Supporting Information: Capturing Subdiffusive Solute Dynamics and Predicting Selectivity in Nanoscale Pores with Time Series Modeling.

Benjamin J. Coscia Michael R. Shirts

April 30, 2020

Contents

S1 Setup and analysis scripts

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

Script Name	Section	Description
`/setup/parameterize.py`	2.1	Parameterize liquid crystal monomers and solutes with GAFF
`/setup/build.py`	2.1	Build simulation unit cell
`/setup/place_solutes_pores.py`	2.1	Place equispaced solutes in the pore centers of a unit cell
`/setup/equil.py`	2.1	Equilibrate unit cell and run production simulation
`/analysis/solute_partitioning.py`	2.1	Determine time evolution of partition of solutes between pores and tails
`/timeseries/msd.py`	2.2	Calculate the mean squared displacement of solutes
`/analysis/sfbm_parameters.py`	2.2	Get subordinated fractional Brownian motion parameters by fitting to a solute's dwell and hop length distributions and positional autocorrelation function.
`/timeseries/ctrwsim.py`	2.2	Generate realizations of a continuous time random walk with the user's choice of dwell and hop distributions
`/timeseries/forecast_ctrw.py`	2.2	Combines classes from `sfbm_parameters.py` and `ctrwsim.py` to parameterize and predict MSD in one shot.
`/analysis/ Markov_state_dependent_dynamics.py`	2.3	Identify frame-by-frame state of each solute, construct a transition matrix and simulate realizations of the MSDDM model.
`/timeseries/mfpt_pore.py`	2.4	Simulate mean first passage time distributions using the AD approach.

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

S2 Solute Equilibration

We collected all data used for model generation after the solutes were equilibrated. We assumed a solute to be equilibrated when the partition of solutes in and out of the pore region stopped changing. The pore region is defined as within 0.75 nm of the pore center. We plot the partition versus time in Figure S1 and indicated the chosen equilibration time points.

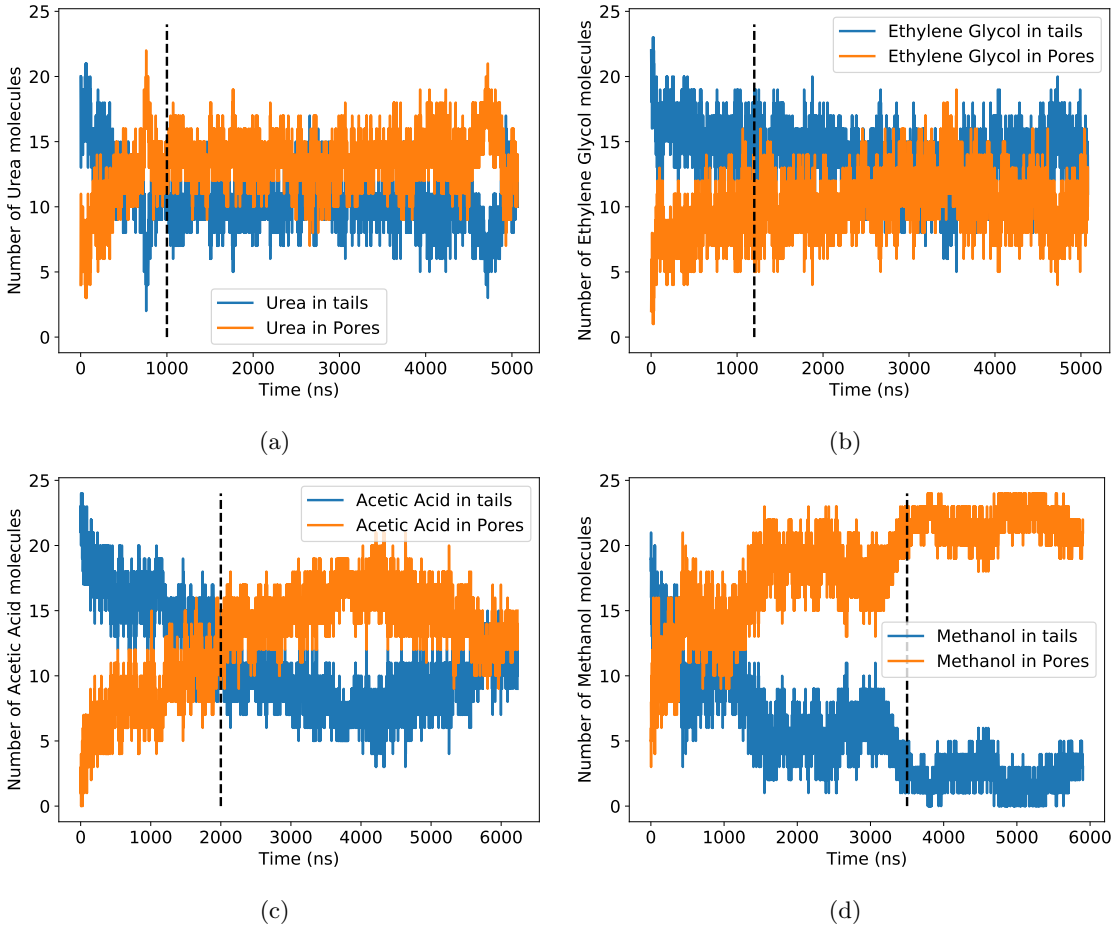


Figure S1: We considered a system to be equilibrated when the partition of solutes between the tails and pore plateaued. Our chosen equilibration point for each solute is indicated by the vertical black dashed line. (a) Urea equilibrates the fastest, after 1000 ns. (b) Ethylene glycol equilibrates after 1200 ns (c) The partition of acetic acid appears oscillate slowly. We considered it to be equilibrated after 2000 ns. (d) We considered methanol to be equilibrated after 3500 ns. Methanol nearly completely partitions into the tails.

S3 Mean Squared Displacement

In this study, we primarily use the MSD as a tool for characterizing the average dynamic behavior of solute trajectories. Rather than using them to calculate diffusion constants or to relate our simulations to experimental measurements, we compare MSDs calculated from MD simulations to those generated from our models in order to validate those models. Therefore, it is only important that we use a consistent definition for calculating the MSD between modeled trajectories and directly observed MD trajectories.

One can measure MSD in two ways. The ensemble averaged MSD measures displacements with respect to a particle's initial position:

$$\langle z^2(t)\rangle = \langle z(t) - z(0)\rangle^2 \tag{1}$$

Fits to the ensemble averaged MSD will always reproduce the form of Equation 1 of the main text. The time-averaged MSD measures all observed displacements over time lag τ:

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

where T is the length of the trajectory.

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 of the main text. [1] 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. [2]

We chose to use just the time-averaged MSD to compare MD trajectories with modeled trajectories, because, compared to the ensemble average, it is a more statistically robust measure of the average distance a solute travels over time. The ensemble MSD of only 24 solute trajectories would have much higher uncertainties.

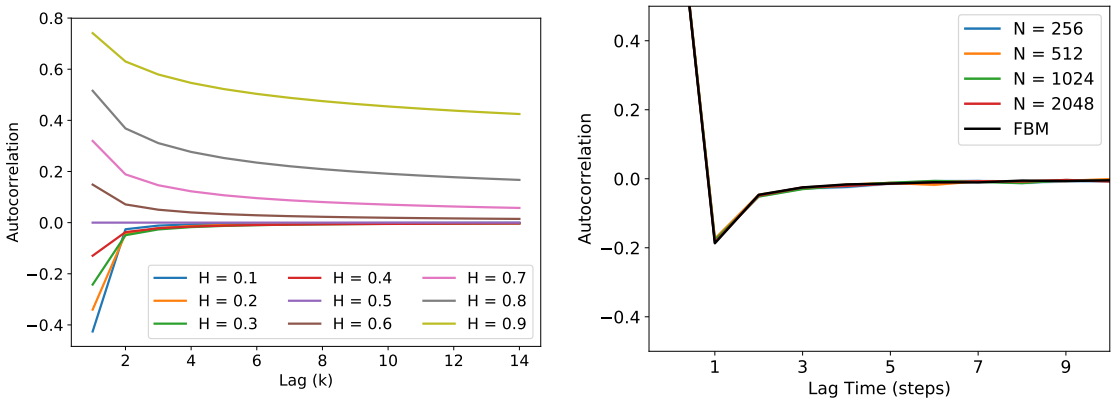
S4 Estimating the Hurst Parameter

We chose to estimate the Hurst parameter, H by a least squares fit to the analytical autocorrelation function for fractional Brownian motion (the variance-normalized version of Equation 6 in the main text):

$$\gamma(k) = \frac{1}{2}\left[|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}\right] \tag{3}$$

In Figure S2a, we plotted Equation 3 for different values of H. When $H > 0.5$, Equation 3 decays slowly to zero meaning one needs to study large time lags with high frequency in order to accurately estimate H from the data. Fortunately, all of our solutes show anti-correlated motion, so most of the information in Equation 3 is contained within the first few lags.

The autocovariance function of fractional Lévy motion is different from fractional Brownian motion (see Equations 6 and 8 of the main text), but their autocorrelation structures are the same. The autocovariance function of FLM is dependent on the expected value of squared draws from the underlying Lévy distribution, $E\left[L(1)^2\right]$. This is effectively the distribution's variance, which is undefined for most Lévy stable distributions due to their heavy tails. As a consequence, one should expect $E\left[L(1)^2\right]$ to grow as more samples are drawn from the distribution with the autocovariance function responding accordingly. However, we are only interested in the autocorrelation function. In order to predict the Hurst parameter from the autocorrelation function, we must show that it has a well-defined structure and is independent of the coefficient in Equation 8 of the main text. In Figure S2b, we plot the average autocorrelation function from an FLM process with an increasing number of observations per generated sequence. For all simulations we set H=0.35 and α=1.4. The variance-normalized autocovariance function, i.e. the autocorrelation function, does not change with increasing sequence length. Additionally, the autocorrelation function of FBM, with the same H, is the same. Therefore we are confident that we can use the same Hurst parameter as an input to both FBM and FLM simulations.



(a) (b)

Figure S2: (a) The analytical autocorrelation function of FBM decays to zero faster when H < 0.5 compared to when H > 0.5. (b) The autocorrelation function of an FLM process does not change with increasing sequence length (N). It shares the same autocorrelation function as fractional Brownian motion (FBM). Note that all lines plotted lie on top of each other. All sequences used to make this plot were generated using H=0.35 and, for FLM, α=1.4.

S5 Simulating Fractional Lévy Motion

S5.1 Truncated Lévy stable hop distributions

Determining where to truncate the hop distribution: A pure Lévy stable distribution has heavy tails which can lead to arbitrarily long hop lengths. Our distribution of hop lengths fits well to a Lévy distribution near the mean, but under-samples the tails. In Figure S4 we compare the empirically measured transition emission distribution of the MSDDM for urea to its maximum likelihood fit to a Lévy stable distribution. The ratio between the two distributions at each bin is nearly 1 close to the center, indicating a near-perfect fit, larger than 1 slightly further from the center, suggesting that we slightly over sample intermediate hop lengths, and below 1 far from the center, indicating undersampling of extremely long hop lengths. Based on the plot, we chose a cut-off of 1 nm in order to compensate for over sampled intermediate hop lengths. We chose the same cut-off for all solutes.

Generating FLM realizations from a truncated Lévy distribution: To generate realizations from an uncorrelated truncated Lévy process, one would randomly sample from the base distribution and replace values that are too large with new random samples from the base distribution, repeating the process until all samples are under the desired cut-off.

This procedure is complicated by the correlation structure of FLM. At a high level, Stoev and Taqqu use Riemann-sum approximations of the stochastic integrals defining FLM in order to generate realizations. [3] They do this efficiently with the help of Fast Fourier Transforms. In practice, this requires one to Fourier transform a zero-padded vector of random samples drawn from the appropriate Lévy stable distribution, multiply the vector in Fourier space by a kernel function and invert back to real space. The end result is a correlated vector of fractional Lévy noise.

We are unaware of a technique for simulating truncated FLM, therefore we devised our own based on the above discussion. If one is to truncate an FLM process, one can apply the simple procedure above for drawing uncorrelated values from the marginal Lévy stable distribution, *but*, after adding correlation, the maximum drawn value is typically lower than the limit set by the user. Additionally, the shape of the distribution itself changes. Therefore, we created a database meant to correct the input truncation parameter (the maximum desired draw). The database returns the value of the truncation parameter that will properly truncate the output marginal distribution based on H, α and σ (the width parameter). Figure S4 shows the result of applying our correction. Note that generating this database requires a significant amount of simulation and still likely doesn't perfectly correct the truncation parameter. The output leads to a somewhat fuzzy, rather

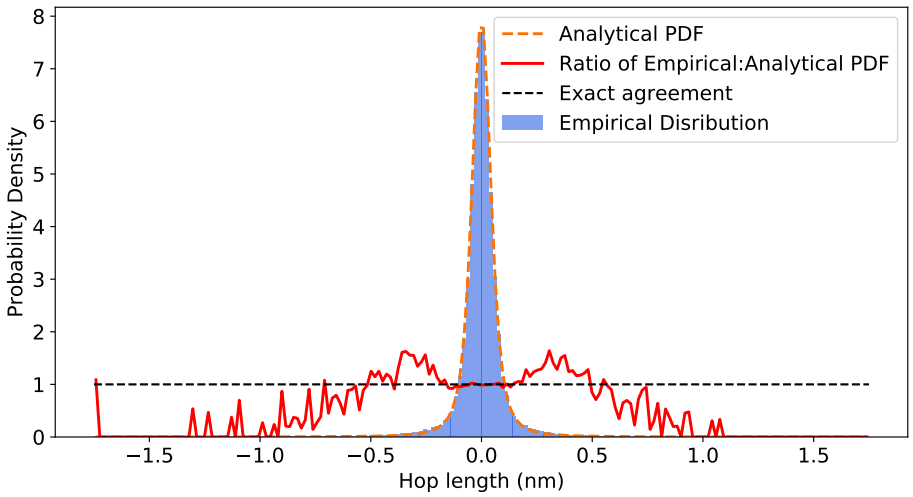


Figure S3: The ratio between the empirical and maximum likelihood theoretical distribution quantifies the quality of fit as function of hop length. The fit is near-perfect close to the mean. Intermediate hop lengths are over sampled, and the tails are undersampled. We used this type of plot to determine the appropriate place to truncate the Lévy stable distributions.

than abrupt, cut-off of the output distribution. This is likely beneficial since we observe a small proportion of hops longer the chosen truncation cut-off. When the cut-off value is close to the Lévy stable σ parameter, as it is in our anomalous diffusion models, we observed that the tails of the truncated distribution tend to be undersampled. In order to maintain the distribution's approximate shape up to the cut-off value we recommend ensuring that the cut-off value is at least 2 times σ. However, this may lead to a slight over-prediction of the MSD.

S5.2 Achieving the right correlation structure

We simulated FLM using the algorithm of Stoev and Taqqu [3]. There are no known exact methods for simulating FLM. As a consequence, passing a value of H and α to the algorithm does not necessarily result in the correct correlation structure, although the marginal Lévy stable distribution is correct. We applied a database-based empirical correction in order to use the algorithm to achieve the correct marginal distribution and correlation structure.

Stoev and Taqqu note that the transition between negatively and positively correlated draws occurs when $H = 1/\alpha$. When $\alpha = 2$, the marginal distribution is Gaussian and the transition occurs at $H = 0.5$ as expected from FBM. We corrected the input H so that the value of H measured based on the output sequence equaled the desired H. We first adjusted the value of H by adding $(1/\alpha - 0.5)$, effectively recentering the correlation sign transition for any value of $1 \leq \alpha \leq 2$. This correction alone does a good job for input H values near 0.5, but is insufficient if one desires a low value of H. The exact correction to H is not obvious so we created a database of output H values tabulated as a function of input H and α values. Figure S5 demonstrates the results of applying our correction. Without the correction, FLM realizations are more negatively correlated. This would result in under-predicted mean squared displacements when applying the model.

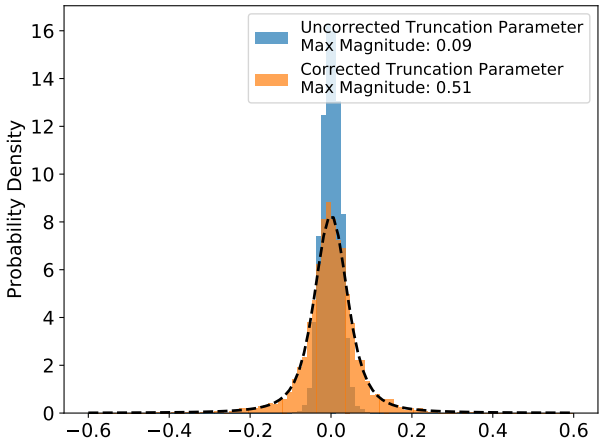


Figure S4: We can accurately truncate the marginal distribution of FLM innovations by applying a correction to the input truncation parameter. We generated FLM sequences and truncated the initial Lévy stable distribution (before Fourier transforming) at a value of 0.5. After correlation structure is added, the width of the distribution of fractional Lévy noise decreases significantly. We corrected the input truncation parameter with our database resulting in a distribution close to the theoretical distribution (black dashed line) with a maximum value close to 0.5.

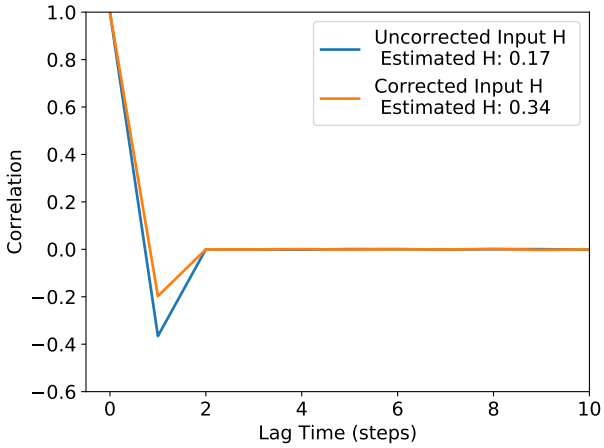
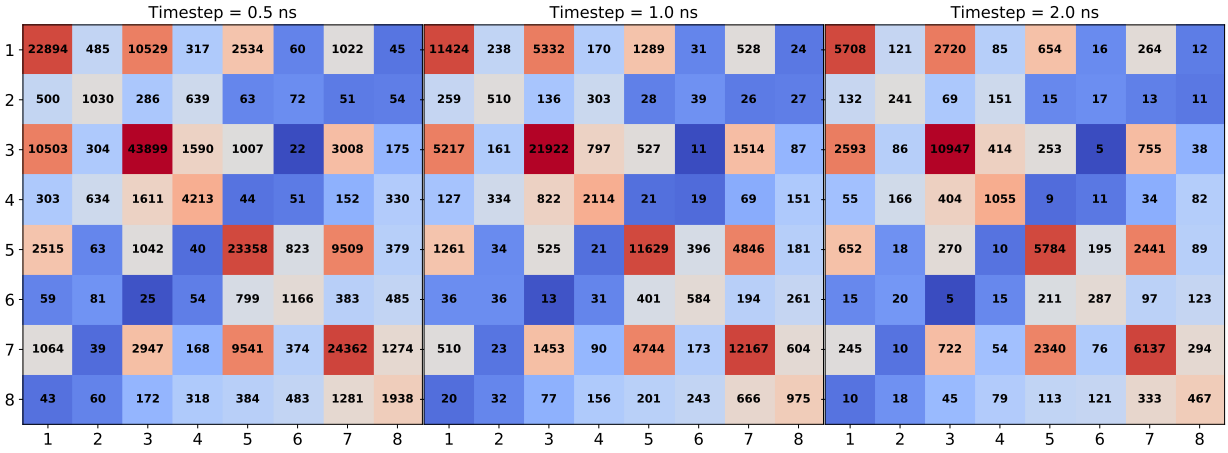
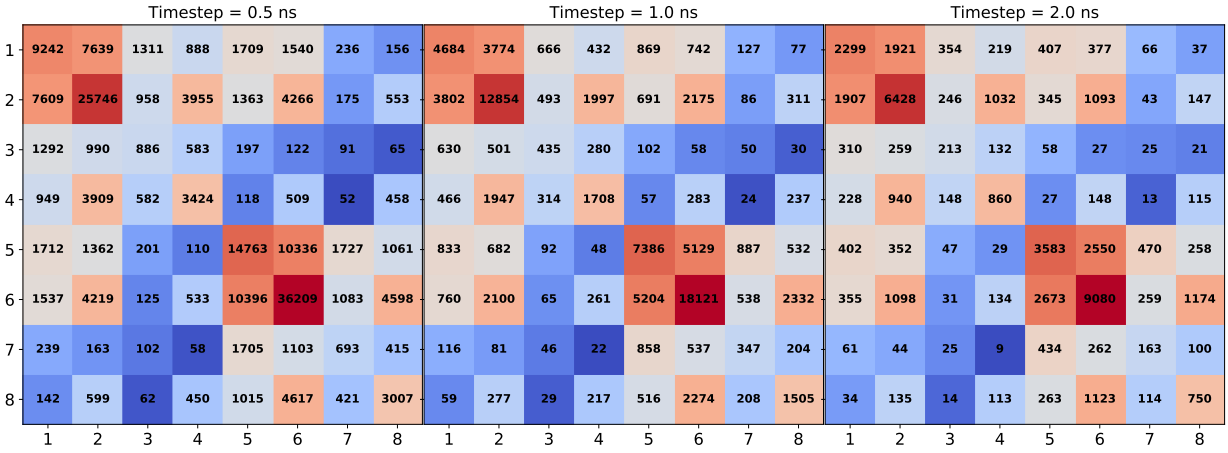


Figure S5: Correcting the Hurst parameter input to the algorithm of Stoev and Taqqu results in an FLM process with a more accurate correlation structure. We generated sequences with an input H of 0.35. We estimated H by fitting the autocorrelation function. Without the correction, H is underestimated, meaning realizations are more negatively correlated than they should be.



(a) Urea



(b) Ethylene Glycol

Figure S6: The number of transitions from state i to j and j to i are very close indicating that our process obeys detailed balance. Detailed balance is conserved for different sized time steps.

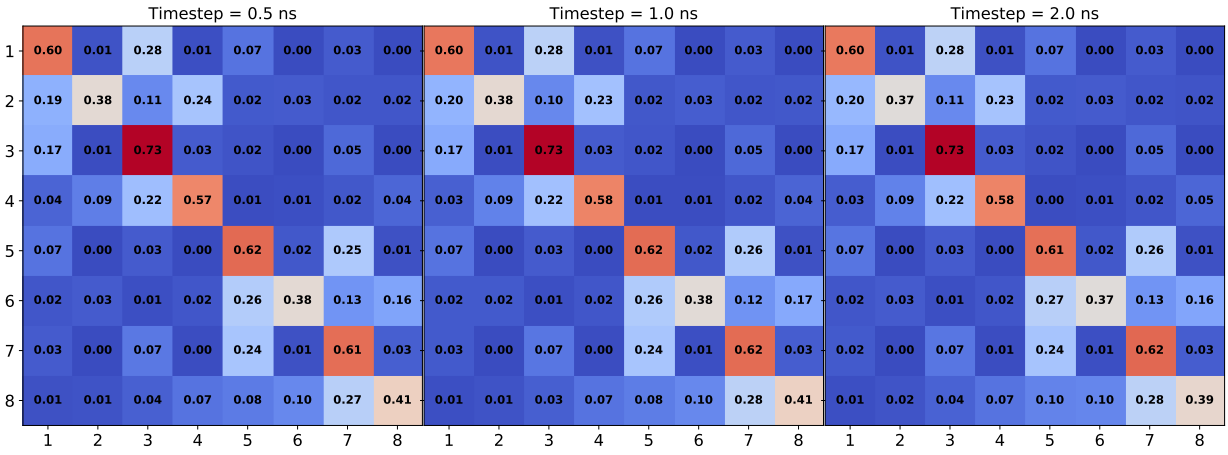
S6 Verifying Markovianity

We verified the Markovianity of our transition matrix, T, in two ways. First we ensured that the process satisfied detailed balance:

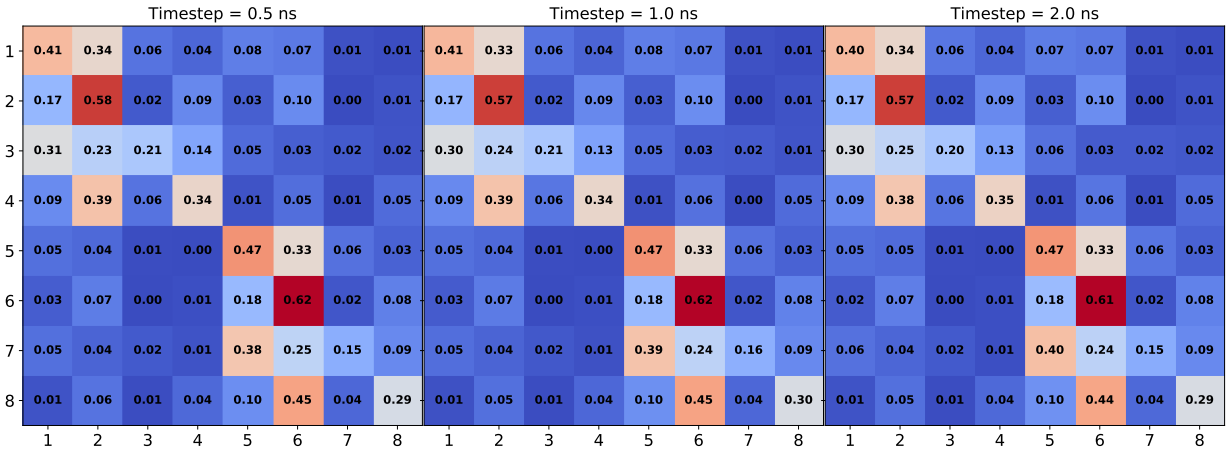
$$T_{i,j}P_i(t=\infty) = T_{j,i}P_j(t=\infty) \tag{4}$$

where P is the equilibrium distribution of states. This implies that the number of transitions from state i to j and from state j to i should be equal. Graphical representations of the count matrices show that this is true in Figure S6.

Second, we ensured that the transition matrix did not change on coarser time scales. In Figures S6 and S7, we show that increasing the length of time between samples does not change the properties of the count or probability transition matrices.



(a) Urea



(b) Ethylene Glycol

Figure S7: As the timestep between observations increases, the probability transition matrix does not change significantly.

S7 Derivation of Passage Time Distributions

To derive an analytical equation for the mean first passage time (Equation 10 of the main text), first consider an initial pulse spreading out over time with a fixed mean. We can solve for the time-dependent probability density of particle positions, p, by solving the one dimensional diffusion equation:

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial z^2} \tag{5}$$

The appropriate initial and boundary conditions are:

$$BC1 : t > 0, z = \infty, p = 0$$

$$BC2 : t > 0, z = 0, \frac{\partial p}{\partial z} = 0$$

$$IC : t = 0, c = \delta(z)$$

It has been shown elsewhere that the solution to this equation is: [4]

$$p(z,t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(\frac{-z^2}{4Dt}\right) \tag{6}$$

We can make the substitution $z = z - vt$, where v represents a constant average velocity, in order to linearly shift the mean as a function of time:

$$p(z,t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(\frac{-(z-vt)^2}{4Dt}\right) \tag{7}$$

One can track the fraction of particles, F, that have crossed the pore boundary by integrating:

$$F(t) = \int_L^\infty p\ dz = \text{erfc}\left(\frac{L-vt}{2\sqrt{Dt}}\right) \tag{8}$$

where L is the pore length. This represents the cumulative first passage time distribution so we take its derivative in order to arrive at the first passage time distribution:

$$P(t) = -\frac{1}{\sqrt{\pi}} e^{-(L-vt)^2/(4Dt)}\left(-\frac{D(L-vt)}{4(Dt)^{3/2}} - \frac{v}{2\sqrt{Dt}}\right) \tag{9}$$

where the only free parameters for fitting are v and D. We calculated the expected value of Equation 9 in order to get the MFPT. Specifically, we used the python package `scipy.integrate.quad` to numerically integrate:

$$E[t] = \int_0^\infty tP(t)dt \tag{10}$$

S8 Solute hopping and trapping behavior

Analogous to Figure 3 of the main text, Figure S8 demonstrates that all solutes exhibit the same kind of anti-correlated hopping and trapping behavior.

S9 AD model MSD Predictions with Pure Power Law Dwell Times

When we use a pure power law distribution to parameterize the dwell time distributions of the one and two mode AD models, the MD MSDs are severely under-predicted because we are incorporating dwell times on the order of the simulation length into simulated trajectories (see Figure S9). The parameters of the pure power law distribution are included in Figure S10.

S10 Stationarity of Solute Trajectories

We observe that in some cases, solute trajectories extracted from our MD simulations display non-stationary behavior. 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). With this definition, we observe evidence of non-stationary solute behavior after the perceived equilibration point, on the μs timescale.

We trained the model parameters on the first half of the equilibrated MD trajectory data and then compared 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. In Figure S11, we show that urea and acetic acid show acceptable stationary behavior while methanol and ethylene glycol do not.

We validated both the one and two mode AD 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 S12. 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 timescales, the MSD of urea is under-predicted for both the one and two 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.

This brief analysis suggests that we may be operating on the border of the minimum amount of data required to accurately parameterize AD approach models. Working with only half of the data we collected ($\sim$ 2 μs post-equilibration) may not always be sufficient for extracting reliable parameter estimates. Therefore, in the main text, we employ parameters fit to the entire equilibrated portion of the solute trajectories. Even doubling the data might not be good enough for molecules with statistical non-stationarity, meaning the predictive and interpretive power of the time series modeling applied to these trajectories will be lower.

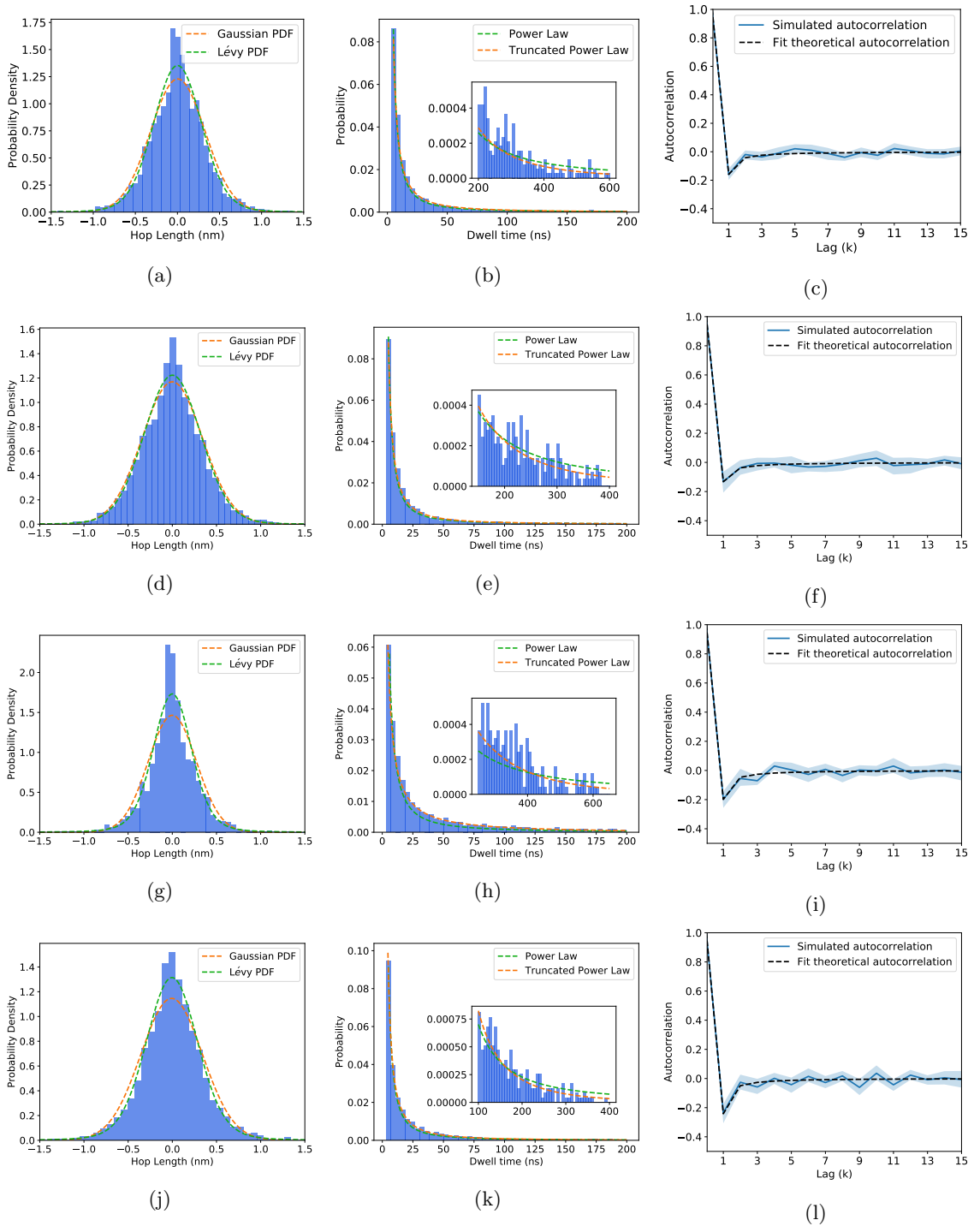
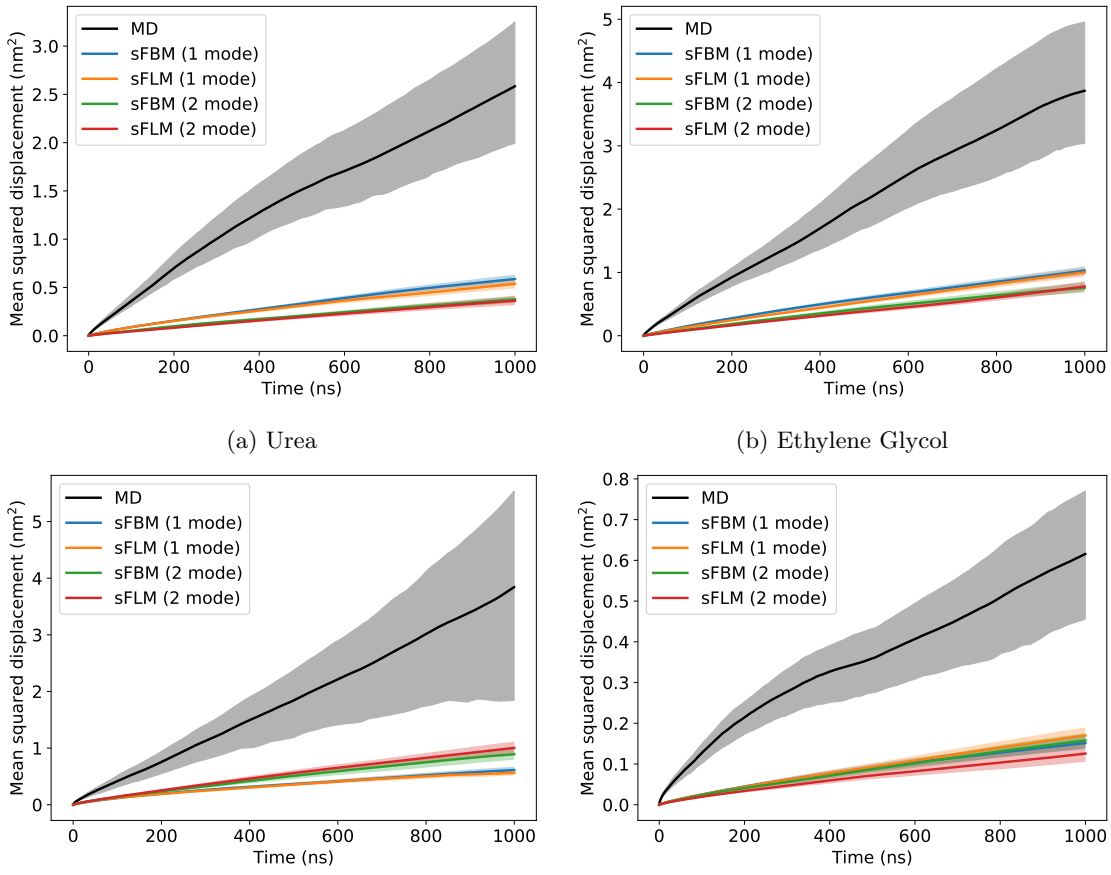


Figure S8: Hop length distributions, dwell time distributions and hop autocorrelation functions respectively for urea (a-c), ethylene glycol (d-f), acetic acid (g-i), and methanol (j-l). Using urea as an example, (a) the distribution of hop lengths can be fit by a Gaussian in addition to a more general Lévy stable distribution, though the Lévy stable distribution does a better job of capturing the heavier tails and increased density near 0. We explore models fit to both distributions, as Gaussian hops are more convenient to model. (b) The distribution of dwell times is fit well by a power law but it over-estimates the probability density at long dwell times. A power law truncated with an exponential cut-off better describes the probability of long dwell times in our simulations. (c) The hops are negatively correlated to their previous hop. In combination, (a) – (c) support modeling solutes as either subordinated fractional Brownian or Lévy motion.



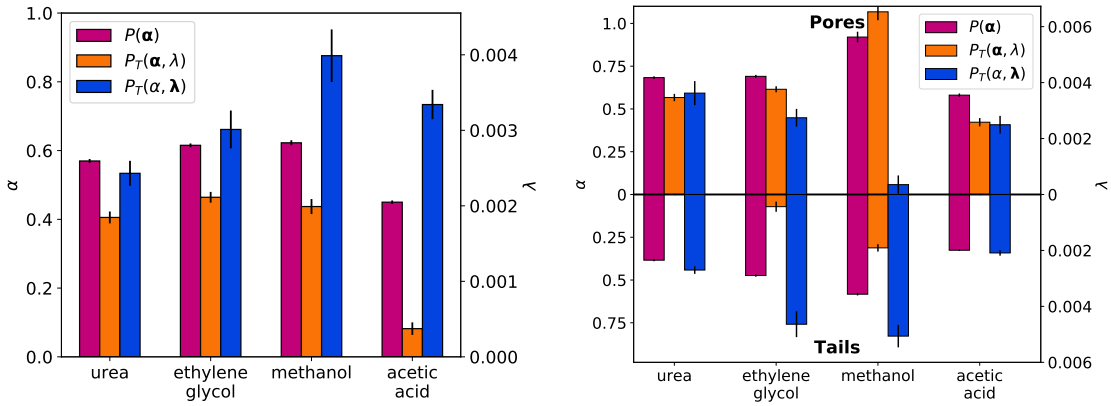
(a) Urea

(b) Ethylene Glycol

(c) Methanol

(d) Acetic Acid

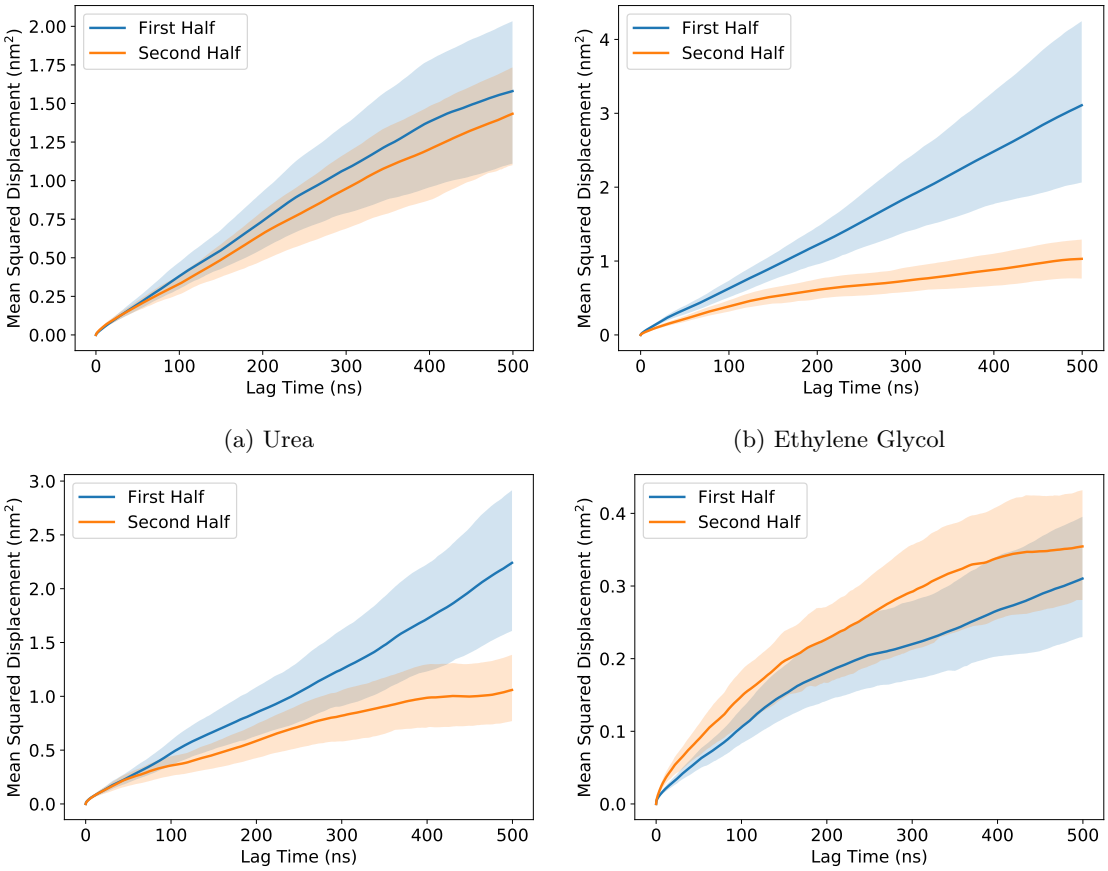
Figure S9: When we do not apply an exponential cut-off to the power law distribution of dwell times, MSDs are consistently under-predicted by the AD model.



(a)

(b)

Figure S10: We can parameterize the dwell time distribution in two ways: as a pure power law (P(α)) and as a power law with an exponential cut-off (P(α, λ)). Pure power laws have an infinite variance which allows extremely long dwell times to be sampled (see Figure S9.



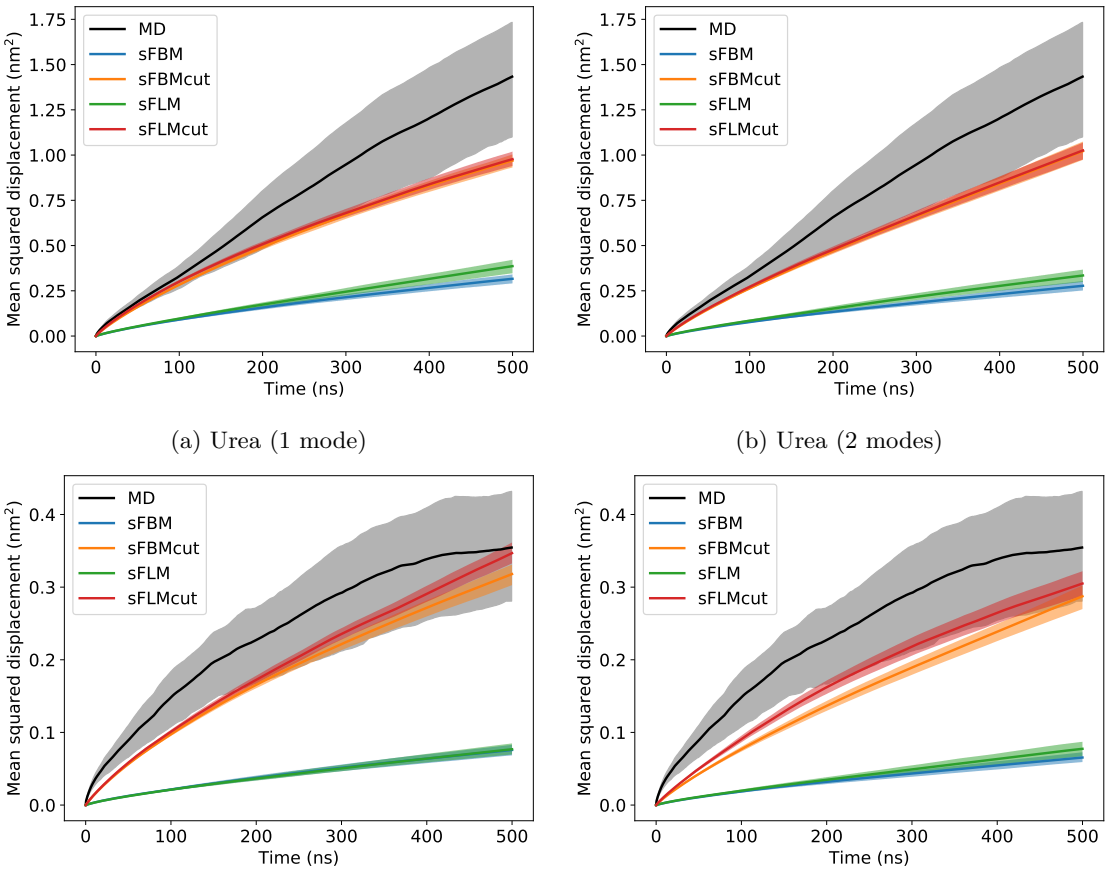
(a) Urea

(b) Ethylene Glycol

(c) Methanol

(d) Acetic Acid

Figure S11: The ensemble of solute trajectories may be stationary if the MSD calculated from different portions of the trajectory are the same. Here we plot the MSD calculated up to a 500 ns time lag of the first and second halves of the equilibrated solute trajectories. Urea and acetic acid have similar MSDs, providing evidence of stationarity, while the MSDs of ethylene glycol and methanol are different suggesting that they are not.



(a) Urea (1 mode)

(b) Urea (2 modes)

(c) Acetic acid (1 mode)

(d) Acetic acid (2 modes)

Figure S12: 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. Over- and under-estimated curvature of the of urea and acetic acid's MSD curves respectively causes the magnitude of urea's predicted MSDs to be under-predicted at long timescales and those of acetic acid to be under-predicted at short timescales. The models which use pure power laws systematically under-predict the MD MSD curves.

S11 Tables of Anomalous Diffusion Parameters

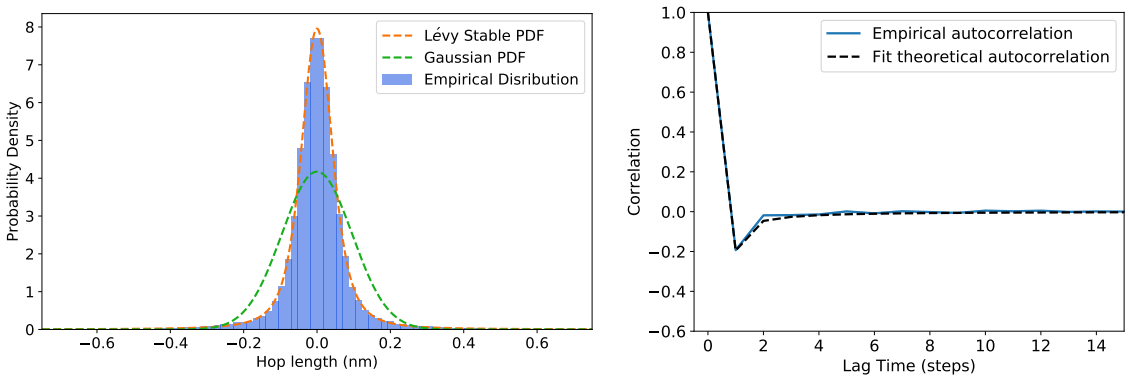
The tables in this section are tabular representations of the parameters depicted in Figures 6 and 7 of the main text and used to generate AD approach realizations whose MSDs are shown in Figure 5 of the main text.

1 Mode Model	Parameters	Urea	Ethylene Glycol	Methanol	Acetic Acid
Dwell Distributions	$P(\alpha_d)$	0.57	0.62	0.62	0.45
	$P_T(\alpha_d, \lambda)$	0.40, 0.0024	0.47, 0.0030	0.44, 0.0040	0.08, 0.0033
Hop Distributions	$\mathcal{N}(\sigma)$	0.33	0.34	0.35	0.27
	$L(\sigma, \alpha_h)$	0.21, 1.84	0.23, 1.92	0.22, 1.80	0.16, 1.72
Correlation	$\gamma(H)$	0.37	0.40	0.30	0.34

Table S2: Parameters of the one mode AD approach models. See the main text for further details.

2 Mode Model						
	Parameters	Mode	Urea	Ethylene Glycol	Methanol	Acetic Acid
Dwell Distributions	$P(\alpha_d)$	1	0.69	0.69	0.90	0.58
		2	0.38	0.48	0.58	0.33
	$P_T(\alpha_d, \lambda)$	1	0.56, 0.0037	0.62, 0.0026	1.04, 0.0006	0.41, 0.0026
		2	0.00, 0.0027	0.06, 0.0049	0.30, 0.0054	0.00, 0.0021
Hop Distributions	$\mathcal{N}(\sigma)$	1	0.35	0.38	0.45	0.32
		2	0.24	0.23	0.32	0.17
	$L(\sigma, \alpha_h)$	1	0.24, 1.91	0.26, 1.99	0.31, 1.97	0.21, 1.91
		2	0.12, 1.50	0.15, 1.90	0.20, 1.85	0.09, 1.50
Correlation	$\gamma(H)$	–	0.37	0.40	0.30	0.34

Table S3: Parameters of the 2 mode AD approach models. See the main text for further details.



(a) (b)

Figure S13: (a) The emission distributions of hop lengths 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.

S12 MSDDM parameters

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 AD model. (see Figure S13a) We therefore did not consider the Gaussian case for the MSDDM. The correlation structure between hops is consistent with that of FLM (Figure S13b).

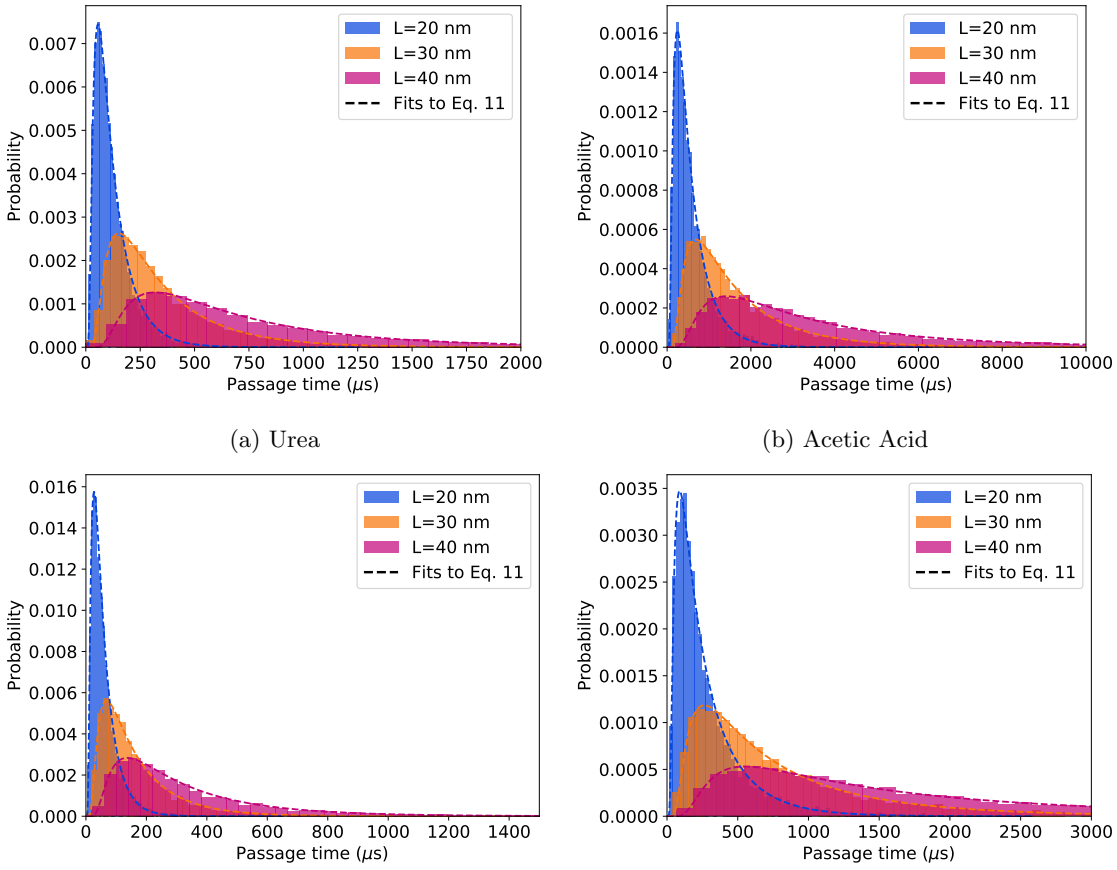
The following table is a tabular representation of the parameters depicted in Figure 10 of the main text.

	Urea			Ethylene Glycol			Methanol			Acetic Acid		
State	H	α_h	σ	H	α_h	σ	H	α_h	σ	H	α_h	σ
1	0.10	1.79	0.034	0.09	1.68	0.045	0.11	1.56	0.052	0.10	1.78	0.035
2	0.06	1.80	0.033	0.09	1.75	0.037	0.07	1.63	0.043	0.08	1.88	0.032
3	0.11	1.88	0.030	0.11	1.86	0.030	0.02	1.80	0.036	0.04	2.00	0.030
4	0.10	1.95	0.027	0.04	1.91	0.028	0.02	1.75	0.036	0.04	2.00	0.027
5	0.19	1.34	0.048	0.15	1.40	0.062	0.10	1.28	0.074	0.13	1.47	0.048
6	0.15	1.45	0.040	0.11	1.52	0.040	0.03	1.50	0.042	0.09	1.70	0.038
7	0.15	1.61	0.032	0.05	1.60	0.040	0.28	1.20	0.043	0.08	1.77	0.031
8	0.11	1.71	0.028	0.05	1.74	0.030	0.04	1.83	0.037	0.01	2.00	0.030
T	0.34	1.42	0.036	0.37	1.44	0.045	0.35	1.45	0.057	0.34	1.54	0.040

Table S4: We calculated values of H, α_h and σ from MD simulation trajectories and used them to generate realizations of our MSDDM model. The states are defined in Table 2 of the main text except state T which describes the transition emissions.

Analytical fits to MFPT distributions

In Figure S14 we demonstrate the high quality of our analytical fits of Equation 9 to the distribution of solute first passage times derived from both the AD and MSDDM models. In Figures S15– S17, we show that one can reliably fit Equation 9 to the passage time distributions with as few as 100 independent trajectory realizations at each pore length. For higher precision, we recommend using at least 1000 trajectories.



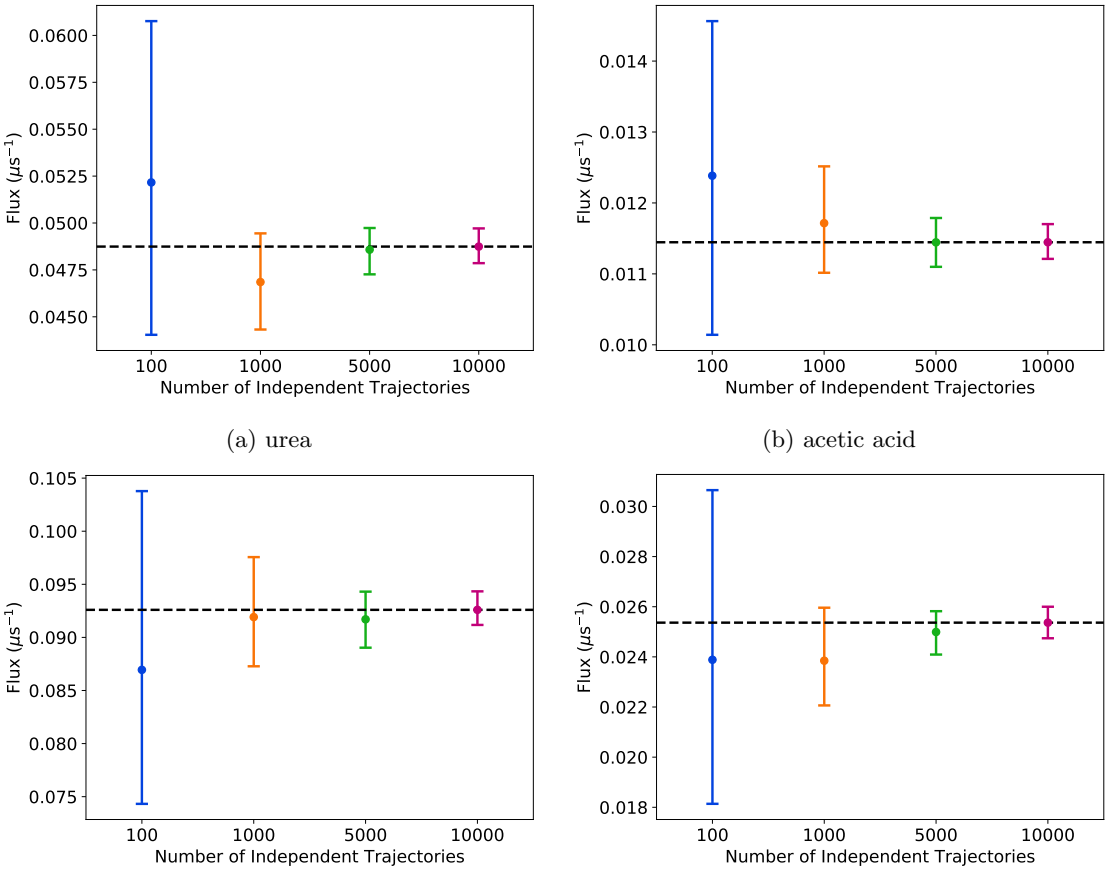
(a) Urea

(b) Acetic Acid

(c) Ethylene Glycol

(d) Methanol

Figure S14: We fit Equation 10 of the main text to the first passage time distributions generated by 10,000 realizations of the anomalous diffusion model.



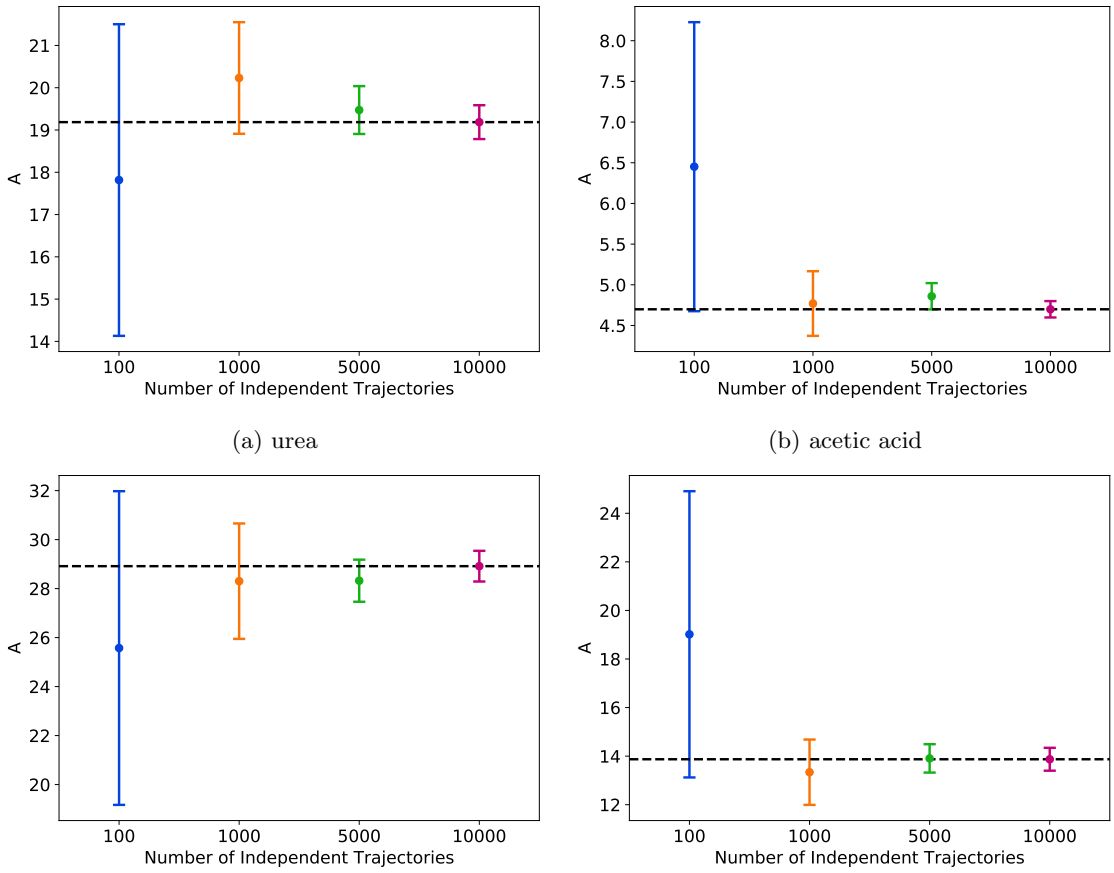
(a) urea

(b) acetic acid

(c) ethylene glycol

(d) methanol

Figure S15: Even using a small number of independent trajectories, one can reliably estimate solute flux across a pore 10 nm long. The uncertainty in the flux values decreases as we add more independent trajectories.



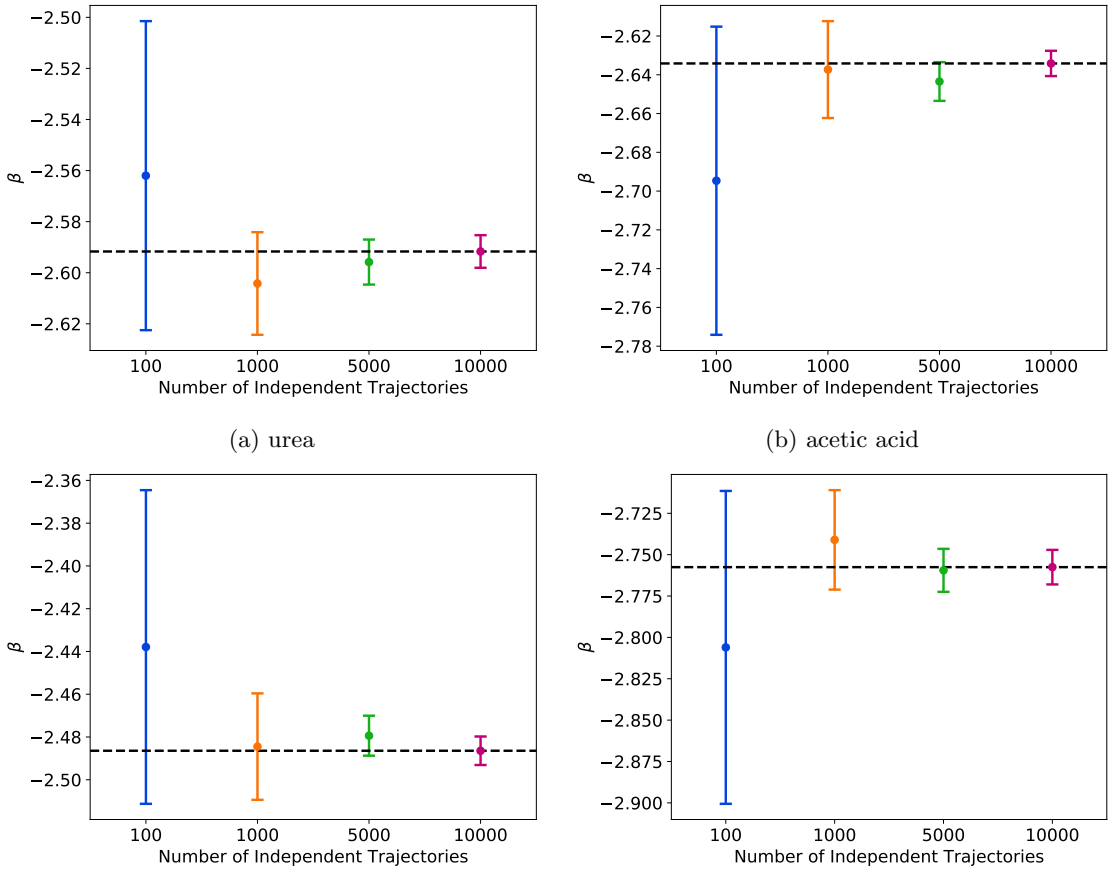
(a) urea

(b) acetic acid

(c) ethylene glycol

(d) methanol

Figure S16: The flux scaling parameter (A) can be reliably estimated using as few as 100 independent realizations of the sFBMcut AD model. To estimate A, we fit Equation 15 of the main text to a series of flux measurements made with 10, 15, 20, 25, 30, 35, 40, 45 and 50 nm pores (see Figure 13b of the main text).



(a) urea

(b) acetic acid

(c) ethylene glycol

(d) methanol

Figure S17: Similar to A, the parameter which describes the scaling of solute flux with pore length, β, can be reliably estimated using as few as 100 independent realizations of the sFBMcut AD model. We estimated β and A simultaneously as described in Figure S16.

References

[1] Y. Meroz and I. M. Sokolov, "A Toolbox for Determining Subdiffusive Mechanisms," *Phys. Rep.*, vol. 573, pp. 1–29, Apr. 2015.

[2] G. Bel and E. Barkai, "Weak Ergodicity Breaking in the Continuous-Time Random Walk," *Phys. Rev. Lett.*, vol. 94, p. 240602, June 2005.

[3] 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.

[4] E. L. Cussler, *Diffusion: Mass Transfer in Fluid Systems*. Cambridge University Press, 3 ed., 2009.