Supporting Information

The non-linear effects of the number of stochastic single-molecule adsorption events on ensemble elution profiles

Ricardo Monge Neria,[1] Rachel A. Saylor[2], and Lydia Kisley,[1, 3]*

[1]Department of Physics and [3]Department of Chemistry, Case Western Reserve University, Cleveland, Ohio, USA 44106

[2]Department of Chemistry and Biochemistry, Oberlin College, Oberlin, OH, USA 44074

*Corresponding author: lydia.kisley@case.edu

S1. Explanation of SM Lévy process model inputs

By applying Lévy process formulation presented in **Eq. 7** and **Eq. 8**, and experimental SM data can be input to generate chromatogram model peaks. The user defined parameters correspond to a fixed r_m value, and a respectively scaling value of n_p, following the recommended $n_p \geq 1.5 \cdot r_m$. The model peak is simulated using time units ($\Delta\tau_S$) of ms, but the resulting time axis (*i.e.* the time delta between model points, n_p) is defined in minutes by unit conversion (*i.e.* $\frac{1\ ms}{100\ sec} \cdot \frac{1\ sec}{60\ min}$). The value of $\Delta\tau_S$ is defined by the SM detector time resolution (32 ms in our case). The SM dwell time distribution exists as a vector with all the unique measured adsorption times ($\tau_{S,i}$), and the corresponding probability ($\Delta F_S(\tau_{S,i})$) for each time, both generated within the code from the input of a vector containing all the measured SM dwell time values, and the variables of the experimental $\Delta\tau_S$, and maximum number of frames per data acquisition (to define the maximum observable adsorption time).

Table S1. Summary of input variables for model.

Variable	Description	Shape and details
$\bar{r}_m$	Average number of SM adsorption events	Single number. Adjustable parameter used to tune model peak elution time and shape.
n_p	Number of data points in the final model chromatogram	Single number. Should scale at least 1.5x with the value of $\bar{r}_m$ to avoid under-sampling errors in the IFFT. Recommend using the same (highest) value for all runs for consistency in peak height or normalize otherwise.
$\Delta\tau_S$	Time increment used in model (same units as $\tau_{S,i}$)	Single number. The time resolution of the SM observations. Will depend on frame rates for the detector. Used in units of ms.
$\tau_{S,i}$	Observed sorption times (limited by detector sampling resolution, *i.e.* case frame rate)	Vector. Array of unique observed adsorption times corresponding to dwell times distribution.
$\Delta F_S(\tau_{S,i})$	Probability distribution for sorption times, in same order as $\tau_{S,i}$	Vector. Array of probability values matching the unique observed adsorption times.

S2. SM data plots and super-resolutions images of tested stationary phase particles.

Single molecule fluorescence microscopy experiments of flowing fluorescent dye rhodamin6G (R6G) were performed on immobilized, isolated stationary phase particles as previously reported[19]. Briefly, single molecules fluorescent signals were identified and analyzed using a home-written, publicly available [33] MATLAB (2022b) code that uses 2D Gaussian fitting strategies, paired with radial symmetry[34] centroid position refinements, to obtain super-resolved, single-molecule locations. From these localizations, to obtain kinetic information individual molecules are identified then grouped as adsorption sites based on their spatial proximity and the cross-correlation factor of their shapes. Based on the individual grouped locations, the kinetics are then extracted by defining the dwell times based on the number of frames during which a molecule is present, and the association times on the number of frames between different molecules at a single site. Based on the time between frames (exposure plus dead/clearing time), the number of frames are converted to ms and stored. From these, the individual site cumulative time distributions can be calculated, as well the ensemble across an entire imaged area by adding the results of all the events.

The resulting single molecule observations yielded measurements of the adsorption (dwell) time cumulative probability distributions for each tested experimental condition of varying flow rate (**Fig. S1A**). Notably, there is significant overlap in dwell times between the increasing flow rate conditions, but rare longer dwell times are also increasingly sampled. Similarly, association time cumulative probability distributions (**Fig. S1C**) can be generated, showing a clear decrease in overall association times (i.e. increasing k_a) with increasing flow. Super-resolution maps of the tested 12 stationary phase particles were also generated (**Fig. S1B**). The particles were imaged as 4 individual representative isolated particles, and two grouped particles (of 3 and 5 respectively), over two different sample cells. As we have previously shown[19] the high degree of surface functionalization present on the cellulose tris(3,5 dimethyl phenyl-carbamate) functionalized silica “fully porous” stationary phases can limit accessibility to the porous network. This pore occlusion can cause apparent flow channeling (**Fig. S1B**), but as we also previously demonstrated, solvent treatment (such as with dimethyl sulfoxide DMSO) decreases the degree of surface functionalization (**Fig. S1D**). This directly leads to improved access to the pore network and reduced flow channeling. This helps to support that our observed analyte-stationary phase interactions are representative of the mass transfer behavior in our on-column HPLC experiments.

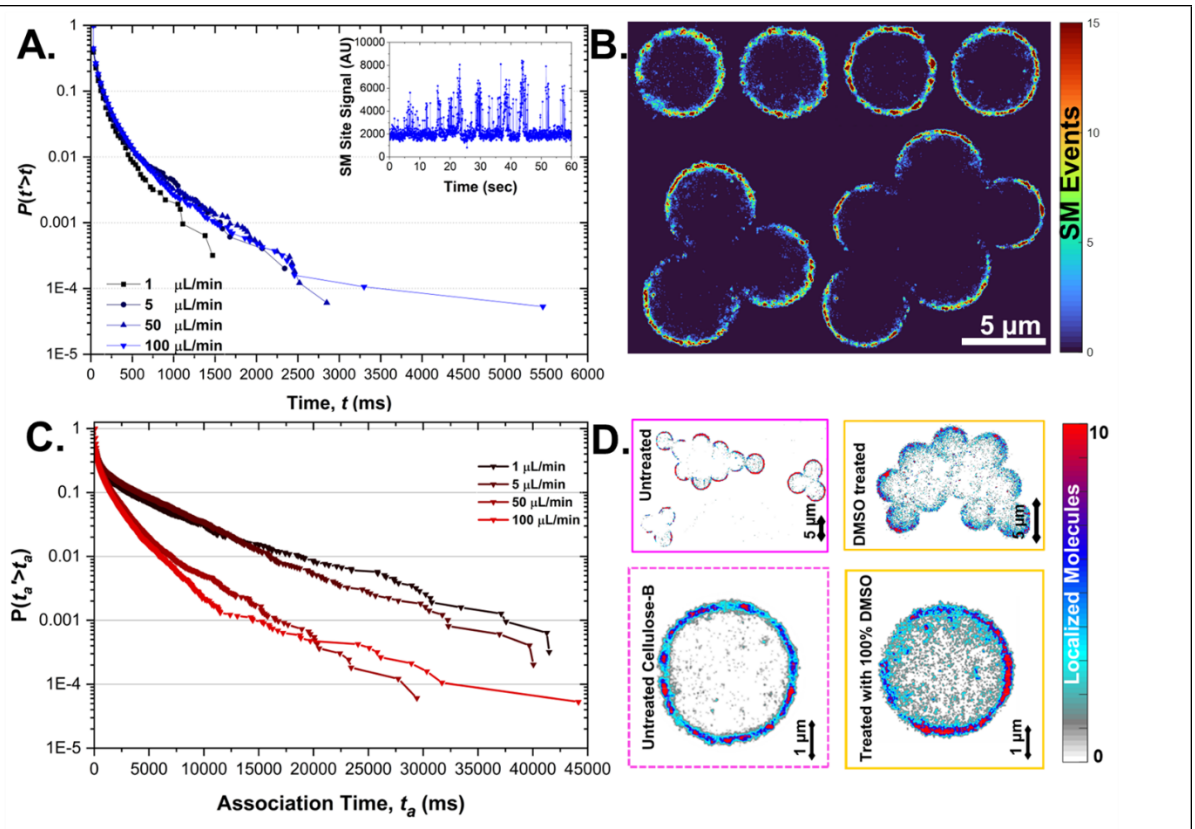


Fig. S1. Obtained SM adsorption time distributions, and super-resolution maps. Dwell time distribution data of 1 nM R6G interacting with commercial cellulose tris(3,5 dimethyl phenyl-carbamate) functionalized silica chromatography particles was obtained by single-molecule fluorescence microscopy. (A, inset) The adsorption (*i.e.* dwell) time of single R6G molecules on individual adsorption sites is obtained by tracking changes in the fluorescence intensity signal that arrives at the EMCCD detector. (A) From SM observations, ensemble dwell time distributions are obtained from at least four individual isolated and eight grouped (packed together; groups of three and five respectively) chromatography particles, measuring thousands of SM adsorption events. This is done for the same samples and mobile phase compositions at varying volumetric flow rates (1, 5, 50, and 100 µL/min), providing both spatial (~25 nm resolution) and kinetic information (32 ms frame rate), useful for stochastic modeling. Solid lines provided as a guide for the eye. (B) The super-resolution maps for all the 12 representative stationary phase particles used in this work. The data corresponds to the 50 µL/min flow rate condition, with the color scale corresponding to the number of observed SM adsorption events.(C) Association time probability distributions are also generated from the SM observations. (D) We also have previously demonstrated that solvent treatment to reduce the degree of surface functionalization can reduce the amount of apparent flow channeling, which we attribute to pore occlusion. Image reproduced from Monge Neria, R. *et al.*.[19]

S3. Unprocessed HPLC on column elution data and plotted by elution volume.

HPLC elution data for the on-column experiments with R6G in HEPES buffer. The raw HPLC detector elution data (Fig. S2A) shows larger apparent gaps between different flow rate conditions due to differences in eluted volume per unit time. Thus, for all subsequent analysis done in this work, the elution time axis was shifted by the void time, defined by the first minimum in the detector signals (Fig. S2B). We also converted x-axis into eluted volume based on the experimental flow rates data used (Fig. S2C) to better showcase the differences in HPLC peak shapes.

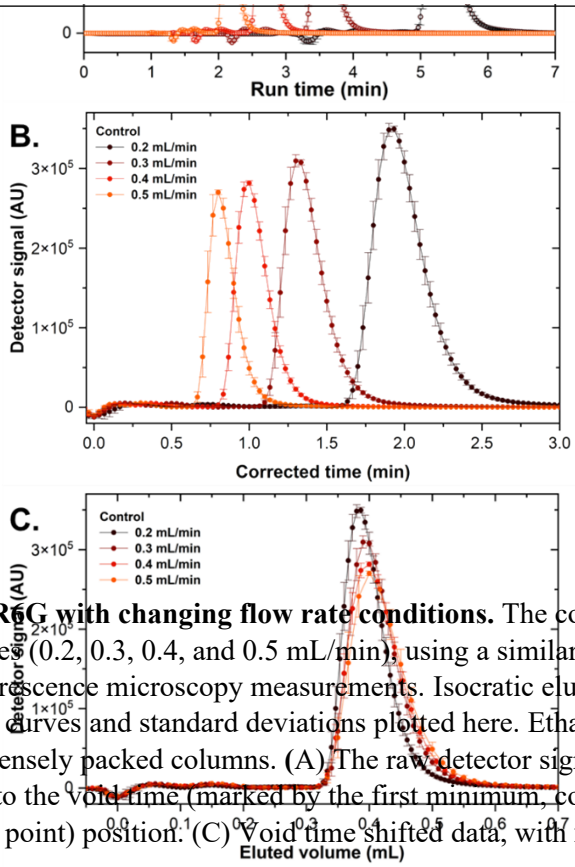


Fig. S2. HPLC elution of R6G with changing flow rate conditions. The column runs were done at varying volumetric flow rates (0.2, 0.3, 0.4, and 0.5 mL/min), using a similar HEPES buffer and R6G composition as the SM fluorescence microscopy measurements. Isocratic elution was done in triplicates, with the average curves and standard deviations plotted here. Ethanol modifier was included to assist elution out of the densely packed columns. (A) The raw detector signal data. (B) Corrected time axis data with respect to the void time (marked by the first minimum, corresponding to the unretained sample injection point) position. (C) Void time shifted data, with x-axis also scaled by volumetric flow rate.

S4. Estimating linear flow velocities in column and microscope flow-cells

We estimate the linear flow velocities our microscopy flow-cells and our chromatography column to help connect the observed behaviors and sample conditions as previously described in Monge Neria et al. (2023) [19]. The flow cell dimensions are cylindrical with 13 mm diameter, 0.15 mm depth (h). There are two outlet and inlet apertures that are ~10 mm in diameter, set 8 mm (inner-to-inner circumference) apart. We image the deposited stationary phases between the inlet and outlet. The flow cross-sectional area is estimated as a range of values defined by vertical slices of the cylindrical well. For a radius (r) of 7.5 mm, and a limited range of deposition (x) defined from the center of the circumference, we can define the cross-sectional radial length (l) as:

$$l = \sqrt{r^2 - x^2}\,; r = 0.65\ cm,\ x = 0\ to\ 0.4\ cm \qquad \textbf{(S1)}$$

from which we define the flow cell cross sectional area, A_X:

$$A_x = h \cdot 2l \qquad \textbf{(S2)}$$

Given that the amount of deposited stationary phase material is very low, we directly estimate that for a given applied flow rate f_r (assuming 1 mL/min = cm^3/min, for solvent density ~1 g/mL), the linear velocity in the Hybriwell cross section (v_x) is defined as:

$$v_x = \frac{f_r}{A_x} \quad \textbf{(S3)}$$

Using **Eq. 15** gives us a range of cross-sectional areas 0.0154 to 0.0195 cm^2. Similarly, we can estimate the linear flow velocity within our HPLC cylindrical column that has dimensions of 5.0 cm length, and 4.6 mm inner diameter (D). For the cylindrical shape of the column, this gives a cross-sectional area (A_c) of ~ 0.17 cm^2. Assuming a 0.6 ± 0.1 occupied volume (V_p, unitless fraction) filled with stationary phase material; we can estimate the linear velocity in the columns (v_c) for a given flow rate (f_r) as:

$$v_c = \frac{f_r}{(A_c \cdot V_p)} \quad \textbf{(S4)}$$

Along these lines, shear rates can also be estimated following the formalism which defines the shear at the walls as a ratio of the volumetric flow rate vs the cross section and linearly scaling with distance from the walls; in other words a gradient flux. For the SM experiment we concern ourselves with the shear at the surface with the immobilized stationary phases, which results in:

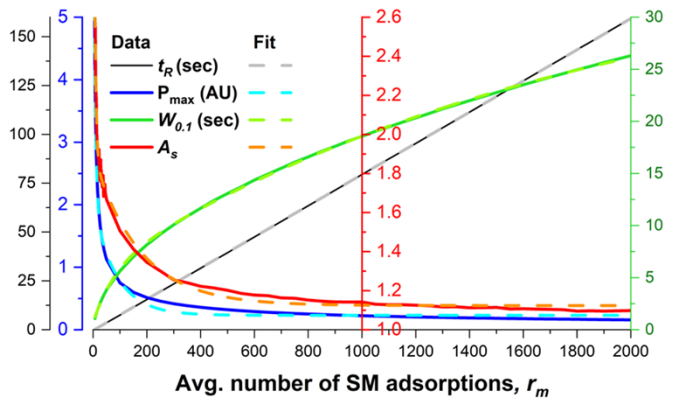
$$\dot{\gamma} = \frac{6 f_r}{A_x \cdot h} \quad \textbf{(S5)}$$

For the cylindrical packed column we can instead estimate the shear rate as:

$$\dot{\gamma} = \frac{32 f_r}{\pi D^3 \cdot V_p} \quad \textbf{(S5)}$$

Table S2. Estimated experimental linear velocities and shear rates.			
Sample	Volumetric flow Rate	Estimated linear velocity (cm/min)	Estimated shear rate (s^{-1})
Hybriwell microscopy flow cell	1 µL/min	0.06 ± 0.01	0.39 ± 0.04
	5 µL/min	0.29 ± 0.05	1.9 ± 0.2
	50 µL/min	2.9 ± 0.5	19 ± 2
	100 µL/min	6.0 ± 1.0	39 ± 4
Regis HPLC column	0.2 mL/min	2.1 ± 0.5	3.0 ± 0.4
	0.3 mL/min	3.1 ± 0.7	15 ± 2
	0.4 mL/min	4.0 ± 1.0	150 ± 20
	0.5 mL/min	5.0 ± 1.0	300 ± 40

S5. Fitting model peak parameter behaviors with respect to r_m.



Parameter	Peak height, Pmax
Equation	y = A1*exp(-x/t1) + A2*exp(-x/t2) + y0
y0	0.2 ± 0.01
A1	7 ± 0.3
t1	6 ± 0.3
A2	2 ± 0.07
t2	90 ± 8
R-Square	0.99

Parameter	Asymmetry, A_s
Equation	y = y0 + A1*exp(-(x-x0)/t1) + A2*exp(-(x-x0)/t2)
y0	1.0
x0	1.4
A1	2.3
t1	5.0 ± 0.3
A2	0.8
t2	170 ± 11
R-Square	0.99

Parameter	Width, w_0.1
Equation	y = y0 + A1*(1 - exp(-x/t1)) + A2*(1 - exp(-x/t2))
y0	0.7 ± 0.04
A1	30 ± 0.4
t1	1800 ± 50
A2	5 ± 0.1
t2	80 ± 3
R-Square	0.99

Parameter	Elution time, t_R
Equation	y = a + b*x
Intercept	-0.3 ± 0.01
Slope	0.08 ± 1E-5
R-Square	1.0

Fig. S3. Fits of the different model elution parameters tested relative to changing r_m**.** Model elution peaks were generated using the 50 µL/min SM experiment dwell time datasets. The average number of adsorptions over the elution time was tuned as the free parameter $\bar{r}_m$. (Blue) The peak height is well fitted by a two-component decaying exponential. (Green) The peak width is well fitted by a two-component exponential association function, and (red) the asymmetry by a two-component decaying exponential, as well. In agreement with theory, the elution time (black/gray) scales linearly with $\bar{r}_m$. Reported uncertainties represent 95% confidence intervals for function fits.

S6. The number of modeling points n_p affects model elution peak shapes and sets a minimum required value to avoid Fourier space under sampling.

We test the dependance of peak shapes and behavior when varying n_p, the number of modeling sampling points, which is also a free parameter in the Lévy process formulation of the stochastic model. The effects that n_p can have on the resulting model peak shapes and elution times were not directly addressed in Pasti et al. We presume this might have been due to the relatively small values ranging $\bar{r}_m = 3$ up to ~900 that were used leading to $n_p$dependances not being readily apparent. For the sake of clarity, here we also test how this variable (n_p) affects the model peaks (**Fig. S4**).

We observe that all the peak parameters show inconsistent behaviors under a threshold n_p value, due to under-sampling effects as we change the value of n_p over a large range of 10-2000 (**Fig. S4**). We use the 50 μL/min SM adsorption kinetics distribution and fix the $\bar{r}_m$ value to 1000 as we vary n_p between 10-2000. We observe erratic behaviors at low n_p values ($n_p < 1500$), ultimately settling at a single elution peak shape with a defined elution time at high n_p values (**Fig. S4A,** $n_p > 1500$). As n_p increases, the model peaks begin as undefined flat (horizontal) lines that first rapidly decrease (**Fig. S4B** and **S4C**, blue regions), followed by peak shapes beginning to form, but shifting to decreasing (*i.e.* faster) elution times with increasing n_p (**Fig. S4B** and **S4C**, green regions). Then, there is a small region that shows discontinuities where the solution for the inverse Fourier transform of the CF results in only non-real numbers (**Fig. S4C**, gray region). After passing a threshold value of approximately 1.5 times $\bar{r}_m$, a consistent peak shape begins to form and settles at set asymmetry, width, and elution times (**Fig. S4B** and **S5C**, orange regions). After this point, the only effect that n_p has on the modeled elution peaks is slowly increasing the total peak height (**Fig. S4B** orange, **Fig. S4C** blue line) linearly with respect to n_p. We find that this behavior is likely the result of numerical rounding that occurs, gradually increasing the overall magnitude of the distribution in this computational method. Thus, we can conclude two important results for the model: 1) n_p values need to be selected above the 1.5 x $\bar{r}_m$ threshold to obtain consistent results, and 2) n_p must be a fixed value to correctly assess changes in peak height in relation to other model parameters.

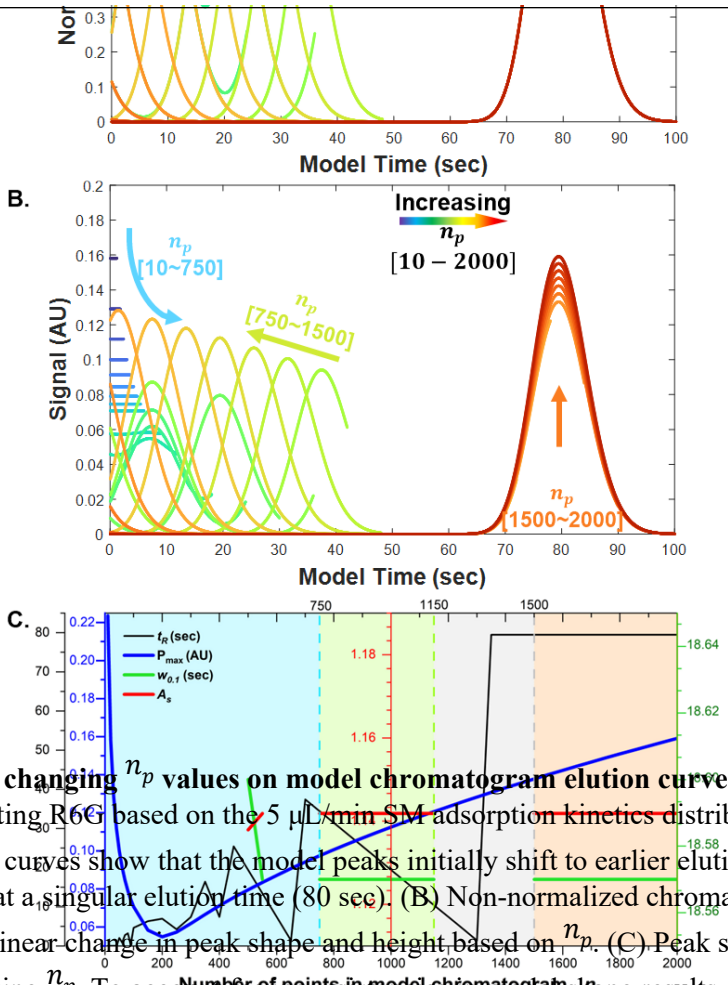


Fig. S4. The effect of changing n_p values on model chromatogram elution curves. Model chromatograms of eluting R6G based on the 5 µL/min SM adsorption kinetics distribution. (A) Peak-maximum normalized curves show that the model peaks initially shift to earlier elution times at low n_p values but then settle at a singular elution time (80 sec). (B) Non-normalized chromatograms show more clearly the non-linear change in peak shape and height based on n_p. (C) Peak shape parameters with respect to increasing n_p. To account for inconsistencies in model shape results, peak height is instead defined in terms of signal maximum here. The value of $\bar{r}_m$ was fixed as 1000 for the shown curves and datasets.

S7. Error minimization fitting by elution time for all HPLC flow rates.

We test how the free parameter r_m can be systematically adjusted to match experimental HPLC elution times with stochastic model data from measured SM adsorption time distributions. HPLC data was collected of the same rhodamine 6g analyte and Cellulose B stationary phase packed within a column as was imaged via SM microscopy. The $\bar{r}_m$ values are determined through our error minimization fitting with respect to the first three experimental chromatogram moments (position, width, and asymmetry) between the model elution and HPLC data as detailed in section 3.3.

Using all the SM datasets of rhodamine 6g solution at varying flow rates through the stationary phases (**Fig. S1**), we compare the results of using **eqn 14,** weighted 2x toward elution time ($W_{time} = 2, W_{width} = 1, W_{asym} = 1$), for error minimization fitting of the model compared to the 3.1 cm/min (0.3 mL/min) experimental HPLC elution (**Fig. S5**, see **Fig. S8** for other flow rates). This flow rate and the SM conditions are the best match amongst the sampled linear velocities between SM and HPLC experiments (**Table S2**). By fitting $r_m$based on elution time, we obtain multiple simulated elution peaks that all match the HPLC elution time (**Fig. S5a**) but differ in shape (**Fig. S5b**).

The obtained $\bar{r}_m$ values from error minimization without *a-priori* knowledge end up closely matching the theoretical values predicted by **eqn 10**, supporting Pasti et al.'s derivation and parameter definitions in relation to elution times (**Fig. S5c**). By adjusting the error minimization weighting, the HPLC peak asymmetry and widths can be better approximated by the simulated chromatograms (**Fig. S7**), but similarly to what was previously observed, the elution time and shape could not be simultaneously matched by adjusting the value of $\bar{r}_m$.

SM data collected at increased flow rates has increasing tailing more similar to the HPLC data. We observe similar asymmetry factors (**Fig. S5d**, red) between SM flow velocities of comparable orders of magnitude (0.06 and 0.29 cm/min), with slightly higher asymmetry for the higher flow (2.9 and 6.0 cm/min). Higher SM flow velocities also result in slightly wider peaks (**Fig. S5d**, green), despite having higher $\bar{r}_m$values. The very distinct tailing visible in the simulated peaks from the higher flow rate datasets (**Fig. S5a and S5b blue lines**) demonstrate the significant influence that SM adsorption times have on the model.

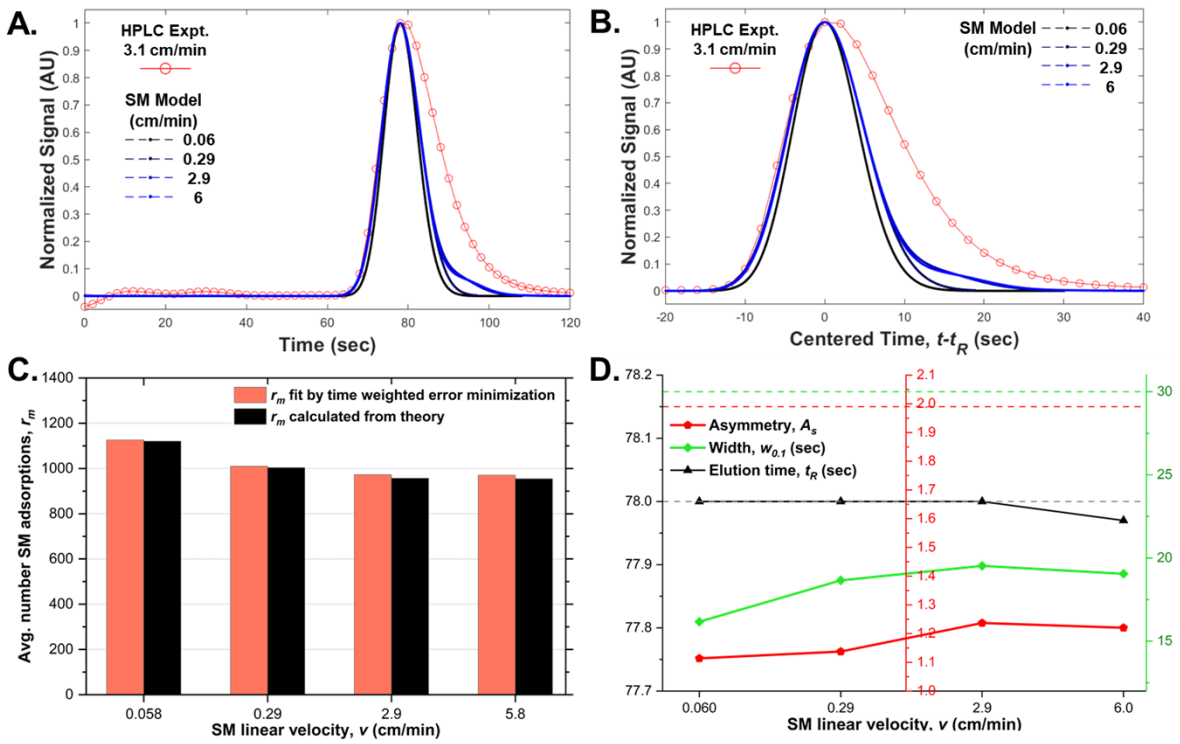


Fig. S5. Elution time-based error minimization fitted $\bar{r}_m$ values for SM data of varying flow rates. Simulated chromatograms of eluting R6G based on SM adsorption kinetics distributions obtained at varying flow rate conditions were fitted to experimental HPLC elution at 0.3 mL/min by error minimization using the free parameter $\bar{r}_m$. (A) The resulting simulated peaks, showing agreement in elution time with the HPLC experimental data. (B) Shifting the time axis to center around each elution peak center shows the increasing simulated peak width and asymmetry with increasing SM flow rate. However, none fully matches the experimental peak shape. (C) The resulting $\bar{r}_m$ values from the elution time fitting show agreement with the predicted theoretical values from **eqn 10**. (D) The elution peak shape parameters show the model peaks closely match the HPLC experimental elution time (dotted gray line), but much less so for the HPLC asymmetry (dotted red) and width (dotted green). Solid lines in (d) provided as a guide for the eye. The value of n_p was fixed to 5000 for all simulated chromatograms, and the weights set as $W_{time} = 2, W_{width} = 1, W_{asym} = 1$.

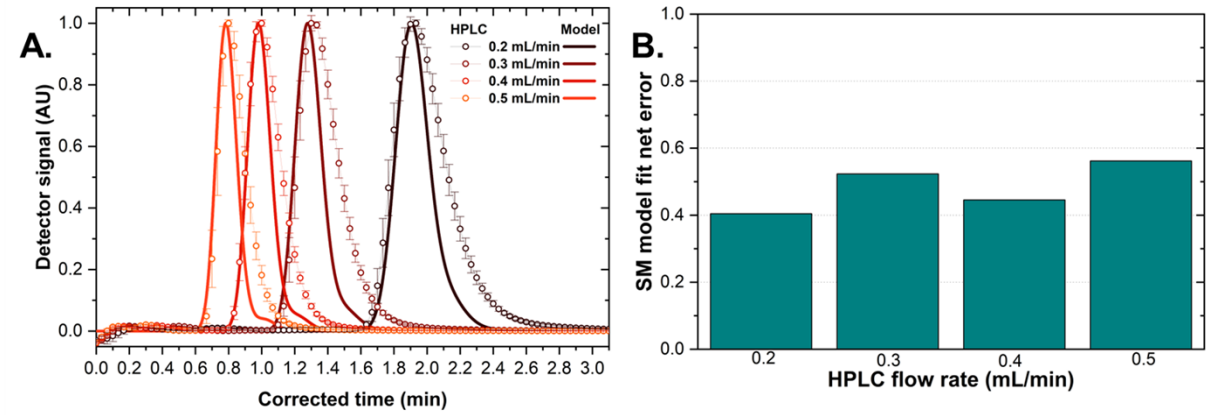


Fig. S6. Error minimization fitting by shape of all the HPLC elution data. The normalized detector signal vs time of experimental HPLC elutions of R6G in HEPES buffer, run at varying volumetric flow rates. (A) SM elution models using the 50 µL/min SM datasets were fitted to the HPLC data with respect to elution time via error minimization using the $\bar{r}_m$ free parameter. (B) The calculated net weighted errors (using **eqn 14**, with weights all equally set to one) after error minimization fitting show comparable relative error values (~0.5 net%, or ~30% in relation to the maximum of $\sqrt{3}$).

S8. Error minimization fitting matches $\bar{r}_m$ based on HPLC peak shape, matching peak width and approaching the right asymmetry, but results in much less accurate elution times.

We ran a second set of simulated peaks to try to match peak shapes rather than elution times, still relying on the free parameter $\bar{r}_m$. This time, the error minimization was weighed toward width ($W_{width} = 2$) and asymmetry ($W_{asymmetry} = 1$), omitting elution time ($W_{time} = 0$) from **eqn. 14**, still using the 0.3 mL/min HPLC dataset as reference. The multiple simulated peaks that do not match the experimental elution time (**Fig. S7A**) but do approach the correct peak shape (**Fig. S7B**). Unlike before, the resulting $\bar{r}_m$ values no longer match the theory description (**Fig. S7C**), all yielding much higher estimated r_m from fitting by peak shape. The overall peak widths closely match the HPLC peak (**Fig S7D**, green) and begin to approach the correct asymmetry (**Fig. S7D**, red). However, asymmetry decreases exponentially with increasing $\bar{r}_m$ values (**Fig. S7C**, red), causing the model error minimization to fail to converge when weighted primarily to asymmetry, since very small $\bar{r}_m$ results in very thin peaks. This results in all fitted values being primarily large $\bar{r}_m$ which can better approach the peak width, with the primary source of asymmetry becoming the SM dwell time distributions. Furthermore, just as in the original work by Pasti et al., the peak shape and elution times cannot be simultaneously matched (**Fig. S7D**, black) between simulation and experiment by manipulating the single free parameter $\bar{r}_m$.

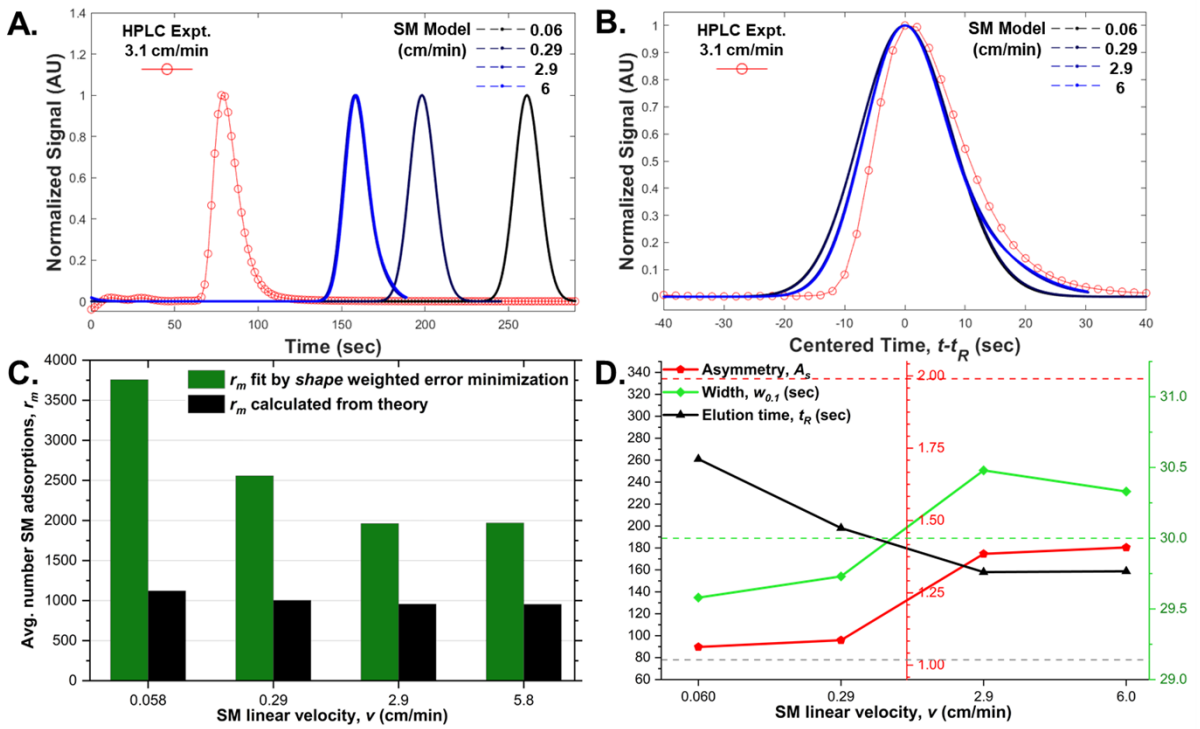


Fig. S7. Peak shape-based error minimization fitted r_m values for SMFM data of varying flow rates. Simulated chromatograms of eluting R6G based on SM adsorption kinetics distributions obtained at varying flow rate conditions were fitted to the experimental HPLC elution peak shape at 0.3 mL/min by error minimization using the free parameter $\bar{r}_m$. (A) Simulated peaks, showing agreement in elution time with the HPLC experimental data. (B) Shifting the time axis to center around each elution peak center shows the increasing simulated peak width and asymmetry with increasing SM flow rate. However, none fully matches the experimental peak shape. (C) The resulting $\bar{r}_m$ values from the elution time fitting show agreement with the predicted theoretical values from **Eq. 14**. (D) The elution peak shape parameters show close matches for the HPLC experimental elution time (dotted gray line), but much less so for the HPLC asymmetry (dotted red) and width (dotted green). Solid lines in D provided as a guide for the eye. The value of n_p was fixed to 5000 for all simulated chromatogram, and weights for error minimization set as $W_{time} = 0, W_{asymmetry} = 1, W_{width} = 2$.

S9. Single-molecule flow rate conditions affect elution model behavior and point toward experimental sampling dependencies.

We observe that the experimental single-molecule flow rate conditions affect the simulated elution peak behavior with respect to the free parameter $\bar{r}_m$, showing trends that differ from HPLC observations. We apply the procedures and concepts developed in section 4.1 to all our available SM datasets for different flow rate conditions and compare how their elution peak parameter curves differ with the changing $\bar{r}_m$ free parameter (**Fig. S8**). All resulting simulations of flow rate data show overall similar parameter dependencies with respect to $\bar{r}_m$, with a few key differences that become more pronounced at large $\bar{r}_m$ values. All elution times (**Fig. S8A**) follow linear behavior, while asymmetries (**Fig. S8B**) and peak heights (**Fig. S8D**) follow decaying exponentials, and peak widths (**Fig. 8C**) follow exponential association curve shapes. However, contrary to what is observed in HPLC measurements (**Fig. S2B**), we see delayed elution times with increasing SM flow rate (**Fig. S8A**), as well as increasing peak widths (**Fig. S8C**), and increasing asymmetries (**Fig. S8B**). The only agreeing trend with HPLC becomes the decreasing peak heights with increasing flow rates (**Fig. S8D**, inset). Furthermore, an interesting jump in the asymmetry parameter curves occurs at large $\bar{r}_m$ values (**Fig. S8B**, and inset, $\bar{r}_m >$ 700), hinting at what might be sampling dependencies from the SM experimental data.

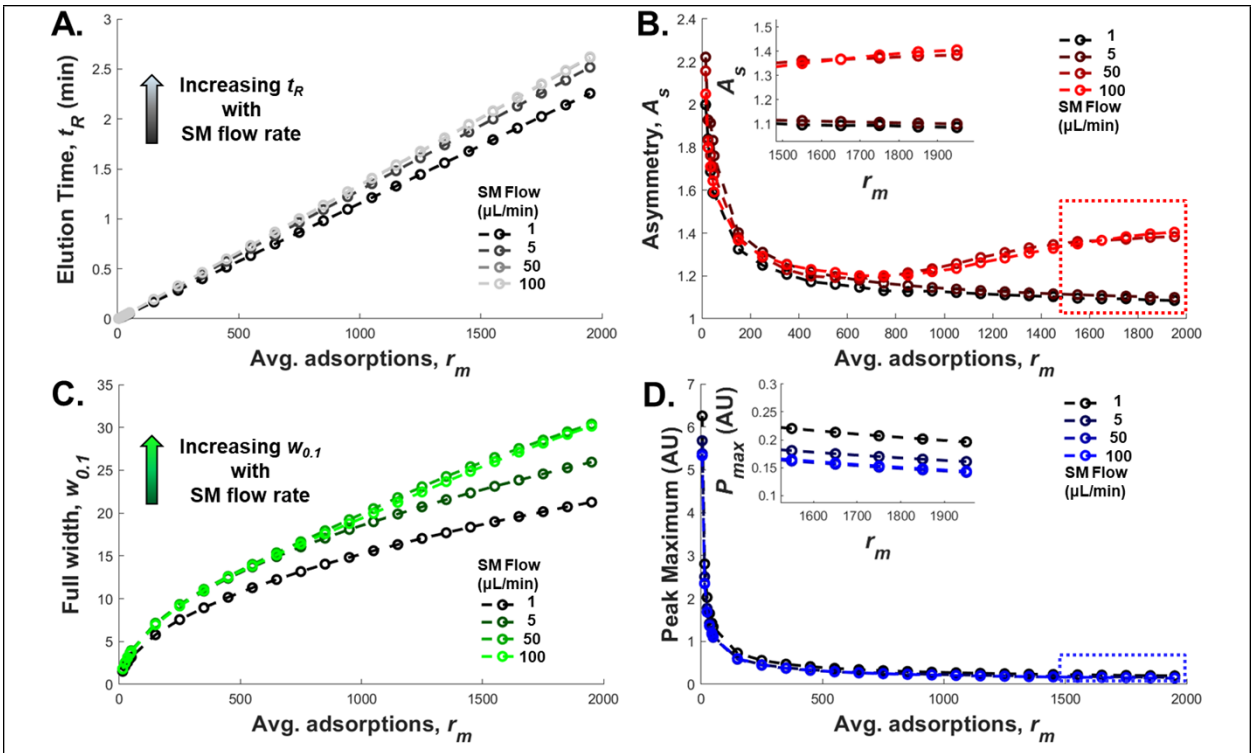


Fig. S2. Influence of SM experimental flow rate on model elution curve behavior with changing $\bar{r}_m$. Model elution curve characteristic parameters tested with changing $\bar{r}_m$ values for four different SM experimental flow rate conditions. (A) All elution times scale linearly with $\bar{r}_m$, but the higher SM flow rate conditions show later elution times. (B) The asymmetry factors for all the modeled curves show an initial decaying exponential behavior with respect to $\bar{r}_m$, but the higher SM flow rate conditions (50 and 100 μL/min) show an inflection (of the form $\sim [1 - \exp(-\bar{r}_m)]$) at large $\bar{r}_m$ values (>700). (C) The peak width with respect to $\bar{r}_m$ for all model curves shows similar two-component exponential association behavior, with wider peaks at higher SM flow rates. (D) All model peak heights show similar exponential decay behavior with increasing $\bar{r}_m$, but higher SM flow rates show slightly faster decay rates. All runs shown used a fixed n_p value of 5000 model points.