Check for updates

OPEN Stochastic modeling of antibody binding predicts programmable migration on antigen patterns

Ian T. Hoffecker ^{1,2}^{III}, Alan Shaw^{1,3}, Viktoria Sorokina¹, Ioanna Smyrlaki¹ and Björn Högberg¹

Viruses and bacteria commonly exhibit spatial repetition of the surface molecules that directly interface with the host immune system. However, the complex interaction of patterned surfaces with immune molecules containing multiple binding domains is poorly understood. We developed a pipeline for constructing mechanistic models of antibody interactions with patterned antigen substrates. Our framework relies on immobilized DNA origami nanostructures decorated with precisely placed antigens. The results revealed that antigen spacing is a spatial control parameter that can be tuned to influence the antibody residence time and migration speed. The model predicts that gradients in antigen spacing can drive persistent, directed antibody migration in the direction of more stable spacing. These results depict antibody-antigen interactions as a computational system where antigen geometry constrains and potentially directs the antibody movement. We propose that this form of molecular programmability could be exploited during the co-evolution of pathogens and immune systems or in the design of molecular machines.

Due to their multiple binding domains, immunoglobulin molecules like the bivalent immunoglobulin G (IgG) antibody exhibit complex interactions with multivalent antigens, that is, clusters of multiple copies of molecules or molecular domains occurring at separation distances of the order of 1-30 nm. Multivalent interactions enhance the stability of binding interactions by enabling the simultaneous attachment of multiple ligands, increasing the magnitude of the apparent affinity or Gibbs free energy of multivalent binding, also called functional affinity in favor of the term 'avidity', and extending the residence times of the bound antibodies¹⁻³.

Many pathogenic surfaces exhibit spatial repetition at length scales relevant to antibody multivalence. Viral capsid proteins undergo self-assembly into periodic patterns⁴, and some neutralizing antibodies achieve their high affinity and neutralization capability through bivalence^{5,6}. Self-assembling crystalline arrays of surface-layer (S-layer) proteins—the outermost structure on many bacteria and archaea—are a major contact point between the pathogen and host⁷ and are implicated as the mediators of innate⁸ and adaptive immunity^{9,10}. Their repetitive organization may be integral to their immunological role, as their removal from bacterial surfaces was seen to reduce the immune response¹¹. Multivalence is also probably an important factor during the affinity maturation of antibodies and thus in vaccine design^{12–14}.

Antibody interaction with patterned surfaces presents a challenge for both experimental control and mathematical modeling as it is a many-bodied problem occurring on the timescales of seconds to minutes. Such systems are too computationally expensive for fullatom molecular simulation. Models treating antibodies and antigens as abstract binding and non-binding units have been the most successful at capturing the relevant dynamics, and have historically treated multivalence as a function of ligand coating density where multivalence emerges statistically as the average nearest-neighbor distance between the ligands decreases¹⁵. More recently, coarsegrained molecular simulations have been fruitfully used to quantify the effects of the cooperative binding of the antibody subunits on binding affinity^{16,17}. Nevertheless, a challenge of precisely calibrating such models remains due to the absence of experimental tools to independently assess monovalent and multivalent binding dynamics as well as a pipeline to connect such data to the models.

The patterned surface plasmon resonance (PSPR) technique enables the measurement of binding kinetics on precise, monodisperse patterns of ligands, achieving a robust control of geometry through the use of DNA origami nanostructures¹⁸ (Fig. 1a). Here 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. 1b). From this basis, the dynamics of more complex patterns of multiple antigens can be reduced (Fig. 1c) to combinations of these elementary states. A causal linkage between the pattern geometry and antibody dynamics could be potentially exploited as a form of spatial programmability by either immunity or pathogens during their adversarial co-evolution. We investigate this possibility and the role of spatial tolerance, that is, the range and impact of antigen separation distances on bivalent binding kinetics (Fig. 1d), in determining the effective binding affinity, walking speed of antibody migration on patterned surfaces and direction of antibody migration. Such control mechanisms might inform the development of vaccines for greater control over the affinity maturation process.

Results

Spatial tolerance model. We developed a model parameterization pipeline based on a progressive fitting of the transient surface plasmon resonance (SPR) profiles first for monovalent and then for bivalent binding processes to reduce the degrees of freedom at

¹Division of Biomaterials, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solna, Sweden. ²Department of Gene Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Solna, Sweden. ³Institute for Quantitative Biosciences, University of California Berkeley, Berkeley, CA, USA. ^{See}-mail: ithof@kth.se; bjorn.hogberg@ki.se



Fig. 1 | Scheme for modeling binding dynamics of antibodies on multiantigen substrates. a, Illustration of patterning concept, where smallmolecule antigens (haptens) are arranged using short, flexible tethers at well-defined locations on DNA origami nanostructures. This enables the multivalent interaction of antibodies with antigen patterns. **b**, Markov model of antibody binding where only the 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 the 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 the antibody structure on the chance of bivalent interconversion.

each stage of fitting. In the first stage, we used either rabbit antidigoxygenin (DIG) IgG or mouse-derived anti-DIG IgG1-targeting DIG-decorated DNA origamis with a single-cycle kinetics program where progressively higher concentrations of antibody were exposed to the immobilized antigen substrate (Fig. 2a,b). This program was performed with a one-antigen configuration (Fig. 2c,e) to parameterize our Markov model ('Markov model of arbitrary antigen pattern geometries' section) by relating the SPR signal to the average occupancy Φ defined as the number of antibodies per structure averaged according to the prevalence of each possible state ('Conversion from SPR signal R_{Ab} to bound antibody n_{Ab} ' section). This yields the respective association and dissociation rates $k_1 = 1.93 \pm 0.05 \times 10^7 \,\mathrm{M^{-1} \, s^{-1}}$ and $k_{-1} = 5.28 \pm 0.07 \times 10^{-4} \,\mathrm{s^{-1}}$ as well as a monovalent dissociation constant $K_{D1} = 2.7 \pm 0.11 \times 10^{-11} \text{ M}$ defined as the ratio of the dissociation to association rates. We parameterized the interconversion between the monovalent and bivalent states by fixing the previously determined monovalent parameters and fitting the model to experiments involving multiple adjacent antigens ('Fitting continuous-time Markov models to PSPR data using autocorrelation of residuals' section). We fitted the model by adjusting K_{D2} or the interconversion constant defined by the ratio of the reverse and forward interconversion rates. For structures configured with two antigens separated by 14.3 ± 1.2 nm, we find $K_{D2} = 8 \pm 6 \times 10^{-3}$ (Fig. 2d,f).

By applying progressive fitting to PSPR runs with structures patterned with two adjacent antigens of varied separation distances, we found the internal conversion process to vary accordingly. Small and large separation distances correspond to reduced bivalence, that is, larger K_{D2} . We constructed a phenomenological equation ('Mathematical description of spatial tolerance' section) modeling the interconversion constant (Fig. 2g). 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, a characteristic that has been substantiated by recent biosensing applications¹⁹. In our model, the interconversion constant is, thus, a function of adjacent antigen separation distance with the form

$$K_{\rm D2} = \frac{K_{\rm D2}^{\rm max}}{1 + e^{-\alpha_{\rm t}(x-\ell_{\rm t})}} + K_{\rm D2}^{\rm max} e^{-\alpha_{\rm c}(x-\ell_{\rm c})},\tag{1}$$

where ℓ_t and ℓ_c are the characteristic lengths defining the scale of the tensile and compressive terms, respectively; 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 the contributions of bivalence to binding dynamics are vanishingly small. We found that when appropriately fitted (Fig. 2h), this model predicts a theoretical min(K_{D2}) located at ~10.6 nm separation distance with an approximately 1.0 nm uncertainty due to an expected random shift in the one-antigen input data (Fig. 2i), whereas the experimental datapoint with the lowest K_{D2} is located at ~15.0 nm. This result is in agreement with another study²⁰ in which the optimal epitope separation distance is estimated using DNA origami and atomic force microscopy, although we note here that the curve is likely to differ between isotype, species and possibly even clones due to angular variation in the epitope–paratope bond.

Steady-state and transient analyses. 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. 3a). This revealed concentration regimes of differing dominant states: empty, bivalent and saturated monovalent at low, medium and high solution-phase concentrations, respectively. The entropic maxima occur at transitions between these domains (Fig. 3b), and the transitions in bivalent and monovalent contributions to chemical potential occur in accordance with the transition from bivalent to saturated monovalent regimes (Fig. 3c; 'Determination of thermodynamic properties' section).

In addition to simulating *de novo* patterns' steady-state properties, the model enables us to simulate the transient dynamics of hypothetical systems with arbitrary geometries and arbitrary timing in the introduction of different solution-phase concentrations. To validate the pipeline, we used the model parameterized with one- and two-antigen data (Fig. 2g) to create a blind a priori prediction of the evolution of a higher-order system with three antigens arrayed in a $7.2 \times 14.3 \times 16.0$ nm right triangle and then check its correspondence with an experimental trajectory (Fig. 3d); additional validation is shown in Supplementary Fig. 7. We found that the experimental trajectories closely conformed to the predictions. The model provides access to the individual contributions of states to the signal through their occupancy (Fig. 3e) and relative proportions (Fig. 3f), enabling us to construct a narrative explanation for the observed dynamics. We see, for example, in the final stage when the concentration was set to zero, as the total occupancy decreased and occupancy contributions from monovalently bound antibodies decreased, bivalent-state contributions counterintuitively increased. This indicates that higher concentrations in the penultimate stage inverted the system to favor monovalent-dominated saturation states that subsequently transitioned into unsaturated bivalent states as the sites became available (Fig. 3g).



Experiments on repetitive antigen patterns. To explore the potential for pattern-based control and programmability of antibody dynamics, we modeled the dynamics of larger systems with greater relevance to periodic pathogenic surfaces. For larger systems, a complete enumeration of states scales poorly with increasing numbers of adjacent antigens. We developed a Markov chain Monte Carlo (MCMC) implementation of the model ('MCMC version of the model' section) to sample the trajectories that converge to state probabilities with large sample numbers. Rather than enumerating all the system states (that is, combinations of antibodies and binding modes on a structure and possible transitions), the system performs a random walk through the large state space, computing its rate of escape into neighboring states at any point in time. We then examined the collections of individual trajectories for such systems **Fig. 2 | A progressive fitting pipeline for obtaining a parameterized model from minimal experimental data. a,b**, Concentration versus time plot of antibody solution exposed to patterned antigen substrates in a single-cycle kinetics PSPR experiment. **c,d**, Experimental binding kinetics data (black line) of one-antigen (**c**) and two-antigen configurations (**d**), superimposed over the occupancy calculated from the parameterized model. Model occupancy is divided and colored according to the state, with the height corresponding to the state's contribution to the total antibody occupancy per structure. **e,f**, For the one-antigen (**e**) and two-antigen (**f**) configurations, the transient-state probabilities stratified according to the model prediction, colored and stacked to satisfy the normalization condition where all the probabilities add to 1. The legends list either all or the five most prevalent states σ_i along with their corresponding occupancy ϕ_i and probabilities $_{pi}$ at the end of their respective run. **g**, Interconversion constants (red points) plotted versus two-antigen configuration distance *x* and the fitted spatial tolerance model (blue line). The blue error bars indicate the model fits due to one standard error of the mean away from a mean input one-antigen run, propagated to K_{D2} values (vertical black error bars). The horizontal black error bars denote spatial uncertainty (defined elsewhere⁴⁴). The vertical black error bars denote uncertainty due to one standard error of the mean variation in the input data. **h**, Goodness-of-fit characterization of the spatial tolerance model. The red points show an apparent random dispersion of K_{D2} points minus a moving average, whereas the blue points show the dispersion of model values subtracted from K_{D2} at each point. **i**, Sensitivity of the model's minimum to shifted values of one-antigen input data by the number of standard error of mean (SEM) away from a mean input run.

to understand their average behavior. Specifically, we examined the role of repetitive antigen spacing in simple one-dimensional (1D) arrays.

Antigens arranged according to a spacing gradient in the range of 10–22 nm separation distances (that is, the interval of the steepest slope in equation (1)) elicit asymmetric accumulation towards the narrow-spaced end of the array (Fig. 4a). This system also exhibited individual walking trajectories that tend towards the narrow-spaced end (Fig. 4b), asymmetric velocity (Fig. 4c) and asymmetric net displacement (Fig. 4d) according to the direction of the gradient. The mechanism for this locomotion is that of a biased random walk, where at any point in time, a bivalently bound antibody has a random chance to dislodge one of its paratopes and then reassociate either with the same epitope or an adjacent one. Differential spacing between the adjacent epitopes leads to a statistical preference for more stable spacings with a lower interconversion ratio K_{D2} .

Antibodies binding to 1D arrays with uniform spacing exhibited divergent residence times, with antibodies spending less cumulative time on 22-nm-spaced arrays (Fig. 4e) than those of narrow 10-nm-spaced arrays (Fig. 4f). The migration speeds of antibodies on high-strain-inducing arrays are greater than those of low-straininducing arrays, and a comparison of the net displacement shows that antibodies moved further from their initial binding location on widely spaced arrays relative to the narrowly spaced ones (Fig. 4g).

Discussion

Repeating epitope patterns are present in many viruses as coat proteins²¹⁻²⁴ and in bacteria as S-layer proteins²⁵⁻²⁷, often with a high degree of symmetry or geometric periodicity. Such repetitive, quasicrystalline patterns have been recognized as a marker of foreignness corresponding to major enhancements of IgG response compared with unorganized substrates²⁸. Investigators have observed both monovalent and bivalent antibody binding to such periodic viral surfaces^{24,29}, and binding enhancement due to bivalence is a recognized factor in determining both immune pathogen recognition and neutralization capability^{30,31}. A high-speed atomic force microscopy study³² captured the real-time bipedal locomotion of antibodies on reconstituted pathogenic surfaces with periodic patterns of epitopes.

The authors of this study proposed that antibody locomotion is enabled by the strain induced during bivalent binding as antibodies accommodate the geometry of their target antigens, weakening the bond and triggering a bipedal step. Our results agree and indicate that precisely tuned spacing on repetitive antigen patterns would have a major impact on the strength of bivalent bonds; furthermore, differences in adjacent antigen spacings statistically drive migration, as antibodies randomly move until becoming immobilized in states with minimal strain.

One limitation of our model is that torsional flexibility of the hinge region⁶ is not considered. This is due to the design of the DNA origami nanostructure substrates in which hapten antigens are tethered by short, flexible spacers with rotational freedom. Future studies could explore rotational spatial tolerance as well as degrees of freedom in the *Z* direction by including additional terms in equation (1) calibrated with the PSPR data that systematically modulate relative epitope orientations or employ structures that incorporate pathogenic protein antigens^{33–36} for more realistic structural complexity and physiological relevance.

An additional limitation of our model is the potential for overextrapolation in increasingly complex systems, with any inaccuracies in spatial tolerance model or parameterization quality subject to propagation. We predicted that long-range gradients of differential spacings could be used to establish the persistent directed migration of antibodies on a surface, and we propose that PSPR and the progressive fitting pipeline presented here should be used in future studies to experimentally test the predictions. First, such designs should be possible using DNA origami; second, the stratification of states obtained by model fitting to convoluted binding data might enable one to measure the spatial distribution of antibodies on gradient structures.

Repetitive antigen arrays have been important in vaccine design^{37,38}. The evolution of protective antibodies against malaria was shown to be dependent on a repetitive motif³⁹, and bacteria are known to interfere with antibody binding such as Fc targeting to

Fig. 3 | *De novo* simulation with parameterized kinetics. **a**, Stationary distributions colored by state of a two-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 the five most abundant states). **d**, Result of a blind test with SPR signal in RU predicted for a trimeric 7.2 × 14.3 × 16.0 nm antigen configuration and known-concentration intervals overlaid with the raw experimental (red) run. **e**, Occupancy of the predicted run stratified by state. **f**, Transient-state probability distribution for the trimeric antigen configuration. **g**, Transient evolution of monovalent (blue) and bivalent (red) contributions of antibodies to the average occupancy or number of antibodies per structure in the trimeric system. The cross between the two lines demonstrates the transition between regimes by changing the concentration. The legends list all or the five most prevalent states σ_i along with their corresponding occupancy ϕ_i and probabilities $_{pi}$ at the end of their respective run or contributions to entropy in the case of **e**.

ARTICLES

prevent opsonization⁴⁰. The apparent importance of spatial organization in immunological signaling suggests a role for non-equilibrium spatial phenomena such as those studied here, and we might

expect antigen organization itself to be under selective pressure during host-pathogen co-evolution. We suggest that the mechanism of stochastic walking predicted here might explain some of the pres-



NATURE COMPUTATIONAL SCIENCE



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 (sample size for **c** and **d** is 200 simulations of 10,000 s each). **e**, Histogram of antibody residence times on uniform 1D antigen array with a 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 marrow (blue) spacings (sample size for **e**, **f** and **g** is 100 simulations of 10,000 s each).

sures guiding the pathogen epitope organization, and such a mechanism might be exploited in the rational design of vaccines.

An inversion of this energy landscape phenomenon pertains to laterally mobile antigens such as the spike proteins in viral lipid envelopes. Mobile antigens would be expected to accommodate bivalent binding via lateral diffusion to achieve the minimum interconversion ratio K_{D2} . Another study derived a theoretical affinity optimum for mobile spike proteins that depends on their surface density, arguing that intermediate densities invoke the greatest immune response and that the low-spike-density characteristic of human immunodeficiency virus is the key to its immune evasion⁴¹. In this respect, a spatial tolerance model and experimental parameterization pipeline could aid vaccine development by informing design choices meant to elicit a precise immune response, for example, immunostimulatory virus-like particles^{35,42}. Our pipeline could also be used to dissect the complex state spaces of bi-, tri- or tetra-specific antibodies that are recently being developed for therapeutic and biosensing applications⁴³. We expect these molecules to exhibit complex binding behaviors, especially as many are engineered with non-Fc-based tethering regions of various flexibilities and lengths.

The capacity for emergent dynamics and programmable behavior makes antibody–antigen interactions a subject of greater potential complexity than previously thought. Experimentally parameterized modeling provides a reality-grounded sandbox for discovery, and we anticipate that future modeling pipelines coupled to other experimental technologies will bear fruit as this subject continues to be explored.

Methods

Statistics and reproducibility. Error assessment (Fig. 2g) was performed using a boostrapping method in which the mean and mean \pm one standard error of the mean input data were used to form the central, upper and lower inputs propagated to obtain the individual output points shown in the figure, with the upper and lower vertical error bars corresponding to the outputs produced by the upper and lower inputs, respectively. Goodness of fit for the phenomenological spatial tolerance function was characterized using an adapted chi-squared metric ('Mathematical description of spatial tolerance' section). The assessment of model robustness and predictive potential (Fig. 3 and Supplementary Figs. 7 and 8) was performed using a blinded test in which LT.H. performed the model parameterization and prediction of experimental SPR curves for an untested three-antigen triangular structure for three independent replicates with different corresponding bound structure amounts, whereas the experimental test data obtained by A.S. and I.S. were withheld until the predictions were submitted for comparison.

Overview of computational methods. Some methods and explanations may be found elsewhere¹⁸. However, in the following, we emphasize the original developments of this work including the following: a minimal parameterization pipeline for Markov models sensitive to arbitrary antigen spacings and its experimental validation; phenomenological model with analytical equation describing spatial tolerance as a continuous function; application of the model and fitting pipeline to two different antibodies: one from rabbit and the other, mouse; a systematic approach to determining the conversion factor between the SPR response units (RUs) to that of the number of bound antibodies per structure; steady-state analysis and determination of thermodynamic quantities from equilibrium-state probability distributions; a random-walk MCMC variant of the model that can be used to simulate larger systems with too many connected antigens to be feasible for enumerative approaches.

Briefly, the pipeline is executed in three parts. The first part requires the empirical estimation of the maximum SPR response due to the saturation of antibody-binding sites, based on the measured signal due to origami structures

binding to the surface and a standard curve constructed to relate structure binding to the maximum SPR response. This information is used as a conversion factor to relate the SPR signal for a given experiment to an average quantity of antibodies bound per structure. The second aspect of the pipeline is the fitting of a continuous-time Markov chain model to the SPR binding data of both oneantigen and two-antigen structures over a range of separation distances. This enables the construction of a parameterized analytical spatial tolerance equation that is used to predict the binding kinetics for arbitrary separation distances and antigen geometry. The third part of the pipeline entails the deployment of this fitted model for predictive purposes. The steady-state properties of a given system geometry are simulated by determining the distribution of states when the net flux between the states is zero. Large systems are simulated using an MCMC simulation that generates many individual trajectories of single antibodies walking on a user-specified pattern geometry. Mathematical explanation can be found in the following sections where this approach is described in more detail.

Model assumptions and constraints. We assume a coarse-grained model of binding states that equates all the physical states in which antibodies are bound by one arm as monovalent and which equates all the physical states in which antibodies bound by two arms as bivalent.

We assume a fixed amount of bound structures that does not change with time *t*, that is,

$$\frac{\mathrm{d}R_{\mathrm{struct}}}{\mathrm{d}t} \approx 0,\tag{2}$$

where R_{struct} is the SPR signal due to the structures and

$$\frac{\mathrm{d}n_{\mathrm{struct}}}{\mathrm{d}t} \approx 0, \tag{3}$$

where n_{struct} is the molar amount of structures.

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

$$\frac{\mathrm{d}c_{\mathrm{Ab}}}{\mathrm{d}t}\approx0,\tag{4}$$

where c_{Ab} is the concentration of the solution-phase antibody.

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.

$$\sigma_{-} \underset{k_{-1}}{\overset{c_{Ab}k_{1}}{\rightleftharpoons}} \sigma_{Ab}, \tag{5}$$

where σ_{\perp} is the state corresponding to an unoccupied structure; σ_{Ab} is the state corresponding to a structure with a bound antibody; and k_1 and k_{-1} are the association and dissociation rates, respectively.

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

$$n_{\rm struct} = n_{\rm -} + n_{\rm Ab}, \tag{6}$$

or we may work in terms of the molar amount of structures, both occupied and unoccupied (units of mol), where n_{Ab} is the number of bound antibody-structure complexes and n_{-} is the number of unoccupied structures; n_{Ab} and n_{-} are functions of *t*.

Therefore, the state probabilities are

$$p_{-} = \frac{n_{-}}{n_{\rm struct}} \tag{7}$$

and

$$p_{\rm Ab} = \frac{n_{\rm Ab}}{n_{\rm struct}},\tag{8}$$

corresponding to the unoccupied and one-antibody-occupied states, respectively.

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

$$\phi_{-} = 0 \tag{9}$$

and

$$\phi_{\rm Ab} = 1, \tag{10}$$

respectively. 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.

ARTICLES

$$\Phi = \sum_{i=1}^{N} p_i \phi_i \tag{11}$$

For the case of the one-antigen structure, this becomes

$$\Phi = p_{\rm Ab} \times 1 + p_{-} \times 0 = \frac{n_{\rm Ab}}{n_{\rm struct}}.$$
 (12)

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

$$R_{\rm Ab} = n_{\rm Ab}\xi_{\rm Ab},\tag{13}$$

where ξ_{Ab} is a conversion factor corresponding to the expected change in the SPR response signal per mole of bound antibodies.

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

$$\frac{\mathrm{d}n_{\mathrm{Ab}}}{\mathrm{d}t} = k_1 c_{\mathrm{Ab}} n_- - k_{-1} n_{\mathrm{Ab}} \tag{14}$$

The SPR signal after the structure-binding step is proportional to the molar amount of bound structures, that is,

$$R_{\rm struct} = n_{\rm struct} \xi_{\rm struct},\tag{15}$$

where ξ_{struct} is a conversion factor corresponding to the expected change in the SPR response signal per mole of structures.

Substituting RU-based expressions of molar amounts into equation (6), we get

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

Substituting RU-based expressions of molar amounts into equation (14) yields

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

which simplifies to

$$\frac{\mathrm{d}R_{\mathrm{Ab}}}{\mathrm{d}t} = \frac{\xi_{\mathrm{Ab}}k_{\mathrm{I}}c_{\mathrm{Ab}}R_{\mathrm{struct}}}{\xi_{\mathrm{struct}}} - k_{\mathrm{I}}c_{\mathrm{Ab}}R_{\mathrm{Ab}} - k_{-1}R_{\mathrm{Ab}}.$$
(18)

Since we have gathered both conversion constants into one term in equation (18), we now define the occupancy signal factor as

$$\xi_* \equiv \frac{\xi_{\rm Ab}}{\xi_{\rm struct}} \tag{19}$$

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

Note, by rearrangement, the relationship to average occupancy—that is, the occupancy signal factor—is the ratio of occupancy (in terms of SPR signal) to the molar quantities.

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

$$=\frac{R_{\rm Ab}}{R_{\rm struct}}\Phi^{-1}$$
(21)

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

$$\frac{\mathrm{d}R_{\mathrm{Ab}}}{\mathrm{d}t} = \xi_* k_1 c_{\mathrm{Ab}} R_{\mathrm{struct}} - k_1 c_{\mathrm{Ab}} R_{\mathrm{Ab}} - k_{-1} R_{\mathrm{Ab}}.$$
(22)

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

$$n_{\rm Ab}^{\rm max} = n_{\rm struct},\tag{23}$$

where n_{Ab}^{max} is the maximum number of moles of antibody that can bind to the system.

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

$$\xi_* = \frac{R_{\rm Ab}^{\rm max}}{R_{\rm struct}} \times 1, \tag{24}$$

where $R_{\rm Ab}^{\rm max}$ is the maximum SPR response signal due to the saturation of antibody-binding sites.

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

$$R_{\rm Ab}^{\rm max} = R_{\rm struct}\xi_*,\tag{25}$$

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

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

where $V_{\rm struct}$ is the valence.

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

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

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 the structure-binding signals and the maximum signals corresponding to saturated monovalent (one-antigen) structures, enabling conversion from SPR signal to occupancy in the absence of a well-understood binding model, provided the structure-binding signal is known.

The structure-bound signal (Supplementary Fig. 1a) is taken to be the difference between the signals before and after the structures are flowed over the chip and allowed to bind. Estimates of the parameters k_1 , k_{-1} and ξ , are supplied to a numerical minimization of the autocorrelation of residuals between the experimental and theoretical curves for the fourth-order Runge–Kutta approximation of equation (22), that is, the function $\frac{dR_{Ab}}{dt} = f(R_{Ab})$ recursively approximated according to the formula

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

where \boldsymbol{h} is a small timestep and the constituent terms have the form

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

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

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

$$\kappa_{n4} = f(R_{Ab,t} + h\kappa_n). \tag{32}$$

For each monovalent run (Supplementary Fig. 1b,c for rabbit and mouse antibodies, respectively) with a unique value of R_{struct} , a projected value of R_{Ab}^{max} is computed using equation (25). By fitting the monovalent models to 1–1 kinetics, we obtain the rate constants that allow the computational prediction of R_{Ab}^{max} (Supplementary Fig. 1d,e for rabbit and mouse antibodies, respectively) in the absence of experimental saturation conditions. 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} (Supplementary Fig. 1f,g for rabbit and mouse antibodies, respectively). We use this value as a conversion factor, enabling us to convert the SPR RUs into the number of antibodies per structure (Supplementary Fig. 1h, ifor rabbit and mouse antibodies, respectively). 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} .

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 the systems across experimental conditions in which the dynamic behavior can vary substantially. 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 the following:

NATURE COMPUTATIONAL SCIENCE

at the steady state,

$$\frac{\mathrm{d}R_{\mathrm{Ab}}}{\mathrm{d}t} = 0, \tag{33}$$

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

Rearranging equation (34) yields

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

and

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

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

$$K_{\rm D} = \frac{k_{-1}}{k_1}.$$
 (37)

Thus, at equilibrium, the SPR signal is

$$R_{\rm Ab^{eq}} = \frac{c_{\rm Ab} R_{\rm Ab}^{\rm max}}{K_{\rm D} + c_{\rm Ab}}.$$
(38)

An 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

$$\frac{1}{R_{\rm Ab^{eq}}} = \frac{K_{\rm D}}{c_{\rm Ab}R_{\rm Ab}^{\rm max}} + \frac{1}{R_{\rm Ab}^{\rm max}},\tag{39}$$

where $R_{Ab^{eq}}$ is the steady-state SPR signal due to the bound antibody.

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

From the dissociation constant, we know the occupancy as

$$\Phi_{\rm eq} = \frac{R_{\rm Ab^{\rm eq}}}{R_{\rm Ab}^{\rm max}} = \frac{c_{\rm Ab}}{K_{\rm D} + c_{\rm Ab}},\tag{40}$$

where Φ_{eq} is the expected occupancy at the steady state.

б

Such a concise description is desirable for complex structures as well. However, the difficulty arises in the case of multivalent structures that 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. In the case of the two-antigen structure, there are N=5 total states: one empty structure (σ_{-}), two states with one monovalently occupied antigen each (σ_{Ab} and σ_{Ab}), one state with both antigens bivalently occupied by one antibody (σ_{Ab}) and a state with both antigens monovalently occupied by antibodies (σ_{AbA}).

First, the reaction system can be represented according to the diagram in Fig. 1 or the set of reactions below:

$$\underbrace{ \overset{c_{Ab}k_1}{\underset{k_{-1}}{\underset{k_{-1}}{\underset{k_{-1}}}{\underset{k_{-1}}{\underset{k$$

$$\sigma_{Ab,_} \stackrel{c_{Ab}k_1}{\underset{k_{_1}}{\overset{c_{Ab}k_1}{\rightleftharpoons}}} \sigma_{Ab,Ab}; \sigma__Ab} \stackrel{c_{Ab}k_1}{\underset{k_{_1}}{\rightleftharpoons}} \sigma_{Ab,Ab}, \tag{42}$$

$$\sigma_{Ab,-} \underset{k_{-2}}{\overset{k_2}{\underset{k_{-2}}{\longrightarrow}}} \sigma_{.Ab,:} \sigma_{_Ab} \underset{k_{-2}}{\overset{k_2}{\underset{k_{-2}}{\longrightarrow}}} \sigma_{.Ab.}.$$
(43)

We have the two dissociation constants for the processes of monovalent (K_{D1}) binding and bivalent (K_{D2}) interconversion as follows.

$$K_{\rm D1} = \frac{k_{-1}}{k_1},\tag{44}$$

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

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

$$\frac{\mathrm{d}p_{\sigma_{-}}}{\mathrm{d}t} = -2c_{\mathrm{Ab}}k_{1}p_{\sigma_{-}} + k_{-1}p_{\sigma_{\mathrm{Ab}_{-}}} + k_{-1}p_{\sigma_{-\mathrm{Ab}}},\tag{46}$$

$$\frac{\mathrm{d}p_{\sigma_{\mathrm{Ab}_{-}}}}{\mathrm{d}t} = c_{\mathrm{Ab}}k_1 p_{\sigma_{-}} - c_{\mathrm{Ab}}k_1 p_{\sigma_{\mathrm{Ab}_{-}}} - k_2 p_{\sigma_{\mathrm{Ab}_{-}}} - k_{-1} p_{\sigma_{\mathrm{Ab}_{-}}} + k_{-1} p_{\sigma_{\mathrm{Ab}\mathrm{Ab}}} + k_{-2} p_{\sigma_{\mathrm{Ab}}},$$
(47)

$$\frac{\mathrm{d}p_{\sigma_{ab}}}{\mathrm{d}t} = c_{Ab}k_1p_{\sigma_{ab}} - c_{Ab}k_1p_{\sigma_{ab}} - k_2p_{\sigma_{ab}} - k_{-1}p_{\sigma_{ab}} + k_{-1}p_{\sigma_{AbAb}} + k_{-2}p_{\sigma_{Ab}},$$
(48)

$$\frac{\mathrm{d}p_{\sigma_{Ab.}}}{\mathrm{d}t} = k_2 p_{\sigma_{Ab_-}} + k_2 p_{\sigma_{Ab}} - 2k_{-2} p_{\sigma_{Ab.}},\tag{49}$$

$$\frac{\mathrm{d}p_{\sigma_{\mathrm{AbAb}}}}{\mathrm{d}t} = c_{\mathrm{Ab}}k_1p_{\sigma_{\mathrm{Ab}}} + c_{\mathrm{Ab}}k_1p_{\sigma_{\mathcal{A}b}} - 2k_{-1}p_{\sigma_{\mathrm{AbAb}}},\tag{50}$$

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

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

Given the 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 the 1–1 kinetics. However, for multiple concentrations, the equilibrium will not shift proportionately; therefore, the apparent binding constant is a function of the concentration from which the equilibrium value is derived, making its value dependent 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, that is,

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

This condition, in addition to 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 the analyte antibody, as follows:

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

$$p_{\sigma_{Ab_{-}}}^{eq} = \frac{c_{Ab}k_1k_{-1}k_{-2}}{c_{Ab}^2k_1^2k_{-2} + c_{Ab}k_1k_2k_{-1} + 2c_{Ab}k_1k_{-1}k_{-2} + k_{-1}^2k_{-2}},$$
(54)

$$\rho_{\sigma_{Ab}}^{eq} = \frac{c_{Ab}k_1k_{-1}k_{-2}}{c_{Ab}^2k_1^2k_{-2} + c_{Ab}k_1k_2k_{-1} + 2c_{Ab}k_1k_{-1}k_{-2} + k_{-1}^2k_{-2}^2},$$
(55)

$$p_{\sigma_{Ab}}^{eq} = \frac{c_{Ab}k_1k_2k_{-1}}{c_{Ab}^2k_1^2k_{-2} + c_{Ab}k_1k_2k_{-1} + 2c_{Ab}k_1k_{-1}k_{-2} + k_{-1}^2k_{-2}^2},$$
(56)

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

We can combine the states according to their corresponding occupancy—that is, the number of antibodies that the state contributes to the overall signal due to the bound antibody, where

$$p_{\sigma_{-}} + p_{\sigma_{Ab_{-}}} + p_{\sigma_{-Ab}} + p_{\sigma_{-Ab.}} + p_{\sigma_{AbAb}} = 1.$$
 (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:

$$\Phi_{\rm eq} = \sum_{i=1}^{N} p_{\sigma_i} \phi_{\sigma_i} = p_{\sigma_-}^{\rm eq} \times 0 + (p_{\sigma_{\rm Ab_-}}^{\rm eq} + p_{\sigma_{\rm Ab}}^{\rm eq} + p_{\sigma_{\rm Ab}}^{\rm eq}) \times 1 + p_{\sigma_{\rm AbAb}}^{\rm eq} \times 2.$$
(59)

Substituting equations (53) through (57), we arrive at

$$\Phi_{\rm eq} = \frac{c_{\rm Ab}k_1 \left(2c_{\rm Ab}k_1k_{-2} + k_{-1} \left(k_2 + 2k_{-2}\right)\right)}{c_{\rm Ab}^2 k_1^2 k_{-2} + c_{\rm Ab}k_1 k_2 k_{-1} + 2c_{\rm Ab}k_1 k_{-1} k_{-2} + k_{-1}^2 k_{-2}}.$$
(60)

Apparent dissociation constant. Taking the equilibrium occupancy of the twoantigen system 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

$$K_{\text{Dapp}} = \frac{c_{\text{Ab}}(1 - \Phi_{\text{eq}})}{\Phi_{\text{eq}}} = \frac{k_{-2} \left(k_{-1}^2 - c_{\text{Ab}}^2 k_1^2\right)}{k_1 \left(2c_{\text{Ab}}k_1 k_{-2} + k_2 k_{-1} + 2k_{-1} k_{-2}\right)}.$$
 (61)

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

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

which is a function of concentration and has a root at the critical value when $c_{Ab}K_{D2}^2 = K_{D1}^2$, that is, the point at which the average equilibrium occupancy greater than 1 is expected in the two-antigen system, and rendering any 1–1 kinetics description as impossible.

The rearrangement of equation (62) enables us to determine the interconversion constant from an apparent dissociation constant provided that we know the monovalent binding constant.

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

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

Mathematical description of spatial tolerance. Spatial tolerance refers to the favourability of bivalent antibody binding according to the spatial distribution of the two 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 the 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 the conversion to bivalent binding and/or favouring unbinding back to the monovalent states. Conversely, at larger separation distances, the molecule must stretch to accommodate the gap, again penalizing the conversion to bivalence and/or favouring 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, sharpness and degree of symmetry of the transitions to monovalent preference at close and far separations. Progressive fitting allows us to obtain K_{D2} for a single two-antigen system provided that we have determined K_{D1} for a one-antigen system, for which we take a mean run of multiple one-antigen runs (n=6) with \pm one and two standard errors of the mean away from the mean run as uncertainty intervals (Supplementary Fig. 3a,i for rabbit and mouse antibodies, respectively). Determining K_{D2} for different antigen separation distances gives us an empirical basis for spatial tolerance. We can phenomenologically model spatial tolerance with an equation for determining the interconversion constant K_{D2} as a function of the separation distance between two antigen x:

$$K_{\rm D2} = K_{\rm D2-tensile} + K_{\rm D2-compression},\tag{64}$$

where $K_{\rm D2-compression}$ (Supplementary Fig. 3b,j for rabbit and mouse antibodies, respectively) and $K_{\rm D2-tensile}$ (Supplementary Fig. 3c,k for rabbit and mouse antibodies, respectively) are the exponential and logistic terms, respectively. These separately model the decrease in interconversion due to the 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.

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

$$K_{\text{D2-tensile}} = \frac{K_{\text{D2}}^{\text{max}}}{1 + e^{-\alpha_{t}(x - \ell_{t})}},$$
(65)

where K_{D2}^{max} is an upper limit of the value of K_{D2} , α_i 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_i is the value of the midpoint of the sigmoidal

NATURE COMPUTATIONAL SCIENCE | VOL 2 | MARCH 2022 | 179-192 | www.nature.com/natcomputsci

<u>ARTICLES</u>

NATURE COMPUTATIONAL SCIENCE

curve that can be thought of as a characteristic length that defines the scale below which favorable interconversion occurs and above which the function approaches minimal interconversion.

The exponential compressive term has the form:

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

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

The combined expression yields equation (1), which predicts the interconversion constant as a function of separation distance (Supplementary Fig. 3d,e for rabbit antibodies and Supplementary Fig. 3l,m for mouse antibodies). Uncertainty represented with vertical error bars is due to variation in the one-antigen input data that bas been propagated to obtain different K_{D2} values fitted with correspondingly different K_{D1} values as constraints. The horizontal error bars represent uncertainty in the separation distance of protruding sites on DNA origami nanostructures, estimated according to the method employed elsewhere⁴⁴.

This can be converted into an effective or apparent dissociation constant as a 1-1 model on the basis of the bivalent model's prediction of equilibrium occupancy (Supplementary Fig. 3f,n for rabbit and mouse antibodies, respectively, and the 'Apparent dissociation constant' section). The propagation of one-antigen input data uncertainties yields slightly different parameterizations of the model due to the shifted K_{D2} values, and thus, we see a corresponding shift in the minimum of the function where bivalent binding is the strongest (Supplementary Fig. 3g,o for rabbit and mouse antibodies, respectively).

To assess the goodness of fit, we employ a chi-squared metric where the expected error is computed by projecting $K_{\rm D2}$ onto a straight axis by subtracting a three-point moving average. Random noise in the data should be approximately Gaussian distributed about the moving average, and we can thus compute a standard error of the mean E(x) from this straightened profile of the data. Supplementary Fig. 3h,p for rabbit and mouse antibodies, respectively, compare the distribution of $K_{\rm D2}$ minus the moving average (red) and model prediction (blue), a dispersion that should be Gaussian/random, if appropriately fitted. The chi-squared metric is

$$\chi^{2} = \sum_{x} \frac{(K_{\text{D2-obs}}(x) - K_{\text{D2-pred}}(x))^{2}}{E(x)},$$
(67)

where $K_{\text{D2-obs}}(x)$ is the observed interconversion value at a distance x and $K_{\text{D2-pred}}(x)$ is that predicted by the model. A good chi-squared metric should be neither much less than 1.000 (indicating overfitting) nor much greater than 1.000 (indicating poor fit).

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, that is, all the 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 as follows:

$$\mathbf{D} = \begin{pmatrix} 0 & d_{1,2} & d_{1,3} & \dots & d_{1,N} \\ d_{2,1} & 0 & d_{2,3} & \dots & d_{2,N} \\ d_{3,1} & d_{3,2} & 0 & \dots & d_{3,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N,1} & d_{N,2} & d_{N,3} & \dots & 0 \end{pmatrix}.$$
 (68)

This matrix can be simplified by applying a cutoff d_{crit} above which the antigens are considered too far apart to be neighbors. This reduces the number of possible states, eliminating those that are so unfavorable that they can be considered negligible.

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

Each state is linked to adjacent states by elementary transitions, that is, 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:

$$\sigma_i \underset{k_{-1}}{\overset{c_{ab}k_1}{\rightleftharpoons}} \sigma_j, \tag{69}$$

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:

$$\sigma_i \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \sigma_j. \tag{70}$$

Not all the states are necessarily connected. An adjacency matrix describes which states are connected by transitions.

$$\mathbf{A}_{i,j} = \begin{cases} 1 & \text{, if } \exists \text{ connection between } \sigma_i \text{ and } \sigma_j \\ 0 & \text{, otherwise} \end{cases}$$
(71)

The system parameters are the set of zero-order transition rates $\{\lambda_1 = c_{Ab}k_1, \lambda_{-1} = k_{-1}, \lambda_2 = k_2, \lambda_{-2} = k_{-2}\}$. A multi-antigen–antibody system is, thus, fully described by the continuous-time Markov model (\mathbb{S} , Λ) defined as the set of states and its corresponding transition rate matrix of the form:

$$\mathbf{\Lambda} = \begin{cases} \lambda(i,j) & \text{, if } A_{i,j} = 1\\ 0 & \text{, if } A_{i,j} = 0 \end{cases}$$
(72)

The automated enumeration of states and their connections in systems of arbitrary antigen pattern geometry are accomplished using an implementation of the breadth-first search algorithm. The algorithm searches for edges between adjacent states and assigns the appropriate elementary rate process. A queue of neighboring states is made on 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 the 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 the 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.

Supplementary Algorithm 1 describes the process by which the states and transitions are discovered starting from a single starting state. Here states are distinguished by the status of each of the sites ζ_k in the pattern, being either empty, monovalently occupied or bivalently occupied, as well as connected to another adjacent site ζ_s . The colored text is used to separate the different classes of transition.

The time complexity of the breadth-first search can be expressed as O(|V| + |E|), where |V| and |E| are the numbers of vertices and edges, respectively. In the case of antigen patterns, the former correspond to the number of antigens in the pattern. The latter correspond to the number of adjacent pairwise connections that are possible between two antigens. This is determined by the bivalent flexibility of the antibody in question; as a rule of thumb, we could say that antigens further than 25 nm apart are not close enough for bivalent bonds to form.

Transient (non-equilibrium) dynamics of enumerative PSPR models. The continuous-time Markov model enables us to compute the transient evolution of a system. The probability distribution

$$\mathbf{p}(t) = \begin{pmatrix} p_0(t) \\ p_1(t) \\ \vdots \\ p_N(t) \end{pmatrix}$$
(73)

is a vector whose elements $_{p_i}(t)$ are the probabilities of the respective system states $\sigma_{0}, \sigma_1, \ldots, \sigma_N$ at time *t*. A uniform probability distribution would, for example, represent equal probabilities of finding a structure in any one of the possible states. 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 the other states.

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

$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) \cdot \mathbf{Q},\tag{74}$$

which uses an infinitessimal generator matrix **Q** obtained from the rate matrix and used to determine the relative rates at which state the probabilities change with incremental time.

$$\mathbf{Q}_{i,j} = \begin{cases} \Lambda_{i,j} &, \text{ for } i \neq j \\ -\sum_{i \neq j} \Lambda_{i,j} &, \text{ for } i = j \end{cases}$$
(75)

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

$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) \cdot \mathbf{e}^{\mathbf{Q} \cdot \Delta t} = \mathbf{p}(t) \sum_{\eta=0}^{\infty} \frac{(\mathbf{Q} \cdot \Delta t)^{\eta}}{\eta!},$$
(76)

where η is the computation's depth of recursion—the higher it is, the more accurate the result is; Δt is an incremental advancement in time. Due to the numerical instability of this solution, we employ the uniformized discrete-time Markov chain method of Fox and Glynn to stably compute equation (76) (ref. ⁴⁵). The continuous-time Markov model (S, Λ) is approximated by a discrete model (S, U) by renormalizing the generator matrix with respect to the fastest outgoing rate or the uniformization rate q:

$$\mathbf{U} := \mathbf{I} + \frac{\mathbf{Q}}{q}, \ q \ge \max_{i} \{ |\mathbf{Q}_{i,i}| \},\tag{77}$$

where I is the identity matrix.

. .

Supplementary Algorithm 2 shows how this matrix is generated in practice. Equation (76) becomes the approximation

$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) \cdot \mathbf{e}^{\mathbf{Q} \cdot \Delta t} = \mathbf{p}(t) \cdot \mathbf{e}^{q(\mathbf{U} - \mathbf{I}) \cdot \Delta t} = \mathbf{p}(t) \cdot \mathbf{e}^{q\mathbf{U}\Delta t} \mathbf{e}^{-q\mathbf{I}\Delta t} = \mathbf{p}(t)\mathbf{e}^{-q\Delta t} \cdot \mathbf{e}^{q\mathbf{U}\Delta t}.$$
(78)

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

$$\mathbf{p}(t+\Delta t) = \mathbf{p}(t) \cdot \mathbf{e}^{q\Delta t} \cdot \sum_{\eta=0}^{\infty} \frac{(q\Delta t\mathbf{U})^{\eta}}{\eta!} = \sum_{\eta=0}^{\infty} \frac{\mathbf{e}^{-q\Delta t}(q\Delta t)^{\eta}}{\eta!} \cdot \mathbf{p}(t) \cdot \mathbf{U}^{\eta}$$
(79)

Using equation (79), we can stably compute the transient evolution of a system from an initial condition.

The system entropy can by computed as

$$S = -k_{\rm B} \sum_{\sigma_i \in \mathbb{S}} p_i(t) \ln[p_i(t)].$$
(80)

where $k_{\rm B}$ is the Boltzmann constant.

Supplementary Algorithm 3 shows how the probability distributions at different timepoints are computed from an initial condition.

Fitting continuous-time Markov models to PSPR data using autocorrelation of residuals. By using equation (79) to compute the transient probability distribution of the system, we are able to also compute the occupancy at each timepoint using the definition from equation (11). The system occupancy is, thus, a function of time of the form

$$\Phi(t) = \sum_{i=1}^{N} p_i(t)\phi_i.$$
(81)

The continuous-time Markov model is fitted to the experimental data by comparing occupancies computed on the basis of equations (79) and (81) with that of the occupancy computed from the normalizing PSPR data via equation (27); evidently, the theoretical curve either correctly or incorrectly fits the experimental data depending on the parameterization (Supplementary Figs. 5a and 6a). The residuals (Supplementary Figs. 5b and 6b) are computed by

$$e(t) = \Phi(R_{\rm Ab}) - \Phi(t, \mathbb{S}, \mathbf{U}).$$
(82)

Although fitting by minimizing the sum of the squared residuals can be used to obtain acceptable model parameterizations, we used residual analysis with autocorrelation to improve the robustness of fitting and reduce the systematic misparameterization 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 as

$$\rho_{e,e}(t,t+k\Delta t) = \frac{\sqrt{(\sum_{k} \mathbf{e} \cdot |\vec{v}|)^{2}}}{k}$$
(83)

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

$$\mathcal{E} = \sum_{t=0}^{t_{\rm f}-k+1} \rho_{e,e}(t,t+k) \sqrt{e(t)^2}$$
(84)

ARTICLES

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

We performed a cross-validation of the Markov model fitting by parameterizing based on the experimental data from various antigen patterns (all with fixed nearest-neighbor separation distances between the 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 the extrapolation of a parameterized model to different antigen pattern geometries. The absolute sums of residuals (Supplementary Fig. 7a) and absolute sums of residuals weighted by autocorrelation (Supplementary Fig. 7b) show that the best-performing models were those that are the most complex and exhibit bivalence such as hexagonal and pentagonal configurations in the last two rows. This suggests that downward extrapolation in pattern complexity is more viable than upward.

To validate the spatial tolerance model, we conducted a blinded test in which an a priori prediction was made using the spatial tolerance parameters (Fig. 2f). This was done with the rabbit anti-DIG IgG antibody. The data used for prediction consisted entirely of the averaged one-antigen data and the series of two-antigen varied-separation-distance data used to parameterize the model and the structure-binding data to determine the conversion factor from RU to occupancy, and vice versa. Thus, no data with structures configured with more than two antigens was used for parameterization. To perform the test, we chose to predict the evolution of a trimeric $7.2 \times 14.3 \times 16.0$ nm configuration with the same single-cycle kinetics protocol of timed known-concentration injections used for the other runs in this study (Supplementary Fig. 7c). Predictions were made first by computing the expected occupancy values on a per-structure basis. These were then converted into SPR RUs by multiplying them with the R_{Ab}^{max} value determined from the structure-binding curve and standard curve (Supplementary Fig. 1). The experimental results were withheld until predictions were made, and were then revealed and compared with the theoretical curves each done in triplicate with the respective structure-binding data used for each one (Supplementary Fig. 7d).

An important question to consider is how sensitive the stratified-state probability distribution predictions are to error in the antibody concentration. As shown earlier, the distribution is dependent on concentration and timing; however, the answer is probably fairly complex. This is because different phases (Fig. 3a-c) are structured in a complex manner in terms of their concentration intervals and shape of transitions between phases. To get only a very rough idea of the sensitivity though, what we have done for the triangle structure is to change each of the concentrations by factors of 0.5, 0.9, 1.1 and 2.0 (Supplementary Fig. 8a) to see how this affects both the resulting transient probability distribution (Supplementary Fig. 8b) and the relative correspondence of the predicted curve to that of the experimental data (Supplementary Fig. 8c). Evidently, the probability distribution is fairly robust, with the five most represented states remaining unchanged in all the five conditions, albeit their relative ranking changes slightly, for example, going from 1.1x to 2.0x.

Determination of thermodynamic properties. We can obtain the equilibrium probabilities from uniformized continuous-time Markov chain first by simulation out to long timescales at a fixed solution concentration until the probabilities cease to change.

$$\mathbf{p}^* = \lim_{t \to \infty} \mathbf{p}(t), \ \left(\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}(t)\right)_{c_{\mathrm{Ab}}} = 0$$
(85)

We can determine the steady-state probability distribution more expediently on the basis of the infinitessimal generator matrix, that is, equation (75), numerically solving for the probability distribution which—when multiplied with the generator matrix—produces a vector of zeros, meaning that there is zero change from one moment to the next, subject to the normalization condition in which all the probabilities must sum to 1. That is, the steady-state probability distribution is the solution to the matrix equation

$$\mathbf{p} \cdot \mathbf{Q} = 0, \ \sum_{\sigma_i \in \mathbb{S}} p_i = 1.$$
(86)

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 the steady state by exchanging antibodies with the bath, obeying the Boltzmann distribution law

$$p_{i}^{*} = \frac{e^{-E_{i}/k_{\rm B}T}}{\sum_{\sigma_{i} \in \mathbb{S}} e^{-E_{i}/k_{\rm B}T}} = \frac{e^{-E_{i}/k_{\rm B}T}}{\mathbb{Z}},$$
(87)

where \mathbb{Z} is a grand canonical partition function that 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 $E_{ipi} = \mu_{Ab} + \mu_{mono} n_{mono} + \mu_{biv} n_{biv}$ are the state energies

NATURE COMPUTATIONAL SCIENCE

determined by the environmental potential due to solution-phase antibody concentration and individual potentials of antibody monovalent and bivalent bonds populating the state; further, n_{mono} denotes the monovalent bonds and $n_{\rm hiv}$ denotes the bivalent bonds with chemical potentials $\mu_{\rm mono}$ and $\mu_{\rm hiv}$ respectively. The value of \mathbb{Z} and the state energies are numerically solved for a given steady-state probability distribution, making it possible for us to determine the thermodynamic quantities.

We can obtain the thermodynamic quantities such as the solutionconcentration-dependent equilibrium grand potential free energy via

$$\Omega^* = -k_{\rm B}T\ln\mathbb{Z}.\tag{88}$$

The equilibrium probabilities enable us to calculate the relative potential differences as

$$\frac{p_i^*}{p_j^*} = e^{-(E_i - E_j)/k_{\rm B}T}.$$
(89)

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

$$\mu_{0-1} = E_{\text{mono}} - E_{\text{empty}} = \left(\frac{\partial \Omega}{\partial N_{\text{mono}}}\right)_{T,P,N_{\text{biv}}} = -k_{\text{B}}T \ln \left[\sum_{\substack{n_{\text{mono}}\\n_{\text{empty}}}}^{\infty} p_{\text{mono}}^{*}\right] + k_{\text{B}}T \ln c_{\text{Ab}} = -k_{\text{B}}T \ln \left[\frac{n_{\text{mono}}}{n_{\text{empty}}K_{\text{D1}}}\right],$$
(90)

$$\mu_{1-2} = E_{\text{biv}} - E_{\text{mono}} = \left(\frac{\partial\Omega}{\partial N_{\text{biv}}}\right)_{T,P,N_{\text{empty}}} = -k_{\text{B}}T \ln\left[\frac{\sum\limits_{\substack{n \neq \text{biv} \\ n \neq \text{biv}}}}{\sum\limits_{\substack{n \neq \text{biv} \\ n \neq \text{mono}}} p_{\text{mono}}^*}\right] = -k_{\text{B}}T \ln\left[\frac{n_{\text{biv}}}{n_{\text{mono}}K_{\text{D2}}}\right],$$
(91)

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

$$\mu_{0-2} = E_{\text{biv}} - E_{\text{empty}} = \mu_{0-1} + \mu_{1-2} = \left(\frac{\partial\Omega}{\partial N_{\text{biv}}}\right)_{T,P,N_{\text{mono}}} = -k_{\text{B}}T\ln\left[\frac{\sum\limits_{n_{\text{biv}}} P_{\text{biv}}^{*}}{P_{\text{empty}}^{*}}\right] + k_{\text{B}}T\ln c_{\text{Ab}}$$
(92)

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

MCMC version of the model. Systems with larger numbers of bivalent connections have many states and transitions, and the fully enumerative continuous-time Markov chain does not scale well. For these systems, we use the MCMC sampling approach in which only the 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 to a higher precision for the enumerative method. Instead of computing the flux of state probabilities over fixed intervals of time, we instead computed Poisson intervals (dwell times) of states according to the rates of processes that each state is subject to here. For any particular state σ_p there is a set of adjacent states $\sigma_j \forall j$ such that $\mathbf{A}_{i,j} \neq \mathbf{0}$.

The exit rate of that state is then the summation of all the outgoing rates:

$$\lambda_{\text{exit}} = \sum_{\forall j} \lambda_{i,j} | \mathbf{A}_{\mathbf{i},\mathbf{j}} \neq \mathbf{0}, \tag{93}$$

and the corresponding dwell time in that state comes from the exponential cumulative distribution function

$$\tau = \frac{\ln[1/p]}{\lambda_{\text{exit}}},\tag{94}$$

where *p* is the probability that a transition to a neighboring state occurs within the time interval and dwell time τ is a random variable that we may sample by choosing random values of *p* from the interval [0, 1].

The choice of state given that a transition occurs is then a matter of the relative rates λ_{ij} for the different states σ_j , with each state having a probability $p_j = \frac{\lambda_{ij}}{\lambda_{exit}}$ of becoming the destination state.

A single iteration, thus, starts with an initial state, followed by the enumeration of each of the adjacent states, and random sampling to determine both dwell time and destination after the transition. The simulation involves performing multiple random walks and keeping track of the occupancy and state distribution over a discretized timeline.

Supplementary Algorithm 4 shows how this random-walk MCMC approach is used to simulate the antibody dynamics by computing the dwell times and local state-to-state transitions.

Experimental methods. Some experimental data used in this work were presented previously¹⁸ and were not analysed 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 a *de novo* simulation that goes beyond the previous work.

Reagents. Oligonucleotides (unmodified and DIG modified) in 96-well plates were purchased from IDT. Chemicals (NaCl, KCl, MgCl₂, Tris-HCl, EDTA, PEG800, NaOH, KOAc, KOH and NaOAc) for buffer preparation were purchased from Sigma-Aldrich. Rabbit anti-DIG IgG (no. 9H27L19) was purchased from Thermo Scientific. Streptavidin was purchased from Sigma-Aldrich and mouse anti-DIG IgG1 (no. ab420) was purchased from Abcam. Phosphate-buffered saline (1 M stock solution) was purchased from Sigma-Aldrich. BIAcore consumables (CM3 sensor chip and HBS-EP (HEPES 10 mM, NaCl 150 mM, EDTA 3 mM, 0.005% Tween 20) running buffer) were purchased from GE Healthcare. Amicon centrifugal filters with 100 kDa molecular weight cutoff were purchased from Merck Millipore.

Origami design and production. The 18-helix bundle DNA origami nanotube was designed with caDNAno⁴⁶ using the honeycomb lattice. This structure has been characterized earlier^{18,33,447}. In short, the p7560 scaffold was extracted from the M13 phage, and the 18-helix bundle DNA nanotube was folded under the following conditions: 20 nM scaffold, 100 nM each staple oligonucleotide, 13 mM MgCl₂, 5 mM Tris at 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 to 60 °C over 20 min and 60 to 24 °C over 14 h. The excess staples were removed by ultrafiltration with Amicon 100 kDa molecular weight cutoff filters. The wash buffer used was 1× phosphate-buffered saline supplemented with 10 mM MgCl₂.

PSPR protocol. A BIAcore t200 instrument (GE Healthcare) was used for the SPR experiments. The running buffer used in all the experiments is HBS-EP supplemented with 10 mM MgCl₂. The flow rate used for all the kinetics experiments is 30 µl min-1. Streptavidin was diluted to a final concentration of 10 µg ml-1 in 10 mM NaOAc at pH 4.5 and chemically attached to the CM3 sensor chip with N-hydroxysuccinimide and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide coupling (using the standard protocol from GE Healthcare). Anchor oligonucleotides containing a 3'-biotin modification were diluted to 200 nM in 1× HBS-EP running buffer and injected over the surface for 20 min followed by the washing of non-specifically bound oligos by injecting 50 mM NaOH for 5 min. The DNA nanostructures were diluted to 5 nM and injected over the surface for 20 mins followed by washing with the running buffer for 10 min. The antibodies were diluted to various concentrations in the running buffer, ranging from 0.025 to 0.500 nM. The single-cycle kinetics injection method was used to sequentially inject the antibody solution over the surface, starting 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 min; then, the surface is ready for the next round of experiments. The t200 evaluation software was initially used to fit the acquired data; for this, we used a 1:1 Langmiur binding model to fit the data and estimate the values of k_{a} , k_{d} , the association and dissociation rates, respectively, and dissociation constant K_{D} and antibody-binding capacity (Supplementary Section 6 details the apparent dissociation constant).

Run design. The parameterization pipeline first requires a dosing scheme for the single-cycle kinetics program, that is, the timing and concentrations of the staged antibody injections into the system. Since different antibody–antigen systems are going to exhibit different kinetic profiles, we believe the following considerations could help inform the initial choice of dosing scheme.

First, an approximate knowledge of $K_{\rm D1}$ is probably the best starting point. Some antibody suppliers report an in-house measured $K_{\rm D}$ ($K_{\rm D1}$ by our terminology) or else report those values published by researchers who used their product, and this value can also be determined with a single ELISA experiment⁴⁸. Knowing this, it is possible to choose a dosing scheme that elicits both monovalent ($K_{\rm D1}$ -dominated) kinetics. $K_{\rm D1}$ -dominated kinetics occur when the magnitude of $K_{\rm D1}/c_{\rm Ab}$ is smaller than $K_{\rm D2}$, that is, at higher

relative concentrations. However, $K_{\rm D2}$ -dominated effects are the most apparent at lower concentrations or during the dissociation phase when the concentration is set to zero. Although $K_{\rm D2}$ is molecule specific, this value is probably subject to less variation among the commonly used isotypes. Therefore, even if $K_{\rm D2}$ is unknown at first, it may be a reasonable starting estimate to assume that it is similar to the values found in our study, that is, of the order of 10^{-2} around the optimal separation distances. Thus, framing a dosing series based on a supplier's reported $K_{\rm D1}$ value and concentrations that span a range where relative $K_{\rm D1}/c_{\rm Ab}$ to $K_{\rm D2}$ goes through an inversion is likely to capture a useful range of kinetics well suited to parameterizing the model. The following concentrations and corresponding injection timings were used for most experiments in our study: timepoints (s): 0, 84, 384, 475, 775, 866, 1,166, 1,257, 1,557, 1,656 and 1,956; concentrations (nM): 0, 0,025, 0, 0.050, 0, 0.100, 0, 0, 250, 0, 0.500 and 0.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw experimental data used to produce the results of this study can be found at https://github.com/Intertangler/spatial_tolerance/tree/master/data_repository and Zenodo⁴⁹ under the subfolder 'Data repository'. Data are licensed under the GNU General Public License. Source data for Figs. 2–4 are provided with this paper. In addition to the raw data, the GitHub/Zenodo repository contains the Jupyter Notebooks detailing the generation of our results including intermediate data and figures and plots are posted in ready-to-run form for reproduction.

Code availability

All the codes used to produce the results of this study including installation, demonstration and result reproduction instructions are available on GitHub (https://github.com/Intertangler/spatial_tolerance) and Zenodo⁴⁹. The code is licensed under the GNU General Public License.

Received: 14 July 2021; Accepted: 18 February 2022; Published online: 24 March 2022

References

- 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, 3213–3219 (2013).
- Macken, C. A. & Perelson, A. S. Aggregation of cell surface receptors by multivalent ligands. J. Math. Biol. 14, 365–370 (1982).
- Rheinnecker, M. et al. Multivalent antibody fragments with high functional affinity for a tumor-associated carbohydrate antigen. *J. Immunol*.157, 2989–2997 (1996).
- Uetrecht, C., Barbu, I. M., Shoemaker, G. K., Van Duijn, E. & Heck, A. J. R. Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nat. Chem.* 3, 126–132 (2011).
- Emini, E. A., Ostapchuk, P. P. & Wimmer, E. Bivalent attachment of antibody onto poliovirus leads to conformational alteration and neutralization. *J. Virol.* 48, 547–550 (1983).
- Thouvenin, E. et al. Bivalent binding of a neutralising antibody to a calicivirus involves the torsional flexibility of the antibody hinge. *J. Mol. Biol.* 270, 238–246 (1997).
- Boot, H. J. & Pouwels, P. H. Expression, secretion and antigenic variation of bacterial S-layer proteins. *Mol. Microbiol.* 21, 1117–1123 (1996).
- Taverniti, V. et al. S-layer protein mediates the stimulatory effect of Lactobacillus helveticus MIMLh5 on innate immunity. Appl. Environ. Microbiol. 79, 1221–1231 (2013).
- Konstantinov, S. R. et al. S layer protein A of *Lactobacillus acidophilus* NCFM regulates immature dendritic cell and T cell functions. *Proc. Natl Acad. Sci.* USA 105, 19474–19479 (2008).
- Kajikawa, A. et al. Mucosal immunogenicity of genetically modified Lactobacillus acidophilus expressing an HIV-1 epitope within the surface layer protein. PLoS ONE 10, e0141713 (2015).
- Suzuki, S., Yokota, K., Igimi, S. & Kajikawa, A. Comparative analysis of immunological properties of S-layer proteins isolated from *Lactobacillus* strains. *Microbiology* 165, 188–196 (2019).
- Veneziano, R. et al. Role of nanoscale antigen organization on B-cell activation probed using DNA origami. Nat. Nanotechnol. 15, 716–723 (2020).
- LoBue, A. D. et al. Multivalent norovirus vaccines induce strong mucosal and systemic blocking antibodies against multiple strains. *Vaccine* 24, 5220–5234 (2006).
- Baschong, W., Hasler, L., Häner, M., Kistler, J. & Aebi, U. Repetitive versus monomeric antigen presentation: direct visualization of antibody affinity and specificity. *J. Struct. Biol.* 143, 258–262 (2003).
- Müller, K. M., Arndt, K. M. & Plückthun, A. Model and simulation of multivalent binding to fixed ligands. *Anal. Biochem.* 261, 149–158 (1998).

- De Michele, C., De Los Rios, P., Foffi, G. & Piazza, F. Simulation and theory of antibody binding to crowded antigen-covered surfaces. *PLoS Comput. Biol.* 12, e1004752 (2016).
- Bongini, L. et al. A dynamical study of antibody-antigen encounter reactions. *Phys. Biol.* 4, 172 (2007).
- Shaw, A. et al. Binding to nanopatterned antigens is dominated by the spatial tolerance of antibodies. *Nat. Nanotechnol.* 14, 184–190 (2019).
- 19. Pfeiffer, M. et al. Single antibody detection in a DNA origami nanoantenna. *iScience* 24, 103072 (2021).
- Zhang, P. et al. Capturing transient antibody conformations with DNA origami epitopes. *Nat. Commun.* 11, 3114 (2020).
- Grant, R. A. et al. Structures of poliovirus complexes with anti-viral drugs: implications for viral stability and drug design. *Curr. Biol.* 4, 784–797 (1994).
- 22. Munoz, N., Castellsagué, X., Berrington de González, A. & Gissmann, L. HPV in the etiology of human cancer. *Vaccine* **24**, S1–S10 (2006).
- 23. Modis, Y., Trus, B. L. & Harrison, S. C. Atomic model of the papillomavirus capsid. *EMBO J.* **21**, 4754–4762 (2002).
- Hewat, E. A., Marlovits, T. C. & Blaas, D. Structure of a neutralizing antibody bound monovalently to human rhinovirus 2. J. Virol. 72, 4396–4402 (1998).
- 25. Carl, M. and Dasch, G. A. The importance of the crystalline surface layer protein antigens of rickettsiae in T-cell immunity. in *T-Cell Activation in Health and Disease* 81–91 (Academic Press, 1989).
- Bahl, H. et al. IV. Molecular biology of S-layers. FEMS Microbiol. Rev. 20, 47–98 (1997).
- Ilk, N., Egelseer, E. M. & Sleytr, U. B. S-layer fusion proteins—construction principles and applications. *Curr. Opin. Biotechnol.* 22, 824–831 (2011).
- Hinton, H. J., Jegerlehner, A. and Bachmann, M. F. Pattern recognition by B cells: the role of antigen repetitiveness versus toll-like receptors. *Specialization* and Complementation of Humoral Immune Responses to Infection 1–15 (2008).
- Hewat, E. A. & Blaas, D. Structure of a neutralizing antibody bound bivalently to human rhinovirus 2. *EMBO J.* 15, 1515–1523 (1996).
- Edeling, M. A. et al. Potent dengue virus neutralization by a therapeutic antibody with low monovalent affinity requires bivalent engagement. *PLoS Pathog.* 10, e1004072 (2014).
- Lee, P. S. et al. Heterosubtypic antibody recognition of the influenza virus hemagglutinin receptor binding site enhanced by avidity. *Proc. Natl Acad. Sci.* USA 109, 17040–17045 (2012).
- Preiner, J. et al. IgGs are made for walking on bacterial and viral surfaces. Nat. Commun. 5, 4394 (2014).
- Shaw, A. et al. Spatial control of membrane receptor function using ligand nanocalipers. *Nat. Methods* 11, 841–846 (2014).
- Shaw, A., Benson, E. & Högberg, B. Purification of functionalized DNA origami nanostructures. ACS Nano 9, 4968–4975 (2015).
- Knappe, G. A., Wamhoff, E.-C., Read, B. J., Irvine, D. J. & Bathe, M. In situ covalent functionalization of DNA origami virus-like particles. ACS Nano 15, 14316–14322 (2021).
- Hellmeier, J. et al. Strategies for the site-specific decoration of DNA origami nanostructures with functionally intact proteins. ACS Nano 15, 15057–15068 (2021).
- Lee, S. Y., Choi, J. H. & Xu, Z. Microbial cell-surface display. *Trends Biotechnol.* 21, 45–52 (2003).
- Mikawa, Y. G., Maruyama, I. N. & Brenner, S. Surface display of proteins on bacteriophage λ heads. J. Mol. Biol. 262, 21–30 (1996).
- Murugan, R. et al. Evolution of protective human antibodies against *Plasmodium falciparum* circumsporozoite protein repeat motifs. *Nat. Med.* 26, 1135–1145 (2020).
- KumraAhnlide, V., de Neergaard, T., Sundwall, M., Ambjörnsson, T. & Nordenfelt, P. A predictive model of antibody binding in the presence of IgG-interacting bacterial surface proteins. *Front. Immunol.* 12, 661 (2021).
- Amitai, A., Chakraborty, A. K. & Kardar, M. The low spike density of HIV may have evolved because of the effects of T helper cell depletion on affinity maturation. *PLoS Comput. Biol.* 14, e1006408 (2018).
- Hellmeier, J. et al. DNA origami demonstrate the unique stimulatory power of single pMHCs as T cell antigens. *Proc. Natl Acad. Sci. USA* 118, e2016857118 (2021).
- Kontermann, R. E. & Brinkmann, U. Bispecific antibodies. Drug Discov. Today 20, 838–847 (2015).
- Reuss, M. et al. Measuring true localization accuracy in super resolution microscopy with DNA-origami nanostructures. *New J. Phys.* 19, 025013 (2017).
- Fox, B. L. & Glynn, P. W. Computing Poisson probabilities. Commun. ACM 31, 440–445 (1988).
- Douglas, S. M. et al. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* 37, 5001–5006 (2009).
- 47. Zhao, Y.-X. et al. DNA origami delivery system for cancer therapy with tunable release properties. *ACS Nano* **6**, 8684–8691 (2012).
- Heinrich, L., Tissot, N., Hartmann, D. J. & Cohen, R. Comparison of the results obtained by ELISA and surface plasmon resonance for the determination of antibody affinity. *J. Immunol. Methods* 352, 13–22 (2010).

ARTICLES

NATURE COMPUTATIONAL SCIENCE

 Hoffecker, I. Intertangler/spatial_tolerance: spatial tolerance initial release. Zenodo https://doi.org/10.5281/zenodo.5961096 (2022).

Acknowledgements

We would like to acknowledge F. Fördős for helpful discussions. We would like to acknowledge support from A. W. Stiftelsen (M19-0547) to I.T.H. and the Knut and Alice Wallenberg foundation (KAW 2017.0114 and KAW 2017.0276) and from the European Research Council ERC (GA no. 724872) to B.H.

Author contributions

I.T.H. conceived and implemented the core model and carried out the computational experiments. A.S. and I.S. obtained and preprocessed the PSPR data. I.T.H. and V.S. developed and tested the fitting pipeline and the MCMC simulation. I.T.H., A.S. and B.H. conceived the study concept. I.T.H., A.S., V.S., I.S. and B.H. wrote the manuscript.

Funding

Open access funding provided by Royal Institute of Technology.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-022-00218-z.

Correspondence and requests for materials should be addressed to Ian T. Hoffecker or Björn Högberg.

Peer review information *Nature Computational Science* thanks Pontus Nordenfelt, Philippe Robert and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Handling editor: Ananya Rastogi, in collaboration with the *Nature Computational Science* team. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long

as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

nature research

Corresponding author(s): Ian Hoffecker, Björn Högberg

Last updated by author(s): Feb 17, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

Statistics

For	all st	atistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.		
n/a	/a Confirmed			
	\boxtimes	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement		
\boxtimes		A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly		
\boxtimes		The statistical test(s) used AND whether they are one- or two-sided Only common tests should be described solely by name; describe more complex techniques in the Methods section.		
\boxtimes		A description of all covariates tested		
\boxtimes		A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons		
	\boxtimes	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)		
\boxtimes		For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted Give <i>P</i> values as exact values whenever suitable.		
\boxtimes		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings		
\boxtimes		For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes		
\boxtimes		Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated		
Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.				

Software and code

Policy information	about <u>availability of computer code</u>
Data collection	Physical data was collected using BIAcore t200 v 3.0 evaluation software was used for the initial preprocessing of data.
Data analysis	Data analysis was performed using custom code written in Python that can be accessed in its entirety at https://github.com/Intertangler/spatial_tolerance including visual workflows for deploying the code and installation instructions.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The raw data used to produce the results of this study are available along with at https://github.com/Intertangler/spatial_tolerance and via Zenodo with DOI: 10.5281/zenodo.5961096. Figures 2,3, and 4 have associated raw data.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>

Life sciences study design

All studies must disclose on these points even when the disclosure is negative. The sample size for Monte Carlo experiments were conducted many times in order to reach an arbitrarily high number such that results would Sample size indicate an obvious qualitative difference between conditions not reliant on statistical comparison. The sample size for each point on our analytical spatial tolerance function is 1 though multiple points spanning a range of separation distances are covered, some very close to each other in distance, so noise in the KD2 values are incorporated into the fit of the analytical spatial tolerance function. This can be observed by the random distribution of roughly equal portions of datapoints lying above and below the analytical function. No data were excluded from analysis. Data exclusions Replication The Methods section entitled "Fitting continuous time Markov models to PSPR data using autocorrelation of residuals" describes our investigation of model robustness to extrapolation and reproducibility using cross validation. Briefly, a training dataset was used to parameterize our model, while a test dataset was used to verify performance on non-training data. The datasets used were also independent from the datasets used to produce the main results of the paper, demonstrating reproducibility of the model workflow. Reproduction was successful. For more information about the limits of model performance and extrapolation, please see the Methods and Supplementary Figure 7 and 8. The model construction pipeline used for Figure 2 was replicated one additional time using mouse monoclonal antibody, shown in Supplementary Figure 5. Monovalent runs used for the initial parameterization in Figure 2 and Supplementary Figure 5 were each performed 6 times as independent physical experiments. Bivalent runs for Figure 2 and Supplementary Figure 5 were performed once for each separation distance. The blind trial presented in Figure 3 was performed in 3 independent trials, shown as a complete set in Supplementary Figure 7. Randomization Randomization in collection of data was achieved by performing SPR experiments in random order, i.e. distances were not tested with any correlation to distance, and spanning the use of multiple chips and reagents. One of the covariates relevant to our study (amount of bound structures) is dealt with through a normalization/standard curve that we describe in Supplementary Fig 1. This way, variation in amount of bound structure, which could systematically vary depending on the batch of SPR chips used, does not systematically influence the signal and parameterization of the model Blinding Blinding is not relevant to the majority of this study due to the high degree of automation built into the process, i.e. results are acquired by the SPR instrument, and any possible covariates/systematic trends in reagents or chips are scrambled away through the random order of sample acquisition. However, a blinded challenge to our model pipeline is included to evaluate the model's predictive capability. The format and result of this test is described in Figure 3. Training data was used to parameterize the model and predictions were made by one author, while experimental test data obtained by the other authors was withheld during this phase. When the predictions were made, they were then compared to the test data to evaluate the prediction quality. This was done to ensure that no bias due to prior knowledge could make its way into the predictions.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems Methods n/a Involved in the study n/a Involved in the study Antibodies \boxtimes ChIP-seq Eukaryotic cell lines \boxtimes \boxtimes Flow cytometry \boxtimes MRI-based neuroimaging Palaeontology and archaeology \boxtimes \boxtimes Animals and other organisms \times Human research participants \times Clinical data \boxtimes Dual use research of concern

nature research | reporting summary

Antibodies

Antibodies used	Rabbit anti-digovygenin IgG 9H27I 19 Sigma-Aldrich Mouse anti-digovygenin IgG1 no. ab420 Abcam
Antibodies used	
Validation	 Rabbit anti-digoxygenin IgG 9H27L19: Validated by manufacturers with western blot: "Western blot analysis of recombinant Digoxigenin conjugated to BSA (1 ng/lane) using a Digoxigenin recombinant rabbit monoclonal antibody (Product # 700772) at a dilution of 0.1 µg/mL. NBT/BCIP was used as the substrate" https://www.thermofisher.com/antibody/product/Digoxigenin-Antibody-clone-9H27L19-Recombinant-Monoclonal/700772 Appears in publications: IHC: Tsai A, Alves MR, Crocker J. Multi-enhancer transcriptional hubs confer phenotypic robustness. Elife. 2019 Jul 11;8:e45325. IF: Brown MW, Kim Y, Williams GM, Huck JD, Surtees JA, Finkelstein IJ. Dynamic DNA binding licenses a repair factor to bypass roadblocks in search of DNA lesions. Nature communications. 2016 Feb 3;7(1):1-2. ELISA: Bruno JG. Application of DNA aptamers and quantum dots to lateral flow test strips for detection of foodborne pathogens wit improved sensitivity versus colloidal gold. Pathogens. 2014 Jun;3(2):341-55. ISH: Canadas I, Rojo F, Taus Á, Arpí O, Arumí-Uría M, Pijuan L, Menéndez S, Zazo S, Dómine M, Salido M, Mojal S. Targeting epithelia to-mesenchymal transition with Met inhibitors reverts chemoresistance in small cell lung cancer. Clinical Cancer Research. 2014 Feb 15;20(4):938-50.
	Mouse anti-digoxygenin IgG1 no. ab420 Abcam: Validation performed by manufacturers: "The Abpromise guarantee Our Abpromise guarantee covers the use of ab420 in the following tested applications. The application notes include recommended starting dilutions; optimal dilutions/concentrations should be determined by the end user. Application Abreviews Notes WB 1/500 - 1/5000. ab419 can be used for the detection of a diverse range of digoxigenin labelled compounds and in a variety of techniques ELISA 1/1000 - 1/10000. IHC-P 1/200 - 1/2000. IHC-Fr 1/200 - 1/2000. In situ hybridization Use at an assay dependent concentration. Southern Blot 1/500 - 1/5000."
	References: Qian F et al. MiR-378a-3p as a putative biomarker for hepatocellular carcinoma diagnosis and prognosis: Computational screening with experimental validation. Clin Transl Med 11:e307 (2021).PubMed: 33634974 Mann EK et al. Associations between DNA Damage and PD-L1 Expression in Ovarian Cancer, a Potential Biomarker for Clinical Response. Biology (Basel) 10:N/A (2021).PubMed: 33946684 Samsonova A et al. Lin28, a major translation reprogramming factor, gains access to YB-1-packaged mRNA through its cold-shock
	domain. Commun Biol 4:359 (2021).PubMed: 33742080 Shibata T et al. Small molecule targeting r(UGGAA)n disrupts RNA foci and alleviates disease phenotype in Drosophila model. Nat Commun 12:236 (2021).PubMed: 33431896
	Cortes-Troncoso J et al. T cell exosome-derived miR-142-3p impairs glandular cell function in Sjögren's syndrome. JCI Insight 5:N/A (2020).PubMed: 32376798