Extending Rule-Based Methods to Model Molecular Geometry

Brittany Hoard, Bruna Jacobson, Kasra Manavi, Lydia Tapia
Department of Computer Science
University of New Mexico
Albuquerque, New Mexico, 87131

***Abstract*—Computational modeling is an important tool for the study of complex biochemical processes associated with cell signaling networks. However, it is challenging to simulate processes that involve hundreds of large molecules due to the high computational cost of such simulations. Rule-based modeling is a computational method that can be used to model these processes with reasonably low computational cost, but traditional rule-based modeling approaches do not include details of molecular geometry. The incorporation of molecular geometry into biochemical models can more accurately capture details of these processes, and may lead to insights into how geometry affects the products that form. Furthermore, geometric rule-based modeling can be used to complement other computational methods that explicitly represent molecular geometry in order to quantify binding site accessibility and steric effects.**

In this work, we propose a novel implementation of rule-based modeling that encodes details of molecular geometry into the rules and the binding rate constant associated with each rule. We demonstrate how the set of rules is constructed according to the curvature of the molecule. We then perform a study of antigen-antibody aggregation using our proposed method. We first simulate the binding of IgE antibodies bound to cell surface receptors FcεRI to various binding regions of the shrimp allergen Pen a 1 using a previously developed 3D rigid-body Monte Carlo simulation, and we analyze the distribution of the sizes of the aggregates that form during the simulation. Then, using our novel rule-based approach, we optimize a rule-based model according to the geometry of the Pen a 1 molecule and the data from the Monte Carlo simulation. In particular, we use the distances between the binding regions of the Pen a 1 molecule to optimize the rules and associated binding rate constants. We perform this procedure for three molecular conformations of Pen a 1 and analyze the impact of conformation on the aggregate size distribution and the optimal rule-based model. We find that the optimized rule-based models provide information about the average steric hindrance between binding regions and the probability that IgE-FcεRI receptor complexes will bind to these regions. In addition, the optimized rule-based models provide a means of quantifying the variation in aggregate size distribution that results from differences in molecular geometry.

Keywords*—*rule-based model; geometric model; allergen-antibody interactions

I. Introduction

Computational methods are widely used to study biomolecular interactions due to their complexity. Models which are constrained by physicochemical principles are useful because they are based on causality and their parameters can be measured independently [1], [2]. Models that allow the incorporation of site-specific details and that can overcome the problem of combinatorial complexity are also highly useful for biomolecular simulations [2]–[4]. One technique that meets all of the above requirements is rule-based modeling. Rule-based modeling is a technique for studying the site dynamics of biomolecular networks [2], [5], which involves representing biomolecular interactions as local rules. With this method, a set of rules, each representing multiple reactions and each associated with a rate law that is assigned to these reactions, is specified. During simulation, a reaction network is created from which a set of coupled ordinary differential equations (ODEs) is derived. These equations characterize the rates of change of observables (such as chemical species). There are several different formalisms that may be used in creating rule-based models [2], [5].

One popular rule-based modeling software is BioNetGen [5], which uses graph rewriting. The biomolecules are represented by graphs, vertices represent molecular components, and edges represent bonds between these components [2], [5]. Biomolecular complexes are represented by connected sets of graphs. Rules are applied to these graphs and sets of graphs, and the graphs are changed according to the results of the biomolecular interactions specified by these rules [2], [5]. One disadvantage of the traditional rule-based model is that it does not capture details of molecular geometry. The traditional rule-based model is based on a set of binding rules that only include the number of binding sites on a molecule and do not incorporate geometric information [6], [7]. This limitation of the rule-based modeling approach results in models that are unable to capture the effects of molecular geometry. In this work, we propose a method to implicitly represent molecular geometry by using simple measurements between regions on the molecule to encode steric effects into the rules themselves.

In order to obtain aggregate size distribution data on which to base our rule-based models, we employ a three-dimensional rigid-body graph-based Monte Carlo method inspired by robotic motions, which has previously been developed [8]. This method can explicitly represent molecular geometry and molecular motions. A graph-based structure defines the molecular interactions; ligands and receptors are represented by vertices, and bonds between ligands and receptors are represented by edges. This structure allows for the easy maintenance of aggregation information throughout the simulation, and for the analysis of aggregate structure [8]. Instead of all-atom molecular structures, 3D isosurface representations of the molecules are used, which reduces the simulation to a rigid-body problem and reduces the computational cost [8]. This method can be combined with rule-based modeling in order to quantify the steric effects between allergen binding regions that affect binding site accessibility and to represent the differences in these steric effects caused by variations in

978-1-4673-6799-8/15/$31.00 ©2015 IEEE

molecular curvature.

In this work, we look at the process of antigen-antibody aggregation; in particular, we study the binding of IgE-FcϵRI antibody-receptor complexes to the antigen Pen a 1. Studying antigen-antibody aggregation and the structure of the aggregates that form during this process is important for understanding how the allergic response is initiated. The allergic response in humans is set into motion by a tyrosine kinase cascade that results from the crosslinking of IgE-FcϵRI receptor complexes via the binding of the IgE antibodies to an antigen [9]. An antigen is a substance that is capable of inducing the production of antibodies and binding to them; it may have multiple possible conformations with differing geometric characteristics, which can affect the size and structure of the aggregates that form. The fold of an allergen is known to play a role in the IgE reactivity of its epitopes [10]. The development of a practical method for aggregate structure prediction based on the geometry of a particular antigen conformation could be useful not only for understanding aggregation, but also for possible manipulation of the antigen geometry to obtain a desired aggregate structure. Various properties of allergens and protein complexes, such as structural stability, flexibility, and dimerization, have been studied using molecular dynamics-based methods [11], [12]; however, such methods would not be useful to study antibody aggregation due to the long timescales involved (on the order of seconds) and the large size and quantity of the aggregates that may form.

In this paper, we introduce a novel rule-based method for modeling molecular geometry, and we apply our method to the modeling of antigen-antibody aggregation. This method involves the construction of a rule set that encodes the steric effects between neighboring binding regions of the antigen. First, we demonstrate our method using three U-shaped molecules with large variations in curvature. Then, we use this method to model the aggregation of IgE-FcϵRI receptor complexes with the shrimp allergen Pen a 1 for three different conformations of the Pen a 1 molecule that have small variations in curvature: native, S-shaped, and U-shaped. We compare the results of this method to the results of the aforementioned Monte Carlo simulations, and we analyze the differences in results between the three Pen a 1 conformations studied.

II. Related Work

Our work builds off of computational methods for modeling molecular geometry in biochemical processes, as well as experimental methods to study IgE antibody binding behavior.

A. Rule-Based Modeling

Biological signaling systems are often comprised of macromolecules that can exist in a large number of functionally distinct states. This number scales exponentially with the amount of modification possibilities [7]. One problem that arises when modeling these systems is the specification problem, i.e. how to specify such a large system.

One solution is implicit specification, which involves the coarse-graining of sets of reactions and parameters into rules; the only explicitly specified features in a reaction rule are those which affect the reaction. Rules define the conditions for molecular transformations and interactions, and are associated with rate laws [2]. Some rules define multiple reactions, which means that all of these reactions are associated with the same rate law. The rules can usually be specified independently. Rule-based specification methods include Kappa-calculus [13], BioNetGen [5], ANC [14], and ML-Rules [15]. The Simmune project and the SSC allow the specification of molecules within spatial regions of arbitrary geometries [16].

The rule-based methods can be population-based, particle-based, or hybrid. Population-based methods include ODE/PDE numerical integration and the stochastic Gillespie algorithm. In these methods, the application of a rule changes the size of one of the populations, each of which consists of all molecules that share the same state and same species. The system state space can be very large, so methods to reduce it have been introduced [7].

Particle-based rule evaluation involves tracking individual particles (molecules and molecular complexes) through the simulation [2]. This is a network-free method; at any time point, only the existing particles, their states, and the possible reactions for the existing particles are necessary. Spatial particle-based methods include an explicit specification of space, and include SRSim [17] and MCell [18].

Our method differs from traditional approaches to rule-based modeling in that the rules used in our method are constrained by molecular geometry.

B. Geometric Molecular Modeling

The spatial simulation software SRSim [17] is a rule-based modeling method that allows for the specification of molecule geometry. SRSim integrates rule-based modeling, molecular dynamics, and a stochastic, diffusing-particle simulator. Molecular geometry is provided by the user via data files. Our proposed method is different in that it is a purely rule-based ODE model that does not require any additional data files to run, as the molecular geometry is encoded into the rules themselves. In addition, our method only requires the BioNetGen software to run.

The stochastic, particle-based Meredys software [19] uses Brownian dynamics to simulate reaction-diffusion systems at the mesoscopic level. It requires the specification of details such as molecule positions, molecular geometry, reaction site positions, and reaction types. Our rule-based method is population-based and only requires the distances between pairs of binding regions on a single antigen molecule to create the model.

Computational methods for modeling two-molecule ligand-receptor docking simulate systems on a smaller scale than those studied using our method. Our method uses more realistic geometric molecular models than do existing methods for self-assembly of molecular structures, such as those employing simple bead models [20].

C. Experimental Methods

Nanoprobe labeling and transmission electron microscopy (TEM) of cell membranes are used to study cell signaling. Methods for the spatial analysis of these nanoprobes, including statistical analysis of clustering, were developed in [21]. Quantum dot (QD)-IgE probes that bind FcϵRI have been used to study the mobility of receptors by single-particle tracking [22]. The kinetics of DNP-BSA binding to IgE has been studied by observing fluorescence quenching [23], and the data was analyzed using a mathematical model in which the IgE binding sites are transiently exposed, allowing binding and cross-linking to occur.

III. METHODS

In this section, we present a rule-based model of the shrimp allergen Pen a 1 that encodes the steric hindrances between the IgE binding regions of the allergen within the set of rules. We first briefly discuss the binding sites and binding regions of the Pen a 1 molecule (Section III-A). Next, we outline the assumptions that were made in the development of this rule-based model (Section III-B). We then describe our methods for determining the steric hindrances between the binding regions (Section III-C), constructing a set of rules based on these steric effects (Section III-C), optimizing the rate constants (Section III-D), and implementing the rule-based simulation (Section III-E). We explain how the probabilities of formation of various aggregate sizes were calculated (Section III-F). Lastly, we describe the Monte Carlo rigid-body model (Section III-H) and how we compared the results of our rule-based model to that of the Monte Carlo simulation (Section III-I).

A. Pen a 1 Structure and Valency

The Pen a 1 allergen possesses a double-stranded coiled structure. All-atom structures of shrimp tropomyosin were obtained from the Protein Data Bank (PDB:1CG1) and the Structural Database of Allergenic Proteins (SDAP Model #284) [24], [25]. The Pen a 1 model used contains 568 amino acids and 4,577 atoms. Experimental studies have predicted that Pen a 1 possesses 16-18 binding sites, which can be grouped into five general binding regions per strand [24], [26], [27]. The amino acid sequences of the binding sites are listed in [24]. The all-atom structure of the IgE-FcϵRI receptor complex was obtained from [28]. It contains 1,709 amino acids and 13,477 atoms.

We split the longest region into two binding regions (E and F) in our rule-based model such that Pen a 1 has six binding regions per strand, with 12 total regions in our model (see Figure 1). This is because the longest region is significantly larger than the other regions, and an analysis of the conditional probabilities of binding of the binding sites in these two regions shows that there is a significant probability that a site in one of these regions can be bound while a site in the other region is also bound in the 3D Monte Carlo simulation. Our preliminary analysis showed that IgE can bind to the same binding region on opposite strands simultaneously.

B. Rule-Based Model Assumptions

In order to simplify our rule-based model and to ensure that the number of rules in the rule set does not become too large for implementation, we make several practical assumptions when constructing our rule sets. Firstly, we assume that an IgE can only bind to a single binding region on the Pen a 1 molecule, and that an IgE cannot be bound to multiple regions simultaneously. For simplicity, and since the Monte Carlo data does not incorporate cross-linking, our rules forbid cross-linking Pen a 1 molecules through IgE binding. As discussed in Section III-A, we simplify our model further by assuming that there are only 12 total binding regions on the Pen a 1 molecule (six per strand) rather than 16-18, and that each IgE can only bind to one of these 12 regions. We make this assumption based on the fact that the individual binding sites within a binding region are very close to each other (< 5 nm) such that an IgE bound to one of these sites is highly likely to block the accessibility to the other sites within the same region. Finally, we assume that each of the two strands in the Pen a 1 molecule binds independently of the other strand. An IgE bound to a region on one strand does not, in any way, affect the probability of an IgE receptor binding to a region on the opposite strand.

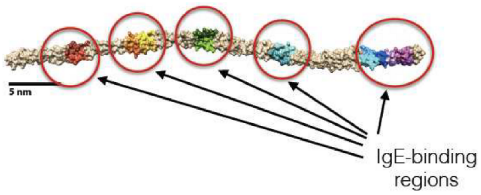

Fig. 1. The all-atom native molecular structure of the shrimp tropomyosin Pen a 1 (tan), with the IgE binding regions circled. The IgE binding regions (various colors) are located in five regions per strand, although for our rule-based model, we have split the longer rightmost region into two separate regions so that there are six binding regions per strand. In this paper, we label the regions (from left to right) as regions A, B, C, D, E, and F.

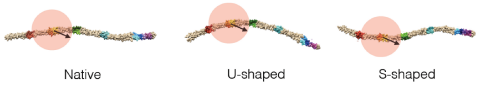

Fig. 2. All-atom structures of three Pen a 1 conformations, with circles representing a possible region of steric hindrance around the yellow/orange binding region where the radius of each circle represents the cutoff distance d_c.

C. Determining Steric Effects and Rules for Various Conformations of Pen a 1

The cutoff distance d_c is an important parameter in this study. In this paper, we use the term "cutoff distance" to specify the maximum distance separating two binding regions on a strand of Pen a 1 at which the two regions have steric effects on each other (Figure 2), meaning that if one of these regions is bound to a receptor, then the probability that the other region can be bound to a receptor is reduced. The cutoff distance determines the rule set of the rule-based model. For each conformation, the cutoff distance is varied and tested to find its optimal value, which is the value that results in a rule-based model that most accurately represents the aggregate size probability data obtained from the Monte Carlo simulation.

The Pen a 1 molecule is flexible and has various possible conformations due to local energy minima. In our model, IgE-FcϵRI receptor complexes are bound to a two-dimensional cell membrane, and the Pen a 1 molecule is constrained to move on a 2D surface. Each Pen a 1 conformation possesses different curvature properties. In this study, we focus on the native [29], S-shaped, and U-shaped conformations (Figure 2). The conformations for our study were designed using standalone Foldit [30] and were energy-minimized using MOIL software [31]. The molecular curvature around a binding region, along with the IgE receptors bound to neighboring binding regions, may cause steric hindrance around the binding region, i.e., IgE receptors may be prevented from binding to the region due to the region being blocked by receptor bound to neighboring regions. A neighboring region is any region that is close enough to the region under consideration to potentially block it if bound to a receptor. There are three general categories of steric effects that we consider in our model, which are illustrated in Figure 3. Firstly, if all neighboring regions are unbound, there is no steric hindrance imposed on a binding region.

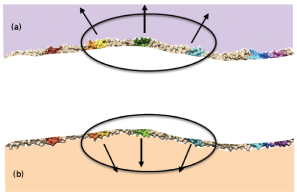

Fig. 3. Types of curvature on the Pen a 1 molecule, visualized split into two strands. The accessibility to the middle (green) region may be affected when neighboring regions are occupied. (a) Positive curvature, and (b) Negative curvature.

Secondly, if the molecular curvature around a binding region is positive (Figure 3 (a)), receptors bound to neighboring regions are unlikely to affect the accessibility to this region since positive curvature around two regions pulls them farther apart (the linear distance between the regions increases). Lastly, if the molecular curvature around a binding region is negative (Figure 3 (b)), receptors bound to neighboring regions may reduce the accessibility to this region since negative curvature around two regions brings them closer together (the linear distance between the regions decreases).

The distances between each pair of binding regions on each strand of the Pen a 1 molecule were measured for each Pen a 1 conformation studied. For region pairs located in an area of negative curvature, the linear distance between the regions was measured. Otherwise, the distance of a free-form path along the molecule between the two regions was used. This difference in distance measurement accounts for the variation in steric effects that results from different types of curvature. If the distance between a pair of regions is less than the specified cutoff distance, then those two regions are considered to exert steric effects on each other, and the steric hindrance is encoded into the binding rules for those regions.

For this study, the first rule construction method we tested used a simple set of rules in which every rule has the same rate constant k_{f1}, and no binding is allowed onto a region that has any steric effects exerted on it by any other region. The second method builds on the first method by allowing binding onto these regions with a reduced, but non-zero, probability set by a separate forward rate constant k_{f2} assigned to rules that specify a steric hindrance between regions.

Given that r is the distance between two binding regions and d_c is the cutoff distance, if one of these two binding regions is occupied and the other region is free, the binding rate constant k_f for a receptor binding to the free region is assigned according to the following:

$$\begin{aligned} &\text{If } r{>}d_c, \text{ then } k_f{=}k_{f1}{=}1.0 \text{ /mol/s} \\ &\text{If } r{\leq}d_c, \text{ then } k_f{=}k_{f2}{<}k_{f1} \end{aligned} \tag{1}$$

The rate constant k_{f2} was determined by performing an optimization and fitting the resulting aggregate size data to that of the 3D Monte Carlo model.

D. Rate Constant Optimization

Because the cutoff distance is unknown, this parameter was varied from 3.0 nm to 20.0 nm in 0.1 nm increments, with the rule set being reconstructed for each cutoff distance. An optimization of k_{f2} was performed for every cutoff distance.

The forward rate constant k_{f2} for each Pen a 1 conformation was optimized using an adaptive algorithm based on the Metropolis-Hastings algorithm. This algorithm finds a minimum of the residual sum-of-squares (RSS) between the RBM data and the Monte Carlo data (see Section III-I). If the RSS value for a new rate constant k_{f2_n} is RSS_n, the current RSS value is RSS_c, and the current rate constant is k_{f2}, then the rate constant is determined according to the following:

$$\begin{aligned} &\text{If } RSS_n{>}RSS_c, \text{ then } RSS_c{=}RSS_n \text{ and } k_{f2}{=}k_{f2_n} \text{ with probability } e^{-\Delta RSS/T} \\ &\text{If } RSS_n{>}RSS_c, \text{ then } RSS_c{=}RSS_c \text{ and } k_{f2}{=}k_{f2} \text{ with probability } 1{-}e^{-\Delta RSS/T} \\ &\text{If } RSS_n{\leq}RSS_c, \text{ then } RSS_c{=}RSS_n \text{ and } k_{f2}{=}k_{f2_n} \text{ with probability } 1 \end{aligned} \tag{2}$$

If RSS_n is higher than RSS_c, then RSS_n (and k_{f2_n}) are accepted with a probability dependent on the difference between the two RSS values ΔRSS and the simulated annealing temperature T. If the new value is accepted, then the rate constant is incremented according to the specified step size, and the new rate constant is tested. Adaptive rate constant step sizes of 0.0001 /mol/s and 0.00001 /mol/s were used (if the RSS is decreasing, the smaller step size is used to find and test a new rate constant; otherwise, the larger step size is used). However, if the new value is rejected, then the algorithm will choose a new rate constant at random from over the entire allowed range. The algorithm was allowed to search over the range 0.00 to 0.40 /mol/s. We can expect from previous scans of k_{f2} that the RSS value for any value of k_{f2} greater than 0.40 /mol/s will be too high to be acceptable.

E. Rule-Based Model Implementation

The rule-based model is specified using the BioNetGen language [5] and implemented with the RuleBender program [32]. RuleBender generates the ordinary differential equations (ODEs) associated with the binding rules and tracks the aggregates that are formed as the IgE receptors bind to the Pen a 1 molecules in an ODE simulation. Each strand of the two-stranded molecules is simulated separately. Each run takes no more than 10 seconds to reach the steady state.

Another version of the rule set with forward binding rate constants proportional to the number of binding sites in the region was also tested, but did not yield aggregate size data that fit better to the Monte Carlo data than the rule sets used.

F. Aggregate Size Probability Calculation

Our model assumes that Pen a 1 is comprised of two strands, which we refer to here as strand I and strand II. Therefore, the probability of formation of an aggregate of a certain size is calculated by combining the independent probabilities of formation of each strand (Figure 4). The probability $P(n)$ to form an aggregate of size n is given by:

$$\begin{aligned} P(n{\leq}6){=}{\textstyle\sum_{m=0}^{n}} P_I(m)P_{II}(n{-}m) \\ P(n{>}6){=}{\textstyle\sum_{m=n-6}^{6}} P_I(m)P_{II}(n{-}m) \end{aligned} \tag{3}$$

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

G. Variation of Rule Set with Curvature

In order to more clearly illustrate how the set of binding rules for a given molecule is affected by molecular curvature using our method of rule construction, we present an example of a U-shaped Pen a 1 molecule with varying curvature. We look at three molecules: the U-shaped molecule seen in

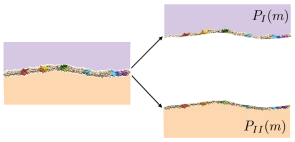

Fig. 4. The all-atom native molecular structure of the shrimp tropomyosin Pen a 1 (left), visualized split into two strands: strand I and strand II (right). The probability of formation of an aggregate of a certain size is calculated by combining the independent probabilities for each strand.

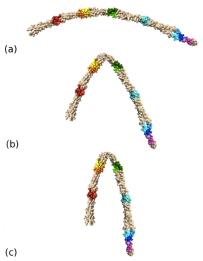

Fig. 5. Visualizations of the (a) U-shaped, (b) 45-degree-rotated, and (c) 60-degree-rotated molecular structures. (It should be noted that the latter two molecules are not energy-minimized conformations and are only presented here for the purpose of demonstrating our rule construction method.)

Figure 2, the same molecule with its two ends rotated inward by 45 degrees, and the same molecule with its two ends rotated inward by 60 degrees (see Figure 5). (It should be noted that the latter two molecules are not energy-minimized conformations and are only presented here for the purpose of demonstrating our rule construction method. The latter two molecules are unlikely to be energetically feasible due to their high degree of curvature.) The rule sets for strand I are shown for the U-shaped Pen a 1 molecule (Table III), the 45-degree-rotated molecular structure (Table I), and the 60-degree-rotated molecular structure (Table II).

TABLE I. RULE SET FOR STRAND I (T_I) OF THE 45-DEGREE-ROTATED MOLECULAR STRUCTURE. LETTERS IN PARENTHESES ARE BINDING SITES. OMITTED LETTERS ARE FREE OR OCCUPIED. THE IGE SUBSCRIPT SHOWS WHICH SITE IT IS BOUND TO.

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

For each of these three molecules, the distances between each pair of binding regions were measured, and the rule sets for each molecule were constructed according to these distances. For the purpose of comparing how molecular curvature affects the rule set, the cutoff distance was fixed at 8.1 nm, and the forward rate constant k_{f2} was fixed at 0.005 mol^{-1}s^{-1}. The rule-based model for each molecule was simulated and the antibody aggregate size probabilities were calculated (see Figure 6). We observe that as the degree of curvature of the molecule increases, the aggregate size distribution shifts towards smaller aggregate sizes, which corresponds to the increase in steric effects between binding regions encoded in the rule sets.

TABLE II. RULE SET FOR STRAND I (T_I) OF THE 60-DEGREE-ROTATED MOLECULAR STRUCTURE. LETTERS IN PARENTHESES ARE BINDING SITES. OMITTED LETTERS ARE FREE OR OCCUPIED. THE IGE SUBSCRIPT SHOWS WHICH SITE IT IS BOUND TO.

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

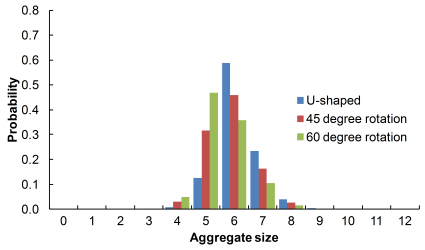

Fig. 6. Comparison of rule-based model aggregate size distributions for the U-shaped, 45-degree-rotated, and 60-degree-rotated molecular structures.

H. Monte Carlo Geometric Simulation

We compared the aggregate size distributions from our rule-based model to those of a previously developed Monte Carlo model [8] that uses 3-D rigid body models of the antigens and receptors. Initially, the molecules are randomly positioned within the bounding volume with no bonds present. Then, during the simulation, at every time step, the positions of the molecules change according to a combination of random sampling and biological constraints such as molecular speeds, binding rates, and unbinding rates (Figure 7). Also, at every time step, any two binding sites (on two separate molecules)

Fig. 7. A visualization of the results of a 3D Monte Carlo simulation of one Pen a 1 molecule (tan) among multiple receptor complexes (blue).

within binding distance of each other will bind with a probability determined by the binding rate. The stability of the number of edges in the graph-based structure can be used to determine when to stop the simulation. Since the simulation models activity taking place on the surface of a cell membrane, the molecules can translate on the XY plane and rotate about the Z axis.

It should be noted that experimental data pertaining to the aggregation of IgE-FcϵRI receptor complexes onto the Pen a 1 molecule is not readily available as of the date this work was written. Therefore, we cannot compare the computational results of this work to experiment. In addition, the 3D Monte Carlo simulation does not include energetics, which limits our understanding of receptor binding onto the binding sites of the Pen a 1 molecule, and is the reason why we cannot derive the forward rate constants directly from the simulation.

The 3D molecular models of the Monte Carlo simulation were created by creating isosurface models of the all-atom molecular structures using UCSF Chimera [33].

I. Comparison Analysis

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

$$RSS = \frac{\sum_{i=1}^{N} (P^i_{MC} - P^i_{RBM})^2}{N} \quad (4)$$

where N is the total number of possible aggregate sizes in a histogram (each histogram has the same number of possible aggregate sizes), P^i_{MC} is the occurrence probability of the ith aggregate size of the Monte Carlo data, and P^i_{RBM} 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 1, and the minimum possible normalized RSS (corresponding to two identical histograms) is zero.

IV. Results

A. Experimental Setup

1) Monte Carlo Simulation: The environment of the Monte Carlo simulations was a 200 nm x 200 nm (40,000 nm^2) discrete membrane with non-periodic boundaries. For each run, one Pen a 1 molecule and 24 IgE-FcϵRI receptor complexes were simulated, such that the receptor density was $\sim$600 receptors/μm^2. Sixty runs were performed for each Pen a 1 conformation. Association and dissociation rates of 1.0 mol^{-1}s^{-1} and 0.01 s^{-1}, from [23] were used for the Pen a 1 antigen. The diffusion coefficient 0.09μm^2/s of IgE-FcϵRI found in [22] was used for all molecules. A time step of 10 μs was used, and every experiment was run for 500,000 time steps, which is long enough for the simulations to reach a steady state. The simulations are rather costly, with one run finishing in approximately 20 hours at full model resolution.

The reduction in speed of aggregates as they increase in size [22], which has not been fully quantified, is included in the simulation by reducing the diffusion coefficient of an aggregate such that it is inversely proportional to the size of the aggregate. For example, the diffusion coefficient of an aggregate containing five receptors would be 1/5 of the original diffusion coefficient. The method of speed reduction implemented only affects aggregation kinetics and does not have a significant effect on the packing structure of aggregates at equilibrium.

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

2) Rule-Based Modeling: The rule-based model was specified in the BioNetGen language, and ODE simulations were conducted on these models using RuleBender [32]. In each experiment, 100 Pen a 1 antigen molecules and 1000 receptors were simulated. Because each of the two strands was simulated individually, the total population included 100 strand I molecules, 100 strand II molecules, and 1000 receptors. Each experiment was run for 1000 time steps, long enough for the simulation to reach a steady state, using a time step of 0.01 s. The simulations were run on the system described in Section IV-A1.

B. Rule Sets

The optimized cutoff distance corresponds to an optimal set of binding rules for each conformation of Pen a 1.

The optimal rule sets were found to be the same for all three conformations, although the optimized rate constant k_{f2} varies based on conformation. These rule sets are shown for strand I (Table III) and strand II (Table IV) of the Pen a 1 molecule.

TABLE III. Rule set for Strand I (T_I) of Pen a 1. Letters in parentheses are binding sites. Omitted letters are free or occupied. The IgE subscript shows which site it is bound to.

Binding Site	Reaction Rule	Binding Rate
A	T_I(A,B) + IgE $\rightarrow$ T_I(IgE$_A$,B)	k$_{f1}$
	T_I(A,IgE$_B$) + IgE $\rightarrow$ T_I(IgE$_A$,IgE$_B$)	k$_{f2}$
B	T_I(A,B,C) + IgE $\rightarrow$ T_I(A,IgE$_B$,C)	k$_{f1}$
	T_I(IgE$_A$,B,C) + IgE $\rightarrow$ T_I(IgE$_A$,IgE$_B$,C)	k$_{f2}$
	T_I(A,B,IgE$_C$) + IgE $\rightarrow$ T_I(A,IgE$_B$,IgE$_C$)	k$_{f2}$
	T_I(IgE$_A$,B,IgE$_C$) + IgE $\rightarrow$ T_I(IgE$_A$,IgE$_B$,IgE$_C$)	k$_{f2}$
C	T_I(B,C,D) + IgE $\rightarrow$ T_I(B,IgE$_C$,D)	k$_{f1}$
	T_I(IgE$_B$,C,D) + IgE $\rightarrow$ T_I(IgE$_B$,IgE$_C$,D)	k$_{f2}$
	T_I(B,C,IgE$_D$) + IgE $\rightarrow$ T_I(B,IgE$_C$,IgE$_D$)	k$_{f2}$
	T_I(IgE$_B$,C,IgE$_D$) + IgE $\rightarrow$ T_I(IgE$_B$,IgE$_C$,IgE$_D$)	k$_{f2}$
D	T_I(C,D,E) + IgE $\rightarrow$ T_I(C,IgE$_D$,E)	k$_{f1}$
	T_I(IgE$_C$,D,E) + IgE $\rightarrow$ T_I(IgE$_C$,IgE$_D$,E)	k$_{f2}$
	T_I(C,D,IgE$_E$) + IgE $\rightarrow$ T_I(C,IgE$_D$,IgE$_E$)	k$_{f2}$
	T_I(IgE$_C$,D,IgE$_E$) + IgE $\rightarrow$ T_I(IgE$_C$,IgE$_D$,IgE$_E$)	k$_{f2}$
E	T_I(D,E,F) + IgE $\rightarrow$ T_I(D,IgE$_E$,F)	k$_{f1}$
	T_I(IgE$_D$,E,F) + IgE $\rightarrow$ T_I(IgE$_D$,IgE$_E$,F)	k$_{f2}$
	T_I(D,E,IgE$_F$) + IgE $\rightarrow$ T_I(D,IgE$_E$,IgE$_F$)	k$_{f2}$
	T_I(IgE$_D$,E,IgE$_F$) + IgE $\rightarrow$ T_I(IgE$_D$,IgE$_E$,IgE$_F$)	k$_{f2}$
F	T_I(E,F) + IgE $\rightarrow$ T_I(E,IgE$_F$)	k$_{f1}$
	T_I(IgE$_E$,F) + IgE $\rightarrow$ T_I(IgE$_E$,IgE$_F$)	k$_{f2}$

TABLE IV. RULE SET FOR STRAND II (T_{II}) OF PEN A 1. LETTERS IN PARENTHESES ARE BINDING SITES. OMITTED LETTERS ARE FREE OR OCCUPIED. THE IGE SUBSCRIPT SHOWS WHICH SITE IT IS BOUND TO.

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

TABLE V. RULE-BASED MODEL PARAMETERS FOR THREE PEN A 1 CONFORMATIONS

Parameter	Native	S-shaped	U-shaped
Cutoff distance (nm)	7.0-8.7	6.8-8.3	6.8-8.6
k_{f1} **(mol^{-1}s^{-1})**	1.00	1.00	1.00
k_{f2} **(mol^{-1}s^{-1})**	0.006595	0.003558	0.007315
k_r **(mol^{-1}s^{-1})**	0.01	0.01	0.01
RSS	0.000703	0.000135	0.000469

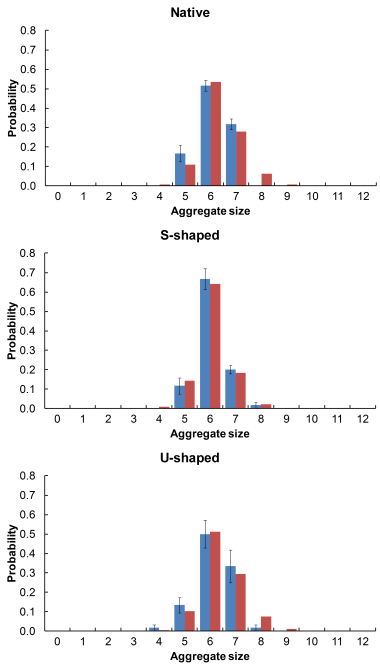

Fig. 8. Comparison of Monte Carlo (blue) and rule-based model (red) aggregate size distributions for the native (top), S-shaped (center), and U-shaped (bottom) Pen a 1 conformations.

C. Aggregate Size Histograms

The forward rate constant k_{f2} and the cutoff distance range were optimized for each of the three energy-minimized conformations of Pen a 1. Table V displays the rate constants, cutoff distances, and RSS values for the native, S-shaped, and U-shaped Pen a 1 conformations.

The Monte Carlo aggregate size probability histogram data is shown along with the optimized rule-based modeling data for the native, S-shaped, and U-shaped Pen a 1 conformations (Figure 8). The error bars for the Monte Carlo data were calculated by binning the 60 runs into 10 sets of six runs each and then calculating the standard error of the mean.

1) Comparison of Conformations: For the 3D Monte Carlo simulation, we observe that there is a significant difference between the aggregate size distribution of the S-shaped conformation and that of the native and U-shaped conformations. The optimal forward binding rate constant k_{f2} shows variation between conformations, particularly between the S-shaped conformation and the other two. However, the optimal set of binding rules was found to be identical for all three conformations, and the optimal cutoff distance range is similar, but not identical, for all three conformations. This indicates that the distance at or below which two binding regions exert significant steric effects on each other is not highly dependent on conformation for the three energy-minimized conformations studied.

The rate constant k_{f2} represents the probability of an IgE-receptor complex binding to a region on the Pen a 1 molecule that is under significant steric hindrance from IgE-receptor complexes that are bound to neighboring binding regions. For the S-shaped molecule, the optimized k_{f2} value is approximately half of the k_{f2} value for the other two molecular geometries, indicating that the curvature of the S-shaped molecule may be reducing the probability that a binding region can bind without being blocked by receptors bound to nearby regions. This difference indicates that molecular geometry may play an important role in antibody aggregation onto the Pen a 1 molecule and should be taken into consideration in future aggregation research.

V. CONCLUSION

We developed a novel implementation of rule-based modeling that encodes molecular geometry into the rules and associated rate constants. We studied the effects of molecular geometry on the rule sets of three U-shaped molecules of varying curvature. We also studied three energy-minimized molecular conformations of the Pen a 1 allergen using this method combined with a 3D rigid-body Monte Carlo simulation. We analyzed the similarities and differences among the rule-based models for each geometry.

In our study of the U-shaped molecule, we found that the degree of curvature of the antigen has a strong effect on the rule set constructed using our proposed method, with individual binding regions becoming dependent on a greater number of neighboring binding regions as the degree of curvature increases. In our study of the three energy-minimized Pen a 1 conformations, we found that our proposed method of rule set construction provides a quantification of the steric effects that affect binding site accessibility and allows us to observe which neighboring binding regions most strongly affect a particular region. Although the set of binding rules will not always be different for different antigen conformations, the binding rate constants associated with the rules provide another means of quantifying the variation in aggregate size distribution based on antigen conformation. We used a rod-shaped molecule in our study, but our method for measuring distances between binding sites could be modified for use with

other molecule geometries. Furthermore, our method could be extended to model experimental data in the future; analysis of probe positions in electron microscopy images allow for the estimation of receptor clustering.

ACKNOWLEDGMENTS

The authors would like to thank Bridget Wilson for her guidance on the antibody assembly problem, William S. Hlavacek for providing advice on rule-based modeling, and Chang-Shung Tung for providing molecular structures. Also, thanks to the UNM Center for Advanced Research Computing for providing computational resources. This work supported by NIH Grant No. 5P50GM085273-07 to the Center for Spatiotemporal Modeling of Cell Signaling, University of New Mexico.

REFERENCES

[1] R. H. Jr, K. Leung, J. Barkinge, M. Ciaccio, C. Chuu, and R. Jones, "Comprehensive binary interaction mapping of SH2 domains via fluorescence polarization reveals novel functional diversification of ErbB receptors." *PLoS One*, vol. 7, no. 9, p. e44471, 2012.

[2] L. A. Chylek, L. A. Harris, C.-S. Tung, J. R. Faeder, C. F. Lopez, and W. S. Hlavacek, "Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems," *WIRESBM*, vol. 6, pp. 13–36, 2014.

[3] M. Blinov, J. Faeder, B. Goldstein, and W. Hlavacek, "A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity," *BioSystems*, vol. 83, no. 2-3, pp. 136–151, 2006.

[4] W. S. Hlavacek, J. R. Faeder, M. L. Blinov, A. S. Perelson, and B. Goldstein, "The complexity of complexes in signal transduction," *Biotechnology and Bioengineering*, vol. 84, no. 7, pp. 783–794, 2003.

[5] M. L. Blinov, J. R. Faeder, B. Goldstein, and W. S. Hlavacek, "BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains," *Bioinformatics*, vol. 20, no. 17, pp. 3289–3291, 2004.

[6] B. Goldstein and A. Perelson, "Equilibrium theory for the clustering of bivalent cell surface receptors by trivalent ligands. Application to histamine release from basophils," *Biophysical Journal*, vol. 45, no. 6, pp. 1109–1123, 1984.

[7] J. Yang, M. I. Monine, J. R. Faeder, and W. S. Hlavacek, "Kinetic Monte Carlo method for rule-based modeling of biochemical networks," *Phys. Rev. E*, vol. 78, no. 3, p. 031910, 2008.

[8] K. Manavi, B. S. Wilson, and L. Tapia, "Simulation and analysis of antibody aggregation on cell surfaces using motion planning and graph analysis," in *Proc. ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB)*, 2012, pp. 458–465.

[9] J. Rivera and A. M. Gilfillan, "Molecular regulation of mast cell activation," *Journal of Allergy and Clinical Immunology*, vol. 117, no. 6, pp. 1214–1225, 2006.

[10] R. Valenta, B. Linhart, I. Swoboda, and V. Niederberger, "Recombinant allergens for allergen-specific immunotherapy: 10 years anniversary of immunotherapy with recombinant allergens," *Allergy*, vol. 66, pp. 775–783, 2011.

[11] M. Garrido-Arandia, C. Gómez-Casado, A. Díaz-Perales, and L. F. Pacios, "Molecular dynamics of major allergens from *Alternaria*, birch pollen and peach," *Molecular Informatics*, vol. 33, no. 10, pp. 682–694, Oct 2014.

[12] B. Orndorff and F. Jagodzinski, "A combined molecular dynamics, rigidity analysis approach for studying protein complexes," in *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, ser. BCB'13. New York, NY, USA: ACM, 2013, pp. 793:793–793:798. [Online]. Available: http://doi.acm.org/10.1145/2506583.2506711

[13] V. Danos and C. Laneve, "Formal molecular biology," *Theoretical Computer Science*, vol. 325, no. 1, pp. 69–110, 2004.

[14] J. F. Ollivier, V. Shahrezaei, and P. S. Swain, "Scalable rule-based modelling of allosteric proteins and biochemical networks," *PLoS Comput. Biol.*, vol. 6, no. 11, Nov. 2010.

[15] C. Maus, S. Rybacki, and A. M. Uhrmacher, "Rule-based multi-level modeling of cell biological systems," *BMC Systems Biology*, vol. 5, p. 166, 2011.

[16] M. Stefan, T. Bartol, T. Sejnowski, and M. Kennedy, "Multi-state modeling of biomolecules," *PLoS Computational Biology*, vol. 10, no. 9, p. e1003844, 2014.

[17] G. Gruenert, B. Ibrahim, T. Lenser, M. Lohel, T. Hinze, and P. Dittrich, "Rule-based spatial modeling with diffusing, geometrically constrained molecules," *BMC Bioinformatics*, vol. 11, p. 307, 2010.

[18] R. Kerr, T. Bartol, B. Kaminsky, M. Dittrich, J. Chang, S. Baden, T. Sejnowski, and J. Stiles, "Fast Monte Carlo simulation methods for biological reaction-diffusion systems in solution and on surfaces," *SIAM J. Sci. Comput.*, vol. 30, no. 6, pp. 3126–3149, 2009.

[19] D. P. Tolle and N. L. Novère, "Meredys, a multi-compartment reaction-diffusion simulator using multistate realistic molecular complexes," *BMC Syst. Biol.*, vol. 4, no. 24, 2010.

[20] H. W. Hatch, J. Mittal, and V. K. Shen, "Computational study of trimer self-assembly and fluid phase behavior," *The Journal of Chemical Physics*, vol. 142, no. 16, 2015.

[21] J. Zhang, K. Leiderman, J. R. Pfeiffer, B. S. Wilson, J. M. Oliver, and S. L. Steinberg, "Characterizing the topography of membrane receptors and signaling molecules from spatial patterns obtained using nanometer-scale electron-dense probes and electron microscopy," *Micron*, vol. 37, no. 1, pp. 14–34, 2006.

[22] N. L. Andrews, J. R. Pfeiffer, A. M. Martinez, D. M. Haaland, R. W. Davis, T. Kawakami, J. M. Oliver, B. S. Wilson, and D. S. Lidke, "Small, mobile FcεRI receptor aggregates are signaling competent," *Immunity*, vol. 31, no. 3, pp. 469–479, 2009.

[23] K. Xu, B. Goldstein, D. Holowka, and B. Baird, "Kinetics of multivalent antigen DNP-BSA binding to IgE-FcεRI in relationship to the stimulated tyrosine phosphorylation of FcεRI," *J. Immunol.*, vol. 160, no. 7, pp. 3225–3235, 1998.

[24] O. Ivanciuc, C. H. Schein, and W. Braun, "SDAP: Database and computational tools for allergenic proteins," *Nucleic Acids Research*, vol. 31, no. 1, pp. 359–362, 2003.

[25] ——, "Data mining of sequences and 3D structures of allergenic proteins," *Bioinformatics*, vol. 18, no. 10, pp. 1358–1364, 2002.

[26] R. Ayuso, S. Lehrer, and G. Reese, "Identification of continuous, allergenic regions of the major shrimp allergen Pen a 1 (tropomyosin)," *International Archives of Allergy and Immunology*, vol. 127, no. 1, pp. 27–37, 2002.

[27] G. Reese, J. Viebranz, S. M. Leong-Kee, M. Plante, I. Lauer, S. Randow, M. S.-M. Moncin, R. Ayuso, S. B. Lehrer, and S. Vieths, "Reduced allergenic potency of VR9-1, a mutant of the major shrimp allergen Pen a 1 (tropomyosin)," *The Journal of Immunology*, vol. 175, no. 12, pp. 8354–8364, 2005.

[28] A. Mahajan, D. Barua, P. Cutler, D. S. Lidke, F. A. Espinoza, C. Pehlke, R. Grattan, Y. Kawakami, C.-S. Tung, A. R. M. Bradbury, W. S. Hlavacek, and B. S. Wilson, "Optimal aggregation of FcεRI with a structurally defined trivalent ligand overrides negative regulation driven by phosphatases," *ACS Chemical Biology*, vol. 9, no. 7, pp. 1508–1519, 2014.

[29] K. Manavi and L. Tapia, "Influence of model resolution on antibody aggregation simulations," in *RSS Workshop on Robotics Methods for Structural and Dynamic Modeling of Molecular Systems (RMMS)*, 2014.

[30] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, and Z. Popovi, "Predicting protein structures with a multiplayer online game." *Nature*, vol. 466, pp. 756–760, 2010.

[31] R. Elber, A. Roitberg, C. Simmerling, R. Goldstein, H. Li, G. Verkhivker, C. Keasar, J. Zhang, and A. Ulitskya, "MOIL: A program for simulations of macromolecules," *Computer Physics Communications*, vol. 91, pp. 159–189, September 1995.

[32] A. M. Smith, W. Xu, Y. Sun, J. R. Faeder, and G. E. Marai, "RuleBender: Integrated modeling, simulation and visualization for rule-based intracellular biochemistry," *BMC Bioinformatics*, vol. 13, no. 8, 2012.

[33] T. D. Goddard, C. C. Huang, and T. E. Ferrin, "Software extensions to UCSF Chimera for interactive visualization of large molecular assemblies," *Structure*, vol. 13, no. 3, pp. 473–482, 2005.