

Peer Review Information

Journal: Nature Computational Science

Manuscript Title: Stochastic modeling of antibody binding predicts programmable migration on antigen patterns

Corresponding author name(s): Ian Hoffecker

Reviewer Comments & Decisions:

Decision Letter, initial version:
--

Dear Dr Hoffecker,

Your manuscript "Stochastic modeling of antibody binding predicts programmable migration on antigen patterns" has now been seen by 4 referees, whose comments are appended below. You will see that while they find your work of interest, they have raised points that need to be addressed before we can make a decision on publication.

The referees' reports seem to be quite clear. Naturally, we will need you to address all of the points raised.

While we ask you to address all of the points raised, the following points need to be substantially worked on:

- As discussed by Reviewer #1, Please use data from more than one monoclonal to showcase the applications of the proposed pipeline.
- Please provide experimental proof of the model's predictions.
- Please discuss the relevancy of migration time and direction for the system.
- Reviewer #3 has indicated the need to better incorporate existing literature in the manuscript. Some relevant studies that should be considered are the ones mentioned by Reviewer #3 and the following:
 1. <https://www.nature.com/articles/s41467-020-16949-4>,
 2. <https://www.sciencedirect.com/science/article/pii/S2589004221010403>
 3. <https://pubs.acs.org/doi/10.1021/acsnano.1c05411>
 4. <https://www.pnas.org/content/118/4/e2016857118>

- Please detail possible applications of the pipeline additionally to the predictions as requested by Reviewer #3.
- Please relate to the work of Amitai et al as discussed by Reviewer #3.
(<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006408>)
- Please include sensitivity analysis in the manuscript as indicated by Reviewer #4.

In addition to these points, it would also be beneficial to address the following concerns:

- Please better compare the technical novelty of this submission with <https://www.nature.com/articles/s41565-018-0336-3> and better explain what the new contributions in terms of modeling and results are. This needs to be clarified and properly discussed in the manuscript.

Please use the following link to submit your revised manuscript and a point-by-point response to the referees' comments (which should be in a separate document to any cover letter):

[REDACTED]

** This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first. **

To aid in the review process, we would appreciate it if you could also provide a copy of your manuscript files that indicates your revisions by making use of Track Changes or similar mark-up tools. Please also ensure that all correspondence is marked with your Nature Computational Science reference number in the subject line.

In addition, please make sure to upload a Word Document or LaTeX version of your text, to assist us in the editorial stage.

To improve transparency in authorship, we request that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit <http://www.springernature.com/orcid>.

We hope to receive your revised paper within three weeks. If you cannot send it within this time, please let us know.

We look forward to hearing from you soon.

Best regards,

Ananya Rastogi, PhD
Associate Editor
Nature Computational Science

Reviewers comments:

Reviewer #1 (Remarks to the Author):

I have read the manuscript entitled "Stochastic modeling of antibody binding predicts programmable migration on antigen patterns" by Hoffecker and coworkers. They describe a model for antibody binding to regularly spaced antigens on two-dimensional surfaces. A rationale for this is to understand better how antigen spacing affects conversion between monovalent and bivalent antibody binding and how that relates to the phenomenon of antibody migration along repeated antigen patterns. They base the model on surface plasmon resonance data of binding of a rabbit monoclonal IgG to digoxigenin as the target antigen. The digoxigenin is bound to DNA origami structures on an SPR chip. The model is designed as a discrete Markov process with several potential states where the transition between these states is possible and described with elementary rates. The authors then test how antigen spacing will affect monovalent and bivalent binding, as well as vectors of antibody migration.

I find the manuscript well written with mostly well-motivated figures and descriptions. The findings on monovalent vs. bivalent binding are the most important, in my opinion, with the results on migration speed and direction of less general interest. I have some concerns with the study that I summarize below:

Major points:

1. All the input data is based on experiments with a single rabbit monoclonal IgG. The internal structure of the antibody will likely affect how flexible the Fab binding is, and in that way, affect the transition and requirements for bivalent binding. This will probably not even be the same between different rabbit IgG monoclonals and definitely not between species and subclasses (which is relevant for human and mouse IgG, for instance). If the model can be described as applying to antibodies in general, the data needs to be coming from more than one monoclonal.
2. I lack confirmation that the model can predict experimental results. A good way would be to a) start with experimental data, b) create a model, c) test predictions from the model with new experiments.

The latter point is missing. The group is capable of doing such experiments. Binding to pathogen proteins with regularly spaced antigens would, of course, be ideal in terms of how the authors frame the relevancy of their model.

3. What is the relevancy of migration time and direction for a system that relatively fast reaches equilibrium? It is a nice way of showing where the antibodies might end up and how well they will bind, but isn't it easier and more relevant to in vivo function to just focus on the equilibrium end state of binding? What would happen if they test an antigen repeated structure where it is possible to have bivalent binding but no walking (i.e., repeated sections of two antigens close enough but with large intersectional distance)?

Minor points:

1. How significant is the contribution of the interconversion constant, KD_2 . It seems to be very small in comparison to binding constants?

2. How computationally efficient is their model? This has implications for the overall useability of the model. Here it could also be interesting to see a comment on a recently published antibody binding model (Kumra Ahnlide et al., 2021, *Frontiers in Immunology*).

3. The final section of the introduction reads as a results section (including a description of their Markov model), and the figure reference to 1a is wrong. It says that it refers to PSPR, but it shows a model of the DNA origami instead. It would be beneficial to carefully describe the experimental system in a panel in Fig 1. This could be placed in the results section instead.

4. Supplementary Fig. 1e looks like a too-perfect fit (overfitting). I might have misunderstood, but this could be an example of the problems with only having experimental data from one monoclonal.

5. The authors write about mono-, bi-, and multi-valency but don't mention avidity. I think it could be good to clarify the definitions in the introduction.

Reviewer #2 (Remarks to the Author):

Hoffecker et al. provide a simple, but powerful modeling approach to study Ab binding and Ab movement on defined Antigen patterns. The importance of specific Ab binding to repeating patterns, of defined distances has become a very hot topic recently and is certainly relevant not only to basic research, but is also extremely important for clinical biomedical applications, esp. with the view of designing new Ab therapeutics, e.g. vaccine design. It's a beautiful paper and I would suggest publication in its current form.

I only have one question, which the authors may want to answer: In Figure 3a it appears as though the steady state probability for o1 and o2 is not quite the same and o1 appears to be slightly more prevalent (from the graph). Why is this? Or does it only appear this way?

Reviewer #3 (Remarks to the Author):

In this manuscript, Hoffecker and coauthors develop a simulation framework for the bivalent binding of monoclonal antibodies to target molecules separated from controlled distances using DNA origami as support to these molecules.

On the experimental side, they lay two molecules of digoxigenin per origami (a plant small steroid molecule) as antigens, modulating the distance between the two antigens, and incubated the antigen constructs with dynamically regulated amounts of a specific monoclonal antibody. The amount of bound antibody conformations is quantified using Surface Plasmon Resonance (SPR) over time, as a read-out for the dynamics of binding.

On the theoretical side, they develop a kinetic simulation model for the dynamics of monovalent or bivalent antibody binding states with multiple antibodies competing to access the two copies of the antigen. They use a monte carlo formalism as a discrete representation of the binding events, instead of the usual Ordinary Differential Equations formalism, which is fine and well justified by the complexity of the states and molecule behavior.

In Figure 2, they show that the model can faithfully replicate the observed dynamics after fitting, but additionally that a 'smartly designed' experiment setup with successive shoots of increasing antibody doses allows to retrieve the KD of the investigated antibody+origami construct. In particular, by fitting experimental curves for different distances between the two antigens, they can retrieve an optimal KD value for a spacing around 13nm between the two antigens.

In the following figures, they show how far the model can be used to simulate more complex origami settings, with 3 antigens separated from different distances or a 1D line of antigens separated by increasing distances. They predict how the SPR would look like in those settings, and make the interesting prediction that the antibodies 'walk' on the 1D line towards the optimal spacing between the antigens. The simulations are not substantiated by experimental data but show how different hypotheses can be tested in advance as to how the SPR would look like, and the model is therefore promising to test such hypotheses on future data.

Overall, the manuscript is very well written and the authors have put effort to contain the technical details in the methods and supplements while keeping only the necessary explanations in the main text, this is well appreciated.

The study is neat and clearly addresses the problem of interpreting SPR and finding optimal antigen distances in Figure 2. It is a little bit less clear which problem the authors aimed to tackle with figures 3 and 4, i.e., whether they aim to make predictions using the model, or if they tried to suggest other applications of their model to future SPR data, as the focus seems to be a useful pipeline for the community. This could be clarified. Perhaps, the authors could detail possible applications of the pipeline additionally to the predictions, since they seem to present the model as a pipeline in the abstract.

Generally, recent papers have used origami as to study antibody binding. A more thorough comparison of recent literature would be helpful to the reader.

Of note, the authors have also published a study showing the use of SPR together with the seemingly same computational model <https://www.nature.com/articles/s41565-018-0336-3> see supplementary pages 25+ and figures 23. They have already shown a fitting to SPR data in that published paper, corresponding to figure 1, and have shown the optimal distance for high affinity, low affinity or different classes of antibodies. This seems fine as the present manuscript provides new simulated predictions.

Can the authors relate to the work of Amitai et al (<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006408>) that derived a theoretical optimum for affinity depending on the distance between antigens. How does it relate to the empirical formula proposed in the present manuscript in Figure 2f?

Several points for the discussion:

- The study uses small steroid molecules as antigens, that are not representative to the protein antigens mentioned as the motivation of the study (viruses or bacteria). Does the origami system supports such protein antigens and what are the related limitations to expect?
- Can the model help to design experiments in general? Here, all the calibration tests with doses of antibody and origami are not shown (which is OK), but how could a new lab use the model to setup the calibration as efficiently as possible to use the pipeline?
- Beyond multivalent binding to the same antigen, several start-ups have invested in bispecific (or up to tetra-specific) antibodies in the concept of antitumor therapy, targetting a different protein antibody by each Fab region (like tumor check-point inhibitor and a FC receptor as to recruit immune cells, as well as agonist IL2R/CD25 binding for instance). How can the present work be used for this application to test lead candidates? Would it work for tri-specific antibodies?

- Can the model be used for antigen design, as to decide spacing between two similar or mutated HIV variants for instance, and would it help to trigger higher affinity antibodies by modulating their distance?

Minor points:

- The middle paragraph in section 2.3 is particularly hard to read, more explanations/descriptions would be welcome.

- Can the authors show the sensitivity of fitting the empirical curve (figure 2F) to noise in the data? Bootstrap? not important.

The supplements/formula are well explained/substantiated.

The authors provided the python jupyter notebook already filled with the results of each step for reproducing the manuscript. This is convenient. It needs anaconda to setup a precise versioned environment, which is also convenient.

Reviewer #4 (Remarks to the Author):

- Overall, I think supplementary material and methods should be cited in the main results of the manuscript to facilitate the reading and connection between results and methods (for example, refer to section 9.7 in section 2.3). Additionally, some values and concepts should be explained in the results section of the manuscript (even though they are detailed in the method): this is also to help reading through the results. For example, KD1/KD2 could be described at the beginning of the section 2.1. The concept of occupancy could also be described in one phrase.

- To me, what is mostly missing in the manuscript is some sensitivity analysis. What is the uncertainty around the estimated values of k_1 , k_{-1} (paragraph 2.1)? How does the variation of these parameters affect the results shown? Figures 2b and 2c seem a bit overfitted to me. To what extent are the estimated values dependent on experimental conditions? Similarly, in paragraph 2.2, how does the concentrations of the solutions introduced affect the state probabilities and the result from fig 3f?

- Although the predictions from the model in figure 3 make sense qualitatively, I wonder how can these be checked/verified? How are they affected by the antigens configuration? How does the result from figure 3f relate to experimental observations (i.e., are there settings that can be cited here that would confirm the predictions from the model)?

- Paragraph 2.3: what is needed to adapt the results to 2D/3D configurations?

Minor comments:

- Figure in methods should be referred to as supplementary figures
- There is a typo in the paragraph after sup fig 7: "Absolute sums of residuals [...] show that the best-performin models were the those" --> remove "the"

Author Rebuttal to Initial comments

Reviewer #1 (Remarks to the Author):

I have read the manuscript entitled "Stochastic modeling of antibody binding predicts programmable migration on antigen patterns" by Hoffecker and coworkers. They describe a model for antibody binding to regularly spaced antigens on two-dimensional surfaces. A rationale for this is to understand better how antigen spacing affects conversion between monovalent and bivalent antibody binding and how that relates to the phenomenon of antibody migration along repeated antigen patterns. They base the model on surface plasmon resonance data of binding of a rabbit monoclonal IgG to digoxigenin as the target antigen. The digoxigenin is bound to DNA origami structures on an SPR chip. The model is designed as a discrete Markov process with several potential states where the transition between these states is possible and described with elementary rates. The authors then test how antigen spacing will affect monovalent and bivalent binding, as well as vectors of antibody migration.

I find the manuscript well written with mostly well-motivated figures and descriptions. The findings on monovalent vs. bivalent binding are the most important, in my opinion, with the results on migration speed and direction of less general interest. I have some concerns with the study that I summarize below:

Major points:

- 1.) All the input data is based on experiments with a single rabbit monoclonal IgG. The internal structure of the antibody will likely affect how flexible the Fab binding is, and in that way, affect the transition and requirements for bivalent binding. This will probably not even be the same between different rabbit IgG monoclonals and definitely not between species and subclasses (which is relevant for human and mouse IgG, for instance). If the model can be described as applying to antibodies in general, the data needs to be coming from more than one monoclonal.*

Response: We have now added data from an additional mouse-derived monoclonal IgG1 (Abcam ab420) antibody specific to digoxigenin and computed its corresponding separation distance-dependent spatial tolerance curve. This data can be seen in SFig 1

and 3 with side-by-side rabbit and mouse antibody data shown in each case, and is discussed in the main text:

“In the first stage, we used either rabbit anti-digoxigenin IgG or mouse-derived anti-digoxigenin IgG1 targeting digoxigenin-decorated DNA origamis with a single-cycle-kinetics program whereby progressively higher concentrations...”

And Supplementary:

“For each monovalent run (Supplementary Fig 1 b and c for rabbit and mouse antibodies respectively) max with a unique value of R_{struct} , a projected value of R_{Ab} is computed with Equation 25. By fitting the monovalent models to 1-1 kinetics, we obtain the rate constants that allows computational prediction max of R_{Ab}

(Supplementary Fig 1 d and e for rabbit and mouse antibodies respectively) in the absence of mono according experimental saturation conditions. This enables us to make a standard curve to adjust R_{Abmax} to R_{struct} in the absence of a 1-1 R_{Ab} (Supplementary Fig 1 f and g for rabbit and mouse antibodies respectively). We use this value as a conversion factor, enabling us to convert SPR response units into

the number of antibodies per structure (Supplementary Fig 1 h and i for rabbit and mouse antibodies respectively). By knowing R_{struct} , we can estimate this conversion factor for non-trivial antigen configurations max where the multivalence influences the ease of reaching a saturation value corresponding to R_{Ab} .”

“Progressive fitting allows us to obtain K_{D2} for a single 2-antigen system provided that we have determined K_{D1} for a 1-antigen system, for which we take a mean run of multiple 1-antigen runs ($n = 6$) with plus or minus 1 and 2 standard error of the mean away from the mean run as uncertainty intervals (Supplementary Fig 3 a and i for rabbit and mouse antibodies respectively).”

“The combined expression yields Equation 1 which predicts the interconversion constant as a function of separation distance (Supplementary Fig 3 d, e and l, m for rabbit and mouse antibodies respectively).”

The results show that the shape of these curves are indeed molecule-dependent, with notable K_{D2} differences in both in magnitude and distribution over spatial separation distance.

- 2.) I lack confirmation that the model can predict experimental results. A good way would be to a) start with experimental data, b) create a model, c) test predictions from the model with new experiments. The latter point is missing. The group is capable of doing such experiments. Binding to pathogen proteins with regularly spaced antigens would, of course, be ideal in terms of how the authors frame the relevancy of their model.

Response: We have now included the results of a blinded test whereby the model was used to predict the SPR binding curve of a 3-antigen triangle configuration. The model was parameterized according to our pipeline with 1-antigen and various 2-antigen configurations of different distances without any information from higher order structures. The prediction is thus made with zero a priori information about the behavior of antibodies on this 3-antigen, closely-spaced triangular configuration which is expected to give rise to multiple complex states containing 1-3 antibodies and composed of different combinations of monovalently and bivalently bound states.

The only new experimental information used in the prediction was the structure-binding curve data for the particular run, which is used in conjunction with the standard curve from SFig 4 to estimate the conversion factor from SPR response units (RU) to those of antibodies/structure. Predictions were made by Ian Hoffecker on the basis of the model and the structure binding data in a blinded format while experimental data gathered by Alan Shaw and Ioanna Smyrlaki was withheld until the predictions had been submitted. We have added the result of this test in Fig 3, replacing panels d,e, and f i.e. the de novo prediction figures for a different triangle structure that was there previously in order to conserve space and reduce redundancy in the manuscript. We have also updated SFig 4 accordingly. We also include 2 additional replicates for this test in the supplementary material in SFig 7 and analysis of state probability distribution sensitivity to uncertainty in concentration in SFig 8.

This result is discussed in the main text:

“To validate the pipeline, we used the model parameterized with 1 and 2-antigen data (Fig 2 f) to create a blind a priori prediction of the evolution of a higher order system with 3 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 3 d), and additional validation is shown in Supplementary Fig 7. We found that experimental trajectories closely conformed to predictions. The model provides access to the individual contributions of states to the signal through their occupancy (Fig 3 e) and relative proportions (Fig 3 f) enabling us to construct a narrative explanation for the observed dynamics. We see for example in the final stage when concentration was set to zero, that while total occupancy falls and that occupancy contributions from monovalently bound antibodies decreased, bivalent state contributions counterintuitively increased. This indicates that high concentrations in the penultimate stage had inverted the system to favor monovalent-dominated saturation states that subsequently transitioned to unsaturated bivalent states as sites became available (Fig 3 g).”

And in the supplementary:

“To validate the spatial tolerance model, we conducted a blinded test whereby an a priori prediction was made using the spatial tolerance parameters from Fig 2 f in the main text. This was done with the rabbit anti-DIG IgG antibody. The

data used for prediction consisted entirely of the averaged 1-antigen data and the series of 2-antigen varied separation distance data used to parameterize the model and the structure-binding data to determine the conversion factor from RU to Ab/struct and vice versa. Thus no data with structures configured with more than 2 antigens was used for parameterization. To perform the test, we chose to predict the evolution of a trimeric $7.2 \times 14.3 \times 16.0$ configuration with the same single-cycle kinetics protocol of timed, known concentration injections used for other runs in this study (Supplementary Fig 7 c). Predictions were made first by computing the expected occupancy values on a per-structure basis. These were then converted into SPR response units (RU) by multiplying them with the RABmax value determined from structure binding curve and standard curve from Supplementary Fig 1.

Experimental

results withheld until predictions were made were then revealed and compared with the theoretical curves each done in triplicate with respective structure-binding data used for each one (Supplementary Fig 7 d). “

“An important question to consider is how sensitive the stratified state probability distribution predictions are to error in the antibody concentration. As we have shown the distribution is dependent on concentration and timing, however the answer is likely quite complex. As different phases, like those shown in Fig 3 a- 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 9.10

a. to see how this affects both the resulting transient probability distribution (Supplementary Fig 9.10 b) and the relative correspondence of the predicted curve to that of the experimental (Supplementary Fig 9.10 c). What we see here is that the probability distribution is fairly robust, with the top 5 most represented states remaining unchanged in all 5 conditions, albeit their relative ranking has changed slightly, e.g. going from 1.1x to 2x.”

We have also added additional emphasis on the cross-validation (S Fig 7) in the original submission, whereby models were parameterized on the basis of 8 different combinations of configurations derived from a hexagonal equi-spaced configuration. These models were then cross-tested for their predictions with each other, e.g. the two-antigen model was tested for its extrapolative predictions of 3, 4, 5, and 6 antigen systems. This system does not test the accuracy of the distance-dependent spatial tolerance model, but does provide validation for single-distance predictions with complex patterns. The additional emphasis on this validation has

been added in the text :

“...arrayed in a $7.2 \times 14.3 \times 16.0$ nm right triangle and then check its correspondence with an experimental trajectory (Fig 3 d), and additional validation is shown in Supplementary Fig 7.”

- 3.) What is the relevancy of migration time and direction for a system that relatively fast reaches equilibrium? It is a nice way of showing where the antibodies might end up and how well they will bind, but isn't it easier and more relevant to in vivo function to just focus on the equilibrium end state of binding?

Response: Our position is that non-equilibrium kinetics are relevant to biological scenarios in which signals might be transmitted by different spatial distributions of multiple antibodies and particular macro-states or ensembles. For example, the potential for antibodies or B-cell receptors to aggregate, dimerize, and form clusters in immunological synapses suggest an important role of spatial self-organization and many-body ensembles. Furthermore, the fact that regularly spaced patterns are indeed prevalent in nature on the surfaces of viral and bacterial pathogens supports the relevance of migratory behavior which might be exploited by immunity for the purposes of exploring states and achieving higher order organization. In answer to the question regarding relevance to in vivo function, we expect that the “end state of binding” may, in physiological contexts, not be as simple as equilibrium with just 1 or 2 antigens. Signalling pathways are not equilibrium phenomenon, but are rather a chain of cause and effect from an initial state to a downstream/later one. A note about this point has been add to the discussion:

“The apparent importance of spatial organization in immunological signalling suggests a role for non-equilibrium spatial phenomena such as that studied here , and we might expect antigen organization...”

With regards to which is easier to study, indeed equilibrium binding is indeed the prevailing approach to antibody characterization, and this is probably the most sensible first pass study that one should do. However we find the “programmability” aspect of antibodies an interesting and mostly overlooked phenomenon. Thus while long tracks of separation distance gradients are indeed an artificial scenario, we chose this system configuration in order to overtly explore the potential for spatial control of antibody movement by the geometry of antigen patterns. Our goal was to establish the existence of such a control parameter, which we view to be suggestive of evolutionary significance, as nearly every form of leverage in a competitive evolutionary context tends to be exploited. Realistically, we would expect spatial control to take a more subtle or complex form in nature. However, without having directly investigated this in natural systems, we choose to keep our language appropriately agnostic to this possibility and propose simply that this kind of spatial molecular programmability is an interesting area for further study.

- 4.) What would happen if they test an antigen repeated structure where it is possible to have bivalent binding but no walking (i.e., repeated sections of two antigens close enough but with large intersectional distance)?

Response: This would be equivalent to the basic 2-antigen structure, as each region would be spatially isolated in the same conceptual sense that we are testing an ensemble of many 2-antigen structures that are spatially isolated from each other. Or more precisely, the occupancy we would see would be a scalar multiple of that of a single 2-antigen structure, with the multiplier being the number of 2-antigen islands.

Minor points:

- 1.) How significant is the contribution of the interconversion constant, $KD2$. It seems to be very small in comparison to binding constants?

Response: $KD2$ is significant in the sense that if it is $>10^{-1}$ then the system is effectively monovalent, i.e. no contribution from the bivalent state and no additional stabilization of binding. Whereas if it is $<10^{-1}$ then stabilization is observed and the system has an additional state. This increases the entropy of the bound macrostate, as there are more states for the system to explore that involve being bound. If $KD2$ is very low, on the order of $KD1/\text{Concentration}$ or less, then this bivalent state is a low-energy sink that effectively traps the system in a bound state. It is a bit misleading to directly compare the magnitude with $KD1$, since the on-binding process is concentration dependent. If one is working with a 1 nM solution concentration, then the dissociation to association ratio ($KD1 / C$) is on the order of $10^{-11} / 10^{-9} = 10^{-2}$ so roughly the same order as $KD2$ in the system that we explored. At lower concentrations, $KD2$ has a greater contribution to system dynamics, whereas at higher concentrations $KD1$ is more significant.

We have added this discussion to the supplementary text in a new section entitled Run Design:

“The parameterization pipeline first requires a dosing scheme for the single-cycle kinetics program, i.e. the timing and concentrations of 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 of all, approximate knowledge of $KD1$ is probably the best starting point. Some antibody suppliers report an in-house measured KD ($KD1$ 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 experiment [48]. Knowing this, it is possible to choose a dosing scheme that elicits both monovalent ($KD1$ dominated) kinetics and bivalent ($KD2$ dominated) kinetics. $KD1$ dominated kinetics occur when the magnitude of $KD1 / c_{Ab}$ is smaller than $KD2$, i.e. at high relative concentrations. Whereas $KD2$ dominated effects are most apparent at low concentrations or during the dissociation phase when

concentration is set to zero. While K_{D2} is molecule-specific, this value is probably subject to less variation among the commonly used isotypes. So even if K_{D2} is unknown at first, it may be a reasonable starting guess to assume that it is similar to the values found in our study, i.e. on the order of 10^{-2} around optimal separation distances. Thus framing a dosing series based on a supplier's reported K_{D1} value and concentrations that span a range where relative K_{D1}/c_{Ab} to K_{D2} goes through an inversion is likely to capture a useful range of kinetics well suited to parameterizing the model. The following concentrations and injection timings were used for most experiments in our study: time points (sec): 0, 84, 384, 475, 775, 866, 1166, 1257, 1557,

1656, 1956 concentrations (nM): 0, 0.025, 0, 0.05, 0, 0.1, 0, .25, 0, .5, 0."

- 2.) How computationally efficient is their model? This has implications for the overall useability of the model. Here it could also be interesting to see a comment on a recently published antibody binding model (Kumra Ahnlide et al., 2021, Frontiers in Immunology).

Response: We added a comment in the supplementary material about the computational complexity of the breadth first search algorithm used to identify states:

"The time complexity of breadth first search can be expressed as $O(|V| + |E|)$ where $|V|$ and $|E|$ are the number of vertices and edges respectively. In the case of antigen patterns, the former correspond to the number of antigens in the pattern. The latter corresponds 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, and as a rule of thumb we could say that antigens further than 25 nm apart are not close enough for bivalent bonds to form."

We added a reference to Kumra Ahnlide et al in the discussion:

"and bacteria are known to interfere with antibody binding such as through Fc targeting to prevent opsonization [41]."

- 3.) The final section of the introduction reads as a results section (including a description of their Markov model), and the figure reference to 1a is wrong. It says that it refers to PSPR, but it shows a model of the DNA origami instead. It would be beneficial to carefully describe the experimental system in a panel in Fig 1. This could be placed in the results section instead.

Response: We have adjusted the language in this paragraph to sound less like a results section, exchanging phrases like "We show that ..." with "We investigate ..." and other adjustments that we think make it more suitable for an introduction by emphasizing the problems/questions addressed by the study rather than the answers.

We have moved the figure reference to the end of this sentence where it makes more sense as a reference to ligand patterned DNA origami instead of PSPR:

“The patterned surface plasmon resonance (PSPR) technique enables measurement of binding kinetics on precise, monodisperse patterns of ligands, achieving robust control of geometry through the use of DNA origami nanostructures (Fig 1a)”

- 4.) Supplementary Fig. 1e looks like a too-perfect fit (overfitting). I might have misunderstood, but this could be an example of the problems with only having experimental data from one monoclonal.

Response: The close conformity to shape is the result of a well-behaved monovalent kinetics and well-controlled conditions in the instrument that are cleanly described by this model, and not due to fitting to the shape of the curve. The noise in this experiment is hard to see in SFig 1e, but if we zoom in (reducing the size of datapoints so that they can be seen without overlap), one can observe random noise and the fact that the model curve is not conforming to it.

The only parameters we are varying for the fitting are the two monovalent rate constants k_1 and k_{-1} and the interconversion constant KD_2 for bivalent structures and only k_1 and k_{-1} for the monovalent structure (e.g the one shown in SFig 1e). As a comparison, consider a simple linear $y=mx + b$ fit, which has two parameters - the same number of parameters that we are changing for our monovalent fit. We might also add that the observed spikes are features of the curve - these are not the result of experimental noise but are rather the result of concentration changes, i.e. fixed constraints that affect the model predictions but are not variable parameters for fitting. These along with their timings are known constraints. In supplementary figures 5 and 6 we show a comparison of good and bad fits - where a bad fit will still exhibit structural features at each of the timed concentration change stages, but will have a curve that diverges systematically from the experimental. We can see such systematic divergences in the good fits as well if we zoom in (below). One can also find similarly closely conforming fits in SPR literature.

- 5.) The authors write about mono-, bi-, and multi-valency but don't mention avidity. I think it could be good to clarify the definitions in the introduction.

Response: The consensus in the antibody community is that the term avidity should be avoided as it is not precisely defined (Rheinhecker, Michael, et al. The Journal of Immunology 157.7 (1996): 2989-2997).

Rather, the terms “functional affinity” referring to the Gibbs free energy due to binding (i.e. defined in terms of KD_{app} in our case) and “intrinsic affinity” referring to the Gibbs free energy of binding due to monovalent kinetics (defined in terms of KD_1 in our case) should be used instead.

However we agree that widespread use of the term “avidity” makes it important to address for clarity and we had anyway included it as a keyword with the abstract, so we have gone ahead and added a note about it.

“...enabling 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 bound antibodies [1,2,3].”

Reviewer #2 (Remarks to the Author):

Hoffecker et al. provide a simple, but powerful modeling approach to study Ab binding and Ab movement on defined Antigen patterns. The importance of specific Ab binding to repeating patterns, of defined distances has become a very hot topic recently and is certainly relevant not only to basic research, but is also extremely important for clinical biomedical applications, esp. with the view of designing new Ab therapeutics, e.g. vaccine design. It's a beautiful paper and I would suggest publication in its current form.

I only have one question, which the authors may want to answer: In Figure 3a it appears as though the steady state probability for o1 and o2 is not quite the same and o1 appears to be slightly more prevalent (from the graph). Why is this? Or does it only appear this way?

Response: These states have the same probability, as you rightly would expect with them being mirror states both with a single monovalent antibody, though it is hard to see in some places due to the noisy interference resulting from numerical sensitivity of the equation solver to small probabilities that sometimes leads to sudden changes or explosive solutions. We have shuffled the coloring in this figure in a way that we think makes this distribution somewhat clearer.

Reviewer #3 (Remarks to the Author):

In this manuscript, Hoffecker and coauthors develop a simulation framework for the bivalent binding of monoclonal antibodies to target molecules separated from controlled distances using DNA origami as support to these molecules.

On the experimental side, they lay two molecules of digoxigenin per origami (a plant small steroid molecule) as antigens, modulating the distance between the two antigens, and incubated the antigen constructs with dynamically regulated amounts of a specific monoclonal antibody. The amount of bound antibody conformations is quantified using Surface Plasmon Resonance (SPR) over time, as a read-out for the dynamics of binding.

On the theoretical side, they develop a kinetic simulation model for the dynamics of monovalent or bivalent antibody binding states with multiple antibodies competing to access the two copies of the antigen. They use a monte carlo formalism as a discrete

representation of the binding events, instead of the usual Ordinary Differential Equations formalism, which is fine and well justified by the complexity of the states and molecule behavior.

In Figure 2, they show that the model can faithfully replicate the observed dynamics after fitting, but additionally that a ‘smartly designed’ experiment setup with successive shoots of increasing antibody doses allows to retrieve the KD of the investigated antibody+origami construct. In particular, by fitting experimental curves for different distances between the two antigens, they can retrieve an optimal KD value for a spacing around 13nm between the two antigens.

In the following figures, they show how far the model can be used to simulate more complex origami settings, with 3 antigens separated from different distances or a 1D line of antigens separated by increasing distances. They predict how the SPR would look like in those settings, and make the interesting prediction that the antibodies ‘walk’ on the 1D line towards the optimal spacing between the antigens. The simulations are not substantiated by experimental data but show how different hypotheses can be tested in advance as to how the SPR would look like, and the model is therefore promising to test such hypotheses on future data.

Overall, the manuscript is very well written and the authors have put effort to contain the technical details in the methods and supplements while keeping only the necessary explanations in the main text, this is well appreciated.

The study is neat and clearly addresses the problem of interpreting SPR and finding optimal antigen distances in Figure 2. It is a little bit less clear which problem the authors aimed to tackle with figures 3 and 4, i.e., whether they aim to make predictions using the model, or if they tried to suggest other applications of their model to future SPR data, as the focus seems to be a useful pipeline for the community. This could be clarified.

Response: We have made changes to Figure 3 to incorporate a prediction followed by experimental validation of the model. Here we hope to emphasize that the pipeline is useful for generating a predictive and explanatory model - with the potential to provide not only accurate trajectories but also mechanistic and narrative descriptions of the underlying dynamics. We have added emphasis on this point in the text:

“The model provides access to the individual contributions of states to the signal through their occupancy (Fig 3 e) and relative proportions (Fig 3 f) enabling us to construct a narrative explanation for the observed dynamics. We see for example in the final stage when concentration was set to zero, that while total occupancy falls and that occupancy contributions from monovalently bound antibodies decreased, bivalent state contributions counterintuitively increased. This indicates that high concentrations in the penultimate stage had inverted the system to favor monovalent-dominated saturation states that subsequently transitioned to unsaturated bivalent states as sites became available (Fig 3 g).”

Our aim with Figure 4 was to connect our pipeline to what we view as more broadly interesting phenomena in the molecular immunology and biomolecular engineering fields - i.e. the potential for programmability of such multi-state molecular systems and the utility of our model for exploring these phenomena. To make our intentions with this part of the study clearer, we have added the following:

To the introduction:

“A causal linkage between pattern geometry and antibody dynamics could potentially be 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”

To the results beginning at section 2.3:

“In order to explore the potential for pattern-based control and programmability of antibody dynamics, we wished to model the dynamics of larger systems with greater relevance to periodic pathogenic surfaces.”

Perhaps, the authors could detail possible applications of the pipeline additionally to the predictions, since they seem to present the model as a pipeline in the abstract.

Response: We expect that it is possible to design DNA origami with both regular periodic antigen arrays as well as gradients in separation distance. A combination of such a structure with a well-parameterized Markov model would allow one to obtain detailed insight into the underlying states of an SPR signal generated by such a structure. Without discussing designs in too much detail in the text, we have proposed in general terms the possibility to tackle such problems using DNA origami and our pipeline in the discussion:

“We predicted that long-range gradients of differential spacings could be used to establish persistent directed migration of antibodies on a surface, and we propose that PSPR and the progressive fitting pipeline presented here might be used in future studies to test predictions experimentally. Firstly, such designs should be possible using DNA origami, and secondly 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.”

Another application may be in terms of vaccine design especially for HIV. Antigen spacing and density appear to be a key factor determining the immune response to HIV and integral to the virus’s strategy for immune evasion. It thus could be considered in the design of particles meant to elicit a specific immune response. We discuss this in the discussion:

"the low spike density characteristic of HIV is key to its immune evasion [42]. 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, e.g. immunostimulatory virus-like particles [36,43]. "

And connected to our comment below, we think the characterization of bispecific antibodies could be another potential application for the pipeline. This is also in the discussion:

"Our pipeline could also be used to dissect the complex state spaces of bi- tri- or tetra- specific antibodies which are recently being developed for therapeutic and biosensing applications [44]. 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."

Generally, recent papers have used origami as to study antibody binding. A more thorough comparison of recent literature would be helpful to the reader.

Response: We have added the following new literature references related to origami and antibody binding along with their discussion at various points in the text:

Hellmeier, Joschka, et al. "DNA origami demonstrate the unique stimulatory power of single pMHCs as T cell antigens." *Proceedings of the National Academy of Sciences* 118.4 (2021).

Hellmeier, Joschka, et al. "Strategies for the site-specific decoration of DNA origami nanostructures with functionally intact proteins." *ACS nano* 15.9 (2021): 15057-15068.

Pfeiffer, Martina, et al. "Single antibody detection in a DNA origami nanoantenna." *Science* 24.9 (2021): 103072.

Zhang, Ping, et al. "Capturing transient antibody conformations with DNA origami epitopes." *Nature communications* 11.1 (2020): 1-9.

Knappe GA, Wamhoff EC, Read BJ, Irvine DJ, Bathe M. In Situ Covalent Functionalization of DNA Origami Virus-like Particles. *ACS nano*. 2021 Sep 7;15(9):14316-22.

Of note, the authors have also published a study showing the use of SPR together with the seemingly same computational model <https://www.nature.com/articles/s41565-018-0336-3> see supplementary pages 25+ and figures 23. They have already shown a fitting to SPR data in that published paper, corresponding to figure 1, and have shown the optimal distance for

high affinity, low affinity or different classes of antibodies. This seems fine as the present manuscript provides new simulated predictions.

Response: In addition to new predictions, in this work we have also significantly refined the parameterization pipeline (reducing it to a minimal set of experiments) and also developed a phenomenological model with analytical equation describing spatial tolerance that was not in our previous work. With this round of revision, we have also added a blind prediction/validation of the model as well as demonstration of the progressive fitting pipeline for an additional monovalent antibody. We have incorporated a new standard curve-based approach to determining the conversion factor from SPR response units to antibodies per structure that utilizes structure binding curve data. We also further developed the steady state analysis, adding thermodynamic data and phase analysis. We also developed a Markov Chain Monte Carlo or “random walk” variant of our model in this work that is able to tackle larger systems that were too complex for our previous enumerative-only program - this was the system used to examine periodic patterns and spacing gradients.

To reinforce the novelty of our present work, we have added an explicit description of these differences to a new section at the beginning of the supplementary.

“Some methods and explanation may be found in Shaw et al [18]. However, in the following supplementary material, we emphasize the novel developments of this work including the following: 1. A minimal parameterization pipeline for Markov models sensitive to arbitrary antigen spacing distances and its experimental validation. 2. A phenomenological model with analytical equation describing spatial tolerance as a continuous function. 3. Application of the model and fitting pipeline to two different antibodies, one derived from rabbit and the other mouse. 4. A systematic approach to determining the conversion factor between SPR response units (RU) to that of the number of bound antibodies per structure. 5. Steady state analysis and the determination of thermodynamic quantities from equilibrium state probability distributions. 6. A Markov Chain Monte Carlo random walk variant of the model which can be used to simulate larger systems with too many connected antigens to be feasible for enumerative approaches.”

Can the authors relate to the work of Amitai et al (<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006408>) that derived a theoretical optimum for affinity depending on the distance between antigens. How does it relate to the empirical formula proposed in the present manuscript in Figure 2f?

Response: This is an interesting paper that is an ideal converse case of what we have looked at, where instead of mobile antibodies, they are looking at mobile antigens. Much of their findings are in agreement with what we would expect based on our findings. We have added a reference to the discussion:

“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 by laterally diffusing to achieve the minimum interconversion ratio K_{D2} . Amitai et al 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 HIV is key to its immune evasion [42].”

Several points for the discussion:

- The study uses small steroid molecules as antigens, that are not representative to the protein antigens mentioned as the motivation of the study (viruses or bacteria). Does the origami system support such protein antigens and what are the related limitations to expect?

Response: Indeed, protein antigens would have greater physiological relevance. Haptens are easier to work with, and for that reason we chose this for our system. However there is plenty of precedent for producing structures with immobilized proteins, and future studies should seek to incorporate them especially due to the added 3-dimensionality that larger antigens with depth may cause.

We have added this point to the discussion along with several references to protein patterning on origami:

“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 PSPR data that systematically modulates relative epitope orientations or employ structures that incorporate pathogenic protein antigens [34,35,36,37] for more realistic structural complexity and physiological relevance.”

- Can the model help to design experiments in general? Here, all the calibration tests with doses of antibody and origami are not shown (which is OK), but how could a new lab use the model to setup the calibration as efficiently as possible to use the pipeline?

Response: We have created a new section in the experimental methods section to address this:

“The parameterization pipeline first requires a dosing scheme for the single-cycle kinetics program, i.e. the timing and concentrations of 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 of all, approximate knowledge of K_{D1} is probably the best starting point. Some antibody suppliers report an in-house measured K_D (K_{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 experiment [48]. Knowing this, it is possible to choose a dosing scheme that elicits both monovalent (K_{D1} dominated) kinetics and bivalent (K_{D2} dominated) kinetics. K_{D1} dominated kinetics occur when the magnitude of K_{D1}/c_{Ab} is smaller than K_{D2} , i.e. at high relative concentrations. Whereas K_{D2} dominated effects are most apparent at low concentrations or during the dissociation phase when concentration is set to zero. While K_{D2} is molecule-specific, this value is probably subject to less variation among the commonly used isotypes. So even if K_{D2} is unknown at first, it may be a reasonable starting guess to assume that it is similar to the values found in our study, i.e. on the order of 10^{-2} around optimal separation distances. Thus framing a dosing series based on a supplier's reported K_{D1} value and concentrations that span a range where relative K_{D1}/c_{Ab} to K_{D2} goes through an inversion is likely to capture a useful range of kinetics well suited to parameterizing the model. The following concentrations and injection timings were used for most experiments in our study: time points (sec): 0, 84, 384, 475, 775, 866, 1166, 1257, 1557,

1656, 1956 concentrations (nM): 0, 0.025, 0, 0.05, 0, 0.1, 0, .25, 0, .5, 0."

- Beyond multivalent binding to the same antigen, several start-ups have invested in bispecific (or up to tetra-specific) antibodies in the concept of antitumor therapy, targetting a different protein antibody by each Fab region (like tumor check-point inhibitor and a FC receptor as to recruit immune cells, as well as agonist IL2R/CD25 binding for instance). How can the present work be used for this application to test lead candidates? Would it work for tri-specific antibodies?

Response: This is an interesting idea - since PSPR can be used to measure both the intrinsic affinities of each Fab arm and rgw functional affinities on controlled patterns of multiple antigens, the PSPR/origami approach + model would be well-suited provide a lot of information about these molecules and their interactions. As with the system we studied, the main advantage to using the model is the ability to create well-described conditions where we have detailed knowledge of the underlying states despite a convoluted SPR signal. Bispecific antibodies can come in different forms such as those with and without Fc regions, and the length and flexibility of their tethering regions appear to vary greatly over the literature. The different states elicited by such design choices and their dependence on the intrinsic affinities could be thoroughly explored with our methods. We have added a discussion about this:

"Our pipeline could also be used to dissect the complex state spaces of bi- tri- or tetra- specific antibodies which are recently being developed for therapeutic and biosensing applications [44]. 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.”

- Can the model be used for antigen design, as to decide spacing between two similar or mutated HIV variants for instance, and would it help to trigger higher affinity antibodies by modulating their distance?

Response: There appears to be a consensus that the low density spatial distribution of HIV spike proteins is important for their immune evasion, so we also see a potential opportunity for a spatial tolerance model and parameterization pipeline to be of value here both in terms of explanation of selection pressures and for rationale design of vaccines. We have added some discussion about this connected to the discussion of Amitai et al:

“Amitai et al derived a theoretical affinity optimum for mobile spike proteins that depends on their surface density, arguing that intermediatedensities invoke the greatest immune response, and that the low spike density characteristic of HIV is key

to its immune evasion [42]. 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,

e.g. immunostimulatory virus-like particles [36,43].”

Minor points:

- The middle paragraph in section 2.3 is particularly hard to read, more explanations/descriptions would be welcome.

Response: We have reworded this section and broken up some of the long sentence to be more readable. It now reads as follows:

“Antigens arranged according to a spacing gradient in the range of 10 - 22 nm separation distances (i.e. the interval of steepest slope in Equation 1) elicit asymmetric accumulation toward the narrow spaced end of the array (Fig 4 a). This system also exhibited individual walking trajectories that tend toward the narrow-spaced end (Fig 4 b), asymmetric velocity (Fig 4 c), and asymmetric net displacement (Fig 4 d) according to the direction of the gradient.”

- Can the authors show the sensitivity of fitting the empirical curve (figure 2F) to noise in the data? Bootstrap? not important.

Response: Sensitivity analysis has been incorporated now in both Fig 2f and Supplementary Fig 3 and is discussed in the supplementary.

The supplements/formula are well explained/substantiated.

The authors provided the python jupyter notebook already filled with the results of each step for reproducing the manuscript. This is convenient. It needs anaconda to setup a precise versioned environment, which is also convenient.

Reviewer #4 (Remarks to the Author):

- Overall, I think supplementary material and methods should be cited in the main results of the manuscript to facilitate the reading and connection between results and methods (for example, refer to section 9.7 in section 2.3).

Response: We have added several additional references to the supplementary methods including this one.

Additionally, some values and concepts should be explained in the results section of the manuscript (even though they are detailed in the method): this is also to help reading through the results. For example, KD1/KD2 could be described at the beginning of the section 2.1. The concept of occupancy could also be described in one phrase.

Response: We have added lines explaining these concepts to the text along with references to the supplementary sections for more details.

“This program was performed with a 1-antigen configuration (Fig 2 b, d) in order to parameterize our Markov model (Supplementary Section 9.8) 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 (Supplementary Section 9.3).”

And:

“This yields respective association and dissociation rates $k_1 = 1.93 \pm 0.05 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$ and $k_{-1} = 5.28 \pm 0.07 \times 10^{-4} \text{ s}^{-1}$ as well as a monovalent dissociation constant $KD1 = 2.7 \pm 0.11 \times 10^{-11} \text{ M}$ defined as the ratio of the dissociation to association rates.”

- To me, what is mostly missing in the manuscript is some sensitivity analysis. What is the uncertainty around the estimated values of k_1 , k_{-1} (paragraph 2.1)? How does the variation of these parameters affect the results shown?

Response: We have produced a plot of the shift in the optimal separation distance as a function of distance to the mean monovalent run to illustrate how error in this input propagates to the end of the fitting pipeline. The points correspond to one and two standard deviations away from the mean input run, or approximately 68 and 95 % confidence intervals assuming Gaussian-distributed behavior. Because runs are performed

independently and exhibit an apparently random shift in RU relative to the mean run, we expect this assumption to be valid.

However, because the subsequent values of KD1, KD2, and the spatial tolerance curve are not linearly related to the input nor related through a concise analytical expression (i.e. they are determined via numerical fitting), we cannot say that the final uncertainties on, for example, the spatial tolerance function are Gaussian confidence intervals. Rather they simply represent the output results corresponding to Gaussian confidence intervals of the input monovalent run data. We expect this heuristic approach to be mostly valid, as small shifts in input data usually result in small adjustments to the final output. However this is not always the case. As an example, consider a monovalent run which saturates close to 1 ab/struct, with a 95% upper confidence bound extending above the 1 ab/struct level. This represents an “impossible” value for a monovalent system, which cannot have more than 1 ab/struct. Thus the corresponding fits are likely to deteriorate drastically and unpredictably as a result of this. To deal with this, our fitting routine incorporates a limited flexibility of the normalization factor, making most fits robust to such hard constraints. This robustness comes at the expense of sensitivity to input data, and there is thus less variation in output result due to changes in input data than one might expect. This again further underscores the fact that the output uncertainties do not represent strict 68% or 95% confidence intervals, but rather the respective intervals of the input data, and only an arbitrary, though possibly reasonable, confidence interval on the output results (spatial tolerance parameters, KD2, etc).

The output of this uncertainty analysis are the following:

1. Uncertainty bounds on k_1 k_{-1} and KD1 for a given set of monovalent 1-antigen repeat runs.
2. Uncertainty bounds of KD2, determined by feeding the KD1 bounds as constraints.
3. Uncertainty estimates of each points of the spatial tolerance curve, parameterized for the resulting set of KD2 values computed with respective KD1 bounds as constraints. To be conservative, we have taken the maximum of the absolute value difference at each point in the curve relative to the curve computed from mean inputs, and applied this bound both above and below the central curve.
4. Uncertainty estimates for each of the spatial tolerance parameters, determined by the max and min among the 95 and 68% input confidence intervals.
5. Sensitivity analysis of the minimum point of the spatial tolerance curve corresponding to the optimal separation distance for bivalent binding. This is done by examining the minimum distance for each of the input bounds

To characterize the goodness of fit for spatial tolerance curves, we devised a chi-squared-like metric that utilizes a heuristic approach to estimate the expected value for each KD2 point that is independent of our model. This is done by using a moving average with a sliding window of 3 datapoints that is then also used to compute an estimator for the standard error of the mean - or in other words a moving standard error of the mean. The metric thus makes an implicit assumption that data points are randomly distributed about this moving mean, a reasonable assumption that aligns with our intuitive judgement of what constitutes a good fit in this case. A good fit should thus have a chi-squared value close to 1, representing a balanced ratio of the deviation of observed points from the model relative to the expected deviation due to random noise. A value larger than 1 indicates a greater difference in the observed model deviation, an indication that the model is wrong or poorly fitted. And conversely, a value much smaller than 1 indicates an unlikely high correspondence between model and observed data, suggesting an overestimate of the expected error or else an overfitting to the data.

The goodness of fit values obtained for the monoclonal anti-digoxigenin rabbit and mouse antibodies respectively were 0.988 and 0.948 respectively, indicating an appropriately fitted model based on the observed noise. To visualize this, we

have also made a figure whereby the individual data points are projected onto the x-axis by subtracting either the model (in blue) or the moving average (in red). An appropriate error metric and model fit should be reflected in this figure by a random (Gaussian) distribution of points above and below the axis.

We have also added sensitivity analysis with respect to concentration and its qualitative effect on state probability distributions. See comment below:

Figures 2b and 2c seem a bit overfitted to me. To what extent are the estimated values dependent on experimental conditions?

Response: We copy here our response from above regarding overfitting:

The close conformity to shape is the result of a well-behaved monovalent kinetics and well-controlled conditions in the instrument that are cleanly described by this model, and not due to fitting to the shape of the curve. The noise in this experiment is hard to see in (Figures 2b and 2c), but if we zoom in (reducing the size of datapoints so that they can be seen without overlap), one can observe random noise and the fact that the model curve is not conforming to it.

The only parameters we are varying for the fitting are the two monovalent rate constants k_1 and k_{-1} and the interconversion constant KD2 for bivalent structures and only k_1 and k_{-1} for the monovalent structure (e.g. the one shown in SFig 1e). As a comparison, consider a

simple linear $y=mx + b$ fit, which has two parameters - the same number of parameters that we are changing for our monovalent fit. We might also add that the observed spikes are features of the curve - these are not the result of experimental noise but are rather the result of concentration changes, i.e. fixed constraints that affect the model predictions but are not variable parameters for fitting. These along with their timings are known constraints. In supplementary figures 5 and 6 we show a comparison of good and bad fits - where a bad fit will still exhibit structural features at each of the timed concentration change stages, but will have a curve that diverges systematically from the experimental. We can see such systematic divergences in the good fits as well if we zoom in (below). One can also find similarly closely conforming fits in SPR literature.

Similarly, in paragraph 2.2, how does the concentrations of the solutions introduced affect the state probabilities and the result from fig 3f?

Response: Here we assume that you are wondering how variation or error in concentration might change the distribution of states. This is indeed a good question as we have shown that the distribution is highly dependent on concentration and timing, however the answer is likely quite complex. As different phases, like those shown in 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 a factor of 0.5 0.9 1.1 or 2.0 to see how the distribution is affected qualitatively. This result is included in the supplementary as a new figure and discussed

"An important question to consider is how sensitive the stratified state probability distribution predictions are to error in the antibody concentration. As we have shown the distribution is dependent on concentration and timing, however the answer is likely quite complex. As different phases, like those shown in Fig 3 a- 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 9.10 a) to see how this affects both the resulting transient probability distribution (Supplementary Fig 9.10 b) and the relative correspondence of the predicted curve to that of the experimental (Supplementary Fig 9.10 c). What we see here is that the probability distribution is fairly robust, with the top 5 most represented states remaining unchanged in all 5 conditions, albeit their relative ranking has changed slightly, e.g. going from 1.1x to 2x."

- Although the predictions from the model in figure 3 make sense qualitatively, I wonder how can these be checked/verified? How are they affected by the antigens configuration?

How does the result from figure 3f relate to experimental observations (i.e., are there settings that can be cited here that would confirm the predictions from the model)?

Response: We have modified figure 3 to include the results of a blind test of the model's predictive ability. This was done by parameterizing the model on the basis of 1 and 2-antigen data, and then using the resulting model to make predictions for a 3-antigen triangular configuration. The results indicate good correspondence between prediction and model, and this was done for 3 replicates. Additionally, SFig 7 shows our earlier work to validate the model by performing cross-validation with different combinations of test and training data.

- Paragraph 2.3: what is needed to adapt the results to 2D/3D configurations?

Response: From a computational, perspective, this is trivial to do, as the only thing that the model needs is the topology or map of which antigens are connected and their separation distances. However to be more realistic, 3D configurations are likely to involve a more complex landscape of excluded volume or steric hindrance in reality, and we think this would need to be incorporated for accurate prediction/description of physical systems. 2D lattices are already possible, and we think the model would be well suited for simulating these systems so long as the antigens in question are not rigidly oriented. An orientation term should be developed for future studies to incorporate this additional realism. We have made a small note about these points in discussion:

"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 PSPR data that systematically modulates relative epitope orientations or employ structures that incorporate pathogenic protein antigens[34,35,36,37] for more realistic structural complexity and physiological relevance. "

Minor comments:

- Figure in methods should be referred to as supplementary figures

Response: We have fixed this now, so all supplementary figure references in the supplementary are called "Supplementary Figure"

- There is a typo in the paragraph after sup fig 7: "Absolute sums of residuals [...] show that the best-performing models were the those" --> remove "the"

Response: We have fixed this line.

Decision Letter, first revision:

Dear Dr. Hoffecker,

Thank you for submitting your revised manuscript "Stochastic modeling of antibody binding predicts programmable migration on antigen patterns" (NATCOMPUTSCI-21-0604A). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Computational Science, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

TRANSPARENT PEER REVIEW

Nature Computational Science offers a transparent peer review option for new original research manuscripts submitted from 17th February 2021. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. **Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't.** Failure to state your preference will result in delays in accepting your manuscript for publication.

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our [FAQ page](https://www.nature.com/documents/nr-transparent-peer-review.pdf).

Thank you again for your interest in Nature Computational Science Please do not hesitate to contact me if you have any questions.

Sincerely,

Ananya Rastogi, PhD
Associate Editor
Nature Computational Science

ORCID

IMPORTANT: Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, please let your co-authors know that if they wish to have their ORCID added to the paper they must follow the procedure described in the following link prior to acceptance:
<https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

Reviewer #1 (Remarks to the Author):

The authors have done an excellent revision. I'm happy with their clarifications and additional data.

Reviewer #2 (Remarks to the Author):

The authors have made substantial improvements to their manuscript, especially the blind experimental data really show the power of this model. In my opinion, the authors have answered all questions/concerns satisfactorily. I am happy to recommend publication.

Reviewer #3 (Remarks to the Author):

The authors have constructively replied to all my points. It was mainly about discussing broader applicability of their method and citing literature, which they did.

Further, they also went beyond my points by checking a prediction experimentally, and by adding information about sensitivity, which is appreciated.

I support publication as it is.

Final Decision Letter:

Dear Dr Hoffecker,

We are pleased to inform you that your Article "Stochastic modeling of antibody binding predicts programmable migration on antigen patterns" has now been accepted for publication in Nature Computational Science.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

Please note that *Nature Computational Science* is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. [Find out more about Transformative Journals](#)

Authors may need to take specific actions to achieve [compliance](#) with funder and institutional open access mandates. For submissions from January 2021, if your research is supported by a funder that requires immediate open access (e.g. according to [Plan S principles](#)) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route our standard licensing terms will need to be accepted, including our [self-archiving policies](#). Those standard licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

You will not receive your proofs until the publishing agreement has been received through our system.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see <https://www.nature.com/natcomputsci/for-authors>). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

If you have queries at any point during the production process then please contact the production team at rjsproduction@springernature.com. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Content is published online weekly on Mondays and Thursdays, and the embargo is set at 16:00 London time (GMT)/11:00 am US Eastern time (EST) on the day of publication. If you need to know the exact publication date or when the news embargo will be lifted, please contact our press office after you have submitted your proof corrections. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number NATCOMPUTSCI-21-0604B and the name of the journal, which they will need when they contact our office.

About one week before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Computational Science. Our Press Office will contact you closer to the time of publication, but if you or your Press Office have any inquiries in the meantime, please contact press@nature.com.

An online order form for reprints of your paper is available at <https://www.nature.com/reprints/author-reprints.html>. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Computational Science as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). We also welcome suggestions for the Hero Image, which appears at the top of our [home page](#); these should be 72 dpi at 1400 x 400 pixels in JPEG format. Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and

white or grayscale images. Please do not try to design a cover with the Nature Computational Science logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

We look forward to publishing your paper.

Best,
Fernando (on behalf of Ananya Rastogi)

--

Fernando Chirigati, PhD
Chief Editor, Nature Computational Science
Nature Portfolio