Supplemental Information

A new metric for relating macroscopic chromatograms to microscopic surface dynamics: the Distribution Function Ratio (DFR)

Logan D.C. Bishop†, Anastasiia Misiura†, Christy F. Landes†‡§ ∥ *

†Department of Chemistry, ‡Department of Electrical and Computer Engineering, §Department of Chemical and Biomolecular Engineering, and ∥Smalley-Curl Institute, Rice University, Houston, Texas 77251, USA

* Corresponding Author: cflandes@rice.edu

Contents

SI Proofs

Definition of $\theta(t)$

The DFR performs a comparison across a span of time normalized to the interval (0, 1], using the function $\theta(t;t_1, t_2)$. The limiting time point t_1 and t_2 are selected such that the signal at that time point is above a percentage height of the chromatogram peak max ($c(t) > x\% * max c(t)$). When looking at a single peak, this should result in two solutions at the beginning and end of the peak. Because these points fluctuate in time, each chromatogram has independent $\theta(t;t_1,t_2)$ of the form[1]:

$$\theta(t;t_1,t_2) = \frac{t - t_1}{t_2 - t_1} \quad \textbf{(S1)}$$

mapping to the same range of (0,1] enabling the DFR to compare any curves that have a calculated $\theta(t)$ normalization function.

DFR Normalizations

Given a chromatogram, the normalized area at any point defined in the range of $0 < \theta \leq 1$ provided by the $\theta(t;t_1, t_2)$, the percent analyte present at any one point in the original domain of time (t) is:

$$f(\theta) = \frac{f((t_2 - t_1)\theta + t_1)}{\int_{t_1}^{t_2} f(t)dt} \quad \textbf{(S2)}$$

where the value of $0 < f(\theta) \leq 1$.

SI Tables

Ratio of Expected Desorption Times $\left(\frac{\langle\tau_2\rangle}{\langle\tau_1\rangle}\right)$	Relative Prevalence (p_2)				
	0.001	**0.003**	**0.005**	**0.007**	**0.01**
50	0.0927	0.0464	0.0235	0.0222	0.0206
55	0.0810	0.0343	0.0165	0.0108	0.0136
60	0.0740	0.0300	0.0087	0.0037	0.0103
65	0.0702	0.0210	0.0013	0.0002	0.0095
70	0.0618	0.0141	-0.0092	-0.0098	0.0044
75	0.0572	0.0100	-0.0109	-0.0159	0.0009

Table S1. Estimating experimental kinetic parameters from simulated data for lysozyme flowed over hydrophobic membranes. The table above tracks the difference in chromatogram shape via the difference in the value of θ^* between the experiment and simulation ($\theta^*_{exp} - \theta^*_{sim}$). The cells closest to zero represent the kinetic values for a two-retention mode model that closely matches the ensemble measurement.

Ratio of Expected Desorption Times $\left(\frac{\langle\tau_2\rangle}{\langle\tau_1\rangle}\right)$	Relative Prevalence (p_2)				
	0.001	**0.003**	**0.005**	**0.007**	**0.01**
50	0.0732	0.0269	0.0040	0.0027	0.0011
55	0.0614	0.0148	-0.0030	0.0087	-0.0059
60	0.0545	0.0105	-0.0108	0.0158	-0.0092
65	0.0507	0.0015	-0.0182	0.0193	-0.0100
70	0.0423	-0.0054	-0.0287	0.0293	-0.0151
75	0.0377	-0.0095	-0.0304	0.0354	-0.0186

Table S2. Estimating experimental kinetic parameters from simulated data for lysozyme flowed over hydrophilic membranes. The table above tracks the difference in chromatogram shape via the difference in the value of θ^* between the experiment and simulation ($\theta^*_{exp} - \theta^*_{sim}$). The cells closest to zero represent the kinetic values for a two-retention mode model that closely matches the ensemble measurement.

SI Figures

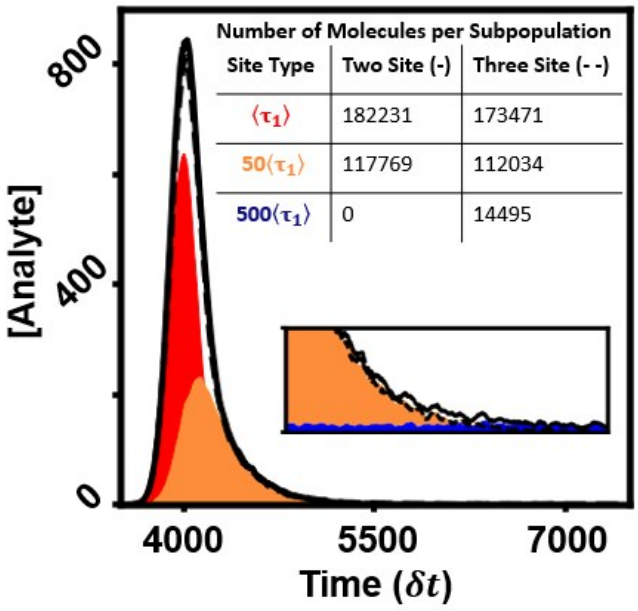

Figure S1. Breaking the macroscopic curve into subpopulations. A simulated chromatogram illustrating that the total profile is the sum of three different underlying subpopulations differentiated by their interactions with two rare retention modes. Here, the two rare retention modes have the following statistics: $\langle\tau_2\rangle = 50\langle\tau_1\rangle, p_2 = 0.001$ and $\langle\tau_2\rangle = 500\langle\tau_1\rangle, p_2 = 0.00001$. The inset is a table of the number of molecules per subpopulation. Direct visualization from the Monte Carlo results avoids the need to calculate the distributions via convolution, providing the subpopulation shape at with less computational power, and enabling extraction of information from the profile shape.

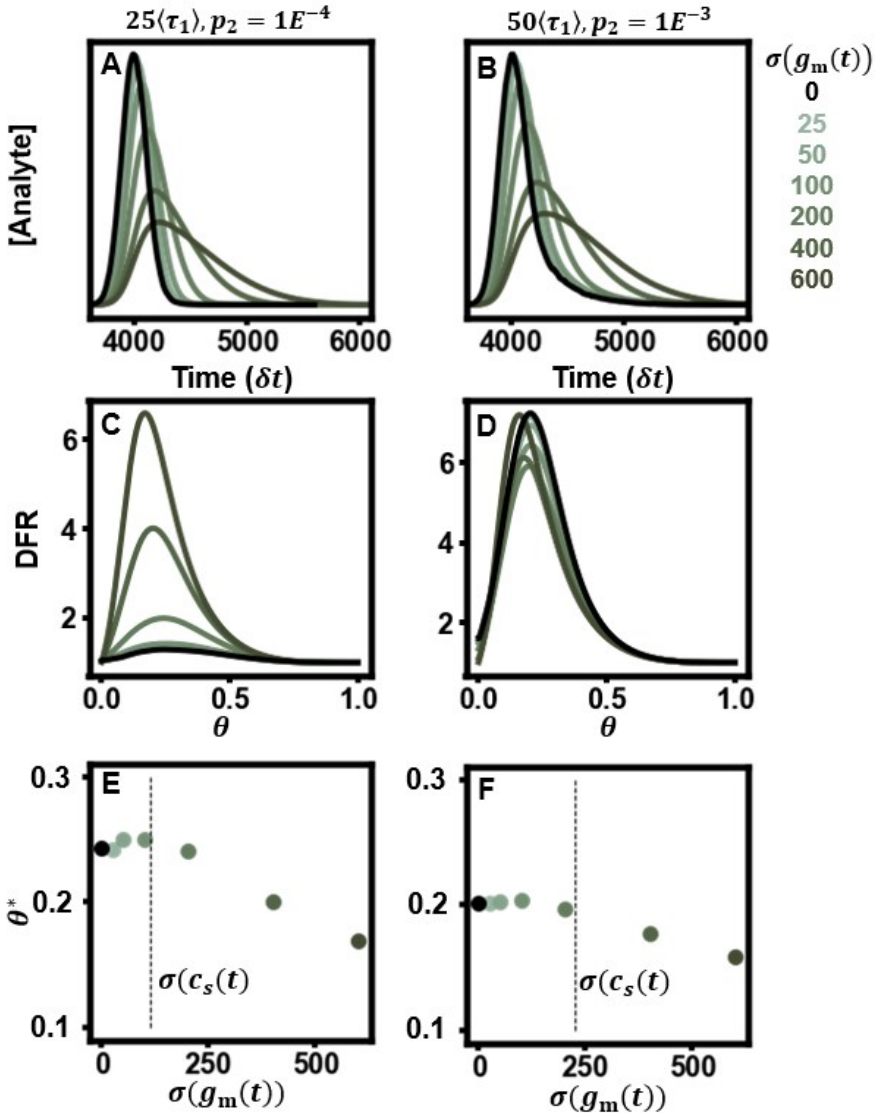

Figure S2. Effects of symmetric mobile phase convolution on the shape of the DFR and value of θ^*. Convolution of the simulated distribution of t_s with Gaussian distributions of t_m with different standard deviations ($\sigma(g_m(t))$ over two heterogeneous surface scenarios (A) $25\langle\tau_1\rangle, p_2 = 1E^{-4}$ and (B) $50\langle\tau_1\rangle, p_2 = 1E^{-3}$. (C,D) DFR graphs corresponding to A, B. (E, F) Values of θ^* versus the standard deviation of the t_m distribution with the standard deviation of t_s ($\sigma(c_s(t))$ annotated with a dotted line. Here, we note that while the distribution of t_m will change the shape of the DFR, the value of θ^* is stable if the standard deviation t_m is less than that of t_s ($\sigma(g_m(t)) < \sigma(c_s(t))$).

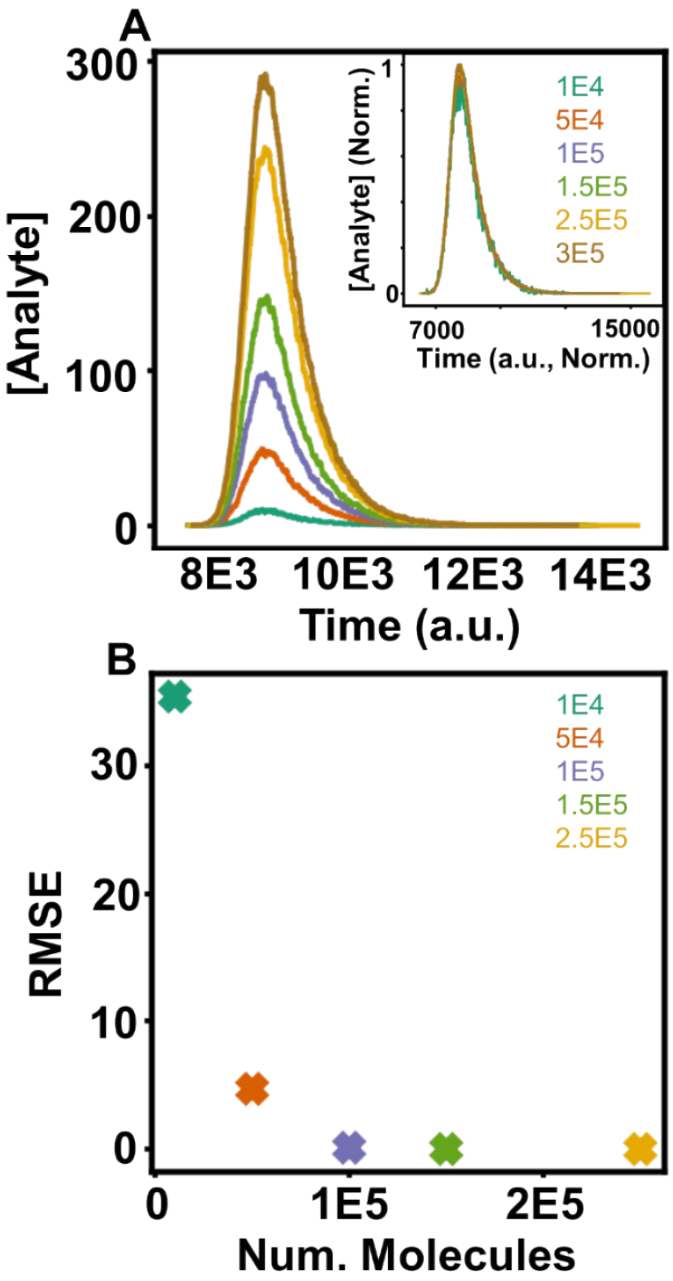

Figure S3. **Comparing the deviations between simulated chromatograms of different particle counts.** (A) Chromatograms with simulated molecule counts ranging from 10,000 to 300,000. (Inset) Normalized view of these same chromatograms. (B) RMSE comparison of each simulation to the highest molecule count, 300,000. RMSE stabilizes ~100,000 particles, establishing lower limit for sampling.

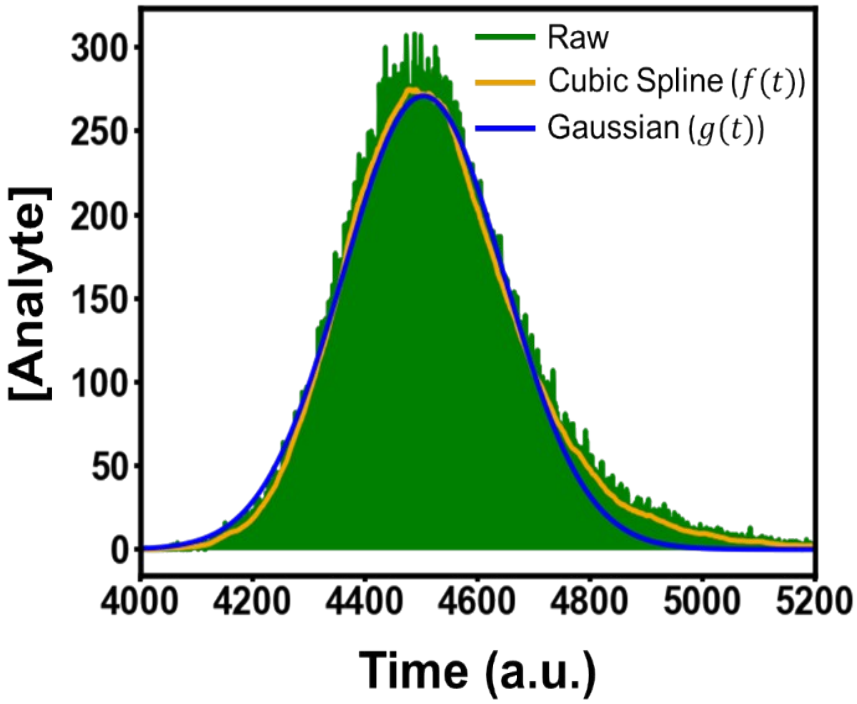

Figure S4. Comparison of the raw chromatogram to the comparative gaussian and the cubic spline of the raw data.

1

2

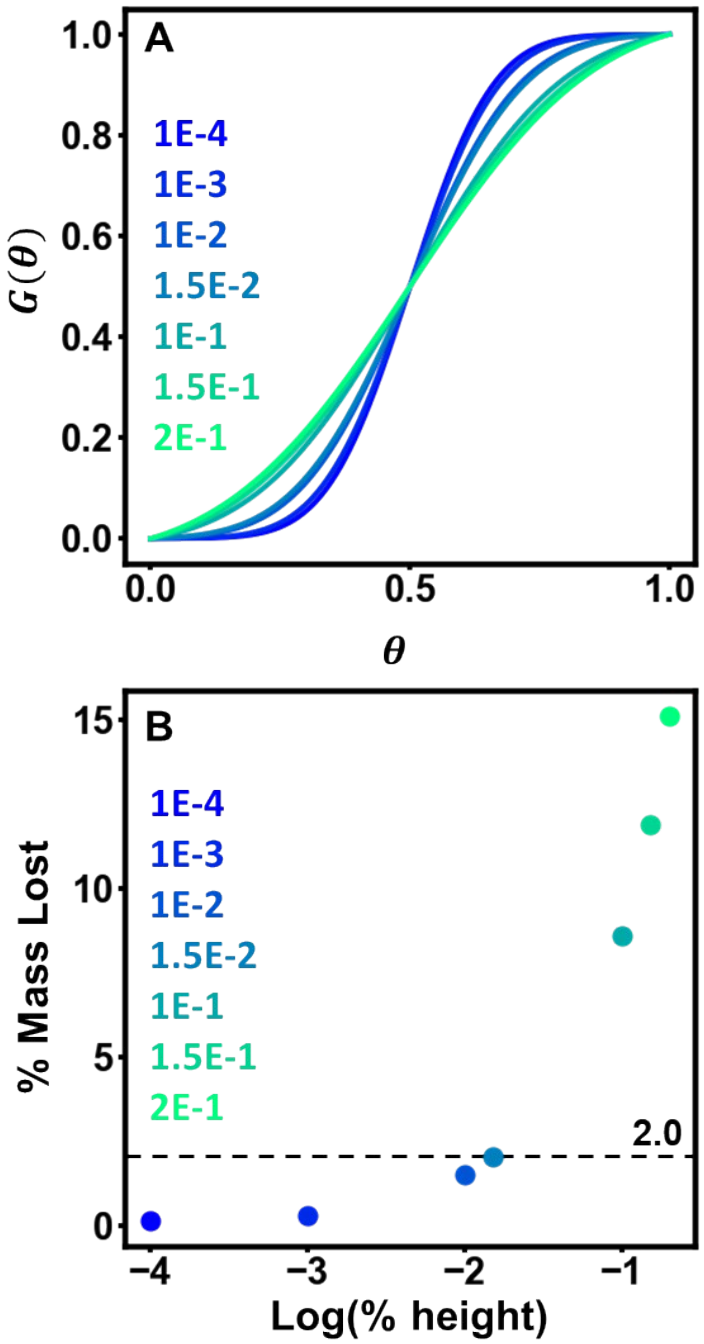

Figure S5. **Analyzing the shifts in the gaussian CDF as different height cutoffs are selected.** (A) The change in shape of the gaussian CDF dependent on height. Higher cutoffs flatten the curve migrating to a linear trend. (B) Analysis of the percent of excluded mass dependent on the selection of the height metric. The limit selected in the text excludes 2% peak mass on either side of the Gaussian.

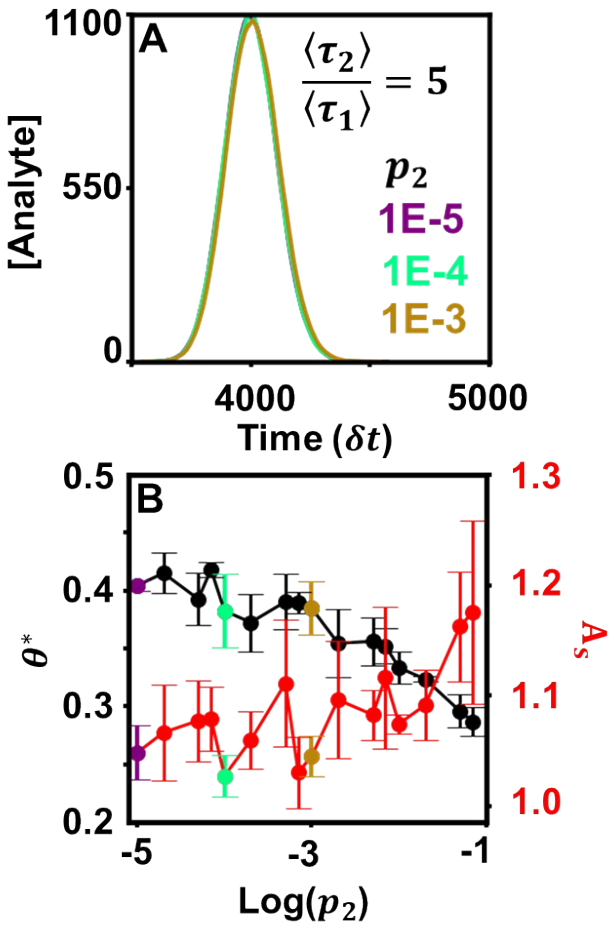

Figure S6. Tracking θ^* as prevalence changes in adsorption conditions that lead to visually symmetrical peaks ($\tau_2 = 5\langle\tau_1\rangle$). (A) Chromatograms generated from a two-mode system where the ratio of expected desorption time is 5. (B) Measurements of θ^* and A_s same ratio of desorption times.

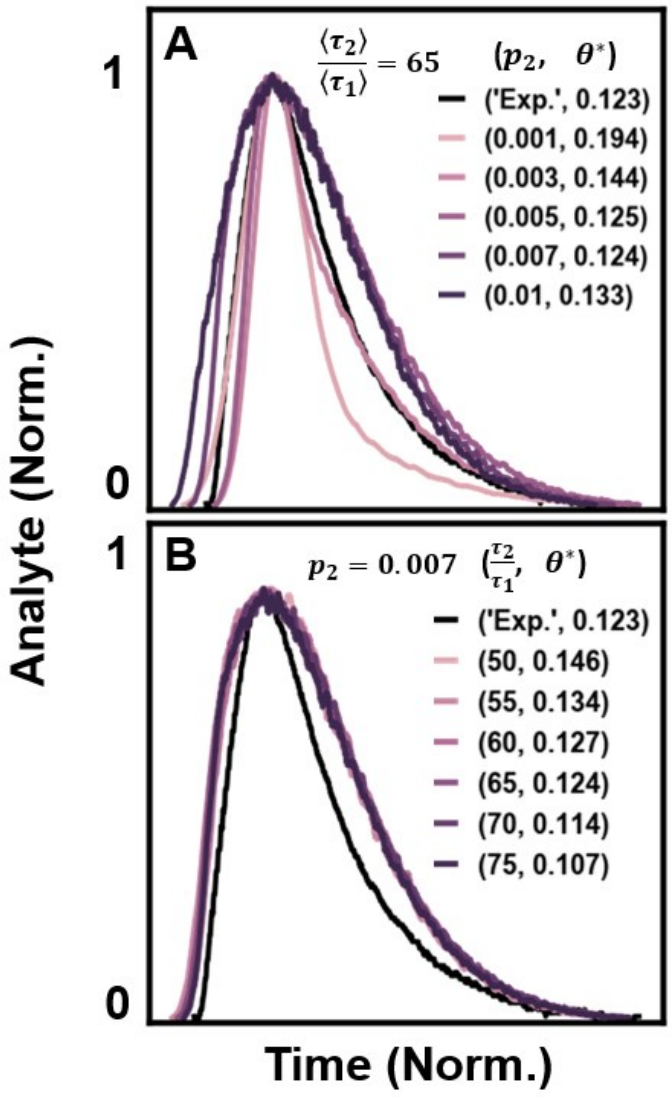

Figure S7. Assessing surface kinetics using a simulated sweep of different surface kinetic scenarios for lysozyme flowed over hydrophobic membranes. (A) Comparing experimental data (black, $\theta^* = 0.123$) to a series of simulated chromatographic curves with varied prevalence and a fixed expected desorption time ($\frac{\langle\tau_2\rangle}{\langle\tau_1\rangle} = 65$). The legend shows the relative prevalence and the value of θ^* for the simulation. (B) Comparing the experimental data (black, $\theta^* = 0.123$) to a series of simulated chromatographic curves with varied desorption ratios and fixed relative prevalence ($p_2 = 0.007$). The simulated chromatogram presented in the main text most closely matches the experimental value of θ^*.

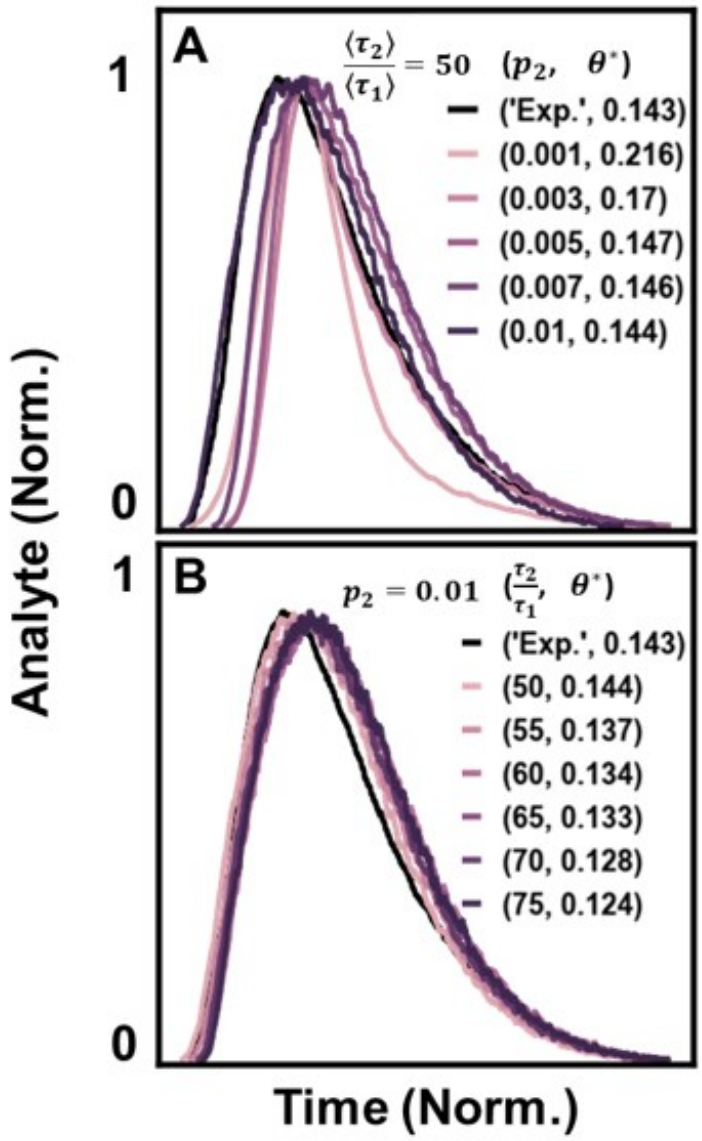

Figure S8. Assessing surface kinetics using a simulated sweep of different surface kinetic scenarios for lysozyme flowed over hydrophilic membranes. (A) Comparing experimental data (black, $\theta^* = 0.143$) to a series of simulated chromatographic curves with varied prevalence and a fixed expected desorption time ($\frac{\langle\tau_2\rangle}{\langle\tau_1\rangle} = 50$). The legend shows the relative prevalence and the value of θ^* for the simulation. (B) Comparing the experimental data (black, $\theta^* = 0.143$) to a series of simulated chromatographic curves with varied desorption ratios and fixed relative prevalence ($p_2 = 0.01$). The simulated chromatogram presented in the main text most closely matches the experimental value of θ^*.

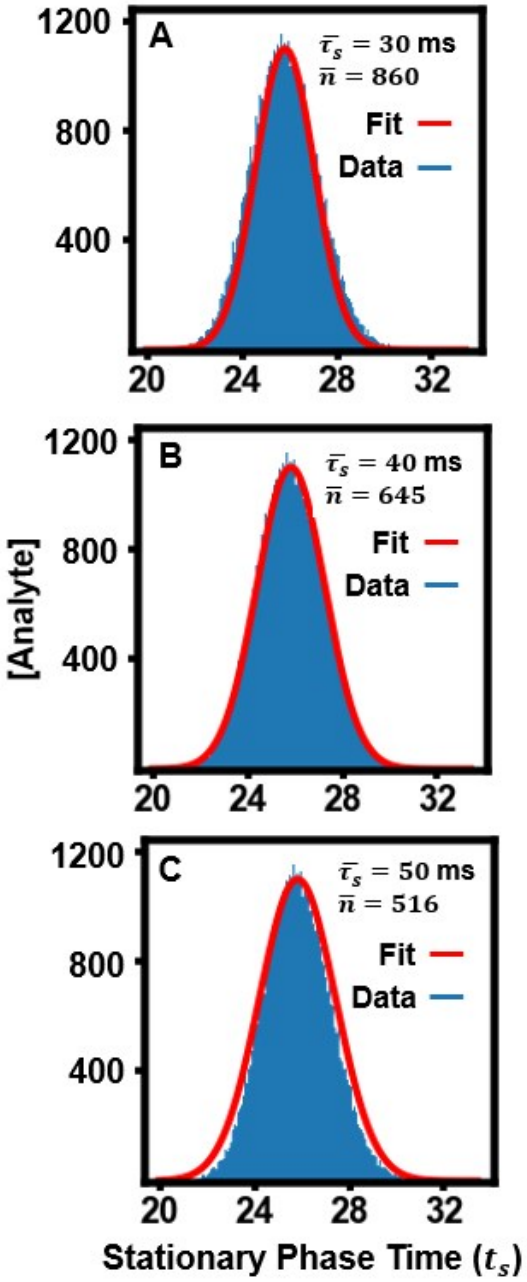

Figure S9. Gaussian fits of the homogeneous population extracted from the simulated chromatogram matching the experimental data. Fit estimates for the number of adsorptions in the column based on the average desorption time of the most common event (A = 30 ms, B = 40 ms, C = 50 ms). Values are extracted from a Gaussian fit user Felinger's method of stochastic analysis.[2]

References

1. Felinger, A., *Data analysis and signal processing in chromatography*. Elsevier: 1998; Vol. 21.
2. Felinger, A., Molecular movement in an HPLC column: A stochastic analysis. *Lc Gc N Am* **2004,** *22* (7), 642-647.