, and O. K. Farha, “Destruction of chemical warfare agents using metalorganic frameworks,” *Nature Materials*, vol. 14, pp. 512–516, May 2015.
- [4] D. L. Gin, J. E. Bara, R. D. Noble, and B. J. Elliot, “Polymerized Lyotropic Liquid Crystal Assemblies for Membrane Applications,” *Macromol. Rapid Commun.*, vol. 29, no. 5, pp. 367–389, 2008.
- [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] 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.
- [7] 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 Mem-

branes with Vertically Aligned 1 nm Pores by Magnetic Field Directed Self-Assembly,” *ACS Nano*, vol. 8, pp. 11977–11986, Dec. 2014.

- [8] 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.
- [9] 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.
- [10] 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.
- [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] W. Zhou, W. Gu, Y. Xu, C. S. Pecinovsky, and D. L. Gin, “Assembly of Acidic Amphiphiles into Inverted Hexagonal Phases Using an l-Alanine-Based Surfactant as a Structure-Directing Agent,” *Langmuir*, vol. 19, pp. 6346–6348, Aug. 2003.
- [13] 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(-acryloxyalkoxy)benzoate Salts,” *Macromol. Chem. Phys.*, vol. 201, no. 11, pp. 1128–1133, 2000.
- [14] B. J. Coscia and M. R. Shirts, “Chemically Selective Transport in a Cross-Linked HII Phase Lyotropic Liquid Crystal Membrane,” *J. Phys. Chem. B*, June 2019.

- [15] H. Vinh-Thang and S. Kaliaguine, “Predictive Models for Mixed-Matrix Membrane Performance: A Review,” *Chem. Rev.*, vol. 113, pp. 4980–5028, July 2013.
- [16] J. Geens, B. Van der Bruggen, and C. Vandecasteele, “Transport model for solvent permeation through nanofiltration membranes,” *Separation and Purification Technology*, vol. 48, pp. 255–263, Mar. 2006.
- [17] S. Darvishmanesh and B. Van der Bruggen, “Mass Transport through Nanostructured Membranes: Towards a Predictive Tool,” *Membranes (Basel)*, vol. 6, Dec. 2016.
- [18] J. G. Wijmans and R. W. Baker, “The Solution-Diffusion Model: A Review,” *J. Membr. Sci.*, vol. 107, pp. 1–21, Nov. 1995.
- [19] D. R. Paul, “Diffusive Transport in Swollen Polymer Membranes,” in *Permeability of Plastic Films and Coatings: To Gases, Vapors, and Liquids* (H. B. Hopfenberg, ed.), Polymer Science and Technology, pp. 35–48, Boston, MA: Springer US, 1974.
- [20] C. Manzo and M. F. Garcia-Parajo, “A review of progress in single particle tracking: from methods to biophysical insights,” *Rep. Prog. Phys.*, vol. 78, p. 124601, Oct. 2015.
- [21] 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],” *Living Journal of Computational Molecular Science*, vol. 1, no. 1, p. 6324, 2018.
- [22] R. Metzler and J. Klafter, “The Random Walk’s Guide to Anomalous Diffusion: A Fractional Dynamics Approach,” *Phys. Rep.*, vol. 339, pp. 1–77, Dec. 2000.
- [23] J.-P. Bouchaud and A. Georges, “Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications,” *Physics Reports*, vol. 195, pp. 127–293, Nov. 1990.

- [24] R. Gorenflo and F. Mainardi, “Fractional Calculus,” in *Fractals and Fractional Calculus in Continuum Mechanics* (A. Carpinteri and F. Mainardi, eds.), International Centre for Mechanical Sciences, pp. 223–276, Vienna: Springer Vienna, 1997.
- [25] R. Klages, G. Radons, and I. M. Sokolov, *Anomalous Transport: Foundations and Applications*. John Wiley & Sons, Sept. 2008. Google-Books-ID: TDch5DfNgWoC.
- [26] Y. Meroz and I. M. Sokolov, “A Toolbox for Determining Subdiffusive Mechanisms,” *Phys. Rep.*, vol. 573, pp. 1–29, Apr. 2015.
- [27] G. T. Morrin and D. K. Schwartz, “Three Regimes of Polymer Surface Dynamics under Crowded Conditions,” *Macromolecules*, vol. 51, pp. 1207–1214, Feb. 2018.
- [28] 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, pp. 24128–24164, Oct. 2014.
- [29] E. W. Montroll and G. H. Weiss, “Random Walks on Lattices. II,” *Journal of Mathematical Physics*, vol. 6, pp. 167–181, Feb. 1965.
- [30] B. Mandelbrot and J. Van Ness, “Fractional Brownian Motions, Fractional Noises and Applications,” *SIAM Rev.*, vol. 10, pp. 422–437, Oct. 1968.
- [31] J.-H. Jeon and R. Metzler, “Fractional Brownian Motion and Motion Governed by the Fractional Langevin Equation in Confined Geometries,” *Phys. Rev. E*, vol. 81, p. 021103, Feb. 2010.
- [32] D. S. Banks and C. Fradin, “Anomalous Diffusion of Proteins Due to Molecular Crowding,” *Biophys. J.*, vol. 89, pp. 2960–2971, Nov. 2005.

- [33] T. Neusius, I. Daidone, I. M. Sokolov, and J. C. Smith, “Subdiffusion in Peptides Originates from the Fractal-Like Structure of Configuration Space,” *Phys. Rev. Lett.*, vol. 100, p. 188103, May 2008.
- [34] C. D. Snow, E. J. Sorin, Y. M. Rhee, and V. S. Pande, “How Well Can Simulation Predict Protein Folding Kinetics and Thermodynamics?,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 34, pp. 43–69, May 2005.
- [35] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, “Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics,” *J. Chem. Phys.*, vol. 126, p. 155101, Apr. 2007.
- [36] J. D. Chodera and F. No, “Markov state models of biomolecular conformational dynamics,” *Current Opinion in Structural Biology*, vol. 25, pp. 135–144, Apr. 2014.
- [37] G. R. Bowman, X. Huang, and V. S. Pande, “Using generalized ensemble simulations and Markov state models to identify conformational states,” *Methods*, vol. 49, pp. 197–201, Oct. 2009.
- [38] 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.
- [39] 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.
- [40] 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.

- [41] 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.
- [42] G. Bel and E. Barkai, “Weak Ergodicity Breaking in the Continuous-Time Random Walk,” *Phys. Rev. Lett.*, vol. 94, p. 240602, June 2005.
- [43] C. Truong, L. Oudre, and N. Vayatis, “Ruptures: Change Point Detection in Python,” *arXiv*, vol. arXiv:1801.00826, Jan. 2018.
- [44] A. Clauset, C. Shalizi, and M. Newman, “Power-Law Distributions in Empirical Data,” *SIAM Rev.*, vol. 51, pp. 661–703, Nov. 2009.
- [45] T. Neusius, I. M. Sokolov, and J. C. Smith, “Subdiffusion in Time-Averaged, Confined Random Walks,” *Phys. Rev. E*, vol. 80, p. 011109, July 2009.
- [46] R. G. Clegg, “A practical guide to measuring the Hurst parameter,” *arXiv:math/0610756*, Oct. 2006. arXiv: math/0610756.
- [47] H. Tikanmki and Y. Mishura, “Fractional L\’evy processes as a result of compact interval integral transformation,” *arXiv:1002.0780 [math]*, Feb. 2010. arXiv: 1002.0780.
- [48] J. P. N. Bishwal, “Maximum Quasi-likelihood Estimation in Fractional Levy Stochastic Volatility Model,” *JMF*, vol. 01, no. 03, pp. 58–62, 2011.
- [49] R. N. Mantegna and H. E. Stanley, “Stochastic Process with Ultraslow Convergence to a Gaussian: The Truncated L\’evy Flight,” *Phys. Rev. Lett.*, vol. 73, pp. 2946–2949, Nov. 1994.
- [50] S. Bacallado, J. D. Chodera, and V. Pande, “Bayesian comparison of Markov models of molecular dynamics with detailed balance constraint,” *J. Chem. Phys.*, vol. 131, p. 045106, July 2009.

- [51] S. Stoev and M. S. Taqqu, “Simulation methods for linear fractional stable motion and farima using the fast fourier transform,” *Fractals*, vol. 12, pp. 95–121, Mar. 2004.
- [52] V. S. Pande, K. Beauchamp, and G. R. Bowman, “Everything you wanted to know about Markov State Models but were afraid to ask,” *Methods*, vol. 52, pp. 99–105, Sept. 2010.
- [53] C. Wehmeyer, M. K. Scherer, T. Hempel, B. E. Husic, S. Olsson, and F. No, “Introduction to Markov state modeling with the PyEMMA software [Article v1.0],” *Living Journal of Computational Molecular Science*, vol. 1, no. 1, pp. 5965–, 2018.
- [54] K. A. Beauchamp, G. R. Bowman, T. J. Lane, L. Maibaum, I. S. Haque, and V. S. Pande, “MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale,” *J. Chem. Theory Comput.*, vol. 7, pp. 3412–3419, Oct. 2011.
- [55] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Prez-Hernndez, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, and F. No, “PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models,” *J. Chem. Theory Comput.*, vol. 11, pp. 5525–5542, Nov. 2015.
- [56] W. C. Swope, J. W. Pitera, and F. Suits, “Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory,” *J. Phys. Chem. B*, vol. 108, pp. 6571–6581, May 2004.
- [57] T. L. Hill, *Free Energy Transduction and Biochemical Cycle Kinetics*. Springer-Verlag, 1989.
- [58] E. L. Cussler, *Diffusion: Mass Transfer in Fluid Systems*. Cambridge University Press, 3 ed., 2009.

- [59] W. F. Guo, T.-S. Chung, and T. Matsuura, “Pervaporation study on the dehydration of aqueous butanol solutions: a comparison of flux vs. permeance, separation factor vs. selectivity,” *Journal of Membrane Science*, vol. 245, pp. 199–210, Dec. 2004.
- [60] O. Kedem and A. Katchalsky, “Permeability of Composite Membranes. Part 1. Electric Current, Volume Flow and Flow of Solute Through Membranes,” *Trans. Faraday Soc.*, vol. 59, no. 0, pp. 1918–1930, 1963.
- [61] H. Al-Zoubi, N. Hilal, N. A. Darwish, and A. W. Mohammad, “Rejection and modelling of sulphate and potassium salts by nanofiltration membranes: neural network and SpieglerKedem model,” *Desalination*, vol. 206, pp. 42–60, Feb. 2007.
- [62] D. Ernst, M. Hellmann, J. Khler, and M. Weiss, “Fractional Brownian motion in crowded fluids,” *Soft Matter*, vol. 8, no. 18, pp. 4886–4889, 2012.
- [63] D. Molina-Garcia, T. Sandev, H. Safdari, G. Pagnini, A. Chechkin, and R. Metzler, “Crossover from anomalous to normal diffusion: truncated power-law noise correlations and applications to dynamics in lipid bilayers,” *New J. Phys.*, vol. 20, p. 103027, Oct. 2018.

TOC Graphic

