Classification: Physical Sciences, Applied Physical Science

SI

Unified superresolution experiments and stochastic theory provide mechanistic insight into protein ion-exchange adsorptive separations

Lydia Kisley[a], Jixin Chen[a], Andrea P. Mansur[a], Bo Shuang[a], Katerina Kourentzi[b], Mohan-Vivekanandan Poongavanam[c], Wen-Hsiang Chen[b], Sagar Dhamane[c], Richard C. Willson[b,c,d,e,1], Christy F. Landes[a,f,1]

[a] Department of Chemistry, Rice University, Houston, TX 77251; [b] Department of Chemical & Biomolecular Engineering, University of Houston, Houston, TX 77004; [c] Department of Biology & Biochemistry, University of Houston, Houston, TX 77004; [d] Houston Methodist Research Institute, Houston, TX, 77030; [e] Centro de Biotecnología FEMSA, Departamento de Biotecnología e Ingeniería de Alimentos, Tecnológico de Monterrey, Monterrey, NL 64849, Mexico; [f] Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251

[1] Corresponding authors: Dr. Christy F. Landes and Dr. Richard C. Willson

Rice University; Department of Chemistry; P.O. Box 1892, MS 60; Houston, TX 77251

(713)-348-4232

cflandes@rice.edu

Department of Chemical & Biomolecular Engineering, University of Houston, Houston, TX 77004

(713) 743-4308

willson@uh.edu

SI information: 30 pages, 6 figures, 3 tables, 2 videos

SI Contents:

Materials and methods: **3-5**

Agarose surface preparation 3-4

Fluorescent dye labeling of α-lactalbumin 4-5

Total internal reflectance fluorescence (TIRF) wide field microscopy setup 5

Data analysis: **5-12**

Identifying protein adsorption events at superresolution level 5-8

Producing superresolution pseudo image and the identification of specific adsorption sites 8-10

Kinetic analysis at individual specific adsorption sites and ensemble level 10-11

Use of the stochastic theory of chromatography to predict elution curves 11-12

Biochemical controls: **13-14**

Agarose structure is temporally stable 13

Protein structure is stable in the presence of peptides 13

Salt studies confirm ion-exchange interactions between pentaargininamide and α-lactalbumin 13-14

Single molecule controls: **14-19**

Aggregation of α-lactalbumin is not the source of different dissociation components 14-16

Power studies rule out any photophysical effects 16

Single molecule peptide functionalization demonstrated with dye-labeled peptides 16-19

Single molecule adsorption events demonstrated with increasing protein concentration 19

Kinetic heterogeneity: **19-20**

Flow rate 19

Agarose percentage 19-20

Additional details: **20-21**

Superresolution pseudo images discussed in main text 20

Electrostatic structure of α-lactalbumin 20-21

Kinetic distribution of single stochastically-clustered monoargininamide sites 21

References: **21-22**

Figures and Tables: **23-29**

Figs S1-S6 23-28

Tables S1-S3 29

Movie Still Images and Captions: **30**

Movies S1-S2 30

Materials and methods:

Agarose surface preparation

Agarose surface preparation and protein labeling have been originally reported in Daniels, *et al.* (1). Glass coverslips (No. 1; VWR, 22 x 22 mm) were cleaned at 80 °C for 90 s in a solution of 4% (v/v) H_2O_2 (Fisher Scientific, Radnor, PA) and 13% (v/v) NH_4OH, then cleaned with an oxygen plasma for 2 min (PDC-32G; Harrick Plasma, Ithaca, NY; medium power). Silicon templates (Grace BioLabs, Bend, OR) were attached to the cover slips, and 1 mL of hot 1% (w/w) agarose solution (US Biological, Swampscott, MA) was spun onto each cover slip at 2000–3000 rpm using a Brewer Science Cee 200CBX precision coat-bake system. A custom flow chamber (Grace BioLabs) was then placed over the agarose thin film.

Charged peptide ligands were attached to the agarose surfaces using aldehyde functionalities introduced by 30 min treatment with 20 mM $NaIO_4$, as described by Afanassiev *et al*. (2). After washing with DI water and drying under a stream of helium, the activated surfaces were treated with a 42 nM solution of the pentaargininamide peptide (NH_2-GGRRRRRamide, Biomatik, Wilmington, DE); the amide form of the peptide was used to eliminate the negative charge on the C-terminus to avoid creating a zwitterionic ligand, as previously described (3, 4) in coupling buffer (100 mM Na_2HPO_4/NaH_2PO_4 [EM Science, Gibbstown, NJ] and 150 mM NaCl [Mallinckrodt Chemical, St. Louis, MO], pH 7.2). The guanidinium side group of arginine (pK_a=12.5) is protonated at this pH, thus we expect it to be unreactive towards the agarose aldehyde reactive groups, and coupling through the N-terminal primary amine to be strongly favored. The grafting density was controlled by varying the offered concentration of the respective peptide, as further discussed below (5). The peptides were immobilized on the surface with several drops of 20 mM $CNBH_4$ (Pierce, Rockford, IL) at 4 °C for 30 min. Excess

uncoupled peptide was removed by rinsing with coupling buffer, and unreacted aldehyde sites were reduced with 66 mM of $NaBH_4$ (in 25% EtOH/75% phosphate-buffered saline) at 4 °C for 5 min. The surfaces were finally rinsed gently with deionized water and stored at 4 °C, if not used immediately. Control agarose surfaces were prepared following the procedure outlined above, but omitting the peptide from the coupling buffer. To investigate the effect of the ligand charge on protein adsorption, the peptide length was varied. Di-, tri- and tetraargininamide peptides (NH_2-GGRRamide, NH_2-GGRRRamide and NH_2-GGRRRR respectively, Biomatik) were attached on agarose supports. The concentration of the peptides used for support modification was 107.5 µM, 71.6 µM, 53.75 µM for the di-, tri- and tetraargininamide respectively, resulting in an identical offered charge concentration (215 µM) in all samples.

Fluorescent dye labeling of α-lactalbumin

For protein labeling, Alexa 555 succinimidyl ester (Life Technologies; abs/em maxima 555/565 nm) was dissolved in dimethylformamide at a concentration of 10 mg/mL, and 0.1 mL of this dye solution was slowly added to 1 mL of 10 mg/mL Ca^{2+}-depleted α-lactalbumin dissolved in 0.1 M $NaHCO_3$ buffer, pH 7.4, to preferentially label the protein amino terminus rather than lysine ϵ-amino groups (6). The reaction was incubated at room temperature for 1 h with constant gentle mixing, and the reaction was then stopped by addition of 0.1 mL of freshly prepared 1.5 M hydroxylamine, pH 8.5. The labeled protein was dialyzed against water for 24 h and then against 10 mM Tris-HCl and 100 mM NaCl at pH 8.0 for 24 h. To ensure complete removal of the unincorporated fluorophore, PD-10 desalting columns and/or gel filtration chromatography were used. Gel filtration was carried out on an AKTA Purifier10 using Sephadex 75 10/300 GL (GE Healthcare, Buckinghamshire, UK) with 10 mM Tris-HCl and 100

mM NaCl at pH 8.0 as running buffer. Fractions with an estimated fluorophore-to-protein ratio of 1.0 ± 0.2 were pooled and used for adsorption studies.

Total internal reflectance fluorescence (TIRF) wide field microscopy setup

Data were collected on a home-built TIRF wide-field microscope, shown schematically in Fig. S1. Either a 532 nm solid state laser (Coherent, Compass 315M-100SL) or a 633 nm HeNe laser (CVI Melles Griot, 25-LHP-925-249) were used for excitation. The beams were passed through separate acousto-optic modulators (IntraAction, 402AF1) controlled by frequency generators (Fluke, 271-U 115V) to synchronize excitation to the detection rate (7) before being combined onto the same path with a long-pass dichroic filter (Chroma, 565dcxt). The beams were expanded prior to focusing at the edge of a 1.45 NA, 100x, oil-immersion objective (Carl-Zeiss, alpha Plan-Fluar) for through-the-objective TIRF microscopy. The TIRF excitation penetrated a 1/*e* depth of ~85 nm at an intensity of ~5 mW/cm^2. Emission was collected in an epi-fluorescence setup and was separated from excitation with a dichroic mirror (Chroma, z532/633rpc) and notch filters (CVI, 623.8-25.0M and Kaiser, HNPF-532.0-1.0). Two dichroics (Chroma, T640lpxr) separated the emission and cross talk was reduced with bandpass filters (Chroma, ET685/70m and ET585/65m). The signal was detected on an electron-multiplied charge coupled device (Andor, iXon 897) at an integration time of 30 ms, frame rate of 16.072 Hz, and electron multiplying gain of 300.

Data analysis:

Identifying protein adsorption events at superresolution level

The data with stochastic bright fluorescent point spread functions on each frame was analyzed for adsorbed proteins. Data analysis was performed using home-written code on MATLAB R2011b with a commonly used 2D Gaussian fitting strategy (8-12). The data

(typically 1,000 frame series) collected by the EMCCD was converted to a MATLAB compatible format as a series of two dimensional images with pixel size 64 nm and frame time 62 ms (30 ms integration time and 32 ms data storage time). Fig. S2 shows an example of how the raw data are treated. On the selected area of this frame there is a bright spot (Fig. S2A). This bright spot is defined as an adsorption event. The event is caused by an adsorbing molecule that may last more than one frame on the substrate surface, in which case more than one event is assigned to the same molecule on multiple frames (the duration of single adsorption events over multiple will be reconstructed later in the analysis). An adsorption event satisfies two criteria:

1. The local maximum is greater than baseline plus 3 times background noise. The baseline is defined as the average of the background, and the background noise is defined as the standard deviation of the intensity histogram of all pixels in all frames (Fig. S2B). This threshold satisfies >99.7% confidence that the fluorescent spot is an adsorbed molecule (13).

2. The sum intensity within 4 pixels radius of this maximum minus baseline is larger than 3 times the total photon counts of an ideal Gaussian peak, which has standard deviation $\sigma \sim$ 2 pixels (128 nm) and amplitude, A, = the background noise. The total photon counts $I = 2\pi A\sigma^2$ for 2D Gaussian peak. This reduces the false positive rate to a negligible level.

The identification was achieved by a method based on a maximum intensity search and subtraction. First, the global maximum of the frame was picked and the pixels within a distance of 4 pixels from the maximum were analyzed with the above two conditions. These pixels were then erased to background baseline. The cycle was repeated until no global maximum was found that satisfies condition 1. A simplified center-of-mass method was used to find the intensity center of the 4 pixel radius of the original image before the erasure of the pixels, similar to the method presented in Sergé, *et al.* (8, 14). The difference between the calculated center and the

location of the maximum was used to judge whether the selected maximum was a leftover tail of a very strong event, and was not the top of a new Gaussian peak. If the calculated center was >3 pixels away from the peak maximum, this maximum would not be considered as an independent adsorption event.

Once the events were identified, the 4 pixel radius of the calculated center was fitted to a 2D Gaussian function to obtain the parameters given in Equation S1 (Fig. S2C).

$$I(x,y) = A \cdot \exp\left[\frac{-(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}\right] + B \quad \text{(S1)}$$

Where *I(x,y)* is the pixel intensities; *A* is the amplitude of the spot, and *B* is the local background; x_0, y_0 are the center of the spot; σ_x and σ_y are the standard deviation of the peak at *x* and *y* dimensions, which are related to the full width at half maximum (FWHM) of the spot with FWHM = $2\sqrt{2\ln 2}\sigma = 2.36\sigma$.

The intensities of the events $2\pi A\sigma_x\sigma_y$ were found typically to be 2×10^4 counts. The background baseline and noise varied with the substrates and the concentrations of the protein in the solution. Typical baseline values were 7.3 ± 0.9 counts per pixel for both agarose and monoargininamide samples, and 47 ± 2 counts per pixel for pentaargininamide sample with 500 pM protein in the solution. Typical noise levels were 26 ± 1 counts per pixel for both the agarose and monoargininamide samples, and 69 ± 2 counts per pixel for the pentaargininamide sample with 500 pM protein in the solution. For the 2D Gaussian peaks, the average standard deviation $\sigma = (\sigma_x + \sigma_y)/2$ = 2.1 ± 0.4 pixel = 140 ± 20 nm and the ratio between σ_x and σ_y was 1.0 ± 0.2 in 2,739 events.

Based on the statistics of photon counts for a stationary Alexa 555 fluorophore measured previously (7) and the instrumental conditions, the temporal resolution of the analysis was

studied, showing proteins exhibiting dissociation times 3 ms and longer were detected. Over one frame, an Alexa 555 molecule will emit $2.7 \pm 1.9 \cdot 10^4$ counts over $2\pi A\sigma_x\sigma_y$ pixels^2. Based on this intensity and the noise level, 2D Gaussian point spread functions were simulated with fractional intensities based on desorption times that were less than one frame. Based on the condition 1, Fig. S2F shows that we can detect proteins that are present for 3 ms or longer.

Producing superresolution pseudo image and the identification of specific adsorption sites

A superresolution pseudo image was generated to help identify the adsorption sites on the surface by correlation. The center of the events on all the frames (Fig. S2D) were used to generate a superresolution image on a pseudo image with 1/10 pixel size (6.4 nm) of the original fluorescent images. Each center is converted to a Gaussian peak $I(x,y) = A \cdot \exp\left[-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2}\right]$ with amplitude A = 1 adsorption event and standard deviation σ ~ 13 nm, FWHM ~ 30 nm (determined to be the resolution of our images based on the spread of centroid locations as shown in Fig. S2D) and summed together on the pseudo image (Fig. S2E).

The adsorption sites were identified from the superresolution pseudo image when the maximum of the 2D Gaussian peak in the superresolution images was larger than 5 adsorption events. This threshold of 5 events satisfies >99% confidence of adsorption site assignment according to the control experiments of the agarose-only samples, which will be discussed later. The same method described above was used to pick the maximum for adsorption sites on the superresolution pseudo image at a size of 21 x 21 pixels around the local maximum. An alternative strategy was used where the original local image was compared to a standard 2D Gaussian peak $I(x,y) = A \cdot \exp\left[-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2}\right]$ with amplitude A = 5 adsorption events, standard deviation σ = 5 pseudo pixels (= 32 nm) at both x and y dimensions, (x_0, y_0) the center

of the local maximum. This was done by calculating the normalized cross-correlation between the selected area and the standard Gaussian peak; both were converted to 1D vectors before the comparison (8, 15). The normalized cross-correlation $CC = \frac{1}{n}\sum_i \frac{(f(i)-\bar{f})(t(i)-\bar{t})}{\sigma_f \sigma_t}$ where n is the number of pixels, f and t are the target vector and the standard reference vector, and σ is the standard deviation of the vector. If the normalized cross-correlation was smaller than 0.6, the identified maximum was not considered as a specific adsorption site. This excluded the adsorption sites with inseparable adsorption sites nearby from consideration of further analysis. The center of the 2D Gaussian peak was then defined as the location of the adsorption site.

The accuracy of the locations (standard deviation to the average center of the group of adsorption events) measured on multiple frames for a same adsorption site is 14 ± 7 nm (20 adsorption sites were analyzed). This value demonstrates that for each fitted center, we have >95% confidence that the real center of the adsorption site is within ~28 nm. This value is a combination of the uncertainties of the localization method, the size of the protein, and the motions of the adsorption sites and the protein.

The threshold chosen for adsorption sites with the pentaargininamide samples was based on the statistics of the desorption times of the molecules on the bare agarose control sample. Because agarose does not have any peptide ligands present, all interactions between the protein and interface are non-specific. The average desorption time of a molecule on the control agarose samples was < 1.01 frame (Table S1). Since the desorption time would be expected to be Poisson distributed, the probability of a molecule that lasted x frames was $f(x;\lambda) = \lambda^x e^{-\lambda} / x!$, where λ = 1.01 is the average dwell time in frame. The probability of a random non-adsorbing molecule with ≥ 5 detected events was < 0.4% based on Poisson statistics (or 0.03% directly calculated

from Table S1). For a typical 1,000-frame measurement on the control samples, the total number of adsorption events was 590 ± 60 in an area of 12 × 12 μm^2 and the adsorption density of events was ~0.0038 per 30 x 30 nm^2 area, using the same Poisson equation with $\lambda = 0.0038$, the probability of 5 random molecules that adsorbed to the same 30 x 30 nm^2 area was negligible. This confidence level gave a predicted average possible false positive of 0.01 adsorption sites in the area of interest shown in the pseudo images.

Kinetic analysis at individual specific adsorption sites and ensemble level

Once the adsorption sites were identified, all the adsorption events within 128 nm radius (the same area for the adsorption-site identification) from the center of an adsorption site were assigned to the adsorption site. From each adsorption site, the desorption time (i.e. dwell time), and adsorption time (i.e. the waiting time between events), were obtained via counting the number and length of events from frame to frame. When events occurred in two (or more) adjacent frames at the same location, it was assumed the molecule stayed throughout the two frames without desorption and re-adsorption during the interval. This assumption is valid because statistically insignificant numbers of very fast events are expected at the utilized protein concentrations and corresponding adsorption/desorption rate constants. Cumulative distributions of the desorption and adsorption times were used, as they have been shown to be more sensitive to rare populations than the commonly used histograms, or probability distributions (16), and were obtained by an integration of the number of counts of the desorption and adsorption times at all locations. For data analysis of data reported in the SI Appendix, ensemble cumulative distributions of all adsorption events at all specific adsorption sites were used due to the time limitations for accurately fitting individual adsorption site kinetics.

Kinetic information was obtained from fitting exponential decays to the cumulative distributions, according to $P(t'>t)=\sum_{i=1}^{N} A_N \exp\left[-t/\tau_N\right]$, where *P(t'>t)* is the probability of observing an event that is *t* long or longer. The data was fit to either a single (N = 1) or 3-component (N =3) decay, for single site (Table S3) and ensemble data, respectively. The number of components for analyzing ensemble data was selected following standard curve fitting considerations (17) and was due to the variability in kinetics from site to site. The amplitude contribution (A_N) and decay time, τ_N, were calculated for each component.

Use of the stochastic theory of chromatography to predict elution curves

To extract the elution curves from the single molecule desorption distributions, the method presented by Pasti, *et al.* (18) was used and is summarized here. The theory is based on a simple equilibrium equation for our system:

$$\alpha LA + I \underset{k_d}{\overset{k_a}{\leftrightarrow}} [\alpha LA - I] \qquad \text{(S2)}$$

where αLA is the protein, I is the interfacial adsorption site, and [αLA-I] is the protein adsorbed to the specific adsorption site. The stochastic theory of chromatography proposes that the probability of a single protein associating with an adsorption site r_m times (*r* number of stochastic adsorption events for a given time, *t*, in the mobile phase, *m*) is given by a Poisson distribution:

$$\psi_m = \frac{(k_a t_m)^{r_m}}{r_m!} \exp\left[-k_a t_m\right] \qquad \text{(S3)}$$

given an adsorption rate constant of k_a and the total time in the mobile phase t_m during the total time the analyte is in the chromatography column. The Poisson distribution is then converted to

the frequency domain for the representative characteristic function formalism (ϕ), a common method for representing stochastic processes (19):

$$\phi(t_s;\omega \mid t_m) = \exp\left[\mu t_m\left(\exp\left\{i\omega\tau_s\right\}-1\right)\right] \quad \text{(S4)}$$

where t_s is the overall time spent in the stationary phase for the analyte and each individual adsorption event lasts for a duration of τ_s, while μ is the frequency of adsorption events. Because the distribution of τ_s is discontinuous for single molecule data, the Lévy formalism is applied to represent the characteristic function (20):

$$\phi(t_s;\omega \mid t_m) = \exp\left[r_m \sum_{i=1}^{i=k}\left(\exp\left\{i\omega\tau_{s,i}\right\}-1\right)\Delta F(\tau_{s,i})\right] \quad \text{(S5)}$$

where k is the index of the discrete set of desorption times given by $\Delta F(\tau_{s,i})$ that are observed in the desorption time distributions extracted from single molecule experiments. By performing an inverse Fourier transform of the Lévy representation of the characteristic function to the time domain:

$$\phi(t_s;\omega \mid t_m) \xrightarrow{FT} f(t_s) \quad \text{(S6)}$$

the chromatographic peak shape, $f(t_s)$ is obtained. The stochastic chromatographic theory was developed for columns, but the presented single molecule experiments do not have a given time within the column that determines r_m in equation S5. Instead, a basis value of $r_m = 100$ was used to predict curves from single molecule data. The value was estimated based on experimental conditions, including observation volume, flow rate, protein concentration, and average number of detected events over time. It is therefore valid to consider the resulting elution curves to be relative to one another but individually unique to the specific single molecule experimental conditions.

Biochemical controls:

Agarose structure is temporally stable

To ensure that changes in agarose structure did not affect the measurements, the stability of the non-crosslinked agarose thin films used was investigated as a function of time. Agarose hydrogels have been shown to have macroscale stability over long periods of time in saline solutions (21), but may undergo conformational changes on short time scales (22). We tracked the number of non-specific events of α-lactalbumin in the agarose as a function of time for two different samples (Fig. S3A). We did not see a significant change in the number events over the course of ~30 min. This time point is the limitation of our experiment, after which the flow chamber can become unstable. Fig. S3A demonstrates that the agarose thin films in our experiment are stable within the timescale of our observations and do not contribute temporal variability in our data.

Protein structure is stable in the presence of peptides

Protein structural stability in the presence of peptides was studied by circular dichroism spectroscopy. A Jasco J-810 spectropolarimeter with a 1 cm path length, 320-250 nm, 10 accumulations, 20 °C, 0.1 nm pitch was used. Solutions of unlabeled α-lactalbumin diluted to 70 μM alone, with 1:1 pentaargininamide, and with 1:1 and 1:5 monoargininamide molar ratios were measured. Fig. S3B shows the resulting circular dichroism spectra, demonstrating that minimal protein tertiary structure perturbation occurs in the presence of the peptide ligands, as demonstrated by minimal changes in the peak at 270 nm (23).

Salt studies confirm electrostatics govern ion-exchange interactions between pentaargininamide and α-lactalbumin

To demonstrate the electrostatic adsorption mechanism, as opposed to hydrophobic or other mechanisms of adsorption of α-lactalbumin to the interface, the ionic strength of the sample was varied. The same area of the sample was analyzed with α-lactalbumin in 10 mM and 1 M NaCl buffer (10 mM Tris-HCl, pH 8). High concentrations of salt are used in ion-exchange chromatography to suppress adsorption. Fig. S3C and D report the superresolution pseudo images of specific adsorption sites, showing there is large decrease in the number of specific adsorption sites with high number of events at increased salt concentration. The decrease in the number of specific adsorption interactions at increased ionic strength solution conditions shows the specific adsorption interaction is electrostatic in nature, and hence, the data throughout this work is an ion-exchange mechanism.

Single molecule controls:

To confirm that our observations were indeed due to single molecule conditions and single molecule adsorption/desorption events, controls were performed as detailed below and reported in Fig. S4.

Aggregation of α-lactalbumin is not the source of different dissociation components

An alternative explanation of the observed adsorption kinetic components at the interface could be that the proteins themselves were aggregating and the different oligomers were the cause of the different populations interacting with the pentaargininamide interface. Oligomerization has been reported to be the cause of different residence times with bovine serum albumin at silicone oil-water interface (24) and fibrinogen at solid interfaces (25). Within the data sets, different intensities of single events were observed that suggested that oligomerization could be a possibility (Fig. S4A). A histogram of the different intensities of all events within a trial shows a minimum of three different intensity populations were observed (Fig. S4B).

There are two possible explanations for the intensity differences; one, the proteins could have been aggregating, where an increase in intensity would indicate multiple single-labeled proteins within each oligomer. The second possibility could have been that the labeling of the protein could not be performed ideally at one dye per protein; instead multiple fluorophores were found on some of the proteins, possible by multiple-labeling at both the N-terminal α-amino group and a lysine ε-amino group. Since labeled protein was purified using gel filtration (Superdex 75) with a single peak observed, we expect a homogeneous population of molecules in the sample without the presence of oligomers or other aggregates or multimeric forms of the labeled protein. To test these two possibilites, the desorption time distributions were extracted separately for each intensity group of proteins at an ensemble level. If aggregation was occurring, the proportion of desorption times would be expected to drastically change due to a change in the electrostatic structure of the protein that is the main interaction between the pentaargininamide peptides and the α-lactalbumin. If dye labeling was the cause, no change would be observed, as the protein structure would be the same for all events.

A fit of the intensity distribution to a three-component Gaussian shows approximately quantized populations, with centroids being located intensities of 2,900 ± 200, 8,100 ± 800, and 20,000 ± 2,000 counts. The deviation from perfect quantized intervals is most likely caused by a wide standard deviation of the normal distributions of the molecular intensities due to excitation intensity variation across the images in objective-based TIRF. From the fit, cutoffs for each intensity were defined as the midpoint between the different centroid locations (orange lines, Fig. S4B). The desorption time distributions were then extracted for each (Fig. S4C), showing near identical behavior for all three intensity groups. Thus, aggregation is not the cause of the long

desorption component, further supporting that clustering of adsorbent charges is required for any type of significant increase in retention time of mobile phase analytes.

Power studies rule out any photophysical effects

In single molecule fluorescence spectroscopy the limited lifetime of individual dye molecules can lead to false observations if photophysical effects occur on the time scale of the events of interest. Dye molecules can enter a temporary (photoblink) or permanent (photobleached) dark state. In the studies reported here, these could lead to incorrect identification of both protein desorption (mimicked by the fluorophore entering the dark state) and adsorption (mimicked by a photoblinked molecule returning to the bright state). These two photophysical effects are dependent on the rate of the molecule entering the excited state (26). Therefore, an increase in incident excitation intensity would decrease the lifetime of the molecules. If the lifetime of the molecules was on the order of the event desorption times, a decrease in the desorption times would be observed at higher excitation intensities. This was not the case, as shown in Fig. S4D, where the ensemble desorption distributions remained consistent at different excitation intensities, showing the times of adsorption event observations were not an effect of photophysics, but instead were due to protein-interface interactions.

Single molecule peptide functionalization demonstrated with dye-labeled peptides

To check that our sample preparation produced single molecule peptide adsorption sites and to calibrate unlabeled peptide densities, we prepared similar samples with varying peptide loading densities with both dye-labeled and unlabeled peptides. In the case of labeled peptides, the number of detected single fluorescent peptide sites was counted directly, whereas with unlabeled peptides, the number of sites undergoing specific dye-labeled protein adsorption was counted, as described earlier. Pentaargininamide was modified with BODIPY650 (Anaspec,

Fremont, CA) and monoargininamide was functionalized with Cy5 (New England Peptide, Gardner, MA). The dye-labeled peptides were immobilized on agarose using the same procedure as for the unlabeled peptides. From the obtained images, single peptide molecules were identified based on intensity in the same manner as α-lactalbumin sorption events described previously. Intensity traces of the identified molecules over time exhibited single step photobleaching (Fig. S4E), confirming we were identifying single fluorophores, and hence single peptides. Representative images of dye-labeled pentaargininamide samples prepared at increasing concentrations are shown in Fig. S4F and identified molecules are shown in Fig. S4G. The number of detected peptides increased linearly with preparation concentration, which, along with single-step photobleaching, (Fig. S4H and I) supports that all experiments, including those of unlabeled peptides, were performed under single ligand conditions.

Because of the small and localized charge on the peptide adsorbents, peptides used in collecting adsorption and desorption dynamics were not dye-labeled, as the presence of a fluorescence functional group on the relatively small peptide (as compared to the large protein) substantially alters the adsorption/desorption kinetics. To show that the adsorption interactions using unlabeled peptides was also at the single molecule level as demonstrated in the labeled studies, the number of specific adsorption sites as identified by fluorescent protein adsorption were studied as a function of peptide concentration (Fig. S4E) and a linear trend was also observed. The slope was not the same for the labeled and unlabeled methods, with a greater number of labeled peptides observed as a function of concentration as compared to unlabeled peptides. This result is consistent with the steric accessibility model discussed in the main text. For the dye-labeled peptide samples, all peptides are observed, regardless of their accessibility to

protein adsorption. In contrast, within a distribution of sterically accessible peptide sites on the agarose support, protein can adsorb to a smaller distribution of peptides than are actually present.

For the concentrations used in experiments with unlabeled monoargininamide ($\geq$ 215 μM), the dye-labeled counterpart showed that the concentration was too high to allow for resolvable single sites. Therefore, stochastic clustering density of monoargininamide below the diffraction limit is not adequate to retain α-lactalbumin. To resolve clustering below this limit would require future work using stochastic techniques with photoswitchable probes/conditions to attain a superresolution image of the arrangement of monoargininamide required for specific protein association (especially for the 1000x benchmark [215 mM] concentration reported in Fig. 2B). At much lower concentrations, dye-labeled monoargininamide also exhibited a linear dependence on peptide concentration offered during immobilization (Fig. S4I), showing that, as for pentaargininamide, the concentration can control single-molecule density of peptides on the surface. The discrepancy between the slopes found between the labeled-penta- and monoargininamide is due to heterogeneity between different areas of the samples, as indicated by the error bars, and also because the different dyes used for labeling the two oligomers have slightly different labeling efficiencies and effectiveness in functionalizing the agarose. The structures of the peptides also vary, as pentaargininamide has an increased number of amine groups compared to monoargininamide. When the respective slopes in Figures S4H and I are converted to immobilization efficiency, we observe efficiencies of $3.4 \pm 1.7 \times 10^{-5}$ and $1.1 \pm 0.6 \times 10^{-5}$ % for penta and mono, respectively. The higher immobilization efficiency of pentaargininamide is likely due to the increased number of amine groups, as the coupling reaction is between a peptide amine and agarose aldehyde group and argininamide side chain amines may participate. Overall, the charge density is similar and controllable using preparation

concentration for comparing samples with different oligomers. The differences between the slopes/efficiencies also do not influence the overall conclusion that clusters of charges are necessary to create detectable adsorption sites, as even at concentrations of monoargininamide at higher values that would overcome the difference in immobilization efficiency, we still do not observe specific adsorption (see Fig. 2B, main text).

Single molecule adsorption events demonstrated with increasing protein concentration

A further control to demonstrate our observations were due to single molecules was through studies of the dye-labeled protein concentration. The number of adsorption events observed increased linearly with increased concentration of Alexa 555-labeled α-lactalbumin (Fig. 4J), supporting single molecule observations of protein adsorption.

Kinetic heterogeneity:

Flow rate

To understand if the shear force of the mobile phase flow rate influenced the kinetic heterogeneity, the flow rate was varied from 2-10 μL/min with data collected at the pentaargininamide interface. Fig. S5A shows the respective cumulative desorption time distributions at an ensemble level of α-lactalbumin adsorption only at specific sites. The identical behavior of adsorption kinetics at the different flow rates reported shows the low flow rate used in our reported experiments and results does not impart any strong forces on the protein that would influence the observed kinetics. Further, the low flow rate ensures the first assumption of the stochastic theory (negligible net diffusion) is upheld (27, 28).

Agarose percentage

To understand if steric or motional properties of the agarose gel play a role in kinetic heterogeneity, agarose percentage was varied from 1-3% in pentaargininamide sample

preparation, as pore size is controlled by agarose percentage (29). Based on agarose percentage, Figure S5B reports the total number of observed specific sorption events over five minutes and Fig. S5C shows the cumulative desorption times on an ensemble level of the kinetics of adsorption only at specific sites. The reduction of the total number of events shown in Fig. S5B as agarose percentage increased demonstrates the influence of the agarose support on adsorption behavior. The cumulative distribution of desorption times shows that at 3% agarose (decreased pore size) there are no longer any long-term events observed above 0.4 s that are present at lower agarose percentages. This suggests that site-to-site heterogeneity is due to steric screening. Steric screening of peptides presents less available charges to the protein, reducing electrostatic affinity, and leading to shorter desorption times. At high agarose percentages the pores are smaller, preventing peptides from being sterically available for adsorption, while at low percentages, the heterogeneity in the pores leads to sites that are more sterically available (with higher affinity and longer desorption times) and also some sites with low steric availability (identical to 3% agarose).

Additional details:

Superresolution pseudo images discussed in main text

Fig. S6A shows the pseudo superresolution image with the individual pentaargininamide ligands for which kinetics and elution curves were reported in Fig. 3. Fig. S6B reports the pseudo superresolution image of dye-labeled α-lactalbumin over the control glass substrate, showing that no interactions between the anionic protein and substrate are observed.

Electrostatic structure of α-lactalbumin

The electrostatic potential distribution for α-lactalbumin (Fig. S6C) shows clustered negatively-charged patches on its surface. Based on our previous results (4) we expected the

anionic patch that is responsible for Ca^{+2} binding between aspartate residues to dominate the interactions between the protein and the argininamide-oligomer peptides, leading to the observed onset of adsorption between di- and tri-valent peptides.

Kinetic distribution of single stochastically-clustered monoargininamide sites

The desorption kinetics at four representative single specific adsorption sites at the 1000x, stochastically-clustered monoargininamide sample is reported in Fig. S6D, showing the heterogeneity observed site-to-site for this sample. Kinetic distributions of desorption times from all the sites within the 165 μm^2 analyzed area of the sample were used to generate the simulated elution profiles in Fig. 3E.

References

1. Daniels CR, *et al.* (2012) Fluorescence correlation spectroscopy study of protein transport and dynamic interactions with clustered-charge peptide adsorbents. *J Mol Recognit* 25(8):435-442.
2. Afanassiev V, Hanemann V, Wölfl S (2000) Preparation of DNA and protein micro arrays on glass slides coated with an agarose film. *Nucleic Acids Res* 28(12):e66.
3. Chen W-h, Fu JY, Kourentzi K, Willson RC (2011) Nucleic acid affinity of clustered-charge anion exchange adsorbents: Effects of ionic strength and ligand density. *J Chromatogr A* 1218(2):258-262.
4. Fu JY, *et al.* (2007) Enhanced Protein Affinity and Selectivity of Clustered-Charge Anion-Exchange Adsorbents. *Anal Chem* 79(23):9060-9065.
5. Cuatrecasas P (1970) Protein purification by affinity chromatography derivatizations of agarose and polyacrylamide beads. *J Biol Chem* 245(12):3059-3065.
6. Asish Xavier K, Willson RC (1998) Association and Dissociation Kinetics of Anti-Hen Egg Lysozyme Monoclonal Antibodies HyHEL-5 and HyHEL-10. *Biophys J* 74(4):2036-2045.
7. Kisley L, *et al.* (2013) Extending single molecule fluorescence observation time by amplitude-modulated excitation. *Methods Appl Fluoresc* 1(3):037001.
8. Cheezum MK, Walker WF, Guilford WH (2001) Quantitative comparison of algorithms for tracking single fluorescent particles. *Biophys J* 81(4):2378.
9. Schmidt T, *et al.* (1996) Imaging of single molecule diffusion. *Proc Natl Acad Sci USA* 93(7):2926-2929.
10. Gordon MP, Ha T, Selvin PR (2004) Single-molecule high-resolution imaging with photobleaching. *Proc Natl Acad Sci USA* 101(17):6462-6465.
11. Sharonov A, Hochstrasser RM (2006) Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proc Natl Acad Sci USA* 103(50):18911-18916.

12. Simonson PD, Rothenberg E, Selvin PR (2011) Single-molecule-based superresolution images in the presence of multiple fluorophores. *Nano Letters* 11(11):5090-5096.
13. Peterson EM, Harris JM (2009) Quantitative Detection of Single Molecules in Fluorescence Microscopy Images. *Anal Chem* 82(1):189-196.
14. Sergé A, Bertaux N, Rigneault H, Marguet D (2008) Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nature Methods* 5(8):687-694.
15. Mashanov G, Molloy J (2007) Automatic detection of single fluorophores in live cells. *Biophys J* 92(6):2199-2211.
16. Walder R, Kastantin M, Schwartz DK (2012) High throughput single molecule tracking for analysis of rare populations and events. *Analyst* 137(13):2987-2996.
17. Zielesny A (2011) *From Curve Fitting to Machine Learning: An Illustrative Guide to Scientific Data Analysis and Computational Intelligence* (Springer, Berlin).
18. Pasti L, *et al.* (2005) Single-Molecule Observation and Chromatography Unified by Lévy Process Representation. *Anal Chem* 77(8):2524-2535.
19. Dondi F, Remelli M (1986) The characteristic function method in the stochastic theory of chromatography. *J Phys Chem* 90(9):1885-1891.
20. Dondi F, Cavazzini A, Pasti L (2006) Chromatography as Lévy Stochastic process. *J Chromatogr A* 1126(1–2):257-267.
21. Shoichet MS, Li RH, White ML, Winn SR (1996) Stability of hydrogels used in cell encapsulation: An in vitro comparison of alginate and agarose. *Biotech Bioeng* 50(4):374-381.
22. Jimenez-Barbero J, Bouffar-Roupe C, Rochas C, Pérez S (1989) Modelling studies of solvent effects on the conformational stability of agarobiose and neoagarobiose and their relationship to agarose. *Int J Biol Macromol* 11(5):265-272.
23. Griko YV, Remeta DP (1999) Energetics of solvent and ligand-induced conformational changes in α-lactalbumin. *Protein Sci* 8(3):554-561.
24. Walder R, Schwartz DK (2010) Single Molecule Observations of Multiple Protein Populations at the Oil−Water Interface. *Langmuir* 26(16):13364-13367.
25. Kastantin M, Langdon BB, Chang EL, Schwartz DK (2011) Single-Molecule Resolution of Interfacial Fibrinogen Behavior: Effects of Oligomer Populations and Surface Chemistry. *J Am Chem Soc* 133(13):4975-4983.
26. Lakowicz JR (2006) *Principles of Fluorescence Spectroscopy* (Springer, New York).
27. Giddings JC (1965) *Dynamics of chromatography* (CRC Press).
28. Giddings JC, Eyring H (1955) A Molecular Dynamic Theory of Chromatography. *J Phys Chem* 59(5):416-421.
29. Pernodet N, Maaloum M, Tinland B (1997) Pore size of agarose gels by atomic force microscopy. *Electrophoresis* 18(1):55-58.

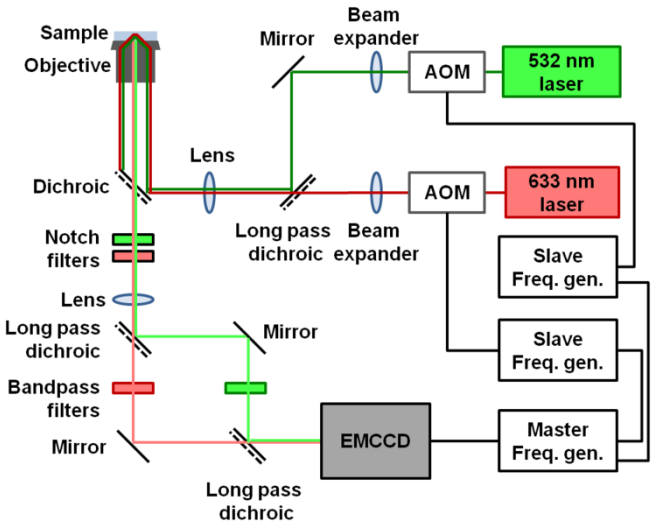


Fig S1. Schematic of wide field microscope, including acousto-optic modulator (AOM) controlled by frequency generators (Freq. gen.) and electron-multiplying charge coupled device (EMCCD) detector.

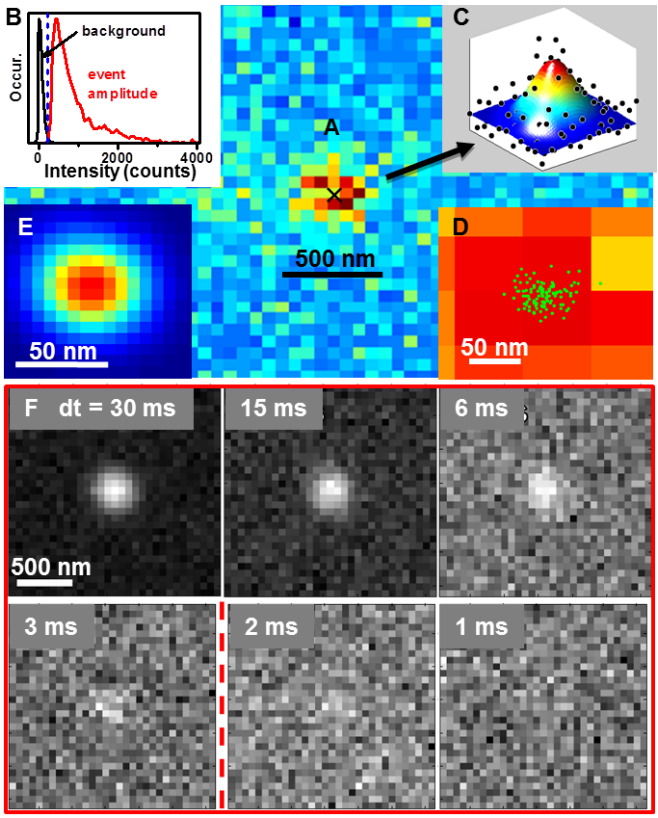


Fig S2. Data analysis and (**A**-**E**) spatial and (**F**) temporal resolution determination. (**A**) On a single frame of the data, an adsorption event is identified, whose intensity (peak amplitude) is larger than the threshold. (**B**) Normalized histograms of all pixel intensities of 1,000 frames (background) and the amplitude of identified adsorption events. The threshold (dashed line) is defined as baseline (average of the background histogram) plus 3 times the background noise level (standard deviation of the background histogram) (13). (**C**) The single adsorption event is fitted with a 2D Gaussian function for its peak center and amplitude. (**D**) A summary of the centers of the adsorption events in all the 1,000 frames at the same area of (**A**). (**E**) A superresolution image (1/10 pixel size) of the same area of (**D**) as sum of the adsorption events in (**D**), with each center converted to a Gaussian peak of amplitude = 1 and standard deviation = 30 nm. (**F**) Simulated PSF of Alexa 555 at different dwell times; the red line shows the cutoff value where our analysis will no longer detect the PSF based on the intensity threshold.

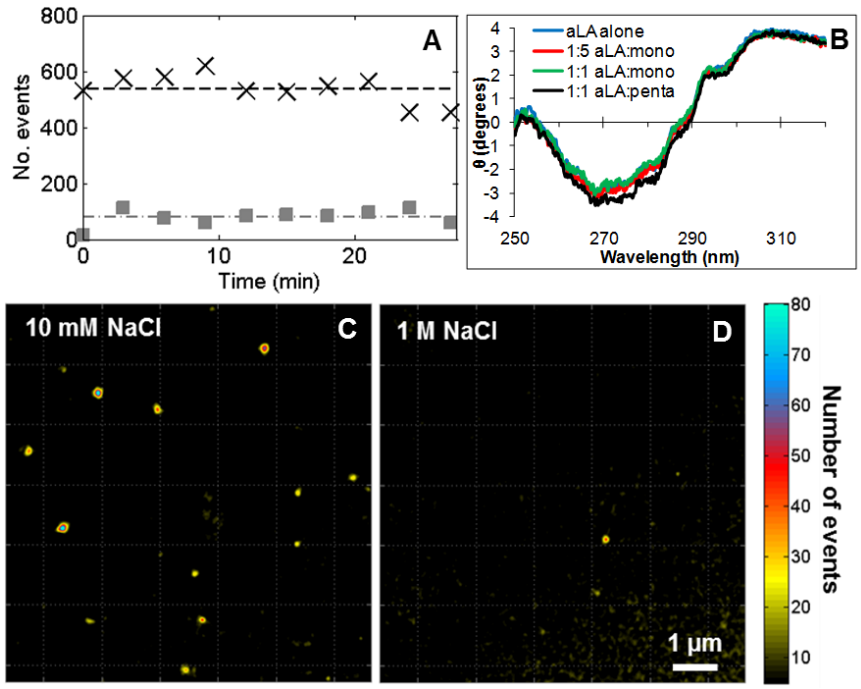


Fig S3. (**A**) Bare agarose control showing numbers of events during 62 s trial over time for two different agarose samples (black x's, grey squares). Dashed lines show average number of events for each sample. (**B**) Circular dichroism spectra of α-lactalbumin in the absence and presence of soluble peptide ligands at different molar ratios listed in the legend. Measured in 10 mM Tris, 10 mM NaCl, pH 8.0. (**C, D**) Superresolution pseudo images of specific adsorption at different electrostatic conditions, (**C**) 10 mM and (**D**) 1 M NaCl.

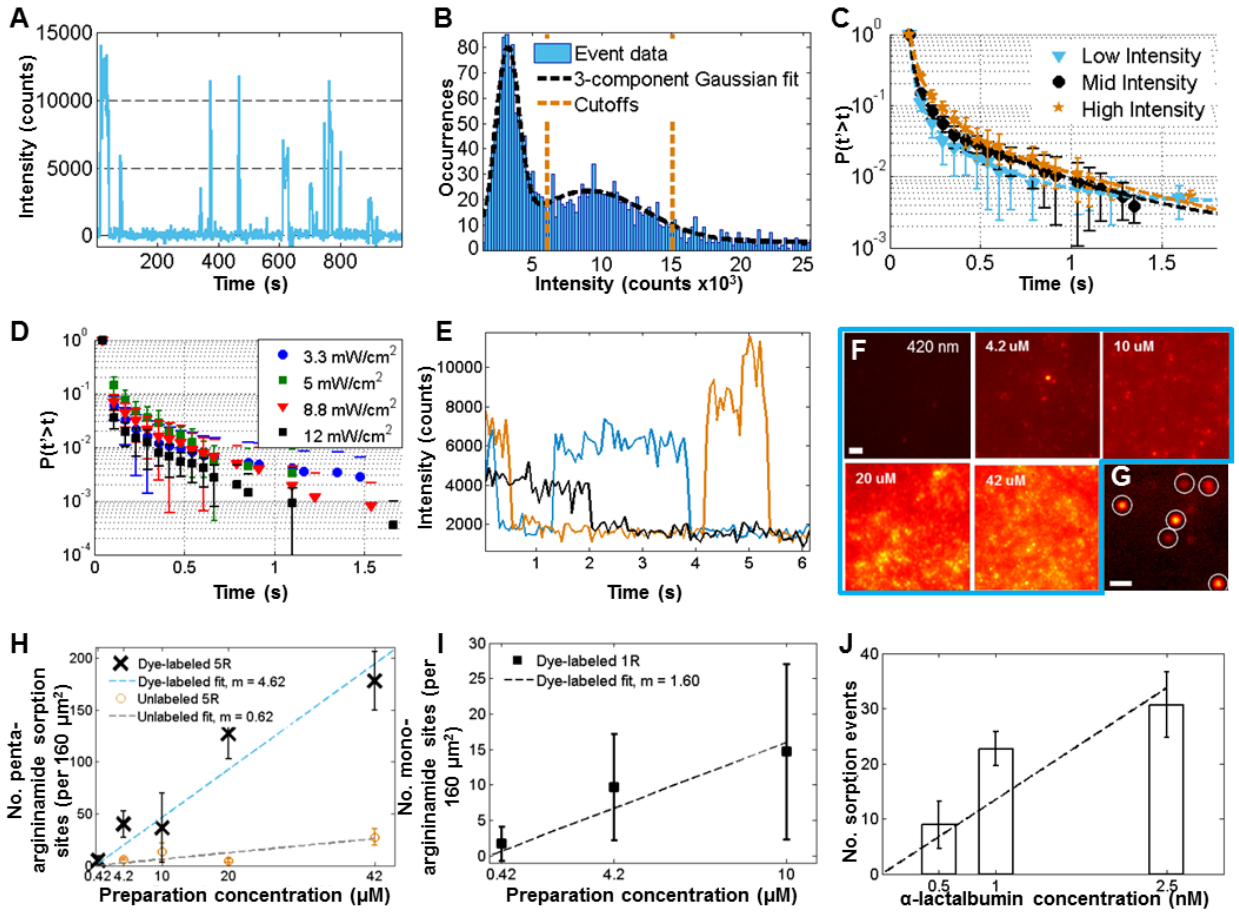


Fig S4. Photophysical controls. (**A-C**) Investigation into protein oligomerization. (**A**) Example intensity trajectory of a single adsorption site exhibiting events at a range of intensities. (**B**) Histogram of adsorption event intensities for all 1,646 events occurring within a single 62 s trial. A three-component Gaussian fit was performed to define the cutoff thresholds for different intensity populations, defined as the midpoint between the Gaussian centroids. (**C**) Resulting desorption time distributions for events separated by intensity. (**D**) Extracted desorption time distributions as a function of excitation intensity. The intensities reported were the incident intensity at the objective. (**E-I**) Dye-labeled peptide identification. (**E**) Example intensity traces from three different BODIPY-labeled pentaargininamide molecules, showing single step photoblinking and bleaching. (**F**, outlined in blue) Summed images (100 frames) of labeled-pentaargininamide at different preparation concentrations. (**G**) Example of identified molecules (white circles) from an image at 10 μM. Scale bars are 640 nm. (**H**) Number of specific adsorption sites at different concentrations identified by dye-labeled pentaargininamide (5R) and α-lactalbumin adsorption to unlabeled pentaargininamide with respective linear fits and slope (*m*) shown. (**I**) Number of monoargininamide (1R) molecules at functionalized agarose interface identified by Cy5-labeled monoargininamide. Number of specific adsorption sites with unlabeled monoargininamide peptide to compare to can be found in the main text, Fig. 2. (**J**) Number of adsorption events as a function of α-lactalbumin concentration demonstrating single molecule conditions. In all plots the resulting fits are plotted as dashed lines. Error bars represent standard deviation.

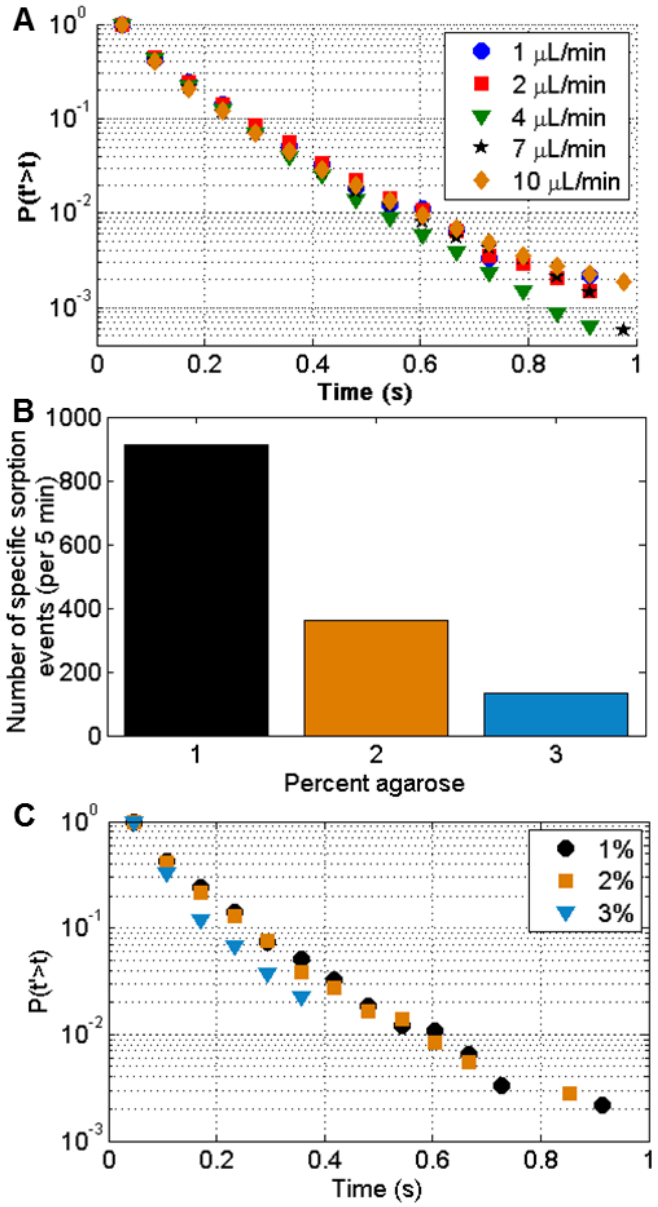


Figure S5. Variation of (**A**) flow rate and (**B, C**) agarose percentage to test possible explanations of kinetic heterogeneity. (**A, C**) Cumulative distributions of desorption times and (**B**) number of specific adsorption events observed.

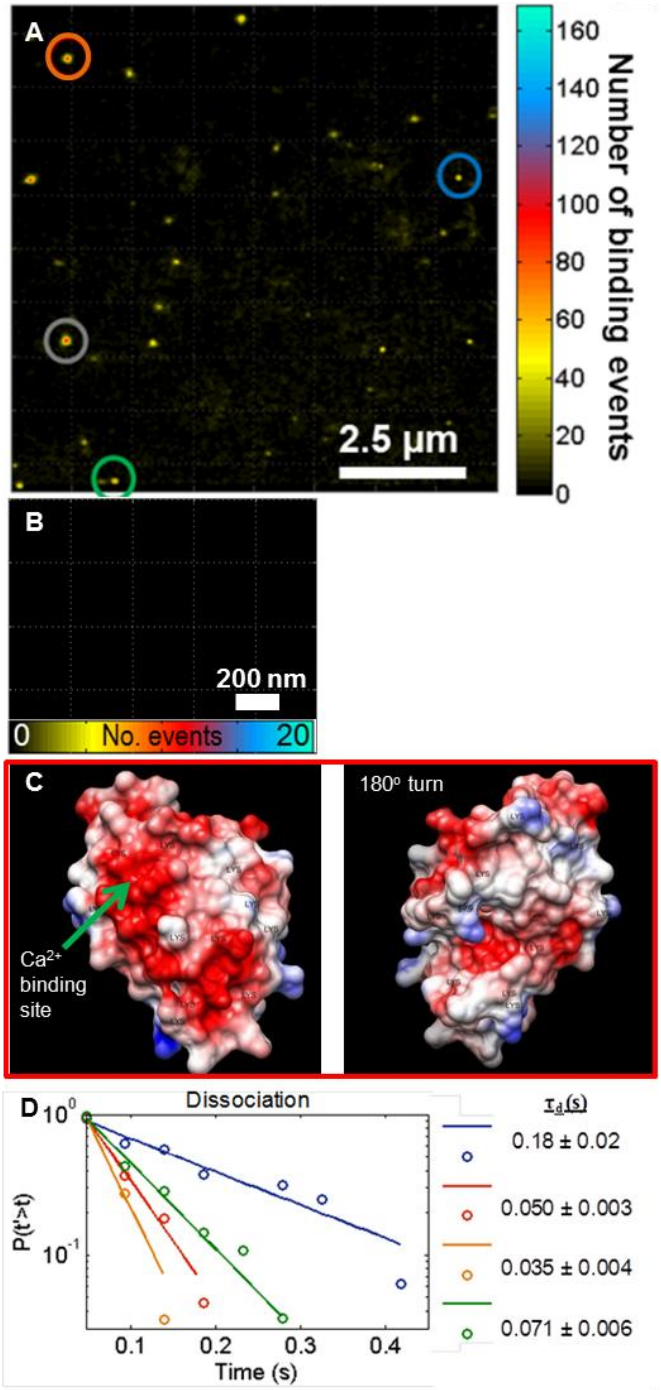


Fig S6. (A) Pseudo superresolution of α-lactalbumin at the pentaargininamide interface with individual ligands used in Fig. 3 noted by colored circles. (**B**) Control pseudo superresolution image of α-lactalbumin at glass interface. (**C**) Electrostatic potential distribution for α-lactalbumin (negatively-charged patches shown in red; positive in blue; UCSF Chimera; structure 1HFX). (**D**) Distribution of τ_d values for four different monoargininamide specific adsorption sites at 1000x loading density.

Table S1. Summary of desorption times of the identified molecules on bare agarose control sample and 215 μM monoargininamide.

Desorption time (frames)	1	2	3	4	5 and above	Average desorption time (frames)
Agarose	6427	30	2	1	2	1.01
Monoargininamide	8007	26	3	4	3	1.01

Table S2. Summary of α-lactalbumin kinetics at individual ligands reported in Figs. 3 and S6.

Color	No. events	k_d (s^{-1})	k_a (s^{-1})
Blue	20	5.6	35.9
Green	32	11.3	14.7
Orange	54	39.0	2.2
Black	58	29.8	3.5

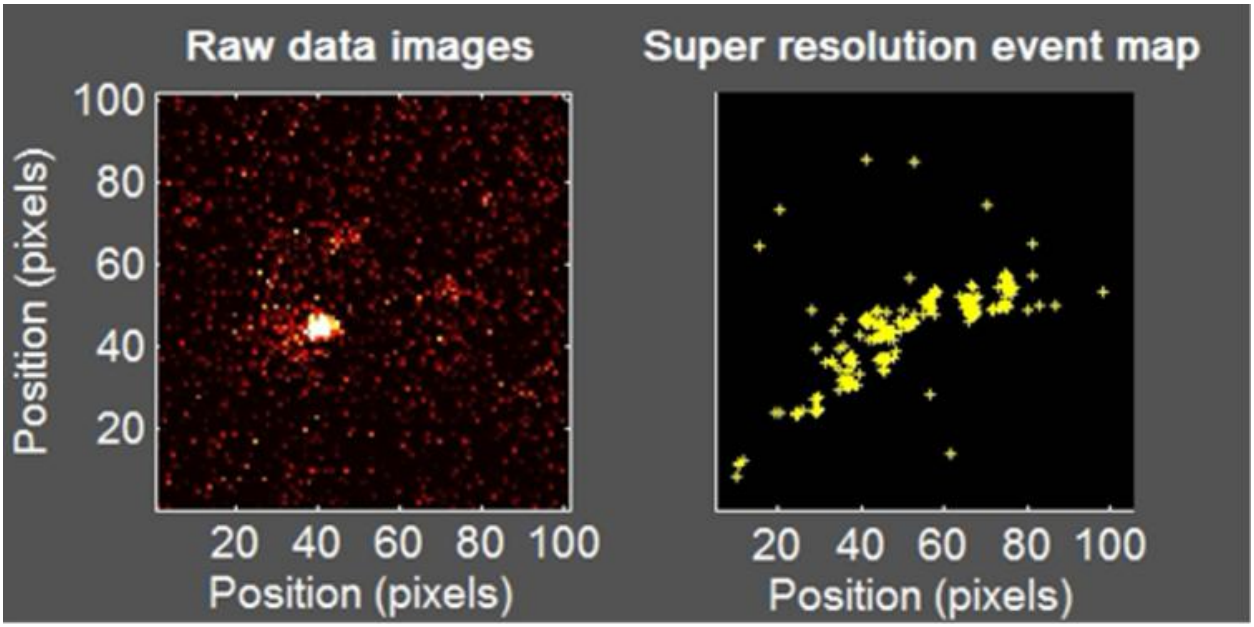
Table S3. Single vs. multi-component* fit to single site kinetics

		1-component	2-component			
	Site	τ (s)	τ1 (s)	τ2 (s)	A1	A2
Dissociation	Blue	0.15	0.15	0.15	NA	
	Green	0.085	0.085	0.085	NA	
	Black	NP				
	Orange	NP				
Association	Blue	NP				
	Green	0.067	0.067	0.067	NA	
	Black	0.33	0.88	0.22	0.25	0.97
	Orange	0.46	0.46	0.46	NA	

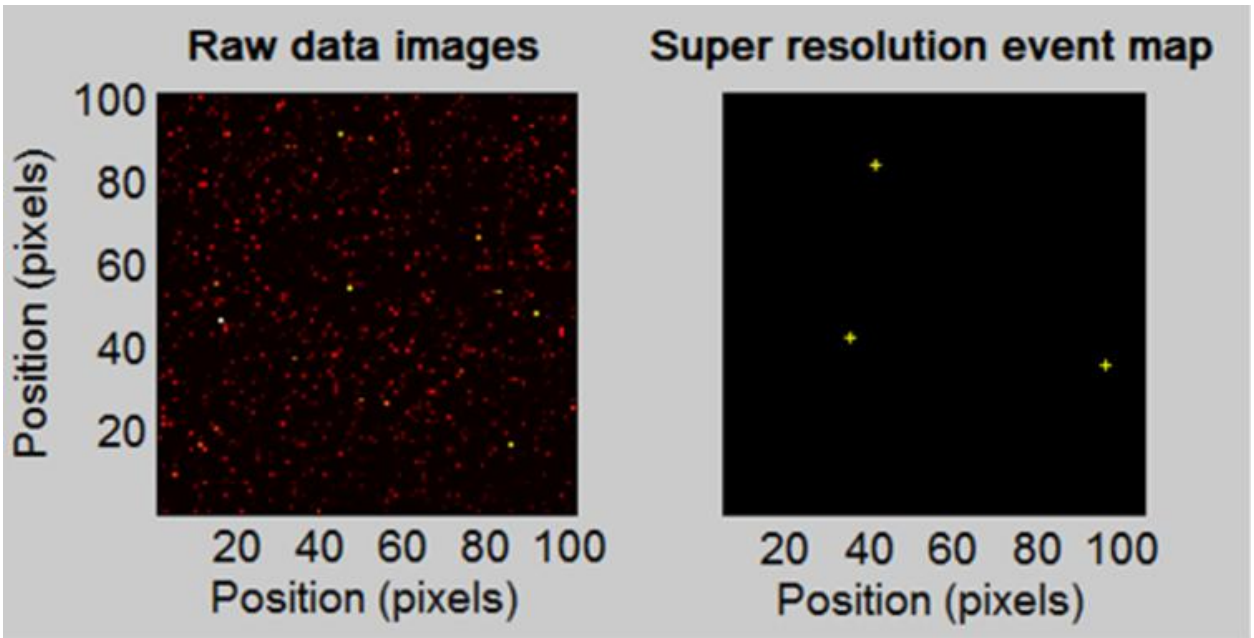
NA - not applicable, since same τ1 and τ2

NP - not possible to try multi-component fit due to number of data points

**See reference17 for considerations on number of components for fitting*



Movie S1. (Left) Real time data of α-lactalbumin interactions at pentaargininamide interface in 10 mM NaCl buffer and (right) the respective superresolution Gaussian centroid positions of spatially mapped events over time. The area is 3.2 x 3.2 μm in size (100 x 100 pixels).



Movie S2. (Left) Real time data of α-lactalbumin interactions at 215 μM monoargininamide interface in 10 mM NaCl buffer and (right) the respective superresolution Gaussian centroid positions of spatially mapped events over time. The field of view is 3.2 x 3.2 μm in size (100 x 100 pixels).