

The geometric determinants of programmed antibody migration and binding on multi-antigen substrates

Ian T. Hoffecker^{1*}, Alan Shaw¹, Viktoria Sorokina¹, Ioanna Smyrlaki¹, and Björn Högberg^{1*}

Dept. of Medical Biochemistry and Biophysics, Karolinska Institutet, S-17177 Stockholm, Sweden

<https://hoffeckerlab.com>

ian.hoffecker@ki.se

<http://hogberglab.net>

bjorn.hogberg@ki.se

Abstract. Viruses and bacteria commonly exhibit spatial repetition of surface molecules that directly interface with the host immune system. However the complex interaction of patterned surfaces with multivalent immune molecules such as immunoglobulins and B-cell receptors is poorly understood, and standard characterization typically emphasizes the monovalent affinity. We developed a mechanistic model of multivalent antibody-antigen interactions as well as a pipeline for constructing such models from a minimal dataset of patterned surface plasmon resonance experiments in which antigen pattern geometries are precisely defined using DNA origami nanostructures. We modeled the change in binding enhancement due to multivalence and spatial tolerance, i.e. the strain-dependent interconversion of bound antibodies from monovalently bound to bivalently bound states at varying antigen separation distances. The parameterized model enables mechanistic *post hoc* characterization of binding behavior in patterned surface plasmon resonance experiments as well as *de novo* simulation of transient dynamics and equilibrium properties of arbitrary pattern geometries. Simulation on lattices shows that antigen spacing is a spatial control parameter that can be tuned to determine antibody residence time and migration speed. We found that gradients in antigen spacing are predicted to drive persistent, directed antibody migration toward favorable spacing. These results indicate that antigen pattern geometry can influence antibody interactions, a phenomenon that could be significant during the coevolution of pathogens and immunity in processes like pathogen neutralization or affinity maturation.

Keywords: avidity · multivalence · biosensor · bivalent · antibody · DNA origami

1 Introduction

Due to their multiple binding domains, immunoglobulin molecules like the bivalent IgG antibody exhibit complex interactions with multivalent antigens, i.e. clusters of multiple copies of molecules or molecular domains occurring on the order of 1-30 nm separation distances. Multivalent interactions enhance the stability of binding interactions by enabling simultaneous attachment of multiple ligands, increasing the statistical ratio of bound to unbound states and extending the residence times of bound antibodies [9, 8].

Many pathogenic surfaces exhibit spatial repetition at length scales relevant to antibody multivalence. Viral capsid proteins undergo self-assembly into periodic patterns [15], and some neutralizing antibodies achieve their high affinity and neutralization capability through bivalence [3, 14]. Self-assembling crystalline arrays of surface-layer (S-layer) proteins, the outermost structure on many bacteria and archaea, are a major contact point between pathogen and host [1] and are implicated as mediators of innate [13] and adaptive immunity [6, 5]. Their repetitive organization may be integral to their immunological role, as their removal from cell surfaces was seen to reduce immune response [12]. Multivalence is also likely an important factor during affinity maturation of antibodies and thus in vaccine design [16, 7].

Antibody interaction with patterned surfaces presents a challenge for mathematical modeling as it is a many-bodied problem occurring on timescales of seconds to minutes. Such systems are too computationally expensive for full-atom molecular simulation. Models treating antibodies and antigens as abstract binding and non-binding units have been the most successful at capturing relevant dynamics, and have historically treated multivalence as a function of ligand coating density whereby multivalence emerges as ligand nearest-neighbor distance statistically decreases with higher densities [10]. More recently, coarse-grained molecular

simulations have been used fruitfully to investigate the role of antibody multivalence on binding kinetics [2]. However a challenge of precisely calibrating such models remains due to the absence of both experimental tools to independently assess monovalent and multivalent binding dynamics as well as a computational pipeline to connect such data to models.

The patterned surface plasmon resonance (PSPR) technique (Fig 1 a) enables measurement of binding kinetics on precise, monodisperse patterns of ligands, achieving robust control of geometry through the use of DNA origami nanostructures [11]. Herein, we demonstrate a pipeline for the automated conversion of PSPR data into a flexible, experimentally parameterized model of antibody interaction with arbitrarily complex multivalent surfaces. The model is based on a coarse-grained simplification of bivalent antibody binding to antigens as a discrete Markov process with distinct states: empty antigen, monovalent antibody-antigen complexes, and bivalent antibody-antigen complexes with transitions between these states governed by elementary rates (Fig 1 b). From this basis, the dynamics of more complex patterns of multiple antigens can be reduced (Fig 1 c) to combinations of these elementary states. We show how the spatial tolerance, i.e. the range and impact of antigen separation distances on bivalent binding kinetics (Fig 1 d), can be exploited in the design of antigen patterns to modulate the effective binding affinity, the walking speed of antibody migration, and even the direction of antibody migration on patterned surfaces. A causal linkage between pattern geometry and antibody dynamics could play a role in the co-evolution of adaptive immunity and pathogens.

2 Results

We developed a model parameterization pipeline based on progressive fitting of the transient SPR profiles for first monovalent and then bivalent binding processes in order to reduce degrees of freedom at each stage of fitting. In the first 1-antigen configuration stage, we used an anti-digoxigenin single-cycle-kinetics program whereby progressively higher concentrations of antibody were exposed to the immobilized antigen substrate (Fig 2 a). This program was performed with a 1-antigen configuration (Fig 2 b, d) in order to parameterize our Markov model yielding respective association and dissociation rates $k_1 = 2.1 \times 10^7 M^{-1} s^{-1}$ and $k_{-1} = 5.3 \times 10^{-4} s^{-1}$ as well as a monovalent dissociation constant $K_{D1} = 2.5 \times 10^{-11} M$. Parameterizing interconversion between monovalent and bivalent states was done by fixing the previously determined monovalent parameters and fitting the model to experiments involving multiple adjacent antigens via adjustment of K_{D2} or the interconversion constant defined by the ratio of reverse and forward interconversion rates. For structures configured with 2 antigens separated by $\approx 14 nm$, we find $K_{D2} = 5.8 \times 10^{-2}$ (Fig 2 c, e).

Applying progressive fitting to PSPR runs with structures patterned with 2 adjacent antigens of varied separation distances, we found the internal conversion process to vary accordingly. Small and large separation distances correspond to reduced bivalence, i.e. larger K_{D2} . We constructed a phenomenological equation modeling the interconversion constant (Fig 2 f). The model is composed of a logistic tension-term representing the reduced bivalence at large separation distances and an exponential compression-term representing the penalty to bivalence observed at extremely close separation distances. The interconversion constant is thus a function of adjacent antigen separation distance with the form

$$K_{D2} = \frac{K_{D2}^{max}}{1 + e^{-\alpha_t(x-\ell_t)}} + K_{D2}^{max} e^{-\alpha_c(x-\ell_c)}, \quad (1)$$

where ℓ_t and ℓ_c are characteristic lengths defining the scale of the tensile and compressive terms, α_t is the sharpness of the tensile penalty, α_c is the decay parameter of compressive penalty to bivalence, and K_{D2}^{max} is the value of K_{D2} at which contributions of bivalence to binding dynamics is vanishingly small.

In order to determine the dependence of system bivalence on solution phase concentration, we used the parameterized model to obtain the steady state probability distributions for a range of solution phase concentrations (Fig 3 a). This revealed concentration regimes of differing dominant states: empty, bivalent, and saturated monovalent at respectively low, medium, and high solution phase concentrations. Entropic maxima occur at transitions between these domains (Fig 3 b), and transitions in bivalent and monovalent contributions to chemical potential occur in accordance with the transition from bivalent to saturated

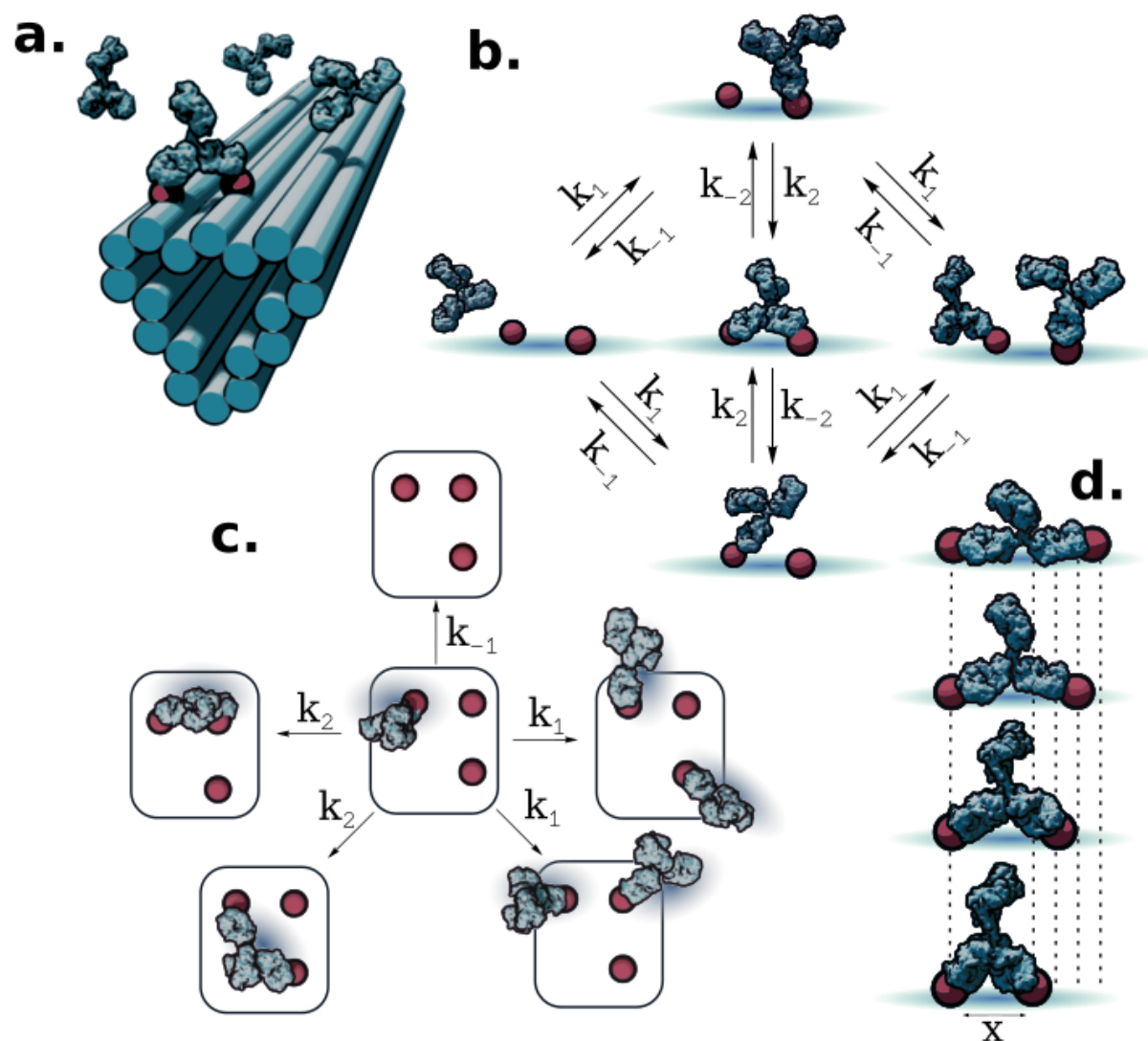


Fig. 1. Scheme for modeling binding dynamics of antibodies on multi-antigen substrates. a: Illustration of patterning concept, whereby small molecule antigens (haptens) are arranged using short, flexible tethers at well-defined locations on DNA origami nanostructures. This enables multivalent interaction of antibodies with antigen patterns. b: Markov model of antibody binding whereby only basic binding/unbinding and bivalent interconversion processes are used to couple discrete monovalent and bivalent binding states. c: Model extension to more complex pattern geometries is accomplished by separating the system into elementary transitions between states comprising different combinations of empty and monovalently or bivalently occupied antigens. d: Pairs of antigens separated by different lengths elicit differing antibody binding kinetics due to the separation distance dependent impact of antibody structure on the chance of bivalent interconversion.

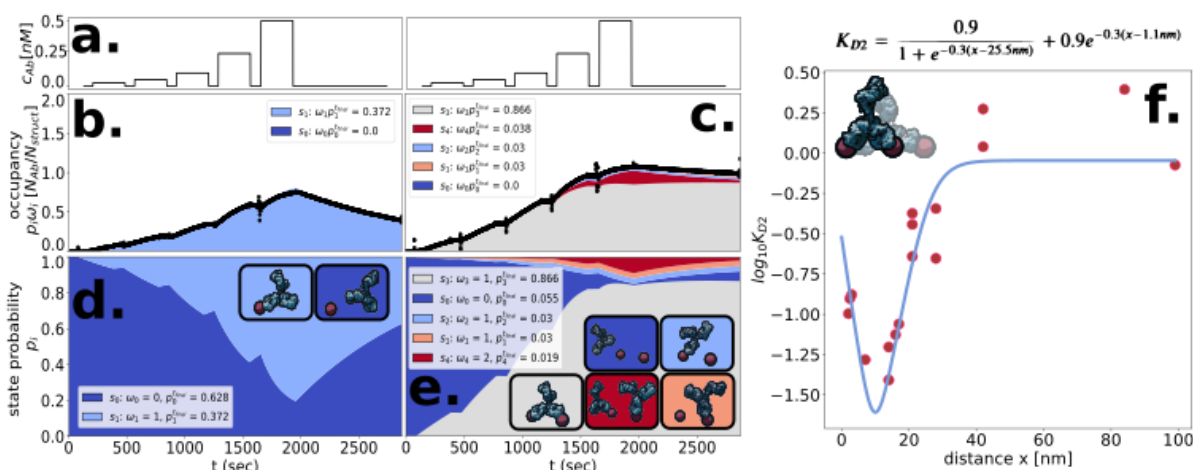


Fig. 2. A progressive fitting pipeline for obtaining a parameterized model from minimal experimental data. a: Concentration versus time plot of antibody solution exposed to patterned antigen substrates in a single cycle kinetics PSPR experiment. b and c: Experimental binding kinetics data (black line) of 1 antigen and 2 antigen configurations respectively, superimposed over the occupancy calculated from the parameterized model. Model occupancy is divided and colored according to state, with height corresponding to the state's contribution to total antibody occupancy per structure. d and e: 1 and 2 antigen configuration respective transient state probabilities stratified according to model prediction, colored and stacked to satisfy the normalization condition whereby all probabilities add to 1. f: Interconversion constants (magenta points) plotted versus 2-antigen configuration separation distance x and the fitted spatial tolerance model (blue line).

monovalent regimes (Fig 3 c). In addition to simulating de novo steady state properties, the model enables us to simulate transient dynamics of hypothetical systems with arbitrary geometries and arbitrary timing in the introduction of different solution phase concentrations. To demonstrate this, we simulated the evolution of state probability for a system with antigens arrayed in a $12 \times 16 \times 20$ nm right triangle (Fig 3 d) with arbitrarily timed introduction of solution phase concentrations: 5×10^{-4} , 6×10^{-2} , 5×10^{-1} , 1 nM at 0, 1000, 2000, and 3000 seconds respectively. This allowed us to track the evolution of entropy (Fig 3 e) as the system approaches steady state, revealing differences in both the total and individual values among states under the influence of different solution concentrations and time evolution. We also saw the time evolution of the number of antibodies bound bivalently or monovalently (Fig 3 f), whose relative values become inverted when a change in concentration triggered a transition from a bivalently dominated to a monovalently dominated regime.

For larger systems, complete enumeration of states scales poorly with larger numbers of adjacent antigens. We developed a Markov Chain Monte Carlo implementation of the model to sample trajectories that converges to state probabilities with large sample numbers. Rather than enumerating all system states (i.e. combinations of antibodies and binding modes on a structure and possible transitions), the system performs a random walk through the large state space, computing at any point in time its rate of escape into neighboring states. We then examined collections of individual trajectories for such systems to understand their average behavior. Specifically, we examined the role of repetitive antigen spacing in simple 1D arrays. Antigens arranged according to a gradient in spacing spanning that of the region of steepest slope in Equation 1 in the range of 10 - 22 nm separation distances elicit asymmetric accumulation toward the narrow spaced end of the array (Fig 4 a), individual walking trajectories that tend toward the narrow-spaced end (Fig 4 b), asymmetric velocity (Fig 4 c), and asymmetric net displacement (Fig 4 d) according to the direction of the gradient. Antibodies binding to 1D arrays of uniform spacing exhibited divergent residence times, with antibodies spending less cumulative time on 22 nm spaced arrays (Fig 4 e) than those of narrow 10 nm spaced arrays (Fig 4 f). A comparison of net displacement also shows that antibodies

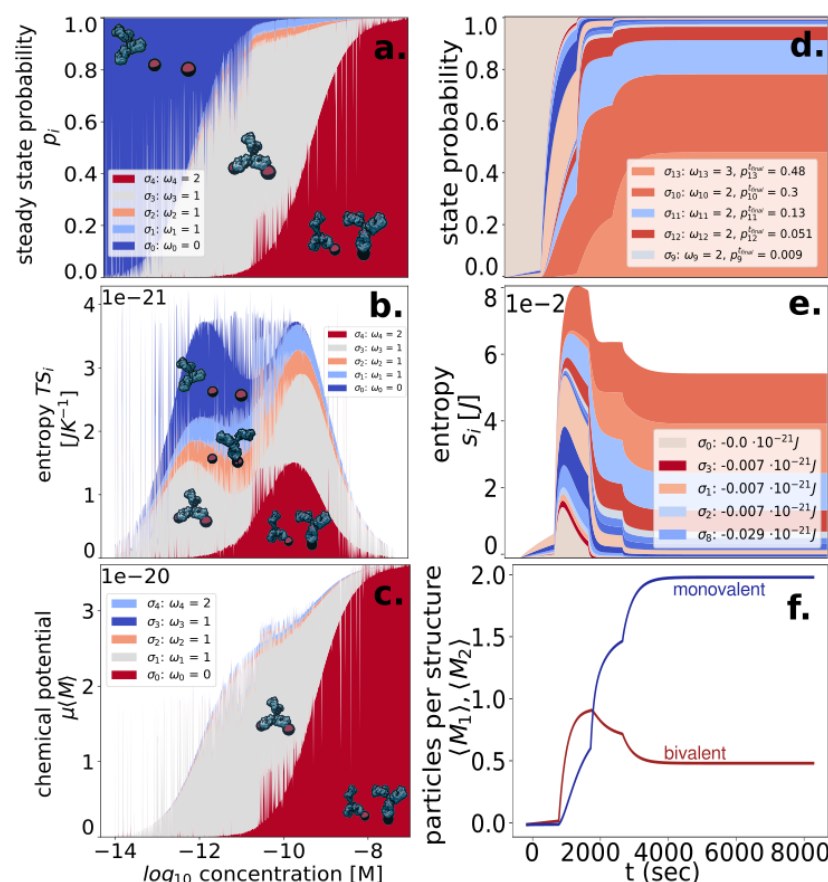


Fig. 3. *De novo* simulation with parameterized kinetics. a: Stationary distributions colored by state of a 2-antigen system (14 nm separation) for a range of solution phase antibody concentrations demonstrating clear regions of predominantly empty, bivalent single antibody occupancy, and monovalent two antibody occupancy regimes connected by smooth transition regions. b: Distribution of state entropic contributions to free energy for a range of solution phase antibody concentrations. c: Chemical potential contributions at equilibrium from each state for a range of concentrations. (legend shows top 5 most abundant states) d: Transient state probability distribution for a simulation of a trimeric $12 \times 16 \times 20$ nm antigen configuration with step changes in concentration. (legend shows top 5 most abundant states) e: The evolution of entropy for the trimeric system stratified by state. f: The transient evolution of monovalent and bivalent contributions of antibodies to the average number of antibodies per structure in the trimeric system. The cross between the two lines demonstrates transition between regimes by changing the concentration.

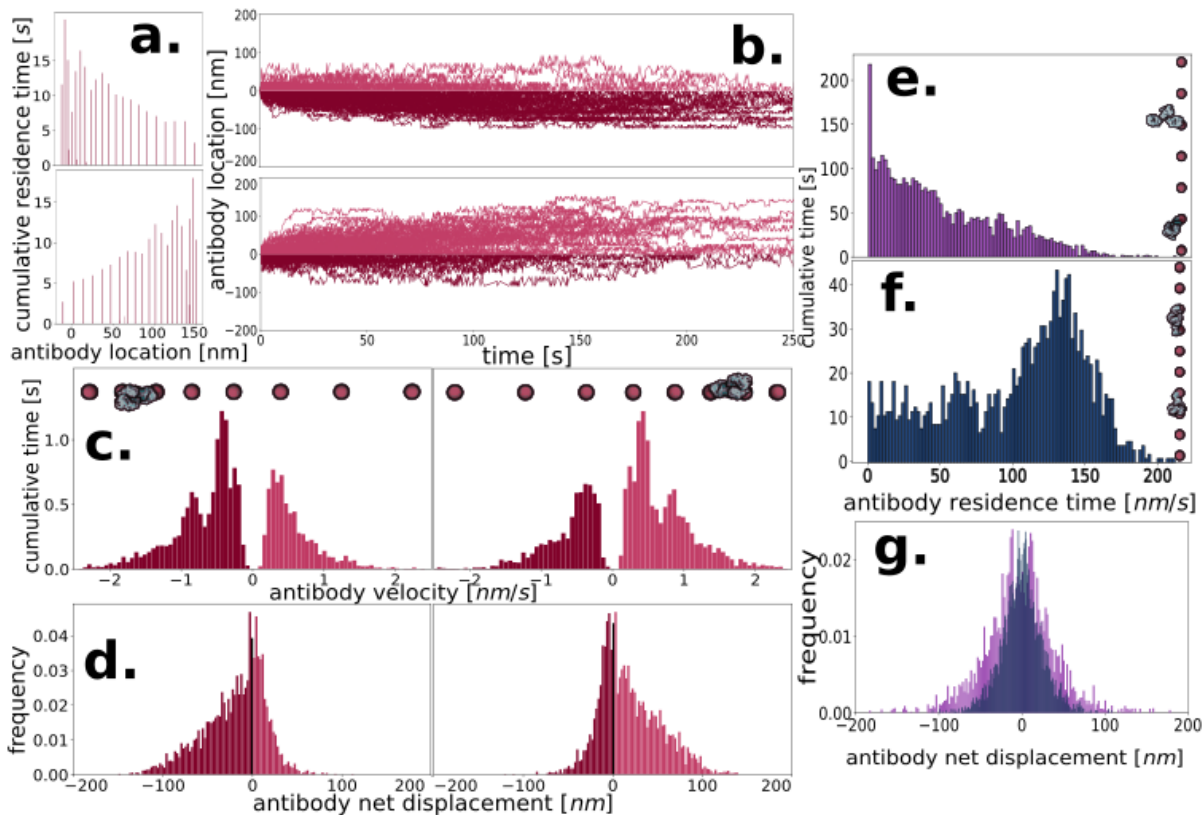


Fig. 4. Manipulation of antibody movement through choice of antigen pattern geometry. a: Cumulative antibody residence times as a function of antibody location on 1D antigen gradients oriented with increasing spacing (top) and decreasing spacing (bottom) b: Random walk trajectories of antibodies tracked from their initial landing locations on 1D antigen gradients with increasing (top) and decreasing (bottom) spacing gradients. c: Histograms of antibody velocity on 1D antigen gradients of increasing (left) and decreasing (right) spacing distance. d: Net displacement of antibodies tracked from their initial binding location on 1D antigen gradients of increasing (left) and decreasing (right) spacing gradients. e: Histogram of antibody residence times on uniform 1D antigen array with wide 22 nm spacing. f: Histogram of antibody residence times on uniform 1D antigen array with narrow 10 nm spacing. g: Histogram of antibody displacements tracked from their initial binding locations for uniform 1D antigen arrays with wide (magenta) and narrow (blue) spacings.

moved further from their initial binding location on widely spaced arrays relative to narrowly spaced (Fig 4 g).

3 Conclusion

In summary, we developed a pipeline for converting a minimal dataset from 2 PSPR experiments into a flexible stochastic model of antibody binding and walking. The model enabled us to determine steady regimes of bivalent and monovalent-dominated states depending on the solution phase concentration. We developed an enhancement to the model in order to simulate arbitrary geometries with differing separation distances between antigens by determining the binding behavior as a function of separation distance from PSPR experiments with varied spacings. This enabled us to develop an analytical expression for the change in interconversion rate as a function of antigen separation that could be applied to arbitrary geometries. The model shows that antigen patterns can influence the global movements of antibodies, with gradients in antigen spacing leading to directed migration as well as residence time control through change in uniform spacing of patterns. Taken together, this work reveals the importance of antigen pattern geometry as a control parameter in the dynamics of antibody binding and migration on multi-antigen substrates and provides a framework for modeling and exploring such phenomena using PSPR-parameterized Markov models.

References

1. Boot, H.J., Pouwels, P.H.: Expression, secretion and antigenic variation of bacterial s-layer proteins. *Molecular microbiology* **21**(6), 1117–1123 (1996)
2. De Michele, C., De Los Rios, P., Foffi, G., Piazza, F.: Simulation and theory of antibody binding to crowded antigen-covered surfaces. *PLoS computational biology* **12**(3), e1004752 (2016)
3. Emimi, E., Ostapchuk, P., Wimmer, E.: Bivalent attachment of antibody onto poliovirus leads to conformational alteration and neutralization. *Journal of Virology* **48**(2), 547–550 (1983)
4. Fox, B.L., Glynn, P.W.: Computing poisson probabilities. *Communications of the ACM* **31**(4), 440–445 (1988)
5. Kajikawa, A., Zhang, L., LaVoy, A., Bumgardner, S., Klaenhammer, T.R., Dean, G.A.: Mucosal immunogenicity of genetically modified lactobacillus acidophilus expressing an hiv-1 epitope within the surface layer protein. *PloS one* **10**(10), e0141713 (2015)
6. Konstantinov, S.R., Smidt, H., de Vos, W.M., Bruijns, S.C., Singh, S.K., Valence, F., Molle, D., Lortal, S., Altermann, E., Klaenhammer, T.R., et al.: S layer protein a of lactobacillus acidophilus ncfm regulates immature dendritic cell and t cell functions. *Proceedings of the National Academy of Sciences* **105**(49), 19474–19479 (2008)
7. LoBue, A.D., Lindesmith, L., Yount, B., Harrington, P.R., Thompson, J.M., Johnston, R.E., Moe, C.L., Baric, R.S.: Multivalent norovirus vaccines induce strong mucosal and systemic blocking antibodies against multiple strains. *Vaccine* **24**(24), 5220–5234 (2006)
8. Macken, C.A., Perelson, A.S.: Aggregation of cell surface receptors by multivalent ligands. *Journal of Mathematical Biology* **14**(3), 365–370 (1982)
9. Martinez-Veracoechea, F.J., Leunissen, M.E.: The entropic impact of tethering, multivalency and dynamic recruitment in systems with specific binding groups. *Soft Matter* **9**(12), 3213–3219 (2013)
10. Müller, K.M., Arndt, K.M., Plückthun, A.: Model and simulation of multivalent binding to fixed ligands. *Analytical biochemistry* **261**(2), 149–158 (1998)
11. Shaw, A., Hoeffcker, I.T., Smyrlaki, I., Rosa, J., Grevys, A., Bratlie, D., Sandlie, I., Michaelsen, T.E., Andersen, J.T., Högberg, B.: Binding to nanopatterned antigens is dominated by the spatial tolerance of antibodies. *Nature nanotechnology* **14**(2), 184–190 (2019)
12. Suzuki, S., Yokota, K., Igimi, S., Kajikawa, A.: Comparative analysis of immunological properties of s-layer proteins isolated from lactobacillus strains. *Microbiology* **165**(2), 188–196 (2019)
13. Taverniti, V., Stuknyte, M., Minuzzo, M., Arioli, S., De Noni, I., Scabiosi, C., Cordova, Z.M., Junttila, I., Hämäläinen, S., Turpeinen, H., et al.: S-layer protein mediates the stimulatory effect of lactobacillus helveticus mihlh5 on innate immunity. *Applied and environmental microbiology* **79**(4), 1221–1231 (2013)
14. Thouvenin, E., Laurent, S., Madelaine, M.F., Rasschaert, D., Vautherot, J.F., Hewat, E.A.: Bivalent binding of a neutralising antibody to a calicivirus involves the torsional flexibility of the antibody hinge. *Journal of molecular biology* **270**(2), 238–246 (1997)

15. Uetrecht, C., Barbu, I.M., Shoemaker, G.K., Van Duijn, E., Heck, A.J.: Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nature chemistry* **3**(2), 126 (2011)
16. Veneziano, R., Moyer, T.J., Stone, M.B., Wamhoff, E.C., Read, B.J., Mukherjee, S., Shepherd, T.R., Das, J., Schief, W.R., Irvine, D.J., et al.: Role of nanoscale antigen organization on b-cell activation probed using dna origami. *Nature Nanotechnology* pp. 1–8 (2020)

4 Supplementary Data

4.1 List of model assumptions, constraints

Assume a fixed amount of bound structures that does not change with time. i.e.

$$\frac{dR_{struct}}{dt} \approx 0 \quad (2)$$

and

$$\frac{dn_{struct}}{dt} \approx 0 \quad (3)$$

The system has an IgG reservoir that is large compared to the available binding surface and thus has an effectively fixed concentration, i.e.

$$\frac{dc_{Ab}}{dt} \approx 0 \quad (4)$$

4.2 Conversion from SPR signal R_{ab} to bound antibody n_{ab}

Consider first the simple 1-1 interaction of an antibody analyte that binds and unbinds to a structure containing a single antigen ligand.

$$\sigma_- \frac{c_{Ab}k_1}{k_{-1}} \sigma_{Ab} \quad (5)$$

We may work in terms of molar quantities rather than concentrations or surface densities, as the dimensions of the system do not change

$$n_{struct} = n_- + n_{Ab}, \quad (6)$$

or the molar amount of structures both occupied and unoccupied [mol], where n_{Ab} is the number of bound antibody-structure complexes and n_- is the number of unoccupied structures and both where n_{Ab} and n_- are functions of t .

State probabilities are therefore:

$$p_- = \frac{n_-}{n_{struct}} \quad (7)$$

and

$$p_{Ab} = \frac{n_{Ab}}{n_{struct}} \quad (8)$$

We define also an occupancy, the number of antibodies that are bound to a single structure for a given state. For simple, 1 antigen structures, this value is zero for the empty state and 1 for the bound state or

$$\omega_- = 0 \quad (9)$$

and

$$\omega_{Ab} = 1 \quad (10)$$

The geometric determinants of programmed antibody migration and binding on multi-antigen substrates 9

The average occupancy is a macroscopic description of the state of the system comprising N states, or the average fraction of bound antibodies per structure.

$$\Omega = \sum_{i=1}^N p_i \omega_i \quad (11)$$

For the case of a 1 antigen structure system, this becomes

$$\Omega = p_{Ab} \cdot 1 + p_- \cdot 0 = \frac{n_{Ab}}{n_{struct}} \quad (12)$$

In a 1-1 binding model, change in SPR signal is proportional to the amount of bound material or in other words the change in molar amount of structures with occupied binding sites n_{Ab} .

$$R_{Ab} = n_{Ab} \xi_{Ab} \quad (13)$$

The rate of change of occupied sites is equal to the rate of conversion of unoccupied sites via binding events minus the rate of conversion of occupied sites via unbinding events.

$$\frac{dn_{Ab}}{dt} = k_1 c_{Ab} n_- - k_{-1} n_{Ab} \quad (14)$$

The SPR signal after the structure binding step is proportional to the molar of amount of bound structure

$$R_{struct} = n_{struct} \xi_{struct} \quad (15)$$

Substitute RU-based expressions of molar amounts into Equation 6:

$$n_- = \frac{R_{struct}}{\xi_{struct}} - \frac{R_{Ab}}{\xi_{Ab}} \quad (16)$$

Substituting RU-based expressions of molar amounts into Equation 14 yields

$$\frac{dn_{Ab}}{dt} = \frac{1}{\xi_{Ab}} \frac{dR_{Ab}}{dt} = k_1 c_{Ab} \left(\frac{R_{struct}}{\xi_{struct}} - \frac{R_{Ab}}{\xi_{Ab}} \right) - k_{-1} \left(\frac{R_{Ab}}{\xi_{Ab}} \right) \quad (17)$$

which simplifies to

$$\frac{dR_{Ab}}{dt} = \frac{\xi_{Ab} k_1 c_{Ab} R_{struct}}{\xi_{struct}} - k_1 c_{Ab} R_{Ab} - k_{-1} R_{Ab} \quad (18)$$

Since we have gathered both conversion constants into one term in Equation 18, we define now the occupancy signal factor

$$\xi_* \equiv \frac{\xi_{Ab}}{\xi_{struct}} \quad (19)$$

or the dimensionless ratio of molar conversion factors: bound-antibody relative to structure.

Note by rearrangement the relationship to average occupancy - i.e. the occupancy signal factor is the ratio of occupancy in terms of SPR signal to that of molar quantities.

$$\xi_* = \frac{R_{Ab}}{n_{Ab}} \frac{n_{struct}}{R_{struct}} = \frac{R_{Ab}}{R_{struct}} \left(\frac{n_{Ab}}{n_{struct}} \right)^{-1} \quad (20)$$

$$= \frac{R_{Ab}}{R_{struct}} \Omega^{-1} \quad (21)$$

Substituting ξ_* we then arrive at the expression for the rate of change in SPR signal with respect to time as a function of structure-binding signal and antibody-binding signal:

$$\frac{dR_{Ab}}{dt} = \xi_* k_1 c_{Ab} R_{struct} - k_1 c_{Ab} R_{Ab} - k_{-1} R_{Ab} \quad (22)$$

In the case of a monovalent structure (1 antigen available for binding) at the point of maximal saturation when average occupancy is unitary ($\Omega = 1$), the molar quantities of bound antibody and structures are equal:

$$n_{Ab}^{max} = n_{struct} \quad (23)$$

Under maximal saturation conditions, the monovalent occupancy signal factor then reduces to

$$\xi_* = \frac{R_{Ab}^{max}}{R_{struct}} \cdot 1 \quad (24)$$

This relationship is then used to produce a standard curve from monovalent structure binding data in order to obtain the linear relationship:

$$R_{Ab}^{max} = R_{struct} \xi_* \quad (25)$$

where an empirically determined ξ_* enables one to estimate the SPR signal corresponding to an occupancy of 1 antibody per structure from the R_{struct} signal. This is useful for structures with valency greater than 1 and whose binding kinetics do not obey simple 1-1 equations. Since R_{Ab}^{max} on a multivalent structure will not resemble that of the monovalent 1-1 system, we refer to this conversion factor obtained from the monovalent R_{Ab}^{max} as R_{Ab}^{mono} , i.e. an SPR signal to antibody number conversion factor:

$$R_{Ab}^{mono} \approx R_{Ab}^{max}, V_{struct} = 1, \quad (26)$$

where V_{struct} is the valency.

In such cases, we obtain the average occupancy using the estimated R_{Ab}^{max} from the linear regression.

$$\Omega = f(R_{Ab}) = \frac{n_{Ab}}{n_{struct}} = \frac{R_{Ab}}{R_{Ab}^{max}} \quad (27)$$

4.3 Empirical estimation of R_{Ab}^{max}

Thus, we obtain a standard curve used to convert the SPR signals for arbitrary structure configurations by empirically determining the correlation between structure binding signals and the max signals corresponding to saturated monovalent (1 antigen) structures, enabling conversion from SPR signal to occupancy in the absence of a well-understood binding model provided knowledge of the structure binding signal.

The structure bound signal (Fig 4.3 a) is taken to be the difference between signals before and after structures are flowed over the chip and allowed to bind.

Guess values of the parameters k_1 , k_{-1} , and ξ_* are supplied to a numerical minimization of the autocorrelation of residuals between experimental and theoretical curves for the 4th order Runge Kutta approximation of Equation 22, i.e. the function $\frac{dR_{Ab}}{dt} = f(R_{Ab})$ recursively approximated according to the formula

$$R_{Ab,t+1} = R_{Ab,t} + \frac{h}{6} (\kappa_{n1} + 2\kappa_{n2} + 2\kappa_{n3} + \kappa_{n4}) \quad (28)$$

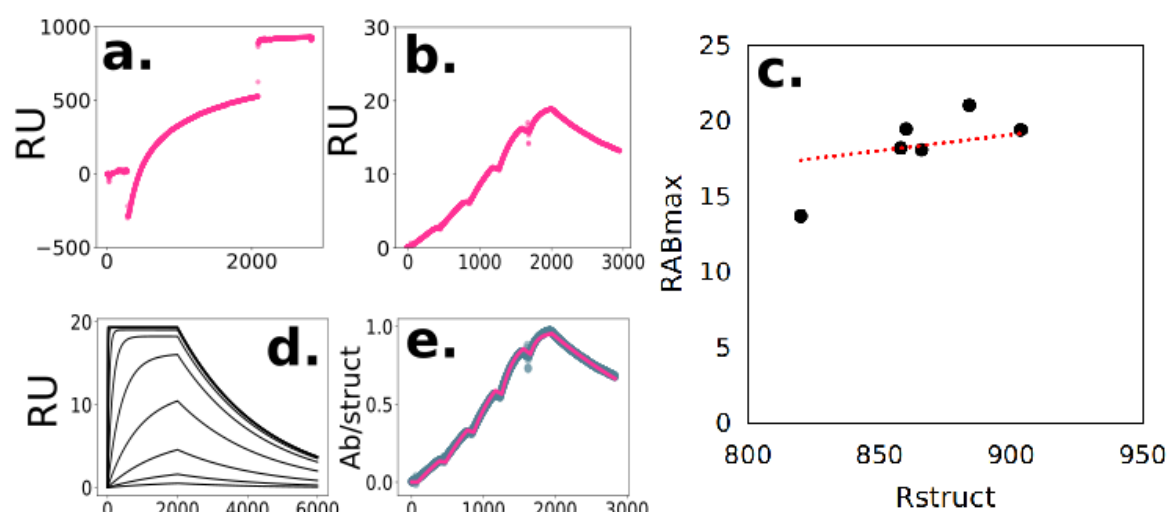
where h is a small timestep and the constituent terms have the form

$$\kappa_{n1} = f(R_{Ab,t}) \quad (29)$$

$$\kappa_{n2} = f(R_{Ab,t} + \frac{h}{2}\kappa_{n1}) \quad (30)$$

$$\kappa_{n3} = f(R_{Ab,t} + \frac{h}{2}\kappa_{n2}) \quad (31)$$

$$\kappa_{n4} = f(R_{Ab,t} + h\kappa_{n3}) \quad (32)$$



Supplementary Figure 1. a: Sensorgram of DNA origami nanostructure binding to chip in SPR response units (RU). b: Sensorgram of antibodies binding to a 1 antigen structure proceeding the structure-binding phase in (a) shown in SPR response units (RU). c: Calculated monovalent R_{Ab}^{max} representing the number of RU corresponding to 1 antibody bound to every structure for various instances of monovalent ODE model fits to experimental binding data plotted against their corresponding structure binding signal R_{Struct} and the linear fit used to estimate the proportionality of the two variables. d: Demonstration of ODE model's ability to predict R_{Ab}^{max} , a point of saturation that the monovalent system converges to as higher concentrations push more structures into the monovalently-bound state. e: Experimental data (blue) and model (magenta) after normalization by R_{Ab}^{max} , with the more useful units of antibodies per structure (AB/struct) instead of RU.

For each monovalent run (Fig 4.3 b) with a unique value of R_{struct} , a projected value of R_{Ab}^{max} is computed with Equation 25. This enables us to make a standard curve to adjust R_{Ab}^{mono} according to R_{struct} in the absence of a 1-1 R_{Ab}^{max} (Fig 4.3 c). This is possible because parameterization of the monovalent models by determining their rate constants enables computational prediction of R_{Ab}^{max} (Fig 4.3 d) in the absence of experimental saturation. We use this value as a conversion factor, enabling us to convert SPR response units into the number of antibodies per structure (Fig 4.3 e). By knowing R_{struct} , we can estimate this conversion factor for non-trivial antigen configurations where the multivalence influences the ease of reaching a saturation value corresponding to R_{Ab}^{max} .

4.4 Equilibrium characterization with dissociation constants

The equilibrium dissociation constant concisely describes the relationship between analyte and ligand, and provides a good basis for comparison between systems across experimental conditions in which dynamic behavior can vary significantly. Given a model of the process, we can derive a formula for the equilibrium dissociation constant by solving the system of equations. For a 1-1 process we have

At steady state:

$$\frac{dR_{Ab}}{dt} = 0 \quad (33)$$

$$k_1 c_{Ab} R_{Ab}^{max} = k_1 c_{Ab} R_{Ab^{eq}} + k_{-1} R_{Ab^{eq}} \quad (34)$$

rearranging yields

$$k_1 c_{Ab} R_{Ab}^{max} = (k_1 c_{Ab} + k_{-1}) R_{Ab^{eq}} \quad (35)$$

and

$$R_{Ab^{eq}} \left(\frac{k_1}{k_{-1}} c_{Ab} + 1 \right) = \frac{k_1}{k_{-1}} c_{Ab} R_{Ab}^{max} \quad (36)$$

For a 1-1 monovalent model, the dissociation constant is

$$K_D = \frac{k_{-1}}{k_1} \quad (37)$$

thus, at equilibrium, the SPR signal is

$$R_{Ab^{eq}} = \frac{c_{Ab} R_{Ab}^{max}}{K_D + c_{Ab}} \quad (38)$$

Empirical measurement of the dissociation constant is obtained by determining the equilibrium binding signals at multiple concentrations and fitting the linearized form of Equation 38 or

$$\frac{1}{R_{Ab^{eq}}} = \frac{K_D}{c_{Ab} R_{Ab}^{max}} + \frac{1}{R_{Ab}^{max}} \quad (39)$$

The equilibrium dissociation constant is a good descriptive parameter which captures the essential dynamics concisely.

From the dissociation constant, we know the occupancy: and the corresponding occupancy is

$$\Omega_{eq} = \frac{R_{Ab^{eq}}}{R_{Ab}^{max}} = \frac{c_{Ab}}{K_D + c_{Ab}} \quad (40)$$

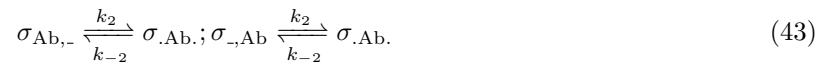
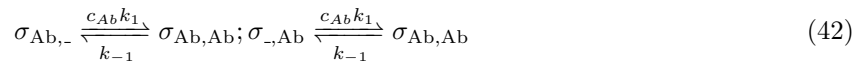
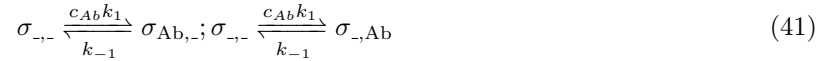
Such a concise description is desirable for complex structures as well. However difficulty arises in the case of multivalent structures which no longer exhibit simple 1-1 dynamics. One approach is to simply approximate the dynamics with a 1-1 model and obtain an apparent dissociation constant.

For the only modestly more complicated bivalent system, we can derive the relationship between an apparent dissociation constant and a complete model with two dissociation constants to describe the multiple processes taking place.

The geometric determinants of programmed antibody migration and binding on multi-antigen substrates 13

In the case of the 2-antigen structure, there are $N = 5$ total states, corresponding to an empty structure (σ_{--}), two states with 1 monovalently occupied antigen each (σ_{Ab-} and σ_{-Ab}), one state with both antigens bivalently occupied by one antibody ($\sigma_{.Ab.}$), and a state with both antigens monovalently occupied by antibodies (σ_{AbAb}).

First, the reaction system can be represented according to the diagram in Figure 1 from the main text or the set of reactions below:



We have two dissociation constants respectively for the processes of monovalent binding and bivalent interconversion.

$$K_{D1} = \frac{k_{-1}}{k_1} \quad (44)$$

$$K_{D2} = \frac{k_{-2}}{k_2} \quad (45)$$

The system can be represented with a system of differential equations:

$$\frac{dp_{\sigma_{--}}}{dt} = -2c_{Ab}k_1p_{\sigma_{--}} + k_{-1}p_{\sigma_{Ab,-}} + k_{-1}p_{\sigma_{-,Ab}} \quad (46)$$

$$\frac{dp_{\sigma_{Ab,-}}}{dt} = c_{Ab}k_1p_{\sigma_{--}} - c_{Ab}k_1p_{\sigma_{Ab,-}} - k_2p_{\sigma_{Ab,-}} - k_{-1}p_{\sigma_{Ab,-}} + k_{-1}p_{\sigma_{AbAb}} + k_{-2}p_{\sigma_{.Ab.}} \quad (47)$$

$$\frac{dp_{\sigma_{-,Ab}}}{dt} = c_{Ab}k_1p_{\sigma_{--}} - c_{Ab}k_1p_{\sigma_{-,Ab}} - k_2p_{\sigma_{-,Ab}} - k_{-1}p_{\sigma_{-,Ab}} + k_{-1}p_{\sigma_{AbAb}} + k_{-2}p_{\sigma_{.Ab.}} \quad (48)$$

$$\frac{dp_{\sigma_{.Ab.}}}{dt} = k_2p_{\sigma_{Ab,-}} + k_2p_{\sigma_{-,Ab}} - 2k_{-2}p_{\sigma_{.Ab.}} \quad (49)$$

$$\frac{dp_{\sigma_{AbAb}}}{dt} = c_{Ab}k_1p_{\sigma_{Ab,-}} + c_{Ab}k_1p_{\sigma_{-,Ab}} - 2k_{-1}p_{\sigma_{AbAb}} \quad (50)$$

where $p_{\sigma_{--}}$, $p_{\sigma_{Ab,-}}$, $p_{\sigma_{-,Ab}}$, $p_{\sigma_{.Ab.}}$, $p_{\sigma_{AbAb}}$ the probabilities of each of the five states in the bivalent systems, subject to the normalization condition:

$$p_{\sigma_{--}} + p_{\sigma_{Ab,-}} + p_{\sigma_{-,Ab}} + p_{\sigma_{.Ab.}} + p_{\sigma_{AbAb}} = 1 \quad (51)$$

Given knowledge of the constituent equilibrium constants, we can in the simple case of the bivalent system, solve for the apparent dissociation constant as a function of the microconstants. This is, in effect, specifying a certain equilibrium value predicted on the basis of the complete bivalent model, and assuming instead that it is the result of 1-1 kinetics. However for multiple concentrations, the equilibrium will not shift proportionately, thus the apparent binding constant is a function of the concentration from which the equilibrium value is derived, making its value depend on the conditions rather than serving as a concise description of the system as a whole.

The bivalent system has, at equilibrium, the condition that the rate of change of each of its states is zero

$$\left(\frac{d\sigma_{--}}{dt}\right)_{eq} = \left(\frac{d\sigma_{Ab,-}}{dt}\right)_{eq} = \left(\frac{d\sigma_{-,Ab}}{dt}\right)_{eq} = \left(\frac{d\sigma_{.Ab.}}{dt}\right)_{eq} = \left(\frac{d\sigma_{AbAb}}{dt}\right)_{eq} = 0 \quad (52)$$

This condition plus the normalization conditions allows us to solve for the equilibrium concentrations of each of the species in terms of rate constants and the fixed solution concentration of analyte antibody.

$$p_{\sigma_{--}}^{eq} = \frac{k_{-1}^2 k_{-2}}{c_{Ab}^2 k_1^2 k_{-2} + c_{Ab} k_1 k_2 k_{-1} + 2c_{Ab} k_1 k_{-1} k_{-2} + k_{-1}^2 k_{-2}} \quad (53)$$

$$p_{\sigma_{Ab-}}^{eq} = \frac{c_{Ab} k_1 k_{-1} k_{-2}}{c_{Ab}^2 k_1^2 k_{-2} + c_{Ab} k_1 k_2 k_{-1} + 2c_{Ab} k_1 k_{-1} k_{-2} + k_{-1}^2 k_{-2}} \quad (54)$$

$$p_{\sigma_{-Ab}}^{eq} = \frac{c_{Ab} k_1 k_{-1} k_{-2}}{c_{Ab}^2 k_1^2 k_{-2} + c_{Ab} k_1 k_2 k_{-1} + 2c_{Ab} k_1 k_{-1} k_{-2} + k_{-1}^2 k_{-2}} \quad (55)$$

$$p_{\sigma_{Ab.}}^{eq} = \frac{c_{Ab} k_1 k_2 k_{-1}}{c_{Ab}^2 k_1^2 k_{-2} + c_{Ab} k_1 k_2 k_{-1} + 2c_{Ab} k_1 k_{-1} k_{-2} + k_{-1}^2 k_{-2}} \quad (56)$$

$$p_{\sigma_{AbAb}}^{eq} = \frac{c_{Ab}^2 k_1^2 k_{-2}}{c_{Ab}^2 k_1^2 k_{-2} + c_{Ab} k_1 k_2 k_{-1} + 2c_{Ab} k_1 k_{-1} k_{-2} + k_{-1}^2 k_{-2}} \quad (57)$$

We can combine states according to their corresponding occupancy - i.e. the number of antibodies that the state contributes to the overall signal due to bound antibody, where

$$p_{\sigma_{--}} + p_{\sigma_{Ab-}} + p_{\sigma_{-Ab}} + p_{\sigma_{Ab.}} + p_{\sigma_{AbAb}} = 1 \quad (58)$$

The probabilistic definition of occupancy is the expectation value of state occupancy. Each state has a corresponding integer occupancy associated with the number of antibodies bound to the structure in that state as well as a respective probability of that state at any point in time. The equilibrium occupancy is thus the average occupancy of all the states weighted by their equilibrium probabilities.

$$\Omega_{eq} = \sum_{i=1}^N p_{\sigma_i} \omega_{\sigma_i} = p_{\sigma_{--}}^{eq} \cdot 0 + (p_{\sigma_{Ab-}}^{eq} + p_{\sigma_{-Ab}}^{eq} + p_{\sigma_{Ab.}}^{eq}) \cdot 1 + p_{\sigma_{AbAb}}^{eq} \cdot 2 \quad (59)$$

substituting Equations 53 through 57, we arrive at

$$\Omega_{eq} = \frac{c_{ab} k_1 (2c_{ab} k_1 k_{-2} + k_{-1} (k_2 + 2k_{-2}))}{c_{ab}^2 k_1^2 k_{-2} + c_{ab} k_1 k_2 k_{-1} + 2c_{ab} k_1 k_{-1} k_{-2} + k_{-1}^2 k_{-2}} \quad (60)$$

4.5 Apparent dissociation constant

Taking the 2-antigen system equilibrium occupancy from Equation 60 and applying it to the equilibrium occupancy in terms of the 1-1 dissociation constant Equation 40 can be used to solve for an apparent equilibrium dissociation constant of the form

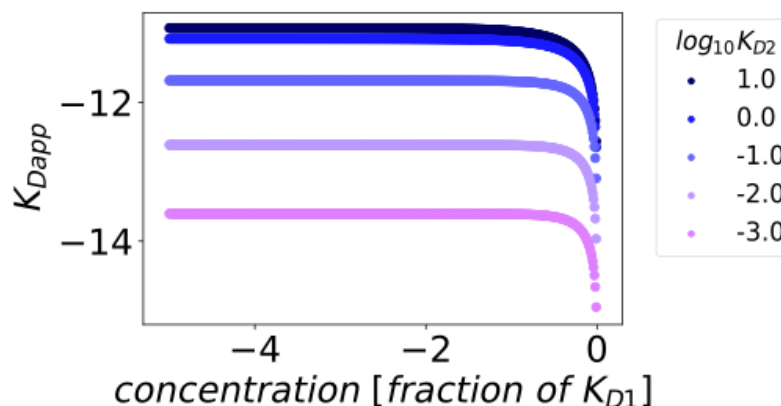
$$K_{Dapp} = \frac{c_{Ab}(1 - \Omega_{eq})}{\Omega_{eq}} = \frac{k_{-2} (k_{-1}^2 - c_{ab}^2 k_1^2)}{k_1 (2c_{ab} k_1 k_{-2} + k_2 k_{-1} + 2k_{-1} k_{-2})} \quad (61)$$

This constant is a value that would be obtained from a 1-1 fit to an equilibrium SPR value that arose from the 2-antigen kinetics. Rearranging and substituting Equations 44 and 45 into Equation 61, the formula simplifies to

$$K_{Dapp} = \frac{-K_{D2}(c_{Ab}^2 - K_{D1}^2)}{2c_{Ab}K_{D2} + K_{D1} + 2K_{D1}K_{D2}} \quad (62)$$

which we may note is a function of concentration, and which has a root at the critical value when $c_{Ab}K_{D2}^2 = K_{D1}^2$, i.e. the point at which average equilibrium occupancy greater than 1 is expected in the 2-antigen system, and rendering impossible any 1-1 kinetic description.

Rearrangement of Equation 62 enables us to determine the interconversion constant from an apparent dissociation constant provided that we know the monovalent binding constant.



Supplementary Figure 2. Log-log (base 10) plot of the apparent dissociation constant as a function of concentration represented as a fraction of the monovalent dissociation constant. For concentrations approaching that of the monovalent dissociation constant, the apparent binding constant changes significantly, whereas low concentrations lead to less concentration dependence.

$$K_{D2} = \frac{K_{Dapp}K_{D1}}{-c_{Ab}^2 + K_{D1}^2 - 2K_{Dapp}c_{Ab} - 2K_{Dapp}K_{D1}} \quad (63)$$

At concentrations where $c_{Ab} \ll K_{D1}$, the relationship between K_{D2} and K_{Dapp}^2 is relatively constant (Fig 2). Note that this is only valid for PSPR data with a 2-antigen topology of a single separation distance.

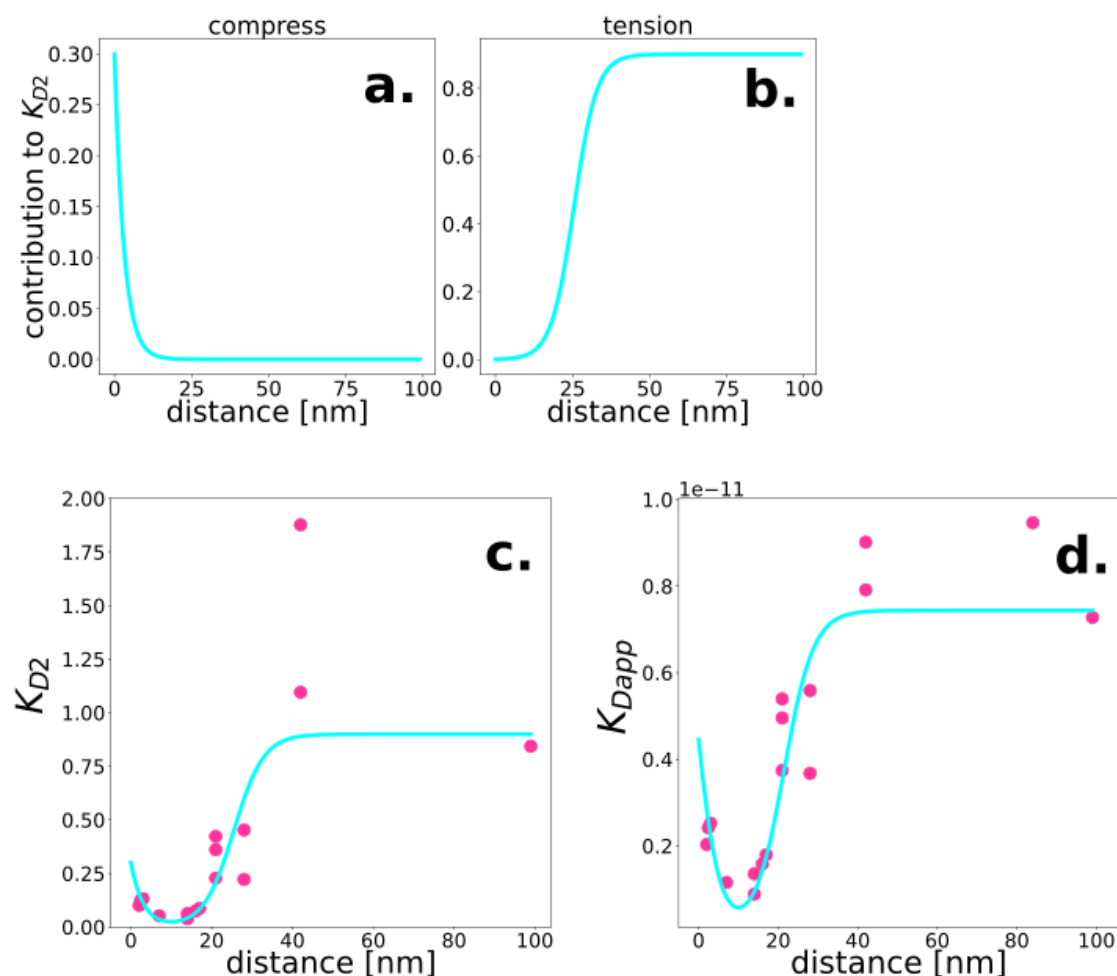
4.6 Mathematical description of spatial tolerance

Spatial tolerance refers the favorability of bivalent antibody binding according to the spatial distribution of the 2 adjoining antigens. Some antibodies stretch and compress more than others leading to a greater chance of entering and remaining in a bivalent state. In our model, we propose that the monovalent binding step occurs separately from the bivalent binding step, and that it is purely dependent on the solution phase concentration and the epitope-paratope binding affinity. Spatial tolerance therefore is a property of the interconversion step from monovalent to bivalent states and the reverse process from bivalent back to monovalent. For antigens separated by very small distances, electrostatic repulsion in response to compression and steric hindrance within the IgG molecule occurs, penalizing conversion to bivalent binding and/or favoring unbinding back to monovalent states. Conversely, at larger separation distances, the molecule must stretch to accomodate the gap, again penalizing conversion to bivalence and/or favoring conversion back to monovalence.

Spatial tolerance is a description of the landscape of this tradeoff - the breadth of the favorable region in between extremes that is conducive to bivalent binding, the sharpness and degree of symmetry of the transitions to monovalent preference at close and far separations. We can model spatial tolerance phenomenologically with an equation for determining the interconversion constant K_{D2} as a function of the separation distance between two antigens x :

$$K_{D2} = K_{D2-tensile} + K_{D2-compression}, \quad (64)$$

where $K_{D2-compression}$ (Fig 3 a) and $K_{D2-tensile}$ (Fig 3 b) are respectively exponential and logistic terms. These model separately the decrease in interconversion due to tensile stretch of the molecule at increasing distances and that due to the onset of excluded volume, electrostatic repulsion, or steric hindrance caused by compression of the molecule to bridge close distances.



Supplementary Figure 3. Measurement and modeling of spatial tolerance of antibody with increasing spatial separation of 2 antigens. a: Exponential compression term of spatial tolerance plotted alone versus distance. b: Sigmoidal tension term of spatial tolerance plotted alone versus distance. c: The spatial tolerance function, experimental interconversion constant values (magenta) and the fitted spatial tolerance model (cyan) both plotted against separation distance of two antigens. d: The apparent dissociation constant assuming 1-1 binding kinetics for what is actually a 2-antigen bivalent system calculated from interconversion constant data on the basis of equilibrium occupancy predicted by the bivalent model in terms of individual rate parameters.

The geometric determinants of programmed antibody migration and binding on multi-antigen substrates 17

The tensile term is built from a logistic function and has the form:

$$K_{D2-tensile} = \frac{K_{D2}^{max}}{1 + e^{-\alpha_t(x-\ell_t)}} \quad (65)$$

where K_{D2}^{max} is an upper limit of the value of K_{D2} , α_t is the logistic growth rate or steepness with which the tensile penalty grows at increasing separation distances and has units of inverse length, and ℓ_t is the value of the midpoint of the sigmoidal curve which can be thought of as a characteristic length that defines the scale below which favorable interconversion will occur and above which the function approaches minimal interconversion.

The exponential compressive term has the form:

$$K_{D2-compression} = K_{D2}^{max} e^{-\alpha_c(x-\ell_c)}, \quad (66)$$

where α_c is the exponential decay rate which has units of inverse length, and ℓ_c is another characteristic length parameter with units of length. The model is subject to the constraint $\ell_c < \ell_t$.

The combined expression yields Equation 1 which predicts the interconversion constant as a function of separation distance (Fig 3 c). This can be converted to an effective or apparent dissociation constant as if a 1-1 model on the basis of the bivalent model's prediction of equilibrium occupancy (Fig 3 d) - see Section 4.5.

4.7 Markov model of arbitrary antigen pattern geometries

For the binding kinetics of multi-antigen patterns of systems of sizes on the order of 2-8 adjacent antigens, we employ a fully enumerative Markov chain model based on a complete transfer matrix, i.e. all possible states and transitions of the system. The antigen pattern itself is modeled as a discrete network of antigen sites with a Euclidean distance matrix

$$\mathbf{D} = \begin{pmatrix} 0 & d_{1,2} & d_{1,3} & \dots & d_{1,N} \\ d_{2,1} & 0 & d_{2,3} & \dots & d_{2,N} \\ d_{3,1} & d_{3,2} & 0 & \dots & d_{3,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N,1} & d_{N,2} & d_{N,3} & \dots & 0 \end{pmatrix} \quad (67)$$

An adjacency matrix is formed by applying a critical distance cutoff to \mathbf{D} , such that long range interactions will be ignored, greatly reducing the number of states and transitions to enumerate.

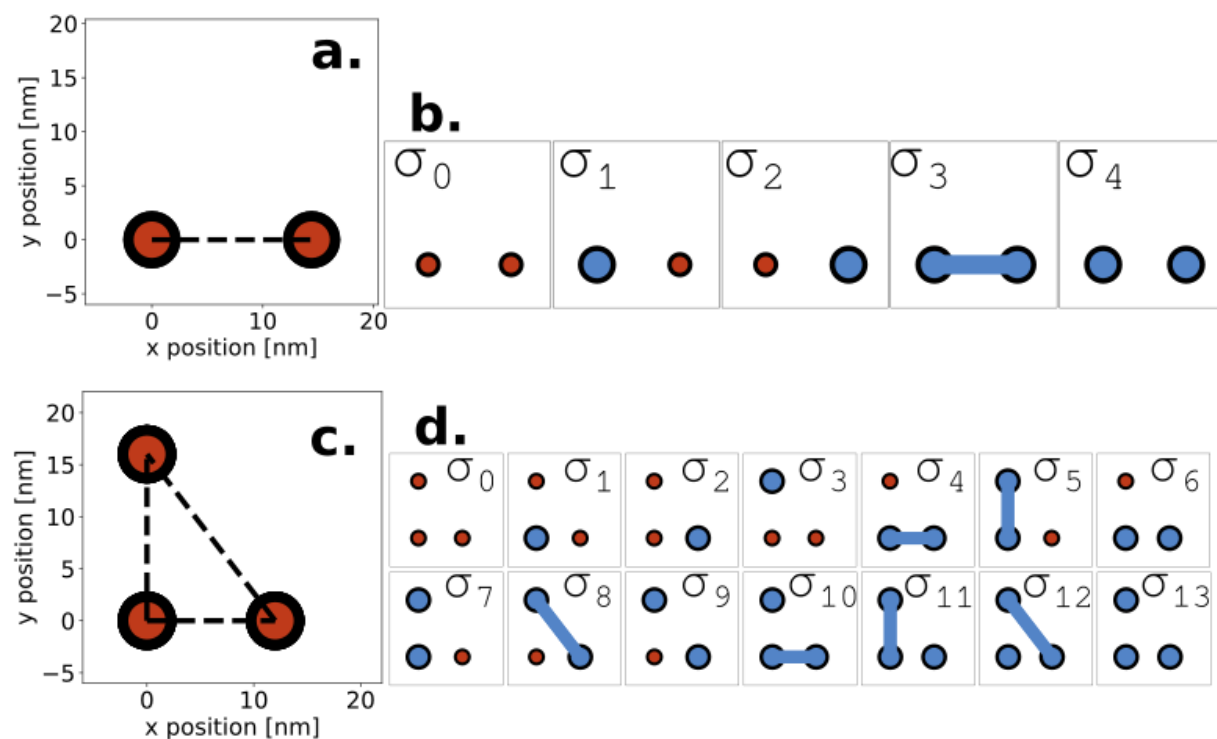
$$\mathbf{A}_{i,j} = \begin{cases} 0 & , \text{ if } D_{i,j} > d_{crit} \\ D_{i,j} & , \text{ if } D_{i,j} \leq d_{crit} \end{cases} \quad (68)$$

A single state σ_i of the system is defined as a set of antigens, their status (empty, monovalently occupied, bivalently occupied) and a pointer indicating of which bivalent-status antigens are linked to each other. The state space of a system is the set of all states that a structure in the system can assume, i.e. $\mathcal{S} := \{\sigma_0, \sigma_1, \dots, \sigma_N\}$. The set of states are thus all the possible configurations of empty, monovalently bound, and bivalently bound antibodies given the constraints of the pattern geometry (Fig 4).

Each state is linked to adjacent states by elementary transitions, i.e. the change in status of individual antibodies comprising the state. Those transitions are either the concentration-dependent addition or the subtraction of a single antibody to the system via monovalent binding or unbinding:



or a bivalent interconversion event where a monovalently bound antibody binds to an adjacent antigen site, changing its status to bivalently bound and vice versa:



Supplementary Figure 4. Enumerated states for basic and arbitrary geometries. a: Dimensions of a basic bivalent antigen pattern with 15 nm separation between antigens. b: The set of possible states for the basic bivalent geometry with blue dots indicating monovalently bound antibodies, blue lines connecting two dots indicating a bivalently bound antibody, and red dots indicating empty antigen sites. c: The dimensions of an arbitrary geometry, a $12 \times 16 \times 20$ nm right triangle. d: The set of states corresponding to the right triangle geometry.

The geometric determinants of programmed antibody migration and binding on multi-antigen substrates 19

$$\sigma_1 \xrightleftharpoons[k_{-2}]{k_2} \sigma_j \quad (70)$$

The system parameters are the set of zero order transition rates $\{r_1 = c_{Ab}k_1, r_{-1} = k_{-1}, r_2 = k_2, r_{-2} = k_{-2}\}$. The multi-antigen-antibody system is thus fully described by the continuous time Markov model $(\mathcal{S}, \mathbf{R})$ defined as its set of states and its corresponding transition rate matrix of the form:

$$\mathbf{R} = \begin{cases} r(i, j) & , \text{ if } A_{i,j} > 0 \\ 0 & , \text{ if } D_{i,j} = 0 \end{cases} \quad (71)$$

Automated enumeration of states in systems of arbitrary antigen pattern geometry is accomplished using an implementation of the breadth-first search (BFS) algorithm. The algorithm searches for edges between adjacent states and assigns the appropriate elementary rate process. A queue of neighboring states is made upon the visitation of any state. One-by-one, the algorithm visits each state in the queue, populating it with additional states when they are discovered, and skipping the addition of states that have already been visited. The algorithm thus is characterized by an initial expansion phase of the queue followed by a systematic reduction of the queue until all states have been visited, and the queue becomes empty. This exhaustive enumeration is deterministic, and enables us to assemble a complete transition matrix regardless of antigen geometry. However as the number of adjacent antigens grows, the number of combinations increases dramatically, thus for larger systems, a sampling based approach must be used instead.

4.8 Transient (non-equilibrium) dynamics of enumerative PSPR models

The continuous time Markov model enables us to compute transient evolution of the system. The probability distribution

$$\mathbf{p}(t) = \begin{pmatrix} p_0(t) \\ p_1(t) \\ \vdots \\ p_N(t) \end{pmatrix} \quad (72)$$

is a vector whose elements $p_i(t)$ are the probabilities of the respective system states $\sigma_0, \sigma_1, \dots, \sigma_N$ at time t . A uniform probability distribution would, for example, represent equal probabilities of finding a structure in any one of the possible states. Or another example is at the start of a single cycle kinetics PSPR run, when the initial condition $\mathbf{p}(t_0)$ is that of a distribution where $p_{-}(t_0) = 1$ for the state σ_{-} corresponding to an empty structure and $p_i(t_0) = 0$ for all other states.

The transient evolution of state probabilities is computed from an initial condition using the linear system of Chapman-Kolmogorov differential equations:

$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) \cdot \mathbf{Q} \quad (73)$$

making use of an infinitesimal generator matrix \mathbf{Q} which is obtained from the rate matrix and used to determine the relative rates at which state probabilities change with incremental time.

$$\mathbf{Q}_{i,j} = \begin{cases} \mathbf{R}_{i,j} & , \text{ for } i \neq j \\ -\sum_{i \neq j} \mathbf{R}_{i,j} & , \text{ for } i = j \end{cases} \quad (74)$$

The infinitesimal generator is then used to compute the change in state probability distribution going from one time point to the next by the matrix exponential formula:

$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) \cdot e^{\mathbf{Q} \cdot \Delta t} = \mathbf{p}(t) \sum_{\eta=0}^{\infty} \frac{(\mathbf{Q} \cdot \Delta t)^{\eta}}{\eta!} \quad (75)$$

where η is the computation's depth of recursion - the higher the more accurate, and Δt is an incremental advancement in time. Due to numerical instability of this solution, we employ the uniformized discrete time Markov chain method of Fox and Glynn in order to stably compute Equation 75 [4]. The continuous Markov model (\mathbb{S}, \mathbf{R}) is approximated by a discrete model (\mathbb{S}, \mathbf{U}) by renormalizing the generator matrix with respect to the fastest outgoing rate or the uniformization rate q :

$$\mathbf{U} := \mathbf{I} + \frac{\mathbf{Q}}{q}, \quad q \geq \max_i \{|\mathbf{Q}_{i,i}|\} \quad (76)$$

where \mathbf{I} is the identity matrix. Equation 75 becomes the approximation

$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) \cdot e^{\mathbf{Q} \cdot \Delta t} = \mathbf{p}(t) \cdot e^{q(\mathbf{U} - \mathbf{I}) \cdot \Delta t} = \mathbf{p}(t) \cdot e^{q\mathbf{U}\Delta t} e^{-q\mathbf{I}\Delta t} = \mathbf{p}(t) e^{-q\Delta t} \cdot e^{q\mathbf{U}\Delta t} \quad (77)$$

The matrix exponential is then approximated with the following Taylor series expansion:

$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) \cdot e^{q\Delta t} \cdot \sum_{\eta=0}^{\infty} \frac{(q\Delta t \mathbf{U})^\eta}{\eta!} = \sum_{\eta=0}^{\infty} \frac{e^{-q\Delta t} (q\Delta t)^\eta}{\eta!} \cdot \mathbf{p}(t) \cdot \mathbf{U}^\eta \quad (78)$$

Using Equation 78, we can stably compute the transient evolution of a system from an initial condition. The system entropy can be computed by

$$S = -k_B \sum_{\sigma_i \in \mathbb{S}} p_i(t) \ln(p_i(t)) \quad (79)$$

4.9 Fitting continuous time Markov models to PSPR data using autocorrelation of residuals

Using Equation 78 to compute the transient probability distribution of the system, we are able to also compute the occupancy at each time point using the definition from Equation 11. The system occupancy is thus a function of time of the form

$$\Omega(t) = \sum_{i=1}^N p_i(t) \omega_i \quad (80)$$

The continuous time Markov model is fitted to experimental data by comparing occupancies computed on the basis of Equations 78 and 80 with that of occupancy computed from normalizing PSPR data via Equation 27 and we can see that the theoretical curve either correctly or incorrectly fits the experimental data depending on the parameterization (Fig 5 a and Fig 6 a). Residuals (Fig 5 b and Fig 6 b) are computed by

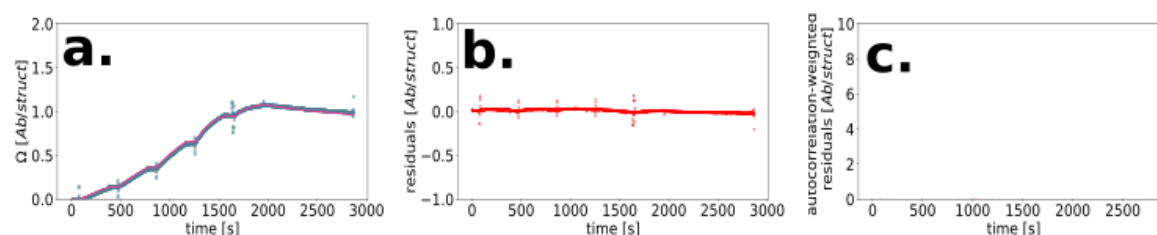
$$e(t) = \Omega(R_{Ab}) - \Omega(t, \mathbb{S}, \mathbf{U}) \quad (81)$$

While fitting by minimizing the sum of squared residuals can be used to obtain acceptable model parameterizations, we used residual analysis with autocorrelation to improve the robustness of fitting and reduce systematic mis-parameterization by making fits more sensitive to divergence in curve shapes. We compute an absolute, average autocorrelation over a fixed interval $k\Delta t$ with $k = 50$ by :

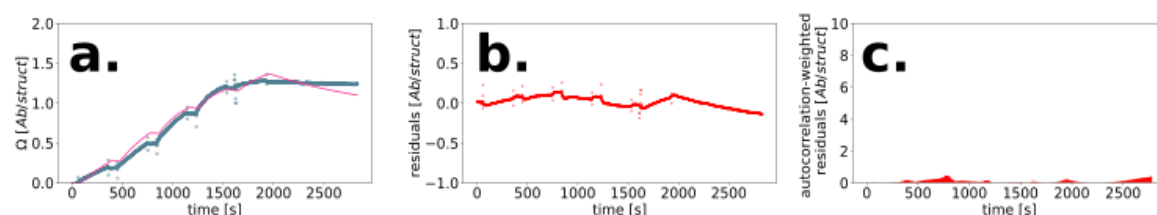
$$\rho_{e,e}(t, t + k\Delta t) = \frac{\sqrt{(\sum_k \mathbf{e} \cdot |\vec{v}|)^2}}{k} \quad (82)$$

where $\mathbf{e}(t, k) = [e(t), e(t + \Delta t), \dots, e(t + k\Delta t)]$, $\mathbf{v} = [0, 1, \dots, k]$, and $|\vec{v}|$ is the conjugate of \mathbf{v} . The objective function numerically minimized $\min(\mathcal{E})$ to obtain fits to experimental PSPR data is then the sum squared residual vector weighted by its autocorrelation vector:

$$\mathcal{E} = \sum_{t=0}^{t_f - k + 1} \rho_{e,e}(t, t + k) \sqrt{e(t)^2} \quad (83)$$

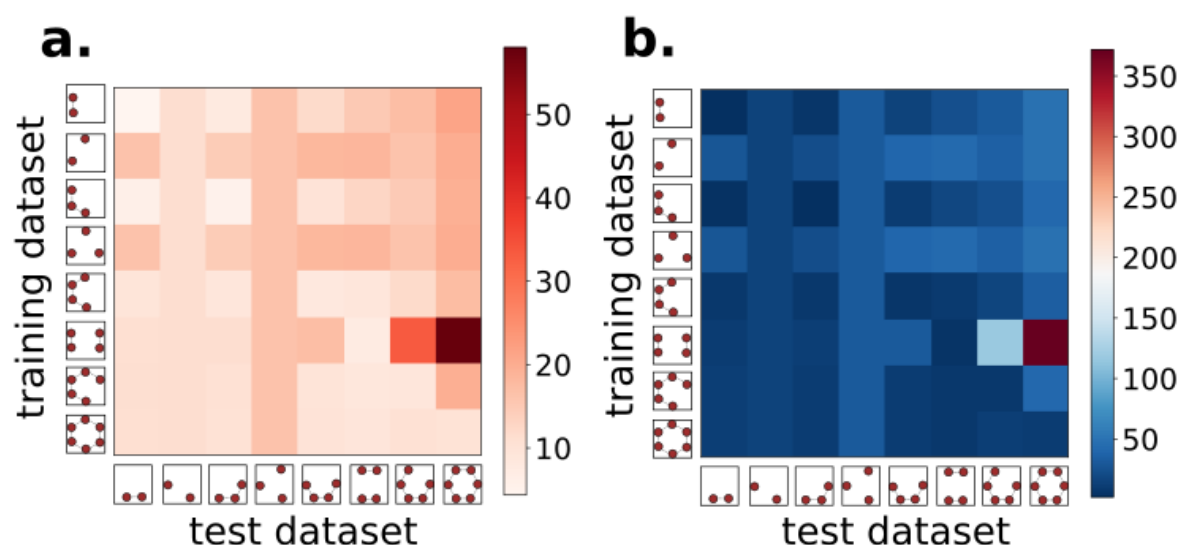


Supplementary Figure 5. Example of a well-parameterized bivalent model, with strong agreement between experimental and theoretical occupancy. a: Markov model (magenta) prediction of average occupancy (antibodies per structure) superimposed on experimental PSPR binding data for a 2 antigen pattern (blue). b: Residuals comparing the theoretical and experimental curves in a. c: Residuals from b weighted by autocorrelation.



Supplementary Figure 6. Example of a misparameterized bivalent model with visual misalignment between theoretical and experimental occupancy curves and corresponding divergence of residuals. a: Markov model (magenta) prediction of average occupancy (antibodies per structure) superimposed on experimental PSPR binding data for a 2 antigen pattern (blue). b: Residuals comparing the theoretical and experimental curves in a. c: Residuals from b weighted by autocorrelation.

This provides an error function sensitive to sustained divergence of model and experimental data (Fig 6 c) even if the two curves cross paths, and like summing the residuals provides a low value when alignment is good (Fig 5 c).



Supplementary Figure 7. a: Matrix depicting the sum of residuals (color map) for Markov models parameterized on experimental binding data from antigen patterns on the vertical axis (training dataset) and performance-tested against other patterns (test dataset). b: Same as (a) except that colormap indicates the residuals weighted by their autocorrelations.

We performed cross validation of Markov model fitting by parameterizing based on experimental data from various antigen patterns (all with fixed nearest-neighbor separation distances between antigens to remove the complication of separation distance dependence of the binding kinetics). The rate parameters derived from these training data were then fixed and the model was applied to other patterns as a limited test of extrapolation of a parameterized model to different antigen pattern geometries. Absolute sums of residuals (Fig 7 a) and absolute sums of residuals weighted by autocorrelation (Fig 7 b) show that the best-performing models were the those that are the most complex and exhibiting bivalence such as the hexagonal and pentagonal configurations in the last two rows. This suggests that downward extrapolation in pattern complexity is more viable than upward.

4.10 Determination of thermodynamic properties

We can obtain equilibrium probabilities from the uniformized CTMC first by simulation out to long time scales at fixed solution concentration until probabilities cease to change:

$$\mathbf{p}^* = \lim_{t \rightarrow \infty} \mathbf{p}(t), \quad \left(\frac{d}{dt} \mathbf{p}(t) \right)_{c_{Ab}} = \mathbf{0} \quad (84)$$

We can determine the steady state probability distribution more expediently on the basis of the infinitesimal generator matrix, Equation 74, solving numerically for the probability distribution which, when multiplied with the generator matrix produces a vector of zeros, meaning that there is zero change

The geometric determinants of programmed antibody migration and binding on multi-antigen substrates 23

from one moment to the next, subject to the normalization condition whereby all probabilities must sum to 1. I.e. the steady state probability distribution is the solution to the matrix equation

$$\mathbf{p} \cdot \mathbf{Q} = \mathbf{0}, \quad \sum_{\sigma_i \in \mathbb{S}} p_i = 1 \quad (85)$$

The multi-antigen structure in the context of a PSPR experiment is an open system, freely allowed to exchange particles with the large external reservoir connected to it. With (T, V, c_{Ab}) held constant, the system (an antigen patterned structure) will approach a minimum free energy at steady state by exchanging antibodies with the bath, obeying the Boltzmann distribution law

$$p_i^* = \frac{e^{-E_i/k_B T}}{\sum_{\sigma_i \in \mathbb{S}} e^{-E_i/k_B T}} = \frac{e^{-E_i/k_B T}}{\mathbb{Z}} \quad (86)$$

where \mathbb{Z} is a grand canonical partition function which predicts equilibrium at a grand potential free energy minimum $d\Phi(T, V, \mu_{Ab}) = 0$ with chemical potential $\mu_{Ab} = -k_B T \ln c_{Ab}$, and $E_i p_i = \mu_{Ab} + \mu_{mono} n_{mono} + \mu_{biv} n_{biv}$ are state energies determined by the environmental potential due to solution phase antibody concentration as well as the individual potentials of antibody monovalent and bivalent bonds populating the state, n_{mono} monovalent bonds and n_{biv} bivalent bonds with respective chemical potentials μ_{mono} and μ_{biv} . The value of \mathbb{Z} and the state energies are solved numerically for a given steady state probability distribution, making it possible for us to determine thermodynamic quantities.

We can obtain thermodynamic quantities such as the solution concentration dependent equilibrium grand potential free energy via:

$$\Phi^* = -k_B T \ln \mathbb{Z} \quad (87)$$

The equilibrium probabilities enable us to calculate relative potential differences via

$$\frac{p_i^*}{p_j^*} = e^{-(E_i - E_j)/k_B T} \quad (88)$$

This makes it possible to compute chemical potentials of monovalent and bivalent bonds, for example from the basic 2-antigen system of a fixed separation, by comparing equilibrium probabilities in states that differ by exactly 1 bond of a given type. This is equivalent to obtaining the change in free energy via the dissociation constant for that process.

$$\mu_{0-1} = E_{mono} - E_{empty} = \left(\frac{\partial \Phi}{\partial N_{mono}} \right)_{T, P, N_{biv}} = -k_B T \ln \frac{\sum_{n_{empty}} p_{empty}^*}{\sum_{n_{mono}} p_{mono}^*} + k_B T \ln c_{Ab} = -k_B T \ln \frac{n_{mono}}{n_{empty} K_{D1}} \quad (89)$$

$$\mu_{1-2} = E_{biv} - E_{mono} = \left(\frac{\partial \Phi}{\partial N_{biv}} \right)_{T, P, N_{empty}} = -k_B T \ln \frac{\sum_{n_{biv}} p_{biv}^*}{\sum_{n_{mono}} p_{mono}^*} = -k_B T \ln \frac{n_{biv}}{n_{mono} K_{D2}} \quad (90)$$

where n_{empty} , n_{mono} , and n_{biv} are the respective degeneracies of empty, monovalently 1-occupied, and bivalently 1-occupied states. For the 2 antigen system, these values are respectively 1, 2, and 1. For the rabbit IgG, anti-digoxygen model and a separation distance of 15 nm, the chemical potentials are $\mu_{0-1} = 1.805 \times 10^{-20}$ J / particle and $\mu_{1-2} = 8.889 \times 10^{-21}$ J/particle.

We can also obtain a stand-alone bivalent binding chemical potential such that tallying the number of monovalent and bivalent molecules in a state would yield the potential of that state.

$$\mu_{0-2} = E_{biv} - E_{empty} = \mu_{0-1} + \mu_{1-2} = \left(\frac{\partial \Phi}{\partial N_{biv}} \right)_{T,P,N_{mono}} = -k_B T \ln \frac{\sum p_{biv}^*}{p_{empty}^*} + k_B T \ln c_{Ab} \quad (91)$$

This gives us $\mu_{0-2} = 2.693 \times 10^{-20}$ J / particle for the 15 nm separation. By this method, the distance-dependent K_{D2} model of Equation 1 can be converted to a chemical potential curve.

4.11 Markov chain Monte Carlo version of model

Systems with larger numbers of bivalent connections have many states and transitions, and the fully enumerative CTMC does not scale well. For these systems, we use a Markov Chain Monte Carlo (MCMC) sampling approach whereby only local states are computed throughout the trajectory of a single system. Multiple trajectories are sampled to approach and approximate the probability distributions that would otherwise be computed exactly for the enumerative method. Instead of computing the flux in state probabilities over fixed intervals of time, we instead computing Poisson intervals (dwell times) of states according to the rates of processes that each state is subject to.

5 Experimental Methods

The experimental data used in this work was presented previously as supplementary control data in [11] and was not analyzed further than a basic fitting to a standard 1-1 model. In the current work, we have used the data to develop a more accurate mechanistic model, a pipeline for constructing such models from a minimal dataset, an in-depth physical characterization framework, and *de novo* simulation that go beyond previous work.

5.1 Reagents

Oligonucleotides (unmodified and digoxigenin-modified) were purchased from IDT (Belgium) in 96-well plates format. Chemicals (NaCl, KCl, MgCl₂, Tris-HCl, EDTA, PEG800, NaOH, KOAc, KOH, NaOAc) for buffer preparation were purchased from Sigma-Aldrich. Streptavidin was purchased from Sigma-Aldrich. Phosphate-buffered saline 1M stock solution was purchased from Sigma-Aldrich. BIAcore consumables (CM3 sensor chip, HBS-EP running buffer) were purchased from GE Healthcare. Amicon centrifugal filters with 100 kDa MWCO were purchased from Merk Millipore.

5.2 Origami design and production

The 18-helix bundle DNA origami nanotube was designed with caDNAno[?], using the honeycomb lattice. This structure has been characterized earlier[11, ?, ?, ?]. In short: the p7560 scaffold was extracted from M13 phage, and the 18-helix bundle DNA nanotube was folded under the following condition: 20 nM scaffold, 100 nM each staple oligonucleotide, 13 mM MgCl₂, 5 mM Tris pH 8.5, 1 mM EDTA. The mixture was subjected to heat denaturation at 80°C for 5 min followed by a slow cooling ramp from 80°C to 60°C over 20 min and 60°C to 24°C over 14 hr. The excess staples were removed by ultrafiltration with Amicon 100 kDa MWCO filters. The wash buffer used was 1× PBS supplemented with 10 mM MgCl₂.

5.3 Pattern surface plasmon resonance protocol

The BIAcore t200 instrument from GE Healthcare was used for surface plasmon resonance experiments. The running buffer used in all experiments is HBS-EP supplemented with 10 mM MgCl₂. The flowrate used for all kinetic experiments is 30 $\mu\text{l}/\text{min}$. Streptavidin was diluted to a final concentration of 10 $\mu\text{g}/\text{ml}$ in 10 mM NaOAc pH 4.5 and chemically attached to the CM3 sensor chip with NHS/EDC coupling, using the standard protocol from GE Healthcare. Anchor oligonucleotides containing a 3' biotin modification

were diluted to 200 nM in $1\times$ HBS-EP running buffer and was injected over the surface for 20 min followed by washing of nonspecifically bound oligos by injecting 50 mM NaOH for 5 mins. The DNA nanostructures were diluted to 5 nM and injected over the surface for 20 mins followed by washing with running buffer for 10 mins. Antibodies were diluted to various concentration in running buffer ranging from 0.025 nM to 0.5 nM. The single cycle kinetics injection method was used to sequentially inject the antibody solution over the surface, started from the lowest concentration, the contact time for each concentration is 5 min. After the final antibody injection, the dissociation curve was recorded for 15 min. The immobilized DNA nanostructures and bound antibodies were removed from the surface by injecting 50 mM NaOH for 5 mins and then the surface is ready for the next round of experiment. The t200 evaluation software was used initially to fit the acquired data, for this we used a 1:1 Langmuir binding model to fit the data and estimate the k_a , k_d , K_D and antibody binding capacity.