Robotica (2016) volume 34, pp. 1754–1776. © Cambridge University Press 2016
doi:10.1017/S0263574716000199

Influence of model resolution on geometric simulations of antibody aggregation

Kasra Manavi, Bruna Jacobson, Brittany Hoard and Lydia Tapia*

Department of Computer Science, University of New Mexico, 87131 Albuquerque, New Mexico.
Emails: kazaz@cs.unm.edu, bjacobson@cs.unm.edu, bhoard@cs.unm.edu

(Accepted March 14, 2016. First published online: May 13, 2016)

SUMMARY

It is estimated that allergies afflict up to 40% of the world's population. A primary mediator for allergies is the aggregation of antigens and IgE antibodies bound to cell-surface receptors, FcϵRI. Antibody/antigen aggregate formation causes stimulation of mast cells and basophils, initiating cellular degranulation and releasing immune mediators which produce an allergic or anaphylactic response. Understanding the shape and structure of these aggregates can provide critical insights into the allergic response. We have previously developed methods to geometrically model, simulate and analyze antibody aggregation inspired by rigid body robotic motion simulations. Our technique handles the large size and number of molecules involved in aggregation, providing an advantage over traditional simulations such as molecular dynamics (MD) and coarse-grained energetic models. In this paper, we study the impact of model resolution on simulations of geometric structures using both our previously developed Monte Carlo simulation and a novel application of rule-based modeling. These methods complement each other, the former providing explicit geometric detail and the latter providing a generic representation where multiple resolutions can be captured. Our exploration is focused on two antigens, a man-made antigen with three binding sites, DF3, and a common shrimp allergen (antigen), Pen a 1. We find that impact of resolution is minimal for DF3, a small globular antigen, but has a larger impact on Pen a 1, a rod-shaped molecule. The volume reduction caused by the loss in resolution allows more binding site accessibility, which can be quantified using a rule-based model with implicit geometric input. Clustering analysis of our simulation shows good correlation when compared with available experimental results. Moreover, collisions in all-atom reconstructions are negligible, at around 0.2% at 90% reduction.

KEYWORDS: Antibody Aggregation; Resolution Study; Rule-Based Modeling.

1. Introduction

The human allergic immune response can be triggered by cell surface molecule aggregation, specifically a tyrosine kinase cascade initiated by the transmembrane signalling caused by the crosslinking of IgE-FcϵRI (receptors) via antigen binding. This crosslinking stimulates mast cells and basophils and induces degranulation, resulting in the release of histamine and other allergic response mediators.[52] The ability to predict the structures of these aggregates is key in determining how the spatial organization of receptors affects transmembrane signaling.

The geometric impacts of molecular aggregation have not been well studied due to the fact that it is presently computationally infeasible to model hundreds of large molecules at an all-atom resolution. The application of robotics-inspired methods to molecular simulations has played a role in improving the computational efficiency of these types of simulations. Molecular interaction using robotic-inspired techniques have been developed to study ligand docking.[8,13,22] However, these methods are generally applied to small scale problems (e.g., two molecule systems). Elastic network models[4,7]

* Corresponding author. E-mail: tapia@cs.unm.edu

have been used to determine molecular modes of motion and have been extended in a variety of manners,[1,47,54] but have not been used to study molecular aggregation. Coarse-grained models have been developed to simulate aggregation that overcome long simulation times while preserving energetics, but they are often either lacking in geometric details or unable to simulate large scale aggregation.[10,45]

To address this gap across different simulation scales, we have previously presented methods to simulate and analyze aggregation using polygon-based models generated from all-atom molecular structures.[40] These polygon-based models retain the geometric information about the molecular structure while reducing the computational footprint by forgoing atom position data. Geometric complexity is an important feature of these molecular models because lower complexity translates into fewer polygon to polygon comparisons and faster run times at the expense of model realism.

The overall goal of this research is to understand the packing of protein complexes and to predict final assembled structures of aggregated molecules. As part of this effort, in this paper we simulate molecular aggregation at different model complexities within our polygon-based model framework. In the scope of this paper, model complexity is analogous to model resolution of the molecules. Different resolutions are obtained by changing the number of polygons composing each molecule's 3-D structure. A central question to be investigated in this analysis is whether we can significantly speed up our simulation run times by lowering the resolution of the model, thus enabling simulation of even larger systems. However, a simulation speed up could come at the cost of generating aggregate structures which are significantly different at various model complexities due to steric effects that may appear at different resolutions. We are interested in quantifying the impact of model resolution on the assembly problem. Preliminary work in molecular multiscale modeling has begun to address these questions,[58] focusing on the unification of different mathematical representations of DNA and RNA molecules.

In this work, we study how model complexity affects molecular aggregation by analyzing its impact via two independent simulation techniques. First, we perform Monte Carlo simulations of 3-D rigid body structures based on all-atom protein structures at different model complexities. Then, we approach this problem with a novel implementation of rule-based modeling, a fast complementary simulation that does not explicitly represent geometry but captures steric interactions of the proteins through ordinary differential equations (ODEs), and look for consistency in the results.

We present results on a synthetic antigen that has been studied experimentally, DF3, and a natural antigen, a common shrimp allergen Pen a 1. These antigens greatly differ in size, structure and number of binding sites. Pen a 1 is a highly relevant allergen due to its allergenic potency and unique structure. We find that impact of resolution is minimal for DF3, but has a larger impact on Pen a 1. A study of clustering of receptor complexes bound to DF3 from our simulations shows strong agreement with experimental data. We find that collisions in all-atom reconstructions, another metric to analyze the simulations, are negligible, at around 0.2% at 90% reduction. Lastly, we use Monte Carlo data to fit the rates of our rule-based model to gain intuition on how reduction impacts packing structure.

2. Related Work

2.1. Macromolecular simulations

In the past 20 years, the biosciences have seen great advances on a diverse range of research topics brought on by computational simulation of biological phenomena. MD simulations can determine individual trajectories for each atom in a solvated protein system. Despite considerable success with small proteins and fast events ($\approx 1\mu s$), simulations of biological processes involving large molecules and over long time scales have proven to be challenging. As an effort to investigate larger systems, a recent paper has shown an all-atom simulation of a chromatophore membrane patch consisting of 20 million atoms. This simulation was run simultaneously using three supercomputers and yielded 150 ns MD trajectories in total.[12] Although this system size is comparable to the antibody–antigen models that we propose to study, the time scales that are feasible with all-atom MD simulations are too short to explore the assembly structures that we aim to investigate.

In order to overcome these computational limitations, techniques that effectively reduce the number of degrees of freedom in the system have been widely used in modeling large biological systems. In a coarse-grained approach, groups of atoms are clustered into coarse-grained charges[29,34,35] or point

https://doi.org/10.1017/S0263574716000199

masses.[43] A variant of MD, coarse-grained MD is a multi-scale approach that simulates trajectories of the coarse-grained units.[6,31,46,53] In general, coarse-grained MD can reduce the number of particles about 10-fold, enabling simulations that contain about 10 million atoms, and generating about 100 ns of MD trajectories in a day.[31] However, for systems where structural conformation is highly relevant, such as our aggregate structures, coarse-graining is disadvantageous as the reverse problem of reconstructing the all-atom structure from the coarse-grained model is non-trivial.

2.2. Design of assembled molecular structures

Methods for designing protein-based assemblies come in two forms: Stochastic (resulting in irregular structures with probability-derived attributes), and deterministic (producing exactly specified geometric features). Principles for the design of ordered protein assemblies are discussed in ref. [33] A majority of the computational design methods have focused on interface construction,[25,28] but new methods go further and fully design self-assembling molecules.[30] These methods are similar in that they generally start outperforming rigid body docking followed by iterative design/minimization steps to refine the interface.

2.3. Molecular aggregation prediction

Many fields from medicine to industrial manufacturing stand to benefit from the use of computational methods to determine possible geometric structures of assembled molecules. Most of the methods developed use lattice models with force fields and focus on the interactions of proteins with both denatured[10] and native[62] conformations. A coarse-grained MD-based approach to study polymer-drug aggregation was done in ref. [45] We note all of these models have energetics included in their computation and would not be feasible to simulate systems of the sizes we study.

2.4. IgE aggregation experiments

Studies using nanoparticles have shown that the size and number of binding sites of the antigen impact degranulation of rat basophilic leukemia cells.[26] Spatiotemporal analysis of IgE aggregation has been done using nanoscale imaging and motion tracking techniques. Methods to analyze clustering of micrograph probes were developed in ref. [61] including Ripleys and Hopkins statistic calculations. The locations of static gold nanoparticle-labeled IgE-FcϵRI have been imaged using transmission electron microscopy (EM).[57] Spatial clustering analysis of IgE-FcϵRI has been done using methods from ref. [61] as well as hierarchical clustering techniques to quantify the numbers of sizes of clusters in ref. [16] Tracking of quantum dot labeled IgE-FcϵRI has determined temporal information such as diffusion rates.[3] While these experimental methods have been able to measure attributes about receptor dynamics, none of them retain information about the aggregate binding patterns. Because of this, distinguishing linked (bound) from simply proximal receptors is challenging.

2.5. Spatial rule-based modeling

Protein complex formation can be difficult to model due to the high number of distinct molecule species possible. Rule-based modeling provides a way to more compactly represent all possible aggregate structures. Languages such as BioNetGen[9] and Kappa[14] provide rule-based modelers a way to cleanly and efficiently represent biological systems. Some methods have extended rule-based modeling to incorporate spatial modeling, including SRSim.[20] These methods include spatial modeling, but still use simple geometric representations of the molecules and have not been applied to reconstruct all atom representations. Rule-based modeling has also been extended to multi-level systems in ML-Rules[41] which uses a nested structure to establish a hierarchical model.

Rule-based modeling has been used to model antigen–antibody interactions. A kinetic rule-based version of the trivalent antigens and bivalent receptors (TLBR) system which accounted for two types of cycles, dimers and heximers, was introduced in ref. [60] Note the antigens in a TLBR system have three binding sites, and the receptors have two binding sites. This method was based on a previous equilibrium theory model.[19] The rule-based model was also extended to consider steric constraints.[43] In this paper, steric constraints are introduced through an external data structure, not in the rules representing the binding events.

https://doi.org/10.1017/S0263574716000199

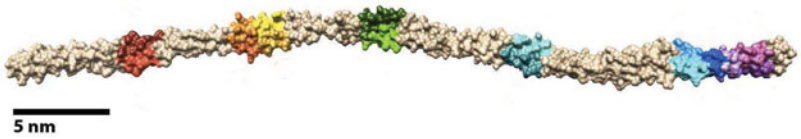


Fig. 1. The molecular structure of Pen a 1, a common shrimp allergen (tan). The binding sites (various colors) are located in five regions on the coiled coil structure.

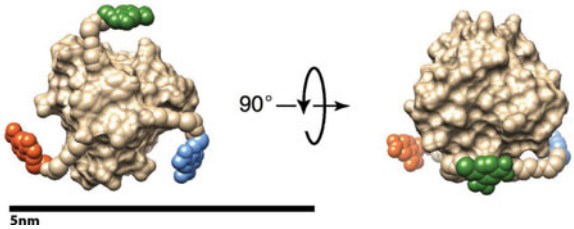


Fig. 2. The molecular structure of DF3 (tan), a synthetic antigen. The trimer has three DNP linkers (various colors) attached to the N-termini of each subunit.

The novelty of our rule-based model lies in the ability to simulate steric constraints by designing rules that incorporate antigen conformational details. The protein structure is a crucial input in deciding how to generate rules associated with ODEs representing protein aggregation.

3. Methods

We present two methods to simulate and analyze the geometric packing of protein complexes, thereby predicting final assemblies. First, we explain the all-atom structures used in our simulations (Section 3.1). Next, we introduce the components of the Monte Carlo simulation, including: polygon-based models and reductions (Section 3.2.1), simulation methods for rotation and translation of geometric structures and aggregate formation (Section 3.2.2), and analysis techniques for the large geometric structures (Section 3.2.3). Finally, we present a novel rule-based model that implicitly represents the geometry of the molecules. This includes: analysis of steric effects of the Pen a 1 all-atom structure (Section 3.3.1), an explanation of the assumptions made during the construction of the model and details of the calculation of aggregate formation probability (Section 3.3.2), a description of how rules are defined and an explanation of the rate constants of the model (Section 3.3.3), and a metric to compare the results from the Monte Carlo and rule-based models (Section 3.3.4).

3.1. Antigen structure and binding sites

There have been numerous studies of natural allergens initiating degranulation. One that has been of particular interest is the common shrimp allergen, Pen a 1. The immune response is triggered by the shrimp tropomyosin molecule, a 40 nm double-stranded coiled coil structure, (Fig. 1), which crosslinks IgE. The allergen has been predicted to have five binding regions in each of the strands of the coiled coil[5] and a total of 16–18 binding sites.[27,51]

Alongside experimentation on natural allergens, synthetic antigens have been generated to provide insights into antibody aggregation.[49,55,59] These synthetic antigens are primed with 2,4-dinitrophenyl (DNP), a linker used to bind to engineered antibodies. We explore the antigen DF3, a molecule with three binding sites that has been characterized in ref. [38] DF3, shown in Fig. 2, is a trimer of peptides representing the C-terminus of the T4 fibritin foldon domain with a DNP linker attached to the N-terminus.

We focus our analysis on Pen a 1 and DF3 for two reasons: known molecular structure and high experimental relevance. There are many natural allergens that have been identified but many do not have molecular structures determined. Along the same lines, there are synthetic antigens that induce a response but do not have well known binding site positions.[59] Synthetic DF3 has known binding sites and available experimental data. Pen a 1 aggregation is currently being studied experimentally and has a known structure and mapped binding sites.

https://doi.org/10.1017/S0263574716000199

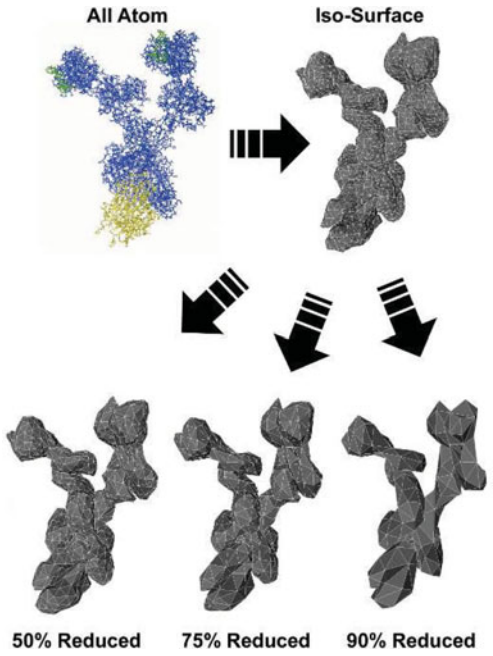


Fig. 3. The model construction process starting with an all-atom model (FcϵRI), generating the isosurface model, and then applying polygon reduction to the isosurface model to generate models with lower resolution.

3.2. Monte Carlo simulation

3.2.1. Model construction. Our method is based on simulating 3-D rigid body models of the receptor complex and antigen molecules. We use our own generated models as well as contributed models. An all-atom structure of the receptor complex was used from ref. [38] The IgE structure, composed of both heavy and light chains, is modeled bound to the α subunit of FcϵRI. The receptor complex is made up of 1709 amino acids totaling 13,477 atoms.

The antigen DF3 was generated by starting with the base fibritin foldon domain (PDB:1RFO) and adding DNP linkers to the N-terminus of each subunit. DF3 is comprised of 81 amino acids with 1365 atoms total. Structural models for shrimp tropomyosin were available in the Protein Data Bank (PDB:1CG1) and in the Structural Database of Allergenic Proteins (SDAP Model #284). The Pen a 1 model used was composed of 568 amino acids totaling 4580 atoms.

Since it would be computationally prohibitive to use these all-atom models at the molecule counts we simulate, we reduce the computational footprint of the model by using a polygon representation (Fig. 3). To construct our models, we begin with an all-atom structure. Using the multiscale model extension of UCSF Chimera 1.9,[18] we generate isosurface models of the molecules at a variety of resolutions, 4Å for DF3 and 6Å for the receptor complex and Pen a 1. The resulting model of the occupied volume, referred to henceforth as the base model, is considered to be the model with the highest resolution, i.e., the most detailed model.

This base model can now be reduced in complexity using standard polygon reduction techniques[11] to produce models with tunable amounts of detail as seen in Fig. 3. Since the amount of detail in the base model potentially hinders performance, we want to observe the costs versus benefits of decreasing the model resolution. We use the polygon reduction feature in Maya,[42] a modeling software package, that allows the generation of models with a reduction specified in the percentage of polygons.

The binding sites are unique to each antigen, and thus had to be modeled accordingly. DF3 with DNP linkers has very flexible binding sites.[38] In order to capture this effect, the model has the linkers compressed to half their length and has a binding site located at the end of the linker. Then, we model a spherical binding volume with a radius of half the DNP linker length (7.5Å) that is centered at the end of the compressed linker.

The binding sites of Pen a 1 are located on its surface. The binding site amino acid sequences for Pen a 1 are documented in ref. [27] Vertices for binding were located on the center of the binding

https://doi.org/10.1017/S0263574716000199

site and exterior of the surface. A binding radius of 3Å was used. The same binding site locations were used regardless of the resolution since the locations of the binding sites on the original all-atom structure do not change.

3.2.2. Simulation methods. As detailed in our previous publications,[39,40] the simulation begins with all molecules randomly placed on a grid in a collision-free state with no molecules bound. At each time step, every molecule gets an updated position and orientation generated via random sampling. Membrane bound receptors rotate about their z-axis and translate on a planar x–y cell surface. During the course of the simulation, all ligand molecules are positioned at an ideal vertical distance from the cell membrane, in which ligand–receptor binding events are possible at all times. The absence of ligand diffusion in the z-direction eliminates the effect of adsorption and desorption of ligands from the system. This assumption maintains the concentration of ligands constant, considerably simplifying the simulations. The updated positions and orientations are generated with the consideration of biological constraints, e.g., diffusion constants[3] and rotational correlation times.[50] Over the course of the simulation, a receptor binding site will end up within the binding volume of the antigen. The probability of a binding event occurring is dependent on binding volume overlap and the association rates specified. Dissociation rates are also used in the simulation, i.e., every bond that exists at a given time step is evaluated for dissociation. Kinetics between DNP specific antibodies and univalent DNP haptens[15] as well as DNP-based synthetic antigen[59] have been well studied and provide values for association/dissociation rates. We use the rates from DNP-BSA in ref. [59] due to the fact that DF3 is multivalent like DNP-BSA, and it is composed of peptide chains as opposed to a small molecules such as the haptens in ref. [15] and DCT2 in ref. [59] These parameters are not known for Pen a 1, thus we cannot determine kinetics from our simulation, only possible packing structures.

3.2.3. Aggregate model and analysis. To analyze the aggregation of molecules in the simulation, we use a graph model, allowing us to utilize a repertoire of algorithms for statistic collection. We define the state of the system as graph $G\{V, E\}$, where V is the set of molecules and E is the set of edges. Given two vertices $\{v_l, v_r\} \in V$, an undirected edge $e\{v_l, v_r\} \in E$ if and only if v_l and v_r are bound. Since antigen only binds to receptors and vice versa, the graph is bipartite. Data about the overall aggregation process can be extracted from the graph.

Due to motion and association, the antigen/receptor positions will become more clustered as aggregates form. These clusters can be observed experimentally, and a theoretical study of clustering can be an instrument to compare our model with experiment. To quantify clustering in our models, we use a geometry-based statistical analysis of clustering tendency, the Hopkins statistic.[24] The Hopkins statistic is a measure of spatial randomness which utilizes nearest-neighbor distance of randomly sampled points and randomly selected probes (known molecule locations). In line with previous work,[16,61] we only use receptor molecule positions in our analysis. We calculate the Hopkins statistic in a similar fashion as in ref. [61] and for nearest neighbor calculations, we use the Euclidean distance between two points. The values calculated for the Hopkins statistic range from [0,1]. The closer the Hopkins statistic value is to 0.5, the more randomly spaced the points are, whereas the closer the value is to 1.0, the more clustered the data is.

To gain further insights into the aggregate structures, we take our modeled aggregate structures and generate all-atom structures. With these all-atom models, we can take measurements of the aggregate structure and analyze features of the aggregate such as steric hindrance. We can also quantify the model construction quality. An example Pen a 1 all-atom aggregate is shown in Fig. 4.

3.3. Rule-based modeling

In addition to the 3-D Monte Carlo simulations to determine aggregate size, we create rule-based models of IgE binding to Pen a 1 that incorporate geometric information in both the rules for aggregate formation and their associated rate constants. The rule-based model is implemented with RuleBender[56] using the BioNetGen language.[9] This method automates the generation of the coupled differential equations associated with the creation of new molecule aggregates as IgE binds to the available binding sites of Pen a 1.

3.3.1. Steric effects in a protein complex. Steric effects of receptors bound to an antigen with multiple binding sites in relation to binding site exclusion has been analyzed in ref. [23] However, they only investigated low dimensional (1D and 2D) shapes with specific geometries (surface or array) and

https://doi.org/10.1017/S0263574716000199

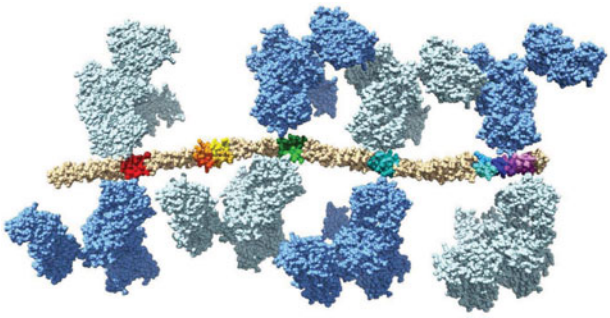


Fig. 4. A resulting aggregate structure generated using our method. The eight IgE-FcϵRI (light/medium blue) are bound to the Pen a 1 antigen (tan) at various binding sites on the antigen (various colors).

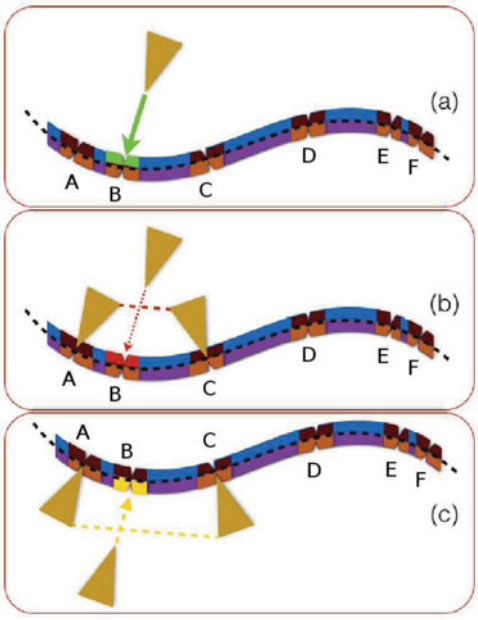


Fig. 5. Steric hindrance induced by neighbor occupation. The six binding regions are labeled A, B, C, D, E, F. (a) No neighbors: receptors are free to bind, (b) Negative curvature reducing binding rate constant, and (c) Positive curvature with possible effect on binding rate.

either ordered or uniformly random binding site distributions. We analyze how the conformation of the antigen affects steric constraints of the system. A description for dependency on neighbor occupancy is shown in Fig. 5. Steric hindrance induced by neighbor occupation can be broken down into three categories. First, IgE can easily bind to a region if neighboring sites are free (Fig. 5(a)). Second, on the strand with negative curvature around a region, occupation of nearest-neighboring regions can reduce accessibility of IgE to this region, effectively reducing the binding rate constant (Fig. 5(b)). Finally, on the opposite side with positive curvature, IgE can still bind to a region even if its nearest neighbors are occupied (Fig. 5(c)). The binding rules (listed in the appendix) are written with explicit neighboring site dependency.

3.3.2. Model construction and calculations. From this geometric analysis, we know that a negative curvature in the coiled coil may introduce hindrances to the accessible surface area for IgE binding and potentially brings binding sites closer. This makes the accessibility of a receptor to a particular binding site dependent on whether its neighboring sites are bound to IgE or not (namely their occupation states). However, because there are 16–18 available binding sites in the antigen, the introduction of geometric effects may lead to the number of rules becoming too overwhelming to implement, as a dependency on the occupation state of neighboring binding sites has to be explicitly added to some

https://doi.org/10.1017/S0263574716000199

of the rules of binding events. Therefore, we make a few assumptions and simplifications to generate the rules and associated ODEs:

- We assume that IgE binds to a single binding site in Pen a 1, i.e., binding events in which IgE binds to two sites on the same Pen a 1 are forbidden.
- To compare with the Monte Carlo simulations, which were carried out for a single Pen a 1 molecule, we do not allow crosslinking through IgE binding to two or more different Pen a 1 molecules.
- We significantly reduce the number of rules (and ODEs) by assuming that each IgE binds to a region on Pen a 1 known to have one or more binding sites. This is a reasonable assumption as binding sites in the same region are close (<5 nm): In the event of IgE binding to one binding site in a particular region, the other binding site(s) in the same region may be automatically blocked.
- Because the binding region on the tail of Pen a 1 is longer than the others (see Fig. 1), for the purpose of our rule-based simulations, we split this longer region into two independent ones, resulting in each strand of the coiled-coil structure having six binding regions.
- We can further decrease the number of rules by considering that each strand in Pen a 1 binds IgE independently of the other, i.e., the occupation state of any binding site on one strand of the coiled coil is independent of the occupation state of any binding site on the other strand. Since each strand has six regions, the maximum number of configurations of IgE binding for each strand is $2^6 = 64$. The aggregate sizes of IgE-Pen a 1 with 12 binding regions is now given by the combined independent probability of the aggregate formation in each strand of the coiled coil.

The probability of finding aggregates of size 0 to 12 is calculated by simulating each strand of the coiled coil separately, with different rules depending on the positive or negative curvature of the strand. The probability $P(n)$ to form an aggregate of size n is given by:

$$P(n \leq 6) = \sum_{m=0}^{n} P_I(m) P_{II}(n-m),$$

$$P(n > 6) = \sum_{m=n-6}^{6} P_I(m) P_{II}(n-m),$$

where $P_{I(II)}(n)$ is the independent probability of forming an aggregate of size n in strand $I(II)$.

3.3.3. Model rate constants. In order to analyze the influence of both rules and rate constants on our rule-based modeling results, we create a set of rules for Pen a 1. The *General* rule set in Tables IV and V takes into account neighboring binding site interactions and employs hierarchical binding rate constants. This means that when IgE binds to any site i on Pen a 1, the associated rate constant depends on the occupation of its nearest (first-order) and next-nearest (second-order) neighbors. We define four hierarchies of binding, thus the binding rate constants are four independent parameters. The unbinding rate constants are equal to 0.01 s^{-1} for all aggregate formation rules. Neighbors are defined by geometry: as the Pen a 1 molecule has a slight S-shaped curvature, binding sites on the concave sides of the molecule (with negative curvature) are closer than sites on the convex side of the molecule (positive curvature).

Resolution changes can be simulated in two ways: by fixing the binding rates and changing the rules for each resolution or by keeping rules fixed and changing the binding rates associated with each rule. The former requires a very careful analysis of the geometry of the receptor/antigen system at each resolution. The latter only needs a set of rules that roughly represents the site interactions given the curvatures of the protein complex, and rates can be changed as binding to particular sites is allowed or disallowed. Therefore, we focus on an implementation of the latter by generating a set of rules which capture the geometric constraints of the molecules and define rate constants that are dependent on the occupation of neighboring sites.

The advantage of using independent rate constants as parameters related to neighbor occupation is that we can simulate high and low resolution studies by turning rules *on* or *off*. The rate constant of a given rule determines whether or not the rule is *on* (non-zero rate constant) or *off* (rate constant set to zero). Rules that favor the formation of large aggregates are turned *on* for low resolution. As our Monte Carlo results indicate, the loss of detail leads to a reduced volume of Pen a 1 and

https://doi.org/10.1017/S0263574716000199

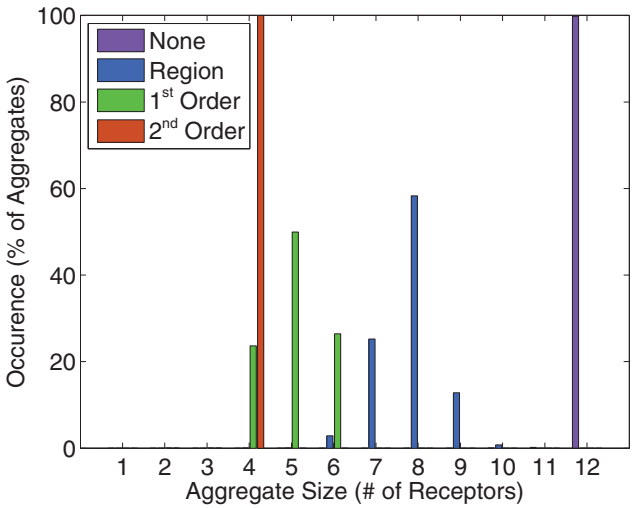


Fig. 6. For the same *General* sets of rules, we selectively turn hierarchic rates *on* and *off*. *On* binding rates have their value fixed at 1.0 molecule^{-1} s^{-1} and *off* binding rate constants are set to zero. As we make the rules more restricted, smaller aggregates are formed. The purple peak at 12 is labeled “None" for having no restrictions on binding due to neighbors (k_{f1}, k_{f2}, k_{f3}, and k_{f4} are *on*). This is why the largest possible aggregates are formed almost 100% of the time. The data labeled “Region" (blue) allows for binding to a site even if nearest and next-nearest neighbor sites are bound (k_{f1}, k_{f2}, and k_{f3} are *on*; k_{f4} is *off*). First order interactions (nearest-neighbors, green) are not allowed for this data set (k_{f1} and k_{f3} are *on*; k_{f2} and k_{f4} are *off*). The peak in red (second-order) does not allow binding to sites if both their nearest and next-nearest neighbors are occupied (k_{f1} is *on*; k_{f2}, k_{f3}, and k_{f4} are *off*).

receptors, thus exposing possible binding sites. At high resolution, the volume of receptors is larger and the extra detail can reduce binding site availability if a number of sites are already occupied. This indicates that the binding rate constants for rules associated with the formation of large aggregates should be turned *on* (for more binding events), and for formation of small aggregates some of these rate constants should be *off*, thus allowing fewer binding events.

The four binding rate constants are assigned to the rules as follows: k_{f1} is assigned to rules that specify that none of the neighboring binding regions are occupied, k_{f2} is assigned to rules that specify that one nearest-neighbor region is occupied, k_{f3} is assigned to rules that specify that one next-nearest neighbor is occupied, and k_{f4} is assigned to rules that specify that two nearest neighbors are occupied. As the tail region in Pen a 1 was split into two independent regions for our model, (regions E and F in the rule set, as seen in Fig. 5) we treat these regions as a special case and assign the rate k_{f4} to the E and F binding rules that specify that nearest-neighbor E or F is occupied.

We demonstrate how the rules were selected using the rules associated with binding region A as an example. Region B is a nearest neighbor to region A, so we specify region B as a region that affects the binding rate of region A in the rules for A binding for both strands. On strand I, the next nearest neighbor to region A, region C, is located in a region of positive curvature (see Fig. 5(c) for an illustration of positive curvature). Therefore, in the A binding rules for strand I, we do not specify region C as a neighbor that affects the binding rate of region A. On strand II, region C is located in a region of negative curvature (see Fig. 5(b) for an illustration of negative curvature) along with region A. Therefore, in the A binding rules for strand II, we specify region C as a neighbor that affects the binding rate of region A.

We use the *General* rule set (Appendix) to illustrate how the influence of neighbor occupancy hierarchy on binding site probability favors particular aggregate sizes. If all rules are *on* with identical binding rates $k = 1.0$ molecule^{-1} s^{-1}, the distribution of aggregate size is skewed to larger aggregates. However, as we turn rules *off* by setting their associated binding rates to zero in hierarchic order (larger to smaller aggregates), we see the progression shown in Fig. 6, until we obtain a single peak at aggregate size 4 if all nearest and next-nearest neighbors to site i need to be empty for a binding event to occur. Note that binding affinities are not known for the binding sites of Pen a 1, so we use values known for DF3. Since our system is finite in size, the association rate unit of molecule^{-1} s^{-1} is used (further discussed in Section 4.1).

https://doi.org/10.1017/S0263574716000199

The analysis shown in Fig. 6 indicates that it is possible to fine tune a particular aggregate distribution by choosing the rule set wisely (based on geometric input) and by setting binding rate constants appropriately. Finding a proper rule set is one of the main difficulties of this method, but PDB structures and feedback from Monte Carlo 3-D rigid body simulations can give crucial input on this step. Binding rate constants can be varied as well. Our Monte Carlo simulations assume a constant binding rate of $k = 1.0$ molecule^{-1} s^{-1} for all sites. However, to mimic loss of accessible volume to a particular binding site on Pen a 1, the rates can be varied to improve fits to data. These rates are considered free parameters of the simulation and they represent physical binding rates qualitatively.

3.3.4. Comparison analysis. In order to quantify the difference between the Monte Carlo and rule-based modeling aggregate sizes for each resolution, the residual sum-of-squares (RSS) normalized by the number of possible aggregate sizes (13) was calculated for each resolution. The equation used to calculate the normalized RSS is:

$$RSS = \frac{\sum_{i=1}^{N}(P_{MC}^{i} - P_{RBM}^{i})^2}{N},$$

where N is the total number of possible aggregate sizes in a histogram (each histogram has the same number of possible aggregate sizes), P_{MC}^{i} is the occurrence probability of the ith aggregate size of the Monte Carlo data, and P_{RBM}^{i} is the occurrence probability of the ith aggregate size of the rule-based modeling data.

Since the data points used in this calculation are probabilities, the maximum possible normalized RSS is one, and the minimum possible normalized RSS (corresponding to two identical histograms) is zero.

4. Results

4.1. Experimental setup

4.1.1. Monte Carlo simulation. For our Monte Carlo simulations, we simulate a discrete patch of membrane 200 nm x 200 nm (40,000 nm^2). In all simulations, we apply reflecting boundary conditions, ensuring that the number of molecules is kept constant as they are not permitted to exit the area representing the membrane patch when a boundary is reached.

We simulate 24 receptors for all experiments, resulting in a density of ∼600 receptors μm^{-2}. In two different experiments, we simulate twelve DF3 and one Pen a 1 antigen molecules at four distinct resolutions, reducing the models of both antigen and receptor by 0%, 50%, 75%, and 90%. Due to the presence of multiple DF3 molecules, it is possible to observe crosslinking in the DF3 simulations. However, no antigen-mediated crosslinking can be observed in the Pen a 1 simulations as only one molecule of Pen a 1 is simulated in each 3-D Monte Carlo experiment. We use the diffusion coefficient 0.09 μm^2 s^{-1} of IgE-FcϵRI found in ref. [3] for all molecules. We use a time step of 10 μs and run experiments for 500,000 time steps, long enough for the simulations to reach a steady state. Association and dissociation rates of 1.0 molecule^{-1} s^{-1} and 0.01 s^{-1}, were used for both antigens. Since our system is finite in size, the association rate unit molecule^{-1} s^{-1} was calculated from the original units of M^{-1} s^{-1} using calculations from ref. [43] The dissociation rate from ref. [59] is used unchanged. Thirty (30) runs of each experiment were performed.

As IgE aggregates get larger, they have been shown to slow down.[3] Since this reduction in speed has not been fully quantified, we use a theoretical basis for modeling the slow down in our system.[17,32] While many slow down models could be employed, we decided to use a model where we use receptor count as a surrogate for size measurement to determine relative slow down. The slow down is incorporated into the simulation by diffusing aggregates inversely proportional to their size, i.e., the diffusion coefficient of an aggregate linking three receptors is 1/3 of the original coefficient. We note that our slow down scheme does not account for physical barriers that exist in the cell membrane that may induce slowing/immobilization (like those in ref. [44]). Also, the geometric analysis in this paper emphasizes packing structure rather than aggregation kinetics, so the choice in slow down method should not impact aggregate structure packing at steady state. Rotations of large aggregates are limited by the diffusion rate of the fastest moving receptor. The receptor furthest from the center

https://doi.org/10.1017/S0263574716000199

Table I. Model reduction statistics including polygon counts and volumes of the molecular models generated at a variety of resolutions.

Molecule name	Model property	Model percent reduction			
		0%	50%	75%	90%
Receptor	Polygons (#)	4876	2438	1216	490
	Volume (nm^3)	234.98	227.90	208.73	162.31
	Volume (%)	100.00	96.99	88.83	69.07
DF3	Polygons (#)	1208	604	302	120
	Volume (nm^3)	15.83	14.90	13.16	9.74
	Volume (%)	100.00	94.13	83.13	61.53
Pen a 1	Polygons (#)	2328	1164	582	234
	Volume (nm^3)	51.60	49.95	44.86	28.80
	Volume (%)	100.00	96.80	86.94	55.81

of the aggregate is limited by its diffusion constant, thus limiting the amount of rotation an aggregate can make. For Hopkins statistic calculations, we set the number of measurements taken to five and repeat the calculation 1000 times. These values are then either averaged or sorted into bins to produce histograms used to quantify clustering.

Monte Carlo simulation code was developed using the Parasol Motion Planning Library (PMPL). Simulations were run on a supercomputer housed at UNM utilizing single cores of Intel Xeon E5645 processors with 4 GB of RAM per processor.

4.1.2. Rule-based modeling. For the rule-based modeling, we specified the rules in the BioNetGen language and conducted ODE simulations using BioNetGen (Version 2.2.6) via the RuleBender software package (Version 2.0.382r3).[56] We simulated 100 Pen a 1 antigen and 1000 receptors in each experiment. Since we simulated strands individually, we ended up with a population of 100 strand I molecules, 100 strand II molecules, and 1000 receptors. These were simulated using a time step of 0.01 s and run for 1000 time steps.

We obtained optimum rate constants using single-parameter and double-parameter scanning. We conducted single-parameter scans for the 0% reduction, varying k_{f2} from 0.0 to 1.0 $molecule^{-1}\,s^{-1}$, and for the 90% reduction, varying k_{f4} from 0.0 to 0.05 $molecule^{-1}\,s^{-1}$. We used a step size of 0.001 for these scans. We also conducted double-parameter scans for the 0%, 50%, and 75% reductions, in which we scanned all possible values of k_{f3} from 0.0 to 1.0 $molecule^{-1}\,s^{-1}$ while scanning all possible values of k_{f2} such that k_{f2} is less than or equal to k_{f3}, using a step size of 0.01.

4.2. Monte Carlo results

4.2.1. Volume and timing. We begin with an analysis of the impact of polygon reduction on the volume of the model. Table I shows the number of polygons and volume for each model. The polygon reduction algorithm works by specifying a percentage of the polygons to reduce, leading to the close correspondence between the reduction percentage and the number of polygons. We find that volumes decrease with increased reduction. Such decrease is expected, and can be quite dramatic (nearly 50% for 90% reduced Pen a 1). We note that volume reduction does not necessarily mean less realistic results; "soft docking" approaches[21] allow a certain degree of inter-protein penetration to approximate flexibility given rigid structures.

As seen in Fig. 7, the reduction in polygons has a clear effect on runtime. We see a linear increase in runtime versus model polygon count. This is due in part to the nature of the rigid body modeling; collision detection is a major factor in computation time and is highly dependent on model complexity. We attribute the calculation of binding site interactions, whose costs depend on valency and molecule size, to the difference in slope between DF3 and Pen a 1 runtimes.

https://doi.org/10.1017/S0263574716000199

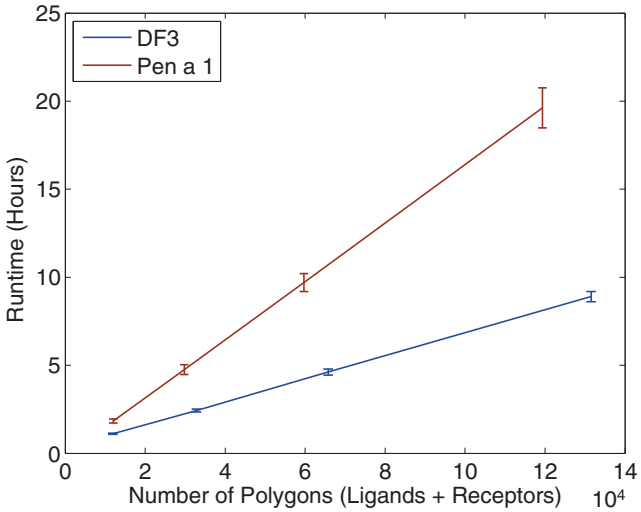


Fig. 7. Runtimes of the different resolutions of the same model. The x-axis is the sum of the number of polygons used to describe all of the models for any given experiment. The y-axis is the runtime in hours.

4.2.2. Impact of resolution on quality of results. Simulations are run until a steady state is reached. To ensure the system is stable, we count the number of bonds between molecules, i.e., the number of edges in G (Fig. 8). Figure 8(a) shows the number of edges in G for DF3 in blue. We see that for DF3, all of the reductions generate similar numbers of connections. The more reduced models produce slightly more connections but all of the average lines are very close. The mean of each reduction is contained in the overlap of the standard deviation of all reductions.

In Fig. 8(b) we see that Pen a 1 model resolution has a higher impact on the number of connections that are made. The 90% reduced model made on average nearly two more connections than the 0% reduced model for a single antigen. This is one of the side effects of reducing model volume. With the reduction, there is more open volume around the binding site, reducing steric hindrance of receptors trying to bind to sites in the same or adjacent regions.

To further analyze the implications of model reduction, we plot histograms of aggregate size versus percentage of occurrence (Fig. 9). We see in Fig. 9(a) that there is minimal impact on aggregate size distributions for the DF3 experiment. The distribution for each model reduction seem to be the same.

This is not the case for Pen a 1, seen in Fig. 9(b). The distribution has the two least reduced models peaking near aggregates of size 7 whereas the two most reduced models peak near aggregates of size 8. This is attributed to the volume reduction of the model which in effect relaxes the steric constraints. With a smaller volume, more free space is available for a molecule to pack into a tight space within a given aggregate.

4.2.3. Clustering analysis of DF3. To quantitatively analyze the clustering of the system, we measure the Hopkins statistic of the receptors over the course of the simulations. We focus on the analysis of clustering for DF3 due to the availability of experimental data.[38] Unfortunately, a similar analysis for Pen a 1 does not result in significant data as there is only one Pen a 1 allergen in each simulation run. To evaluate clustering, the Hopkins statistic values were calculated over the course of the simulation and the results are an average of the simulations for each experiment as seen in Fig. 10. These values are then plotted and compared to the values obtained experimentally in ref. [38] as shown in Fig. 10.

For a baseline, we performed a Hopkins statistic calculation for a simulation with only receptors and no antigen and produced the plot in Fig. 10(a). We find that for the *no antigen* simulation, the value does not change and is hovering at around 0.5, indicating that the receptors are essentially randomly distributed. We compare our *no antigen* simulation with the 0 nM experiment in ref. [38] (Fig. 10(a), red dashed line). Our value (mean 0.5) differs from what is seen experimentally (mean 0.74); however, this difference can be attributed to the fact that cell membranes present topological inhomogeneities that result in natural receptor organization.[2,36,37,48] These features are not incorporated in our simulations for simplicity. Therefore, we observe no clustering instead of the slightly clustered distribution observed experimentally.

https://doi.org/10.1017/S0263574716000199

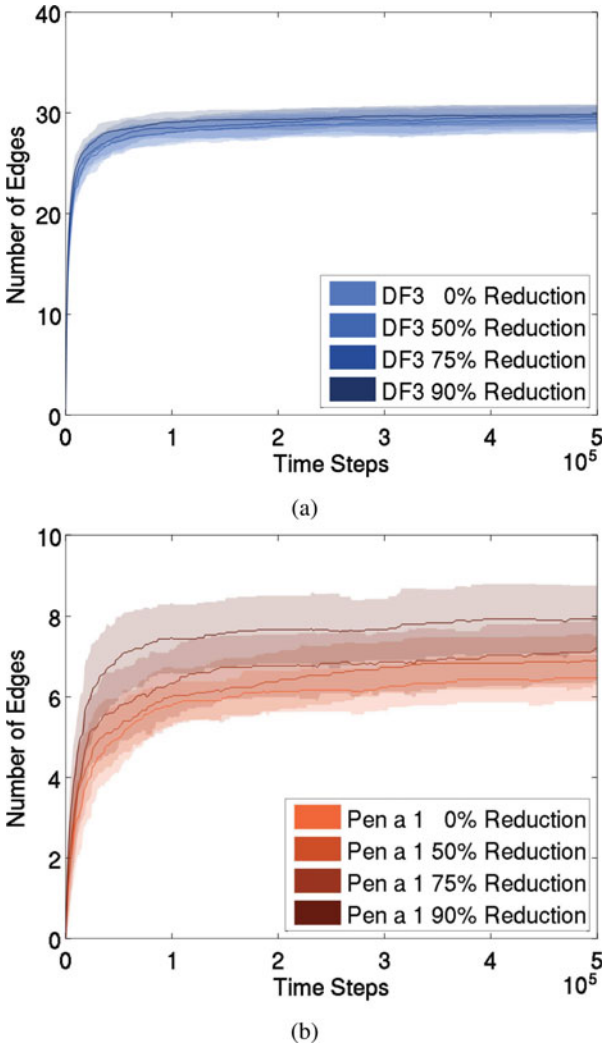
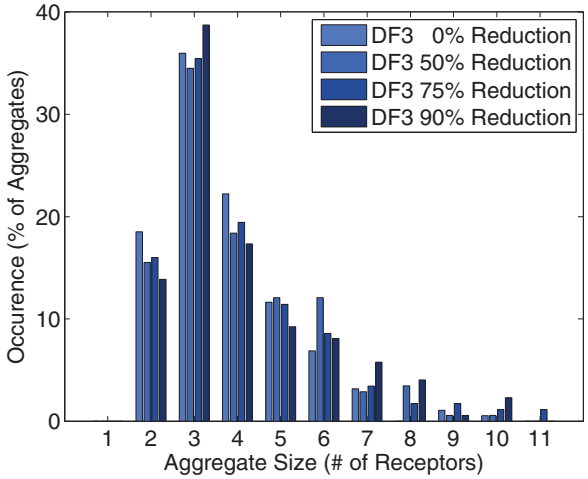


Fig. 8. Influence of model resolution on the number of connections made during a given simulation. For these plots, the x-axis is simulation time step and the y-axis is the number of edges in G. In (a), DF3 connections do not seem to be affected much by model resolution. In (b), Pen a 1 binding is affected by the use of different resolutions; lower resolutions generate more connections than higher resolutions.

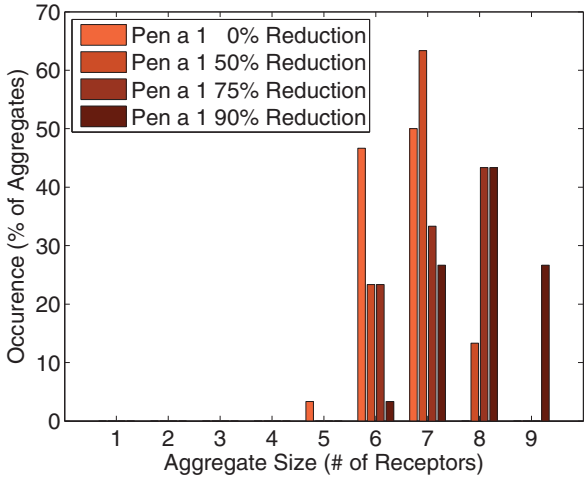
Based on the conclusion of ref. [38] where the authors imply that the ligand-induced aggregate state can influence cell signaling, we assume that final ligand–receptor aggregate size can be associated with cellular degranulation. Thus, to analyze clustering in the presence of antigen, we compare the Hopkins statistic results from the experiment in ref. [38] that resulted in optimal histamine secretion (10 nM of DF3) to the Hopkins statistic obtained from our Monte Carlo simulations. We find that the Hopkins statistic value at equilibrium from experiment has a mean value of 0.85 with error bars between 0.82 and 0.89 (Fig. 10(b), red-dashed line), while our simulations give a mean of 0.88 ± 0.05. This overlap indicates similar clustering observed in both our Monte Carlo simulation and experimentally derived results. In addition, we observe that model resolution does not impact the amount of clustering that occurs for DF3, as all of the values converge to the same result, as seen in Fig. 10.

To verify whether the averaged values are representative of the underlying clustering, we plot the histograms of the Hopkins statistic values calculated at the beginning and end of each experiment. These histograms are plotted against a normal distribution representing uniformly random distributed data to provide an intuition of the amount of clustering. For each experiment, we performed 30 runs, each contributing 1000 calculations, resulting in a total of 30,000 measurements per histogram. The beginning histograms are taken 1000 steps into the simulation to ensure the molecules move away from their initial grid state, which brings bias into the calculation. We see in Figure 11 that in the *no*

https://doi.org/10.1017/S0263574716000199



(a)



(b)

Fig. 9. Influence of model resolution on the size of aggregates generated during a given simulation. Histograms show aggregate size vs. percentage of aggregates of that size. For DF3 seen in (a), the different resolutions do not affect the distribution of aggregate sizes. However, for Pen a 1 seen in (b), aggregate size seems to be dependent on model resolution, with lower resolution Pen a 1 models producing larger aggregates.

antigen experiments, there is no change between the histograms at the beginning of the experiments (Fig. 11(a)) to the histograms at the end of the experiments (Fig. 11(b)), and both histograms are very close to the normal distribution (red line) indicating no clustering.

However, for DF3, we see (Fig. 12) that there is a significant shift in the histogram from start to end. The beginning of the experiment starts off as a random distribution (Fig. 12(a)). By the end of the experiment, we see a shift in the histogram away from the red normal distribution line (Fig. 12(b)), indicating clustering in the simulations. We note this shift is consistent for all resolution models of DF3.

4.2.4. Analysis of model quality. We also investigate the impact of model reduction on all-atom aggregate structures. In order to evaluate this, after aggregates are constructed with low-resolution polygon models, we construct the corresponding all-atom structure.

However, since the polygon models are much simpler than the all-atom structures, there may be unintended interactions. For example, when non-bonded atoms are too close, repulsion may occur due to van der Waals interactions. In order to evaluate the possible effects of transitioning between polygon and all-atom models, we counted the number of C_α atoms and DNP linker carbon rings within 7Å for IgE-FcϵRI and DF3. For Pen a 1, distances were calculated between C_α atoms for the

https://doi.org/10.1017/S0263574716000199

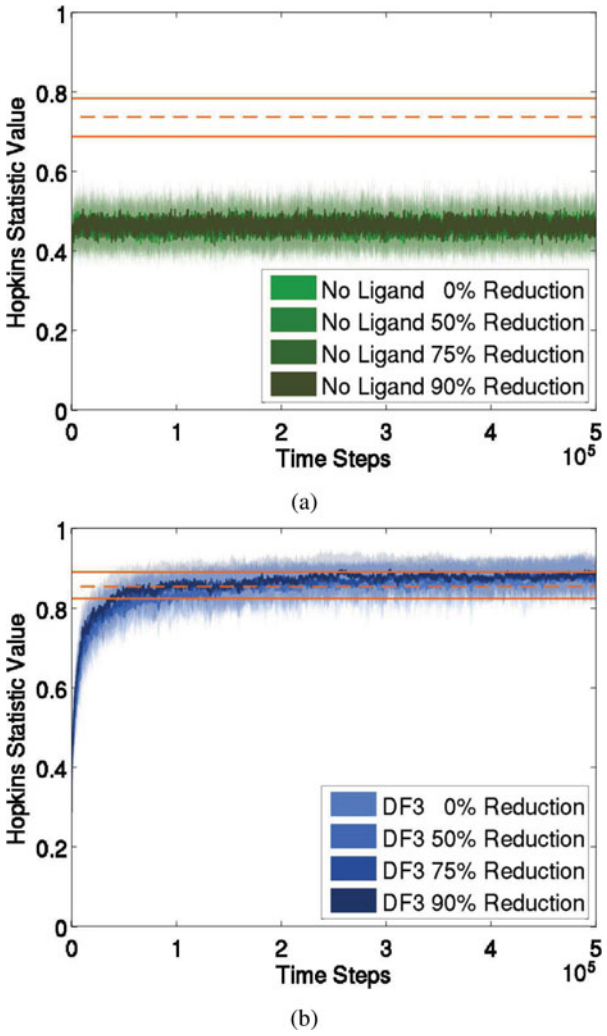


Fig. 10. Influence of model resolution on the clustering of receptors during a given simulation. Hopkins Statistic values (measure of clustering) are plotted over the course of the simulation without antigen in (a) and with DF3 in (b). We see that for both experiments, all resolutions converge to the same results. Comparison with experiment ([37]) are shown as the red lines (mean-dashed, variance-solid).

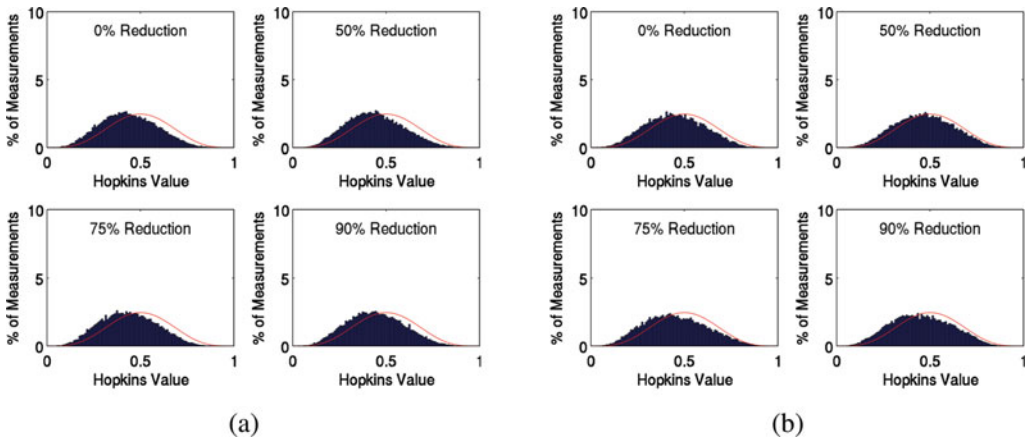


Fig. 11. Hopkins statistic for an experiment with *no antigen*. Histograms of Hopkins values for the beginning of the experiment are seen in (a) and for the end of the experiment shown in (b). Red curve indicates a normal distribution, i.e., no clustering.

https://doi.org/10.1017/S0263574716000199

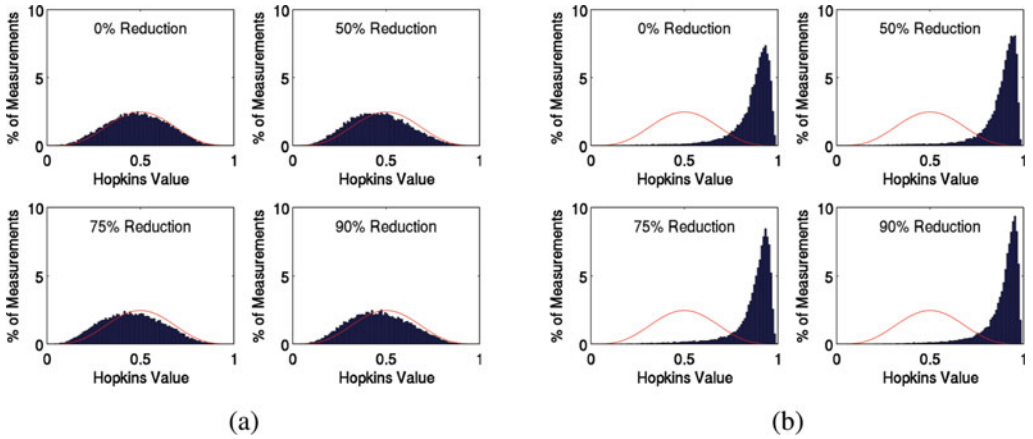


Fig. 12. Hopkins statistic for an experiment with antigen DF3. Hopkins statistic histograms for the beginning of DF3 simulations (a) and for the end of DF3 simulations (b) are shown. Red curve indicates a normal distribution, i.e., no clustering.

Table II. Percentage (%) of residues that exhibit a *potential collision*. Antigen residues involved in binding are not included.

Antigen simulated	Model percent reduction			
	0%	50%	75%	90%
DF3	0.0074%	0.0122%	0.0215%	0.1016%
Pen a 1	0.0158%	0.0350%	0.0816%	0.2238%

aggregated molecules. In order to indicate these proximal non-binding residues, we refer to them as *potential collisions*. Also, antigen binding sites are not included in the enumeration.

We can see from the results in Table II that model resolution has an impact on the number of *potential collisions* that exist in aggregate structures. *Potential collision* residues increase as resolution decreases. We see that DF3 is not significantly impacted by model resolution up to 75%. However, at 90%, there is an order of magnitude increase in *potential collisions*. We see that Pen a 1 model reduction generally has a higher percentage of residues in *potential collision* compared to DF3. This is attributed to the flexibility of the DF3 binding site. The DNP linker has a large, relatively open volume that can be bound. The binding sites of Pen a 1 are smaller in volume since they are on the molecular surface and are partially occupied by the molecular volume. Therefore, the antibodies have to be closer to the allergen surface in Pen a 1.

We note that overall, the number of residues in *potential collision* is minimal. Even at 90% reduced Pen a 1, aggregates generated have about 0.2% of residues in *potential collision*. These interactions could be first addressed through locally evaluated energetics and perturbations.

4.3. Rule-based modeling results

So far, we have presented the results of our Monte Carlo simulations for different resolutions, which explicitly include geometric effects in aggregate formation. Rule-based models of aggregate formation, on the other hand, need to encode all geometric information in the rules of antibody–antigen binding and their binding/unbinding rate constants.

Our approach is to vary the binding rates to reproduce the aggregate size distribution at a particular resolution for the *General* rule set. One way to achieve this is by doing a multi-parameter optimization of the four binding rate constants. This is especially useful when fitting the model to experimental data. In Fig. 13, we compare the aggregate size distribution for the rule-based model with the values found from the Monte Carlo simulation.

Most of the results shown in Fig. 13 were obtained by analyzing the rates via single and two-parameter scans. For the 0% and 50% resolutions, we fixed k_{f1} to unity in all runs and performed

https://doi.org/10.1017/S0263574716000199

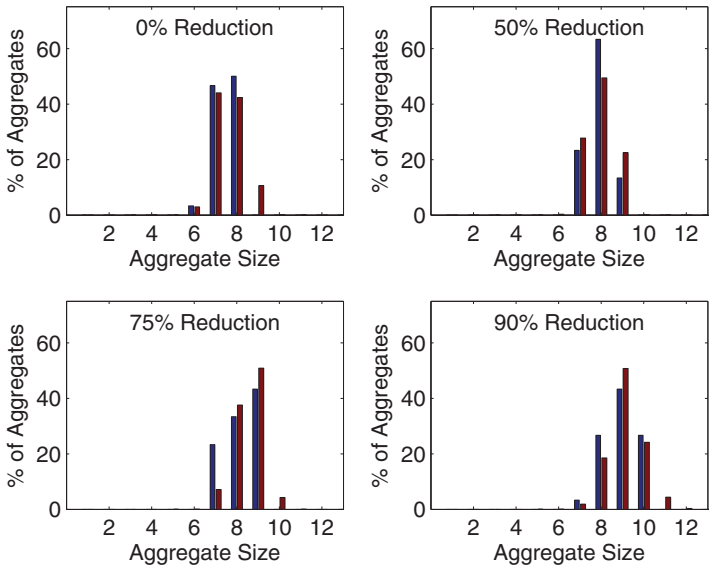


Fig. 13. Comparison of Monte Carlo (Blue) and rule-based model (Red) results for the *General* rule set with variable rates for different resolutions. Rates and RSS values are shown in Table III.

scans of k_{f3} from 0.0 to 1.0 molecule^{-1} s^{-1} and k_{f2} from 0.0 molecule^{-1} s^{-1} to k_{f3}, both at 0.01 intervals. For the 90% resolution, we fixed $k_{f1} = k_{f2} = k_{f3} = 1.0$ molecule^{-1} s^{-1} and varied k_{f4} from 0.0 to 0.05 molecule^{-1} s^{-1} by 0.001 increments. In all cases, the value chosen for the variable parameter was the one that resulted in the smallest RSS from the Monte Carlo data. However, it is important to highlight that even though the RSS from the Monte Carlo data is a reasonable measure of comparison between the two methods, our Monte Carlo data was obtained from 30 independent runs, and additional runs could change the overall distribution of the histograms, thus changing the RSS significantly. For this reason, we use this number mostly as a guide and avoid fitting rates to exactly reproduce the Monte Carlo data.

The rates listed in Table III for the 75% resolution were fitted from the trends of $k_{f1} = k_{f3} = 1.0$ molecule^{-1} s^{-1} and increasing k_{f2}, k_{f4} for lower resolutions, and they show a sufficiently close result to the Monte Carlo data (Figure 13). The trend of increasing rate values of k_{f2} (binding with one nearest neighbor occupied) and k_{f4} (binding with two nearest neighbors occupied) reinforce the intuition that the volume loss due to resolution reduction impacts the binding of neighboring regions.

At the 75% resolution, a two-parameter scan for k_{f2} and k_{f3} yields a slightly better RSS for $k_{f2} = k_{f3} = 0.12$ molecule^{-1} s^{-1} at fixed $k_{f1} = 1.0$ molecule^{-1} s^{-1} and $k_{f4} = 0$, which deviates from the parameter trends. This particular result also seems misleading because at higher resolutions one would expect that $k_{f1} \leq k_{f3}$, due to restrictions on neighbor interactions. However, it is expected that multi-parameter scans may lead to numerous minima of RSS, thus the best-fit solution may not be unique.

5. Conclusion

In our effort to study geometric packing of large protein complexes, we have investigated the impact of model resolution on aggregate formation and clustering. The interest in working with lower resolution models is to improve computational performance while preserving packing structures. In the case of antigen/receptor binding, it is believed that larger clusters are associated with a stronger allergic reaction. The focus on Pen a 1 in this work is due to its particular shape and its importance as a strong allergen.

We have shown that model resolution has negligible impact on DF3, and minimal impact on Pen a 1. This may be due to the geometric nature of the proteins: While DF3 is small and globular, Pen a 1 is long and rod-like, and therefore more affected by volume reduction. While volume reduction is quite dramatic as resolution is lowered, potential collisions among bound receptors and antigens are quite small across all resolutions (highest value of 0.2% for 90% reduced Pen a 1), with reasonable

https://doi.org/10.1017/S0263574716000199

Table III. Binding and unbinding rate constants and RSS differences for the rule-based model to capture the aggregate size distributions of different resolutions.

Rate value	Model percent reduction			
	0%	50%	75%	90%
k_{f1} (molecule^{-1} s^{-1})	1.00	1.00	1.00	1.00
k_{f2} (molecule^{-1} s^{-1})	0.07	0.12	0.50	1.00
k_{f3} (molecule^{-1} s^{-1})	1.00	1.00	1.00	1.00
k_{f4} (molecule^{-1} s^{-1})	0.00	0.00	0.00	0.01
k_r (s^{-1})	0.01	0.01	0.01	0.01
RSS	0.001367	0.002283	0.002731	0.001135

variation among resolutions of about one order of magnitude. These results will guide future work by informing us how to tailor the amount of reduction to the type of molecular shape. The basic shapes of DF3 and Pen a 1 are very simple, and we are able to determine the well-defined characteristics of their geometric reductions. Information reported here alongside additional future surface/volumetric analysis will be necessary to ensure our model construction/reduction appropriately capture molecular topology when dealing with more complex molecular structures.

We performed a clustering analysis of the DF3 and receptor system at different resolutions to compare with experimental data from ref. [38] Our results were able to reproduce the experimental Hopkins statistic metric of cluster formation at a non-saturating concentration of DF3 and receptors. More importantly, our analysis shows that a loss in resolution does not affect the results. This implies that our method can efficiently compare an experimental observable to simulated data for a very large system of molecules. This is currently infeasible for state-of-the-art MD and coarse-grained simulations. Granted, our method is not yet able to calculate energetics of the interactions between molecules. All physical information about the system is included in the binding/unbinding rates, diffusion coefficient (Monte Carlo), and aggregation rules (rule-based method). As our focus is on geometric packing structures, this level of theory has been shown to be sufficient by the agreement with experimental results.

Our study of aggregate formation with a rule-based model of Pen a 1 is a novel contribution of this paper. The ODE-generating rules are based on the inherent geometry of Pen a 1: a slight S-shaped coiled coil, double-stranded protein. Binding sites in regions of negative curvature are affected differently than binding sites in regions of positive curvature during binding events. Our rule set was built to reflect these steric constraints. The assumption of hierarchical binding rates that are varied depending upon the accessibility of sites on the Pen a 1 surface allows us to investigate differences in aggregate formation across resolutions. We see that the rule-based model acts as a complementary method to our original Monte Carlo 3-D rigid body model. Where one is fast (a single run lasts about 10 seconds) but does not present explicit consideration of geometry (rule-based model), the other is still fast considering the number of molecules and detail of the system (a single run takes 2–20 hours for Pen a 1), but is advantageous as geometry is explicitly modeled and geometric packing can be easily tracked throughout the simulation.

For our future work, we aim to include joint flexibility in the rigid body models for the Monte Carlo simulation. This will allow for the generation of distinct packing structures where overall volume becomes another observable of interest. Accessible volume may change even for the same number of receptors bound to the antigen. More importantly, the incorporation of flexibility will allow us to expand our analysis to the results of other experimental methods, including EM. Fitting EM density data of molecules like IgE would require jointed models for fitting due to the amount of flexibility in the molecule. Determining the structure of large heterogeneous aggregates is a difficult problem in EM, and we see the potential for using our method for fitting.

We also are interested in running more simulations for more accurate statistical analysis. A simulation of multiple Pen a 1 antigens crosslinking through receptor binding is also of interest.

https://doi.org/10.1017/S0263574716000199

There is much potential in developing our version of the geometric rule-based model, as it can be applied to a variety of protein shapes to study how conformation affects aggregate formation.

Acknowledgments

This work is supported in part by the National Institutes of Health (NIH) Grant P50GM085273 supporting the New Mexico Spatiotemporal Modeling Center and NIH Grant P20RR018754 supporting the Center for Evolutionary and Theoretical Immunology. Special thanks to Bridget Wilson, Director of the New Mexico Spatiotemporal Modeling Center, Chang-Shung Tung from Los Alamos National Laboratory for providing structures, and to William S. Hlavacek from Los Alamos National Laboratory for fruitful discussions and for providing guidance on conducting multi-parameter optimizations. Also, thanks to Alan Kuntz and Erica Lopez for their early work on model resolution. Authors BDJ and KMM would like to acknowledge the organizers of the Eighth q-bio Summer School in Albuquerque, NM, where part of this work (rule-based modeling) was started. We also thank the UNM Center for Advanced Research Computing for providing the computational resources used in this work.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0263574716000199.

References

1. I. Al-Bluwi *et al.*, "Modeling protein conformational transitions by a combination of coarse-grained normal mode analysis and robotics-inspired methods," *BMC Struct. Biol.* **13**(1), S2 (2013).
2. N. L. Andrews *et al.*, "Actin restricts FcϵRI diffusion and facilitates antigeninduced receptor immobilisation," *Nature Cell Biol.* **10**(8), 955–963 (2008).
3. N. L. Andrews *et al.*, "Small, mobile FcϵRI receptor aggregates are signaling competent," *Immunity* **31**(3), 469–479 (2009).
4. A. R. Atilgan *et al.*, "Anisotropy of fluctuation dynamics of proteins with an elastic network model," *Biophys. J.* **80**(1), 505–515 (2001).
5. R. Ayuso, S. B. Lehrer and G. Reese, "Identification of continuous, allergenic regions of the major shrimp allergen pen a 1 (Tropomyosin)," *Int. Arch. Allergy Immun.* **127**(1), 27–37 (2002).
6. M. Baaden and S. J. Marrink, "Coarse-grain modelling of proteinprotein interactions," *Curr. Opin. Struct. Biol.* **23**(6), 878–886 (2013).
7. I. Bahar and A. J. Rader, "Coarse-grained normal mode analysis in structural biology," *Curr. Opin. Struct. Biol.* **15**(5), 586–592 (2005).
8. O. B. Bayazit, G. Song and N. M. Amato, "Ligand Binding with OBPRM and Haptic User Input: Enhancing Automatic Motion Planning with Virtual Touch," *International Conference on Robotics and Automation (ICRA)*, IEEE, New York, NY (2001) pp. 954–959.
9. M. L. Blinov *et al.*, "BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains," *Bioinformatics* **20**(17), 3289–3291 (2004).
10. D. Bratko *et al.*, "Molecular simulation of protein aggregation," *Biotechnol. Bioeng.* **96**(1), 1–8 (2007).
11. P. Cignoni, C. Montani and R. Scopigno, "A comparison of mesh simplification algorithms," *Comput. Graph.* **22**(1), 37–54 (1998).
12. D. E. Chandler, J. Strümpfer, M. Sener, S. Scheuring and K. Schulten, "Light harvesting by lamellar chromatophores in Rhodospirillum photometricum," *Biophysical journal* **106**(11), 2503–2510 (2014).
13. J. Cortes *et al.*, "Simulating ligand-induced conformational changes in proteins using a mechanical disassembly method," *Phys. Chem. Chem. Phys.* **12**, 8268–8276 (2010).
14. V. Danos and C. Laneve, "Formal molecular biology," *Theor. Comput. Sci.* **325**(1), 69–110 (2004).
15. L. A. Day, J. M. Sturtevant and S. J. Singer, "The kinetics of the reactions between antibodies to the 2,4 dinitrophenyl group and specific haptens. *Ann. New York Acad. Sci.* **103**(2), 611–625 (1963). ISSN: 1749-6632.
16. F. A. Espinoza *et al.*, "Using hierarchical clustering and dendrograms to quantify the clustering of membrane proteins," *Bull. Math. Biol.* **74**(1), 190–211 (2012).
17. Y. Gambin *et al.*, "Lateral mobility of proteins in liquid membranes revisited," *Proc. Natl. Acad. Sci. USA* **103**(7), 2098–2102 (2006).
18. T. D. Goddard, C. C. Huang and T. E. Ferrin, "Software extensions to UCSF chimera for interactive visualization of large molecular assemblies," *Structure* **13**(3), 473–482 (2005).
19. B. Goldstein and A. S. Perelson, "Equilibrium theory for the clustering of bivalent cell surface receptors by trivalent ligands. Application to histamine release from basophils," *Biophys. J.* **45**(6), 1109–1123 (1984).

20. G. Gruenert *et al.*, "Rule-based spatial modeling with diffusing, geometrically constrained molecules," *BMC Bioinform.* **11**, 307 (2010).
21. I. Halperin *et al.*, "Principles of docking: An overview of search algorithms and a guide to scoring functions," *Proteins: Struct. Funct. Bioinformat.* **47**(4), 409–443 (2002).
22. I. Hashmi and A. Shehu, "HopDock: A probabilistic search algorithm for decoy sampling in protein-protein docking," *Proteome Sci.* **11**(Suppl 1), S6 (2013). ISSN: 1477–5956.
23. W. S. Hlavacek, R. G. Posner and A. S. Perelson, "Steric effects on multivalent ligand-receptor binding: Exclusion of ligand sites by bound cell surface receptors," *Biophys. J.* **76**(6), 3031–3043 (1999).
24. B. Hopkins and J. G. Skellam, "A new method for determining the type of distribution of plant individuals," *Ann. Botany* **18**(70), 213–227 (1954).
25. P.-S. Huang, J. J. Love and S. L. Mayo, "A de novo designed proteinprotein interface," *Protein Sci.* **16**(12), 2770–2774 (2007).
26. Y.-F. Huang *et al.*, "Nanoparticlemediated IgEreceptor aggregation and signaling in RBL mast cells," *J. Am. Chem. Soc.* **131**(47), 17328–17334 (2009).
27. O. Ivanciuc, C. H. Schein and W. Braun, "SDAP: Database and computational tools for allergenic proteins," *Nucleic Acids Res.* **31**(1), 359–362 (2003).
28. R. K. Jha *et al.*, "Computational design of a PAK1 binding protein," *J. Mol. Biol.* **400**(2), 257–270 (2010).
29. Y. C. Kim and G. Hummer, "Coarse-grained models for simulations of multiprotein complexes: Application to ubiquitin binding," *J. Mol. Biol.* **375**(5), 1416–1433 (2008).
30. N. P. King *et al.*, "Computational design of self-assembling protein nanomaterials with atomic level accuracy," *Science* **336**(6085), 1171–1174 (2012).
31. M. L. Klein and W. Shinoda, "Large-scale molecular dynamics simulations of self-assembling systems," *Science* **321**(5890), 798–800 (2008).
32. J. D. Knight *et al.*, "Single molecule diffusion of membrane-bound proteins: Window into lipid contacts and bilayer dynamics," *Biophys. J.* **99**(9), 2879–2887 (2010). ISSN: 0006-3495.
33. Y.-T. Lai, N. P. King and T. O. Yeates, "Principles for designing ordered protein assemblies," *Trends Cell Biol.* **22**(12), 653–661 (2012).
34. Y. Li and Q. Huang, "Influence of protein self-association on complex coacervation with polysaccharide: A Monte Carlo study," *J. Phys. Chem. B* **117**(9), 2615–2624 (2013).
35. Y. Li *et al.*, "Monte Carlo simulation on complex formation of proteins and polysaccharides," *J. Phys. Chem. B* **116**(10), 3045–3053 (2012).
36. B. F. Lillemeier *et al.*, "Plasma membrane-associated proteins are clustered into islands attached to the cytoskeleton," *Proc. Natl. Acad. Sci.* **103**(50), 18992–18997 (2006).
37. D. Lingwood and K. Simons, "Lipid rafts as a membrane-organizing principle," *Science* **327**(5961), 46–50 (2010).
38. A. Mahajan *et al.*, "Optimal aggregation of FcϵRI with a structurally defined trivalent ligand overrides negative regulation driven by phosphatases," *ACS Chem. Biol.* **9**(7), 1508–1519 (2014).
39. K. Manavi, A. Kuntz and L. Tapia, "Geometrical Insights into the Process of Antibody Aggregation," *AAAI Workshop on Artificial Intelligence and Robotics Methods in Computational Biology (AIRMCB)*, AAAI Press, Menlo Park, CA (2013) pp. 26–31.
40. K. Manavi, B. S. Wilson and L. Tapia, "Simulation and Analysis of Antibody Aggregation on Cell Surfaces using Motion Planning and Graph Analysis," *Proceeding of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB)*, ACM, New York, NY (2012) pp. 458–465.
41. C. Maus, S. Rybacki and A. M. Uhrmacher, "Rule-based multilevel modeling of cell biological systems," *BMC Syst. Biol.* **5**, 166 (2011).
42. Autodesk Maya (2014) URL: http://www.autodesk.com/.
43. M. I. Monine *et al.*, "Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates," *Biophys. J.* **98**(1), 48–56 (2010).
44. D. V. Nicolau, J. F. Hancock and K. Burrage, "Sources of anomalous diffusion on cell membranes: A Monte Carlo study," *Biophys. J.* **92**(6), 1975–1987 (2006).
45. L. X. Peng *et al.*, "Aggregation properties of a polymeric anticancer therapeutic: a coarse-grained modeling study," *J. Chem. Inform. Model.* **51**(12), 3030–3035 (2011).
46. J. R. Perilla *et al.*, "Molecular dynamics simulations of large macromolecular complexes," *Curr. Opin. Struct. Biol.* **31**, 64–74 (2015).
47. X. Periole *et al.*, "Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition," *J. Chem. Theory Comput.* **5**(9), 2531–2543 (2009).
48. L. J. Pike, "Rafts defined: A report on the Keystone symposium on lipid rafts and cell function," *J. Lipid Res.* **47**(7), 1597–1598 (2006).
49. R. G. Posner *et al.*, "Simultaneous cross-linking by two nontriggering bivalent ligands causes synergistic signaling of IgE FcϵRI complexes," *J. Immunology* **155**(7), 3601–3609 (1995).
50. N. A. Rahman *et al.*, "Rotational dynamics of type I Fc epsilon receptors on individually-selected rat mast cells studied by polarized fluorescence depletion," *Biophys. J.* **61**(2), 334–346 (1992).
51. G. Reese *et al.*, "Reduced allergenic potency of VR9-1, a mutant of the major shrimp allergen Pen a 1 (Tropomyosin)," *J. Immunology* **175**(12), 8354–8364 (2005).
52. J. Rivera and A. M. Gilfillan, "Molecular regulation of mast cell activation," *J. Allergy Clin. Immunology* **117**(6), 1214–1225 (2006).

53. M. G. Saunders and G. A. Voth, "Coarse-graining of multiprotein assemblies," *Curr. Opin. Struct. Biol.* **22**(2), 144–150 (2012).
54. A. D. Schuyler and G. S. Chirikjian, "Normal mode analysis of proteins: A comparison of rigid cluster modes with C (alpha) coarse graining," *J. Mol. Graph. Modelling* **22**(3), 183–193 (2004).
55. D. Sil *et al.*, "Trivalent ligands with rigid DNA spacers reveal structural requirements for IgE receptor signaling in RBL mast cells," *ACS Chem. Biol.* **2**(10), 674–684 (2007).
56. A. M. Smith *et al.*, "RuleBender: Integrated modeling, simulation and visualization for rule-based intracellular biochemistry," *BMC Bioinformat.* **13**(8), S3 (2012).
57. B. S. Wilson, J. M. Oliver and D. S. Lidke, "Spatio-temporal signaling in mast cells," *Adv. Exp. Med. Biol.* **716**, 91–106 (2011).
58. K. C. Wolfe *et al.*, "Multiscale modeling of double-helical DNA and RNA: A unification through lie groups," *J. Phys. Chem. B* **116**(29), 8556–8572 (2012).
59. K. Xu *et al.*, "Kinetics of multivalent antigen DNP-BSA binding to IgE-FcϵRI in relationship to the stimulated tyrosine phosphorylation of FcϵRI," *J. Immunology* **160**(7), 3225–3235 (1998).
60. J. Yang *et al.*, "Kinetic Monte Carlo method for rule-based modeling of biochemical networks," *Phys. Rev. E* **78**(3), 031910 (2008).
61. J. Zhang *et al.*, "Characterizing the topography of membrane receptors and signaling molecules from spatial patterns obtained using nanometer-scale electrondense probes and electron microscopy," *Micron* **37**(1), 14–34 (2006).
62. L. Zhang, D. Lu and Z. Liu, "How native proteins aggregate in solution: A dynamic Monte Carlo simulation," *Biophys. Chem.* **133**, 71–80 (2008).

https://doi.org/10.1017/S0263574716000199

Appendices

Table A1. Rule Set for Strand I (T_I) of Pen a 1 in pseudo BioNetGen language format. Letters in parentheses represent free binding sites. "IgE" in parentheses represent occupied binding sites with the subscript indicating which site is occupied. Omitted letters represent binding sites not included in the rule (can be free or occupied). Dissociations are addressed with complementary rules (not shown), with rates $k_r = 0.01\ s^{-1}$.

Binding site	Reaction rule	Binding rate
A	T_I(A,B) + IgE → T_I(IgE$_A$,B)	k_{f1}
	T_I(A,IgE$_B$) + IgE → T_I(IgE$_A$,IgE$_B$)	k_{f2}
B	T_I(A,B,C,D) + IgE → T_I(A,IgE$_B$,C,D)	k_{f1}
	T_I(IgE$_A$,B,C) + IgE → T_I(IgE$_A$,IgE$_B$,C)	k_{f2}
	T_I(A,B,IgE$_C$) + IgE → T_I(A,IgE$_B$,IgE$_C$)	k_{f2}
	T_I(A,B,C,IgE$_D$) + IgE → T_I(A,IgE$_B$,C,IgE$_D$)	k_{f3}
	T_I(IgE$_A$,B,IgE$_C$) + IgE → T_I(IgE$_A$,IgE$_B$,IgE$_C$)	k_{f4}
C	T_I(B,C,D,E,F) + IgE → T_I(B,IgE$_C$,D,E,F)	k_{f1}
	T_I(IgE$_B$,C,D) + IgE → T_I(IgE$_B$,IgE$_C$,D)	k_{f2}
	T_I(B,C,IgE$_D$) + IgE → T_I(B,IgE$_C$,IgE$_D$)	k_{f2}
	T_I(B,C,D,IgE$_E$) + IgE → T_I(B,IgE$_C$,D,IgE$_E$)	k_{f3}
	T_I(B,C,D,E,IgE$_F$) + IgE → T_I(B,IgE$_C$,D,E,IgE$_F$)	k_{f3}
	T_I(IgE$_B$,C,IgE$_D$) + IgE → T_I(IgE$_B$,IgE$_C$,IgE$_D$)	k_{f4}
D	T_I(B,C,D,E,F) + IgE → T_I(B,C,IgE$_D$,E,F)	k_{f1}
	T_I(IgE$_C$,D,E,F) + IgE → T_I(IgE$_C$,IgE$_D$,E,F)	k_{f2}
	T_I(C,D,IgE$_E$) + IgE → T_I(C,IgE$_D$,IgE$_E$)	k_{f2}
	T_I(C,D,E,IgE$_F$) + IgE → T_I(C,IgE$_D$,E,IgE$_F$)	k_{f2}
	T_I(IgE$_B$,C,D,E,F) + IgE → T_I(IgE$_B$,C,IgE$_D$,E,F)	k_{f3}
	T_I(IgE$_C$,D,IgE$_E$) + IgE → T_I(IgE$_C$,IgE$_D$,IgE$_E$)	k_{f4}
	T_I(IgE$_C$,D,E,IgE$_F$) + IgE → T_I(IgE$_C$,IgE$_D$,E,IgE$_F$)	k_{f4}
E	T_I(C,D,E,F) + IgE → T_I(C,D,IgE$_E$,F)	k_{f1}
	T_I(IgE$_D$,E,F) + IgE → T_I(IgE$_D$,IgE$_E$,F)	k_{f2}
	T_I(IgE$_C$,D,E,F) + IgE → T_I(IgE$_C$,D,IgE$_E$,F)	k_{f3}
	T_I(E,IgE$_F$) + IgE → T_I(IgE$_E$,IgE$_F$)	k_{f4}
F	T_I(C,D,E,F) + IgE → T_I(C,D,E,IgE$_F$)	k_{f1}
	T_I(IgE$_D$,E,F) + IgE → T_I(IgE$_D$,E,IgE$_F$)	k_{f2}
	T_I(IgE$_C$,D,E,F) + IgE → T_I(IgE$_C$,D,E,IgE$_F$)	k_{f3}
	T_I(IgE$_E$,F) + IgE → T_I(IgE$_E$,IgE$_F$)	k_{f4}

https://doi.org/10.1017/S0263574716000199

Table A2. Rule Set for Strand II (T_{II}) of Pen a 1 in pseudo BioNetGen language format. Letters in parentheses represent free binding sites. "IgE" in parentheses represent occupied binding sites with the subscript indicating which site it occupies. Omitted letters represent binding sites not included in the rule (can be free or occupied). Dissociations are addressed with complementary rules (not shown), with rates $k_r = 0.01\ s^{-1}$.

Binding Site	Reaction Rule	Binding Rate
A	T_{II}(A,B,C) + IgE → T_{II}(IgE$_A$,B,C)	k_{f1}
	T_{II}(A,IgE$_B$) + IgE → T_{II}(IgE$_A$,IgE$_B$)	k_{f2}
	T_{II}(A,B,IgE$_C$) + IgE → T_{II}(IgE$_A$,B,IgE$_C$)	k_{f3}
B	T_{II}(A,B,C,D) + IgE → T_{II}(A,IgE$_B$,C,D)	k_{f1}
	T_{II}(IgE$_A$,B,C) + IgE → T_{II}(IgE$_A$,IgE$_B$,C)	k_{f2}
	T_{II}(A,B,IgE$_C$) + IgE → T_{II}(A,IgE$_B$,IgE$_C$)	k_{f2}
	T_{II}(A,B,C,IgE$_D$) + IgE → T_{II}(A,IgE$_B$,C,IgE$_D$)	k_{f3}
	T_{II}(IgE$_A$,B,IgE$_C$) + IgE → T_{II}(IgE$_A$,IgE$_B$,IgE$_C$)	k_{f4}
C	T_{II}(A,B,C,D) + IgE → T_{II}(A,B,IgE$_C$,D)	k_{f1}
	T_{II}(IgE$_B$,C,D) + IgE → T_{II}(IgE$_B$,IgE$_C$,D)	k_{f2}
	T_{II}(B,C,IgE$_D$) + IgE → T_{II}(B,IgE$_C$,IgE$_D$)	k_{f2}
	T_{II}(IgE$_A$,B,C,D) + IgE → T_{II}(IgE$_A$,B,IgE$_C$,D)	k_{f3}
	T_{II}(IgE$_B$,C,IgE$_D$) + IgE → T_{II}(IgE$_B$,IgE$_C$,IgE$_D$)	k_{f4}
D	T_{II}(B,C,D,E,F) + IgE → T_{II}(B,C,IgE$_D$,E,F)	k_{f1}
	T_{II}(IgE$_C$,D,E,F) + IgE → T_{II}(IgE$_C$,IgE$_D$,E,F)	k_{f2}
	T_{II}(C,D,IgE$_E$) + IgE → T_{II}(C,IgE$_D$,IgE$_E$)	k_{f2}
	T_{II}(C,D,E,IgE$_F$) + IgE → T_{II}(C,IgE$_D$,E,IgE$_F$)	k_{f2}
	T_{II}(IgE$_B$,C,D,E,F) + IgE → T_{II}(IgE$_B$,C,IgE$_D$,E,F)	k_{f3}
	T_{II}(IgE$_C$,D,IgE$_E$) + IgE → T_{II}(IgE$_C$,IgE$_D$,IgE$_E$)	k_{f4}
	T_{II}(IgE$_C$,D,E,IgE$_F$) + IgE → T_{II}(IgE$_C$,IgE$_D$,E,IgE$_F$)	k_{f4}
E	T_{II}(D,E,F) + IgE → T_{II}(D,IgE$_E$,F)	k_{f1}
	T_{II}(IgE$_D$,E,F) + IgE → T_{II}(IgE$_D$,IgE$_E$,F)	k_{f2}
	T_{II}(E,IgE$_F$) + IgE → T_{II}(IgE$_E$,IgE$_F$)	k_{f4}
F	T_{II}(D,E,F) + IgE → T_{II}(D,E,IgE$_F$)	k_{f1}
	T_{II}(IgE$_D$,E,F) + IgE → T_{II}(IgE$_D$,E,IgE$_F$)	k_{f2}
	T_{II}(IgE$_E$,F) + IgE → T_{II}(IgE$_E$,IgE$_F$)	k_{f4}

https://doi.org/10.1017/S0263574716000199