2-2-2016

Modeling steric effects in antibody aggregation using rule-based methods

Brittany Hoard

Brittany Hoard

Candidate

Nanoscience and Microsystems

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Lydia Tapia , Chairperson

Trilce Estrada

William Hlavacek

Bruna Jacobson

Modeling Steric Effects in Antibody Aggregation Using Rule-Based Methods

by

Brittany Rafaela Hoard

B.S., Physics, Pennsylvania State University, 2012

THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Nanoscience and Microsystems

The University of New Mexico

Albuquerque, New Mexico

December, 2015

Dedication

This work is dedicated to S.P., a continual source of inspiration.

Acknowledgments

First and foremost, I would like to thank my research advisor, Dr. Lydia Tapia, for her guidance, expertise, and encouragement throughout this entire research project. I would also like to thank Dr. Bruna Jacobson for her regular feedback on this work as well as her research guidance. Thanks to Kasra Manavi and Torin Adamson for fruitful discussions and technical assistance.

I am grateful to our collaborators Dr. Bridget Wilson and Dr. Avanika Mahajan for providing experimental cell secretion data. Additional thanks to Dr. Wilson for permission to incorporate said data in this thesis and for her assistance with the antibody assembly problem. Thanks also to Dr. William Hlavacek for providing advice on rule-based modeling and to Dr. Trilce Estrada for her suggestions regarding predictive modeling.

I would also like to acknowledge Chang-Shung Tung for providing molecular structures for use in our simulations and to Alan Kuntz for permission to use figures from his undergraduate honors thesis.

Finally, thanks to the UNM Center for Advanced Research Computing for providing computational resources. This work was supported by NIH Grant No. 5P50GM085273-07 to the Center for Spatiotemporal Modeling of Cell Signaling, University of New Mexico.

Modeling Steric Effects in Antibody Aggregation Using Rule-Based Methods

by

Brittany Rafaela Hoard

B.S., Physics, Pennsylvania State University, 2012
M.S., Nanoscience and Microsystems, University of New Mexico, 2015

Abstract

The allergic response is produced by the release of immune mediators by mast cells and basophils. This process, in turn, is initiated by the aggregation of antigens and IgE-FcϵRI antibody-receptor complexes. Computational modeling of antibody-antigen aggregate formation as well as the size and structure of these aggregates is an important tool for greater understanding of the allergic response. In addition, the incorporation of molecular geometry into aggregation models can more accurately capture details of the aggregation process, and may lead to insights into how geometry affects aggregate formation. However, it is challenging to simulate aggregation due to the computational cost of simulating large molecules. Methods to geometrically model antibody aggregation inspired by rigid body robotic motion simulations have previously been developed; however, high computational cost mandates that the resolution of the 3D molecular models be reduced, which affects the results of the simulation. Rule-based modeling can be used to model aggregation with low computational cost, but traditional rule-based modeling approaches do not include details of molecular geometry.

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 combined with a previously developed 3D rigid-body Monte Carlo simulation. We first simulate the binding of IgE antibodies bound to cell surface receptors FcϵRI to various binding regions of the allergen Pen a 1 using the aforementioned 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. 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.

We perform this procedure for seven resolutions and three molecular conformations of Pen a 1. We then analyze the impact of resolution and conformation on the aggregate size distribution and on the optimal rule-based model. In addition, we develop a predictive model by first fixing the rule set and varying only the binding rate constant for each resolution, and then fitting the resulting data to a function. This model is intended to enable the prediction of the aggregate size distribution for higher resolutions while requiring only data for lower resolution Monte Carlo models, thus enhancing computational efficiency. Finally, we use a simple rule-based model to fit experimental cell degranulation data for various concentrations of the shrimp allergen Pen a 1 and the IgE antibody.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

The allergic response in humans is at least partially initiated by a tyrosine kinase cascade that results from the crosslinking of IgE-FcϵRI receptors via the binding of the IgE antibodies to an antigen. Studying both the process of antigen-antibody aggregation and the structure of the aggregates that form is important for understanding the allergic response. An antigen 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 [Valenta et al., 2011]. 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.

Computational methods are widely used to study biomolecular interactions, due to their complexity. One example of such a method is molecular dynamics, which

is useful for simulating atomic motion but is computationally costly for studying longer timescales (on the order of seconds or longer) at which cell signaling events occur [Chylek et al., 2014]. Other methods, including constraint-based modeling and Boolean modeling, tend to only loosely be connected to physicochemical principles (or not at all). Models which are constrained by physicochemical principles are useful due to causality and the ability to measure parameters independently. 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. One technique that meets all of the above requirements is rule-based modeling. Rule-based modeling is a technique in chemical kinetics for studying the site dynamics of biomolecular networks [Chylek et al., 2014], 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 used in creating rule-based models. In this study, the rule-based modeling software BioNetGen is employed, which uses graph rewriting. The biomolecules are represented by graphs, vertices represent molecular components, and edges represent bonds between these components. 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. A disadvantage of the rule-based model is that molecular geometry is not explicitly represented by the model itself, although some software with rule-based modeling features allows the user to input geometrical details into the model via external data files [Gruenert et al., 2010]. However, some aspects of the geometry of a molecule may be implicitly represented by encoding steric effects into the rules.

Besides rule-based modeling, there are other methods available to study antigen-antibody aggregation. In this study, we employ a three-dimensional rigid-body graph-based Monte Carlo method inspired by robotic motions, which has previously been developed [Manavi et al., 2012]. 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 facilitates maintenance of aggregation information and the analysis of aggregate structure. One problem that this method faces is the high computational cost of simulating 3D molecular structures, which are geometrically complex. 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. This problem can be ameliorated even further through the use of polygon reduction, which allows this model to be run at reduced resolutions, significantly improving its computational efficiency. However, a reduction in resolution changes the volume of the molecular models and affects the resulting aggregate size distribution [Manavi and Tapia, 2014].

In this study, we employ the methods described above to study the aggregation of antibodies and the shrimp allergen tropomyosin. Shellfish are responsible for many food allergies, and most shellfish species responsible for allergic reactions are crustaceans. Tropomyosin is a protein present in all eukaryotic cells; it is an important allergen found in many crustaceans, including crab (Cha f 1), squid (Tod p 1), lobster (Pan s 1, Hom a 1), oyster (Cra g 1), and snail (Tur c 1) [Reese et al., 2006]. Although there are strong similarities between the various types of tropomyosin, the tropomyosins found in vertebrates are often non-allergenic, and the reason for this difference is not fully understood [Reese et al., 2006]. There is some evidence that differences between IgE binding regions in vertebrate and invertebrate tropomyosins may play a role [Ayuso et al., 2002]. Pen a 1 is a type of tropomyosin molecule found in shrimp. Previous studies have found that tropomyosin elicits an allergic response

in at least 80% of shrimp-allergic subjects, and that tropomyosin binds about 75% of shrimp-specific IgE [Reese et al., 2006].

1.2 Contributions

The contributions of this work, which have been presented in this thesis, [Manavi et al., 2015], and [Hoard et al., 2015], are as follows:

1. We develop a novel geometric rule-based method involving the construction of a rule set that encodes the steric effects between neighboring binding regions of a molecule.

2. We fit rule-based models developed using our novel method to antigen-antibody aggregate size distributions of a 3-D rigid-body graph-based Monte Carlo simulation at seven model resolutions for three energy-minimized conformations of the shrimp allergen Pen a 1. We use our rule-based method to understand and quantify the variation in steric effects that results from variation in model resolution and molecular geometry.

3. We develop models for the prediction of aggregate size distributions for high-resolution Monte Carlo models without requiring high-resolution Monte Carlo simulations to be run.

4. We extend our novel rule-based method to the prediction of aggregate size distributions for three U-shaped Pen a 1 molecules with various degrees of curvature.

Chapter 2

Related Work

2.1 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 [Yang et al., 2008]. 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 [Chylek et al., 2014]. 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 [Danos and Laneve, 2004], BioNetGen [Blinov et al., 2004], ANC [Ollivier et al., 2010], and ML-Rules [Maus et al., 2011]. The Simmune project and

the SSC allow the specification of molecules within spatial regions of arbitrary geometries [Stefan et al., 2014].

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 [Yang et al., 2008].

Particle-based rule evaluation involves tracking individual particles (molecules and molecular complexes) through the simulation [Chylek et al., 2014]. 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 [Gruenert et al., 2010] and MCell [Kerr et al., 2009].

2.2 Geometric Molecular Modeling

The spatial simulation software SRSim [Gruenert et al., 2010] 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 [Tolle and Novère, 2010] 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.

2.3 Experimental Methods

Nanoprobe labeling and transmission electron microscopy (TEM) of cell membranes are used to study aggregation. Methods for the spatial analysis of these nanoprobes, including statistical analysis of clustering, were developed in [Zhang et al., 2006]. Quantum dot (QD)-IgE probes that bind FcϵRI have been used to study the mobility of receptors by single-particle tracking [Andrews et al., 2009]. The kinetics of DNP-BSA binding to IgE has been studied by observing fluorescence quenching [Xu et al., 1998], 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.

Chapter 3

Preliminaries

In this chapter, we introduce the molecules used in our simulations, and we discuss the structure and binding sites of these molecules. We then analyze the conditional probabilities of binding for the 18 binding sites and six binding regions of Pen a 1.

3.1 Molecules

The two types of molecules used in our simulation, which are both proteins, are the shrimp antigen Pen a 1 and the antibody-receptor complex IgE-FcϵRI. A protein is a macromolecule that consists of chains of amino acids. An antigen is a substance that triggers an immune response, and an antibody is a protein that identifies and binds to pathogens such as antigens, bacteria, and viruses. FcϵRI is a type of cell surface receptor that is present on the surface of mast cells and basophils and that can bind to IgE antibodies.

3.2 Antibody-Receptor Complex (IgE-FcϵRI)

The antibody-receptor complex IgE-FcϵRI has two binding sites that can bind to IgE-specific binding sites on the Pen a 1 molecule. The all-atom structure of the IgE-FcϵRI receptor complex was obtained from [Mahajan et al., 2014]. It contains 1,709 amino acids and 13,477 atoms.

3.3 Pen a 1

The Pen a 1 allergen possesses a double-stranded coiled structure. The all-atom structure of Pen a 1 used in this study was obtained from the Structural Database of Allergenic Proteins (SDAP Model #284) [Ivanciuc et al., 2003, Ivanciuc et al., 2002]. This model contains 568 amino acids and 4,577 atoms.

3.3.1 Binding Regions

Experimental studies have predicted that Pen a 1 possesses 16-18 binding sites, which can be grouped into five general binding regions per strand [Ayuso et al., 2002, Ivanciuc et al., 2003, Reese et al., 2005]. The amino acid sequences of the binding sites are listed in [Ivanciuc et al., 2003]. 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 4.1).

3.3.2 Conformations

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 2-D surface. Each Pen a 1 conformation possesses different curvature properties. In this study, we focus on three energy-minimized conformations of Pen a 1, which we call the native, S-shaped, and U-shaped conformations (Figure 4.2). The conformations for our study were designed using standalone Foldit [Cooper et al., 2010] and were energy-minimized using MOIL software [Elber et al., 1995].

The 18 individual binding sites on the Pen a 1 molecule are shown in Figure 3.1 for the native type, Figure 3.2 for the S-shaped type, and Figure 3.3 for the U-shaped type.

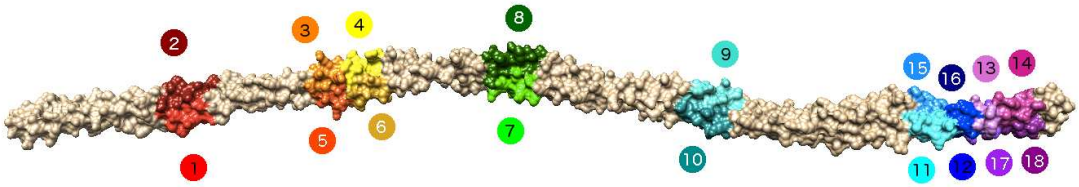

Figure 3.1: The native Pen a 1 conformation, with numbers labeling each binding site. In our model, binding site 1 (2 on the opposite strand) belongs to region A, sites 5 and 6 (3 and 4 on the opposite strand) belong to region B, site 7 (8 on the opposite strand) belongs to region C, site 10 (9 on the opposite strand) belongs to region D, sites 11 and 12 (15 and 16 on the opposite strand) belong to region E, and sites 17 and 18 (13 and 14 on the opposite strand) belong to region F.

Each of the 18 binding sites on the Pen a 1 molecule can be bound to an IgE-

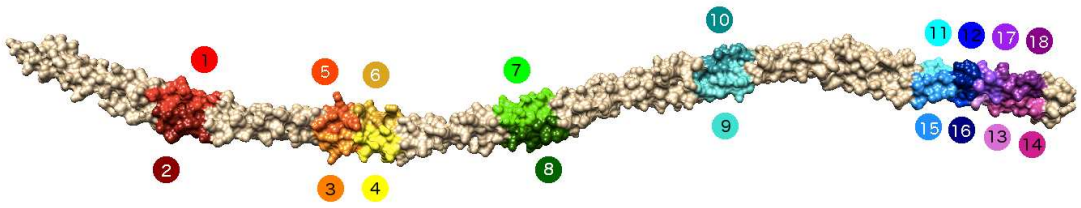

Figure 3.2: The S-shaped Pen a 1 conformation, with numbers labeling each binding site. In our model, binding site 1 (2 on the opposite strand) belongs to region A, sites 5 and 6 (3 and 4 on the opposite strand) belong to region B, site 7 (8 on the opposite strand) belongs to region C, site 10 (9 on the opposite strand) belongs to region D, sites 11 and 12 (15 and 16 on the opposite strand) belong to region E, and sites 17 and 18 (13 and 14 on the opposite strand) belong to region F.

receptor complex, forming an aggregate with one Pen a 1 molecule and multiple receptors. The binding at each of the 18 sites can be impacted due to steric effects between the sites. In order to make several simplifying assumptions that will be applied to our rule-based models, we conduct an analysis of the conditional probabilities of geometric binding for each of the 18 sites, and for the six general binding regions.

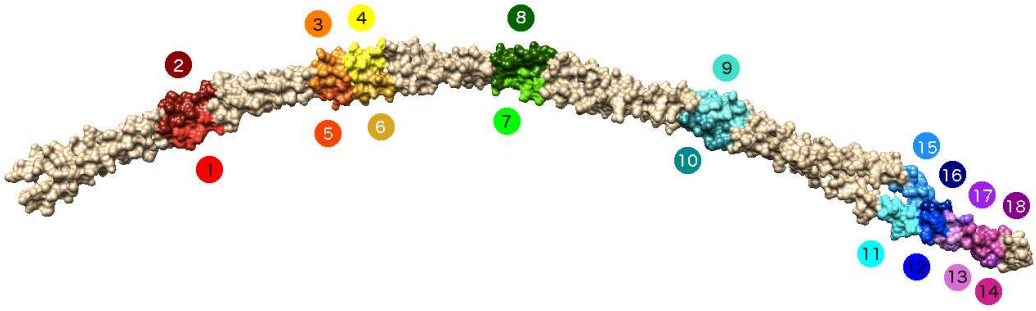

Figure 3.3: The U-shaped Pen a 1 conformation, with numbers labeling each binding site. In our model, binding site 1 (2 on the opposite strand) belongs to region A, sites 5 and 6 (3 and 4 on the opposite strand) belong to region B, site 7 (8 on the opposite strand) belongs to region C, site 10 (9 on the opposite strand) belongs to region D, sites 11 and 12 (15 and 16 on the opposite strand) belong to region E, and sites 13 and 14 (17 and 18 on the opposite strand) belong to region F.

3.3.3 Steric Hindrance

In this section, we present an analysis of the packing structures of the antigen-antibody aggregates formed in our 3D Monte Carlo simulation. We do this by presenting the conditional probabilities of binding, which were calculated by a collaborator in our research group [Kuntz, 2014], for each of the 18 binding sites on the Pen a 1 molecule. (These probabilities are derived only from aggregate structures and not from energetics.) We look specifically at the 90% reduced resolution Monte Carlo model, which we created by reducing the polygons in the 3D isosurface molecule models by 90%. The analysis was calculated using 300 runs of the Monte Carlo simulation for each conformation.

The conditional binding probabilities (the probability that a specific site is bound if another site is bound) are shown in Figure 3.4 for the native type, Figure 3.5 for the S-shaped type, and Figure 3.6 for the U-shaped type.

We make the assumption in our rule-based model that a single binding region of the Pen a 1 molecule cannot be bound to more than one IgE antibody. In other words, no more than one binding site per binding region can be bound at the same time. We use the conditional binding probabilities to justify this assumption. For the native Pen a 1 conformation (Figure 3.4), we see that there is a strong correlation between sites 3 and 4 (region B on strand II) such that the probability of both sites binding is zero. There is also a zero probability for site pairs 5 and 6 (region B on strand I), 11 and 12 (region E on strand I), 15 and 16 (region E on strand II), 13 and 14 (region F on strand II), and 17 and 18 (region F on strand I). We make similar observations for the S-shaped (Figure 3.5) and U-shaped (Figure 3.6) conformations.

We also observe that there are some strong correlations between binding sites that do not belong to the same binding region. For the native conformation (Figure 3.4), sites 6 and 7 (regions B and C on strand I) have a very strong correlation. This may be due to these two sites being neighboring sites, and they are located in a region of negative curvature, which increases the steric effects between the two sites. The same effect is observed for the U-shaped conformation (Figure 3.6). In Figure 3.4, we see that there are strong correlations between site pairs 11 and 17 (regions E and F on strand I), 12 and 17 (regions E and F on strand I), and the other sites in regions E and F. Similar correlations are observed for the S-shaped (Figure 3.5) and the U-shaped (Figure 3.6) conformations.

The following analysis explains why we split the long binding region into two separate regions for our rule-based model. In Figure 3.4, we observe that there is a significant probability that site 18 (which belongs to region F on strand I) will bind if site 11 (which belongs to region E on strand I) is bound. There is also a significant

probability that sites 15 and 16 (which belong to region E on strand II) will bind if site 14 (which belongs to region F on strand II) is bound. Similar observations can be made for the S-shaped type in Figure 3.5. For the U-shaped type in Figure 3.6, we observe that site 14 (which belongs to region F on strand I) can bind if site 11 (which belongs to region E on strand I) is bound. We also observe that site 18 (which belongs to region F on strand II) can bind if site 15 (which belongs to region E on strand II) is bound.

The conditional binding probabilities of the native type Pen a 1 using the model of six binding regions on each strand are shown in Figure 3.7 for strand I and Figure 3.8 for strand II. From these figures, we observe that for almost all pairs of binding regions, there are significant probabilities that a region can bind if any other region is bound, which helps to justify our division of the binding sites into six (and not more) regions. Regions E and F are part of one large binding region, but we observe that there is a small, but non-zero, probability that region F can bind if region E is bound, which is why we split the larger region into the smaller regions E and F for our rule-based model. We observe for strand I that there is a very small, almost zero, probability that region C can bind if region B is bound, but for strand II, the probability is much higher.

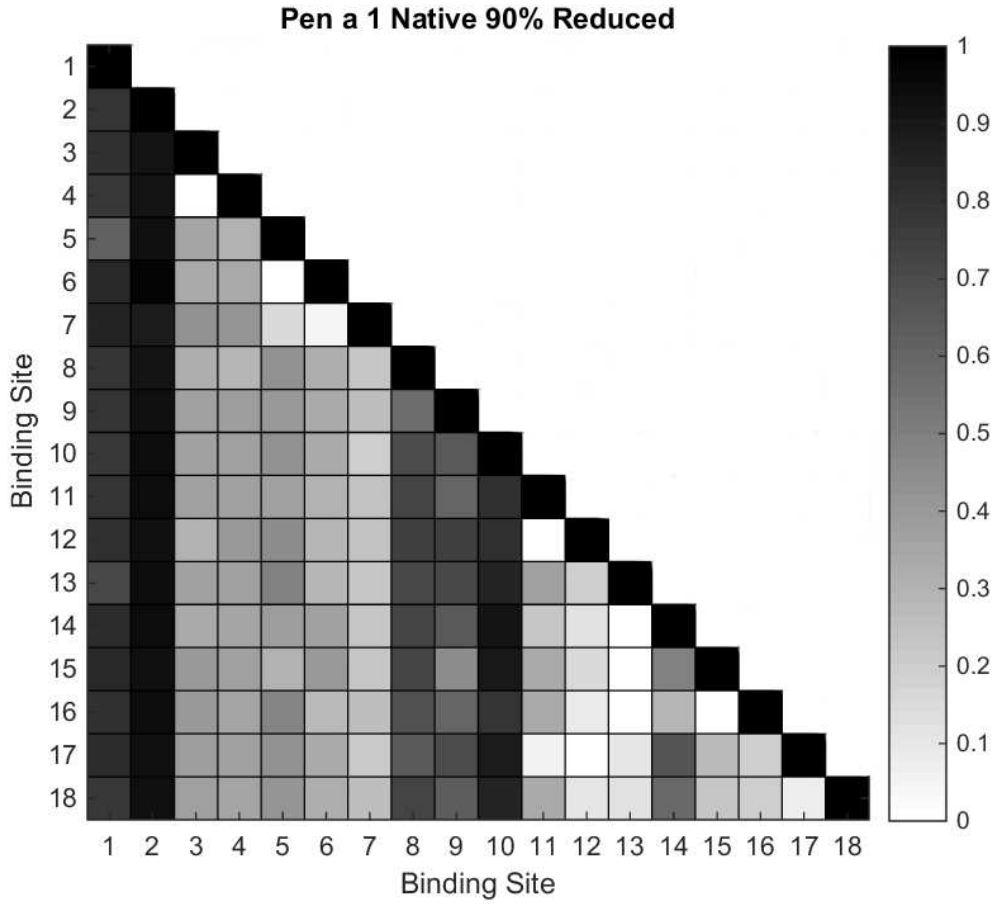

Figure 3.4: The conditional binding probabilities of the native Pen a 1 conformation, with binding site numbers corresponding to those of Figure 3.1. In our model, binding site 1 (2 on the opposite strand) belongs to region A, sites 5 and 6 (3 and 4 on the opposite strand) belong to region B, site 7 (8 on the opposite strand) belongs to region C, site 10 (9 on the opposite strand) belongs to region D, sites 11 and 12 (15 and 16 on the opposite strand) belong to region E, and sites 17 and 18 (13 and 14 on the opposite strand) belong to region F.

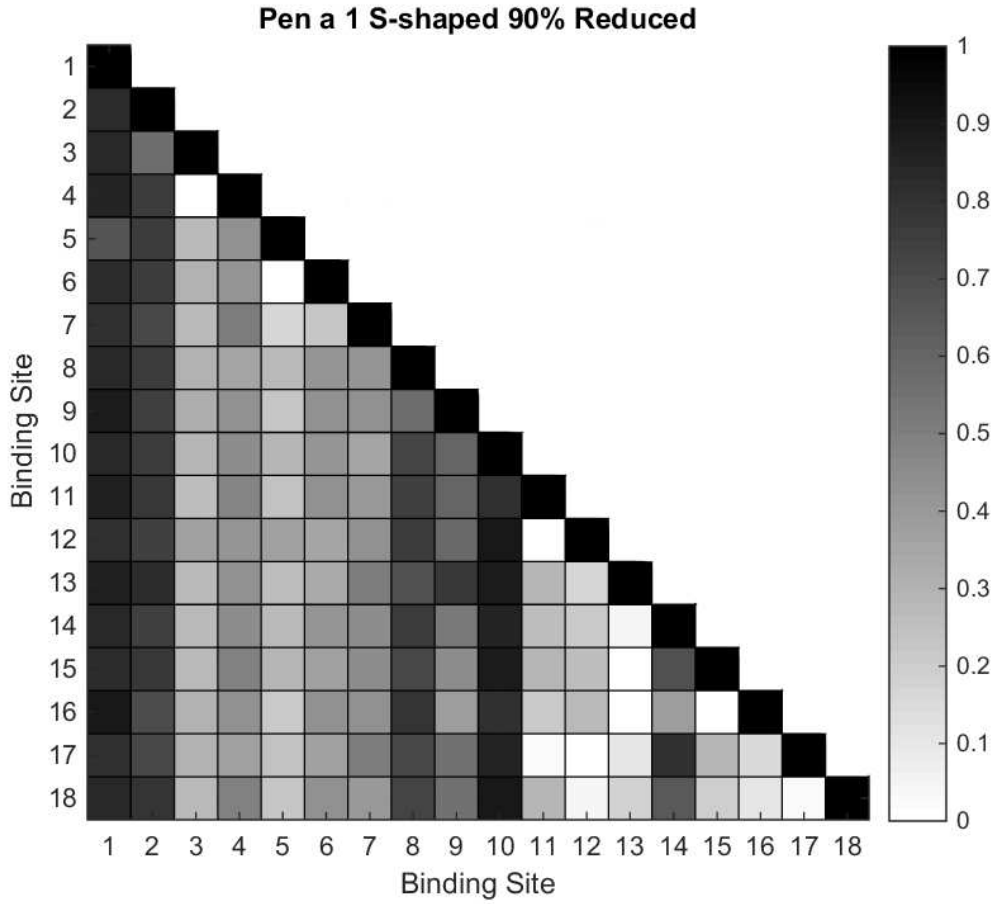

Figure 3.5: The conditional binding probabilities of the S-shaped Pen a 1 conformation, with binding site numbers corresponding to those of Figure 3.2. In our model, binding site 1 (2 on the opposite strand) belongs to region A, sites 5 and 6 (3 and 4 on the opposite strand) belong to region B, site 7 (8 on the opposite strand) belongs to region C, site 10 (9 on the opposite strand) belongs to region D, sites 11 and 12 (15 and 16 on the opposite strand) belong to region E, and sites 17 and 18 (13 and 14 on the opposite strand) belong to region F.

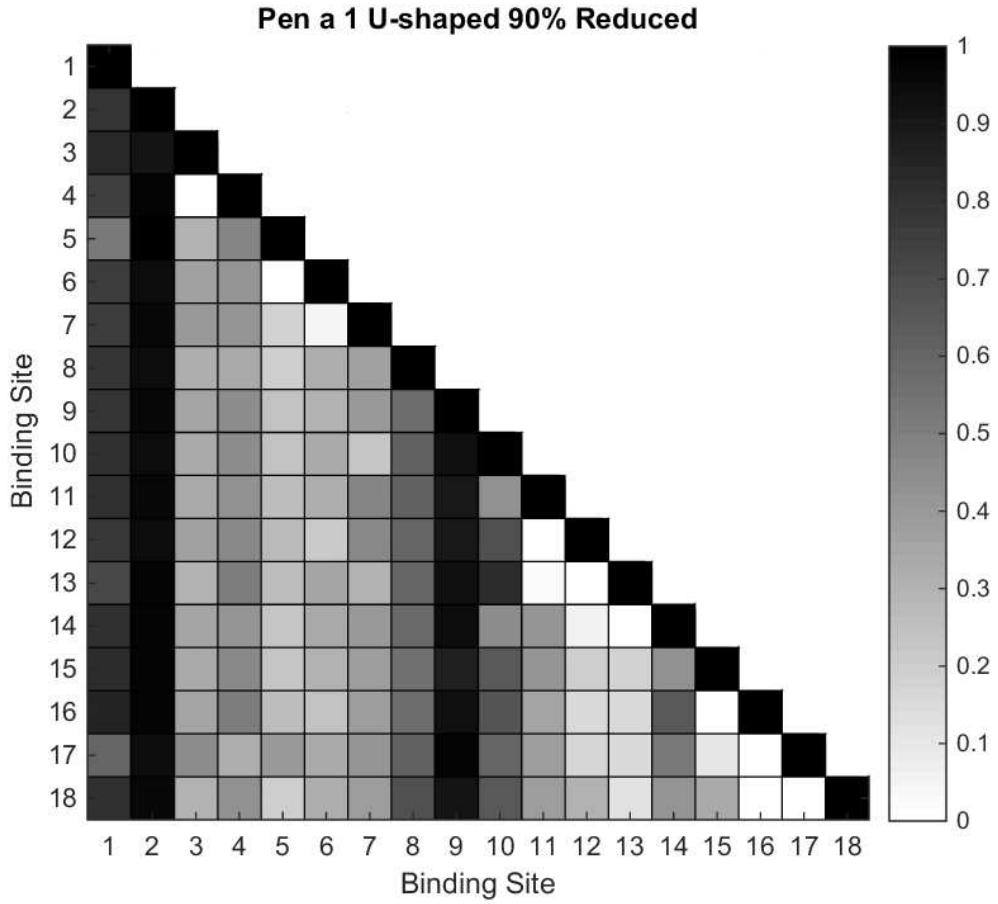

Figure 3.6: The conditional binding probabilities of the U-shaped Pen a 1 conformation, with binding site numbers corresponding to those of Figure 3.3. In our model, binding site 1 (2 on the opposite strand) belongs to region A, sites 5 and 6 (3 and 4 on the opposite strand) belong to region B, site 7 (8 on the opposite strand) belongs to region C, site 10 (9 on the opposite strand) belongs to region D, sites 11 and 12 (15 and 16 on the opposite strand) belong to region E, and sites 13 and 14 (17 and 18 on the opposite strand) belong to region F.

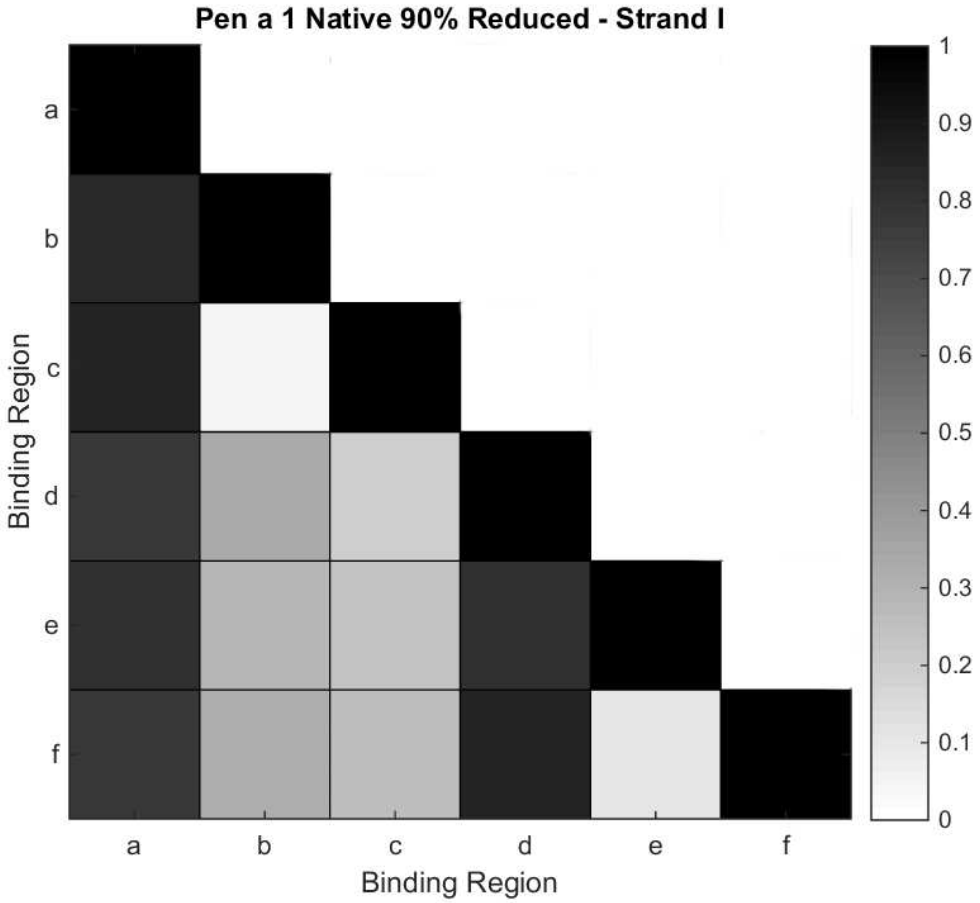

Figure 3.7: The conditional binding probabilities of strand I of the native Pen a 1 conformation using the six-region model.

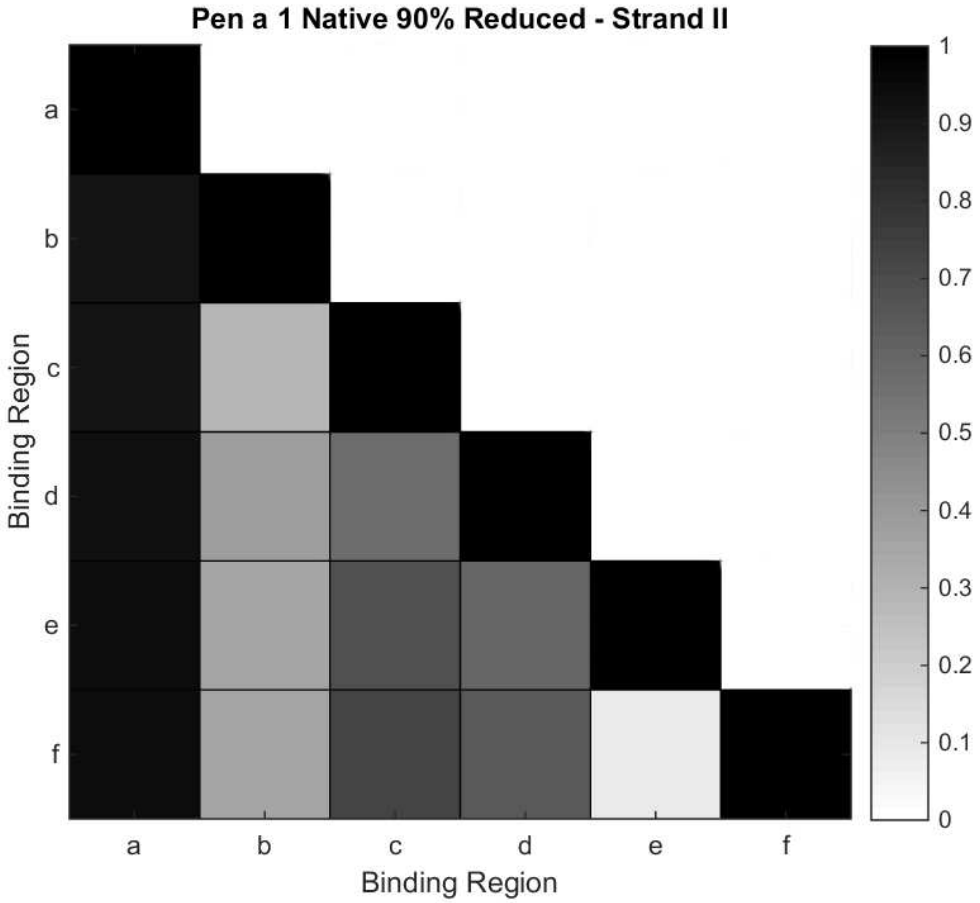

Figure 3.8: The conditional binding probabilities of strand II of the native Pen a 1 conformation using the six-region model.

Chapter 4

Geometric Rule-Based Modeling

In this chapter, 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 outline the assumptions that were made in the development of this rule-based model. We then describe our methods for determining the steric hindrances between the binding regions, constructing a set of rules based on these steric effects, and optimizing the rate constants. We describe the Monte Carlo rigid-body model and how we compared the results of our rule-based model to that of the Monte Carlo simulation. We explain how the probabilities of formation of various aggregate sizes were calculated. Finally, we demonstrate how the rule set varies with molecular curvature.

4.1 Designing Geometric Rules

4.1.1 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:

- We assume that IgE can only bind to a single binding region on the Pen a 1 molecule. An IgE cannot be bound to multiple regions simultaneously. We make this assumption based on the significantly high conditional binding probabilities for most pairs of binding regions (see Figures 3.7 and 3.8).
- In accordance with our Monte Carlo simulations, our rules forbid cross-linking Pen a 1 molecules through IgE binding.
- As previously mentioned (Section 3.3.1), we simplify our model by assuming that there are only 12 total binding regions on the Pen a 1 molecule (six per strand) rather than 16-18 total, and that each IgE can only bind to one of these 12 regions. 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.
- 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 affect the probability of an IgE receptor binding to a region on the opposite strand. We make this assumption based on the significantly high conditional binding probabilities for most pairs of binding sites on opposite strands (see Figures 3.4, 3.5, and 3.6).

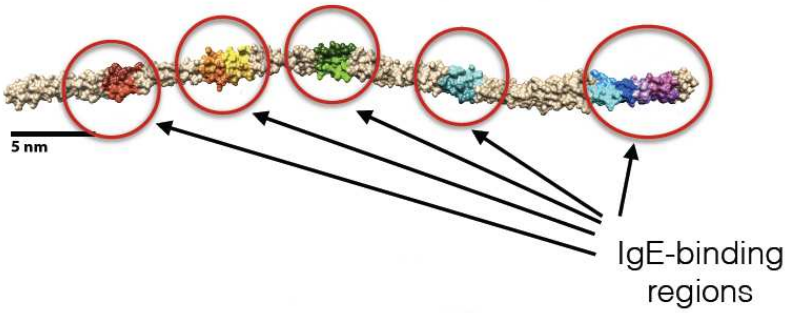

Figure 4.1: The all-atom native molecular structure of the shrimp tropomyosin Pen a 1 (tan), with the IgE binding regions circled. The IgE binding regions (various colors) are located in five regions per strand, although for our rule-based model, we have split the longer rightmost region into two separate regions so that there are six binding regions per strand. 2015 IEEE.

4.1.2 Rule Construction

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 4.2), 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 is an important parameter in this study as it determines the rule set of the rule-based model. For each resolution and 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 molecular curvature around a binding region, along with the IgE receptors bound to neighboring regions, may cause steric hindrance, i.e., IgE receptors may be prevented from binding to the region due to the region being blocked by receptors bound to neighbors. A *neighbor* 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

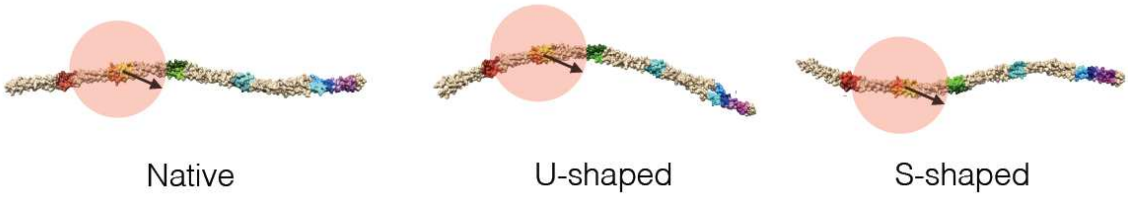

Figure 4.2: All-atom structures of three Pen a 1 conformations, with circles representing a possible region of steric hindrance around the yellow/orange binding region. These are the three conformations used in this study. 2015 IEEE.

categories of steric effects that we consider in our model, which are illustrated in Figure 4.3:

- There is no steric hindrance imposed on a binding region if all neighbors are unbound (Figure 4.3 (a)).
- When the molecular curvature around a binding region is negative, receptors bound to neighbors may reduce the accessibility to this region since negative curvature around two regions brings them closer together (the linear distance between the regions decreases) (Figure 4.3 (b)).
- When the molecular curvature around a binding region is positive, receptors bound to neighbors 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) (Figure 4.3 (c)).

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

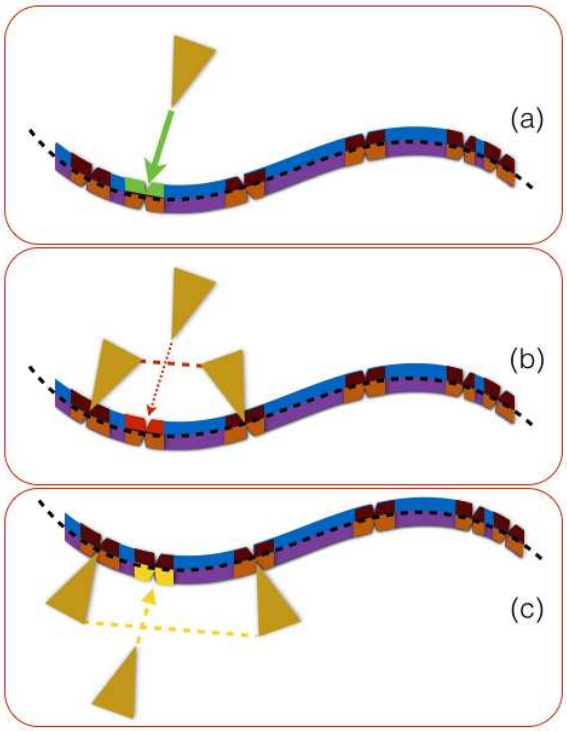

Figure 4.3: Steric hindrance induced by neighbor occupation. (a) No neighbors bound, (b) Negative curvature, and (c) Positive curvature.

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. This rate constant was determined by performing an optimization and fitting the resulting aggregate

size data to that of the 3D Monte Carlo model.

Because the cutoff distance is unknown, this parameter was varied, with the rule set being reconstructed for each cutoff distance. An optimization of k_{f2} was performed for every cutoff distance. (See Table A.1 in the Appendices for an example set of rules for strand I of the native Pen a 1 conformation; in this example, the cutoff distance is 7.0 nm, so there are steric effects between binding regions A and B, B and C, C and D, D and E, and E and F.)

In our previous work [Manavi et al., 2015], we constructed a rule set for the native conformation Pen a 1 molecule by encoding steric hindrances between the binding regions based only on the curvature of the surrounding region of the molecule. No distance data or cutoff distance was used, and the rule set was the same for all of the Monte Carlo resolutions. Each rule was assigned one of four forward rate constants based on nearest neighbor occupation, and three of these four rate constants were optimized for the purpose of fitting the aggregate size data to the Monte Carlo data. This earlier method differs from the new method used in this study in that the new method involves the optimization of the rule set itself, which changes based on the cutoff distance and the conformation of the molecule, and there are only two forward rate constants, only one of which is optimized.

4.2 Analysis

4.2.1 Rate Constant Optimization

The forward rate constant k_{f2} for each resolution of each Pen a 1 conformation was optimized using a simulated annealing 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 4.2.2). If the RSS

value of a new rate constant is higher than the current RSS value, then the new RSS value (and its associated rate constant) are accepted with a probability dependent on the difference between the two RSS values. 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 molecule^{-1}s^{-1} and 0.00001 molecule^{-1}s^{-1} 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 molecule^{-1}s^{-1}. We can expect from previous scans of k_{f2} that the RSS value for any value of k_{f2} greater than 0.40 molecule^{-1}s^{-1} will be too high to be acceptable.

4.2.2 Monte Carlo Comparison

Monte Carlo Method Description

We compared the aggregate size distributions from our rule-based model to those of a previously developed Monte Carlo model [Manavi et al., 2012] 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 4.4). Also, at every time step, any two binding sites (on two separate molecules) 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.

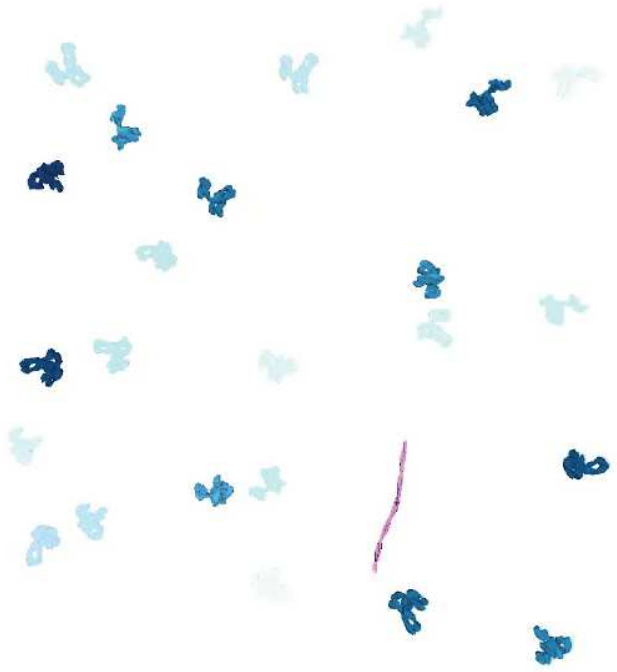

Figure 4.4: A 3D Monte Carlo simulation of one Pen a 1 molecule (violet) among multiple receptor complexes (blue).

The various resolutions of the Monte Carlo simulation were created by first creating isosurface models of the all-atom molecular structures using UCSF Chimera [Goddard et al., 2005] and then performing polygon reduction using Maya [Maya, 2014] (Figure 4.5).

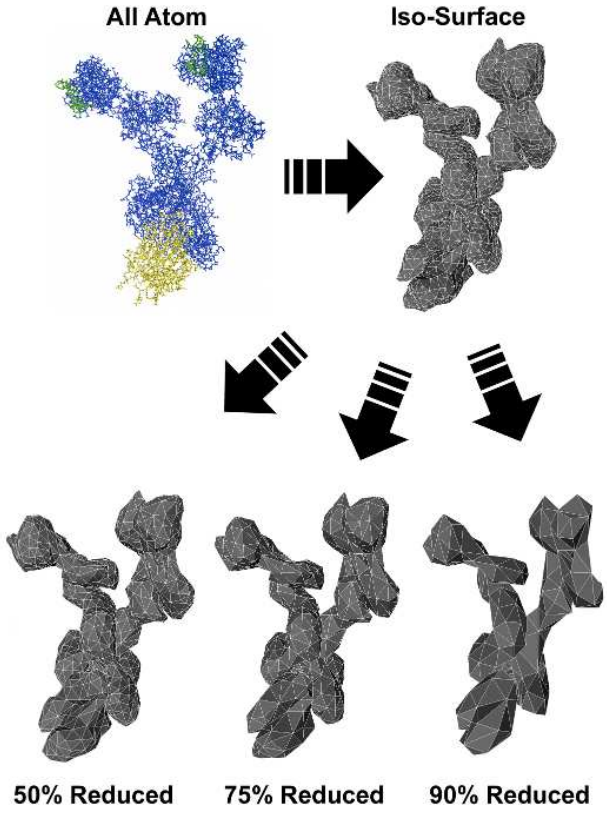

Figure 4.5: The creation of 3D molecular models at various resolutions using isosurface model creation and subsequent polygon reduction.

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 4.6). The probability $P(n)$ to form an aggregate of size n is given by:

$$\begin{aligned} P(n \leq 6) &= \sum_{m=0}^{n} P_I(m) P_{II}(n-m), \\ P(n > 6) &= \sum_{m=n-6}^{6} P_I(m) P_{II}(n-m), \end{aligned} \tag{4.1}$$

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

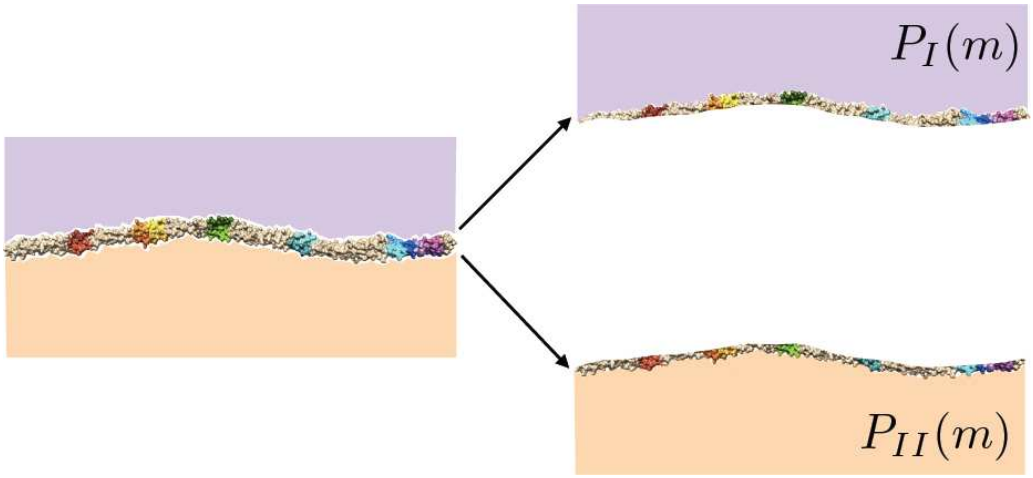

Figure 4.6: The all-atom native molecular structure of the shrimp tropomyosin Pen a 1 (left), visualized split into two strands: strand I and strand II (right). The probability of formation of an aggregate of a certain size is calculated by combining the independent probabilities for each strand. 2015 IEEE.

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 1, and the minimum possible normalized RSS (corresponding to two identical histograms) is zero.

4.3 Demonstration: Impact of Molecular 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 Pen a 1 molecule with dramatic changes in curvature. We look at three molecules: the U-shaped molecule seen in Figure 4.2 (which we refer to here as the U-shaped molecule), the same molecule with its two ends rotated inward by 45 degrees (which we name the 45-degree molecule), and the same molecule with its two ends rotated inward by 60 degrees (which we name the 60-degree molecule) (see Figure 4.7). (It should be noted that the latter two molecules are not energy-minimized conformations.) The rule sets for strand I are shown for the U-shaped molecule (Table A.17), the 45-degree molecule (Table B.1), and the 60-degree molecule (Table B.2).

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

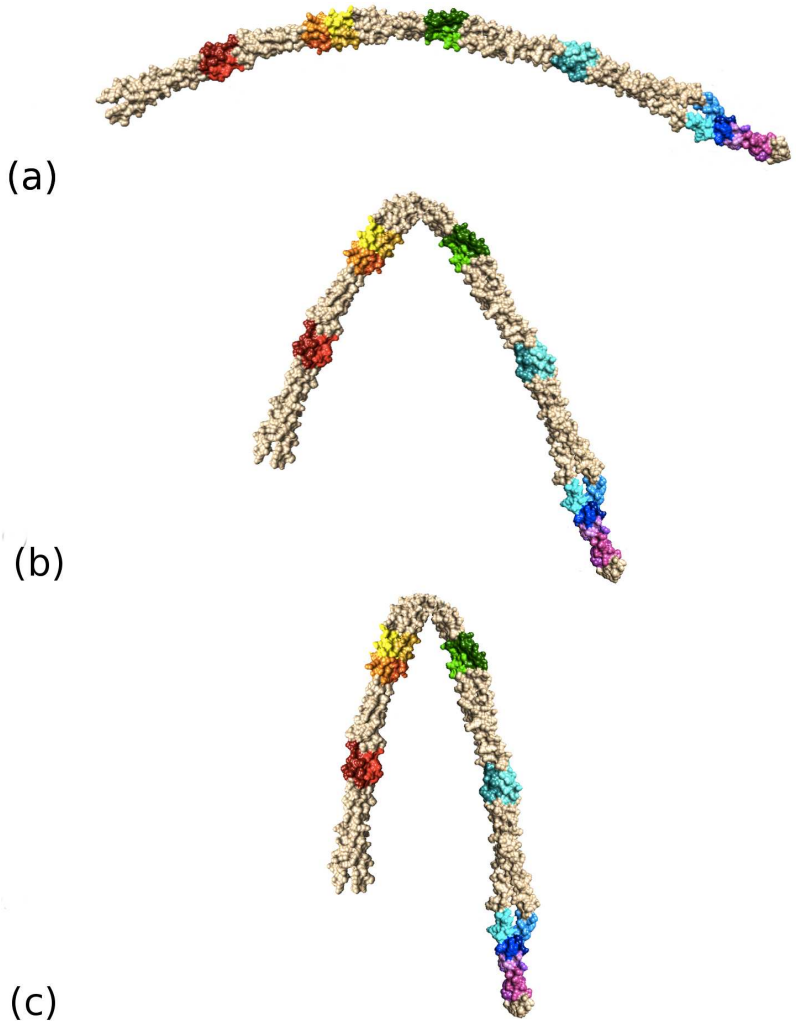

Figure 4.7: Visualizations of the (a) U-shaped molecule, (b) 45-degree molecule, and (c) 60-degree molecule. (It should be noted that the latter two molecules are not energy-minimized conformations and are only presented here for the purpose of demonstrating our rule construction method.) 2015 IEEE.

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 molecule^{-1}s^{-1}. The rule-based model for each molecule was simulated and the antibody aggregate size probabilities were calculated (see Figure 4.8). 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.

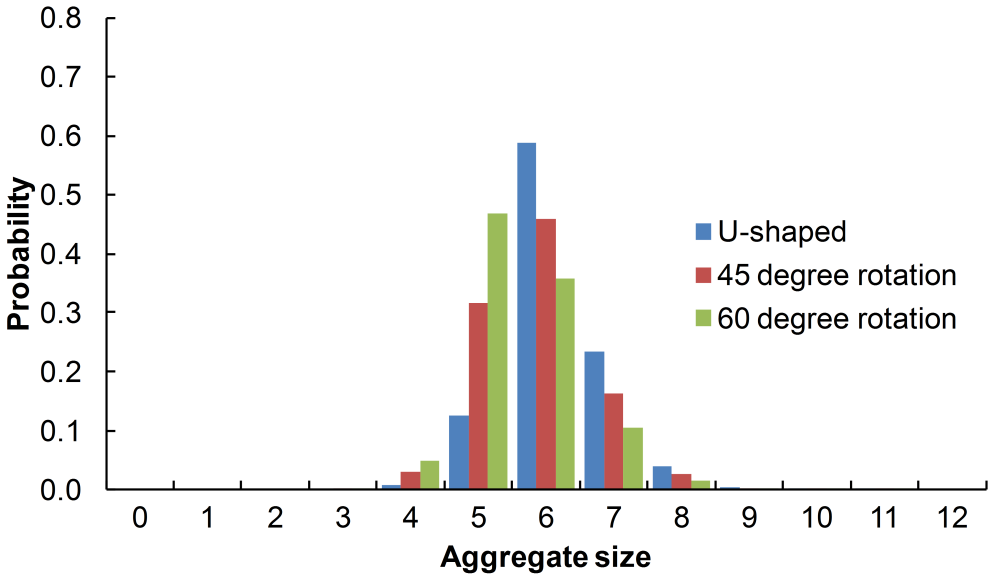

Figure 4.8: Comparison of rule-based model aggregate size distributions for the U-shaped molecule, 45-degree molecule, and 60-degree molecule. 2015 IEEE.

Chapter 5

Results and Discussion

The computational experiments we conducted are as follows. Firstly, in order to better understand how model resolution affects the Monte Carlo results and why, we quantified the variation in steric effects with Monte Carlo model resolution. To achieve this, we generated aggregate size distributions for seven resolutions of the Monte Carlo model, and we then fitted our geometric rule-based model to each of these seven resolutions, using the optimized parameters of our rule-based model to quantify variation in steric effects with resolution. Secondly, we created models to enable the prediction of Monte Carlo results at high model resolution by only running the simulation at lower resolutions, thus decreasing the time required for simulation. To achieve this, we analyzed trends in the variation of the parameters of our optimized rule-based models with resolution. Thirdly, we wanted to understand how steric effects vary with molecular geometry. To achieve this, we conducted the aforementioned resolution experiments and analysis for three conformations of Pen a 1: the native, S-shaped, and U-shaped molecules.

5.1 Experimental Setup

5.1.1 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. The simulations were run at seven different resolutions for each of the three Pen a 1 conformations, reducing the models of both antigen and receptor by 0%, 25%, 50%, 65%, 75%, 90%, and 95%. Sixty runs were performed for each resolution. Association and dissociation rates of 1.0 $molecule^{-1}s^{-1}$ and 0.01 s^{-1}, from [Xu et al., 1998] were used for the Pen a 1 antigen. The diffusion coefficient $0.09\mu m^2 s^{-1}$ of IgE-FcϵRI found in [Andrews et al., 2009] 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 reduction in speed of aggregates as they increase in size [Andrews et al., 2009] 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 Monte Carlo simulation code was developed using the Parasol Motion Planning Library (PMPL). The simulations were run on a supercomputer housed at UNM, utilizing single cores of Intel Xeon E5645 processors with 4 GB of RAM per processor.

5.1.2 Rule-Based Modeling

The rule-based model was specified in the BioNetGen language [Blinov et al., 2004], and ODE simulations were conducted on these models using RuleBender [Smith et al., 2012]. 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. 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 rule sets used in this study (see the Appendices) were chosen after testing several different rule-based models.

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.

5.2 Aggregate Size Histograms

The forward rate constant k_{f2} and the cutoff distance range were optimized for each resolution of each of the three conformations of Pen a 1. The tables display the rate constants, cutoff distances, and RSS values for each resolution of the native (Table 5.1), S-shaped (Table 5.3), and U-shaped (Table 5.5) Pen a 1 conformations.

The Monte Carlo aggregate size probability histogram data for each resolution of the Monte Carlo simulation is shown along with the optimized rule-based modeling

Table 5.1: Binding and unbinding rate constants, cutoff distances, and RSS values for the rule-based model for various resolutions of the native Pen a 1.

Parameter	Model Percent Reduction						
Value	0%	25%	50%	65%	75%	90%	95%
Cutoff distance (nm)	7.0-8.7	5.6-6.2	5.5	5.5	4.0-5.4	4.0-5.4	4.0-5.4
k_{f1} (molecule^{-1}s^{-1})	1.00	1.00	1.00	1.00	1.00	1.00	1.00
k_{f2} (molecule^{-1}s^{-1})	6.60e-03	7.58e-03	3.86e-03	4.90e-04	1.27e-03	1.17e-02	1.64e-02
k_r (s^{-1})	0.01	0.01	0.01	0.01	0.01	0.01	0.01
RSS	7.03e-04	3.62e-04	6.30e-05	1.62e-03	6.31e-04	1.29e-04	9.25e-04

Table 5.3: Binding and unbinding rate constants, cutoff distances, and RSS values for the rule-based model for various resolutions of the S-shaped Pen a 1.

Parameter	Model Percent Reduction						
Value	0%	25%	50%	65%	75%	90%	95%
Cutoff distance (nm)	6.8-8.3	5.8-6.0	5.8-6.0	5.8-6.0	5.4-5.7	4.5-5.3	4.5-5.3
k_{f1} (molecule^{-1}s^{-1})	1.00	1.00	1.00	1.00	1.00	1.00	1.00
k_{f2} (molecule^{-1}s^{-1})	3.56e-03	1.03e-02	8.67e-03	7.84e-03	4.12e-03	6.38e-03	5.58e-03
k_r (s^{-1})	0.01	0.01	0.01	0.01	0.01	0.01	0.01
RSS	1.35e-04	1.65e-03	1.48e-04	3.78e-04	9.40e-05	6.77e-04	4.00e-05

Table 5.5: Binding and unbinding rate constants, cutoff distances, and RSS values for the rule-based model for various resolutions of the U-shaped Pen a 1.

Parameter	Model Percent Reduction						
Value	0%	25%	50%	65%	75%	90%	95%
Cutoff distance (nm)	6.8-8.6	6.8-8.6	5.3	5.3	5.3	4.1-5.2	3.9-4.0
k_{f1} (molecule^{-1}s^{-1})	1.00	1.00	1.00	1.00	1.00	1.00	1.00
k_{f2} (molecule^{-1}s^{-1})	7.32e-03	1.07e-02	1.94e-03	1.18e-03	1.01e-02	1.04e-02	4.26e-03
k_r (s^{-1})	0.01	0.01	0.01	0.01	0.01	0.01	0.01
RSS	4.69e-04	1.41e-03	5.97e-04	2.77e-04	5.23e-04	3.18e-04	2.59e-04

data for the native (Figure 5.1), S-shape (Figure 5.2), and U-shape (Figure 5.3) Pen a 1 conformations. The error bars for the Monte Carlo data were calculated by dividing the 60 runs into 10 sets of six runs each and then calculating the standard error of the mean.

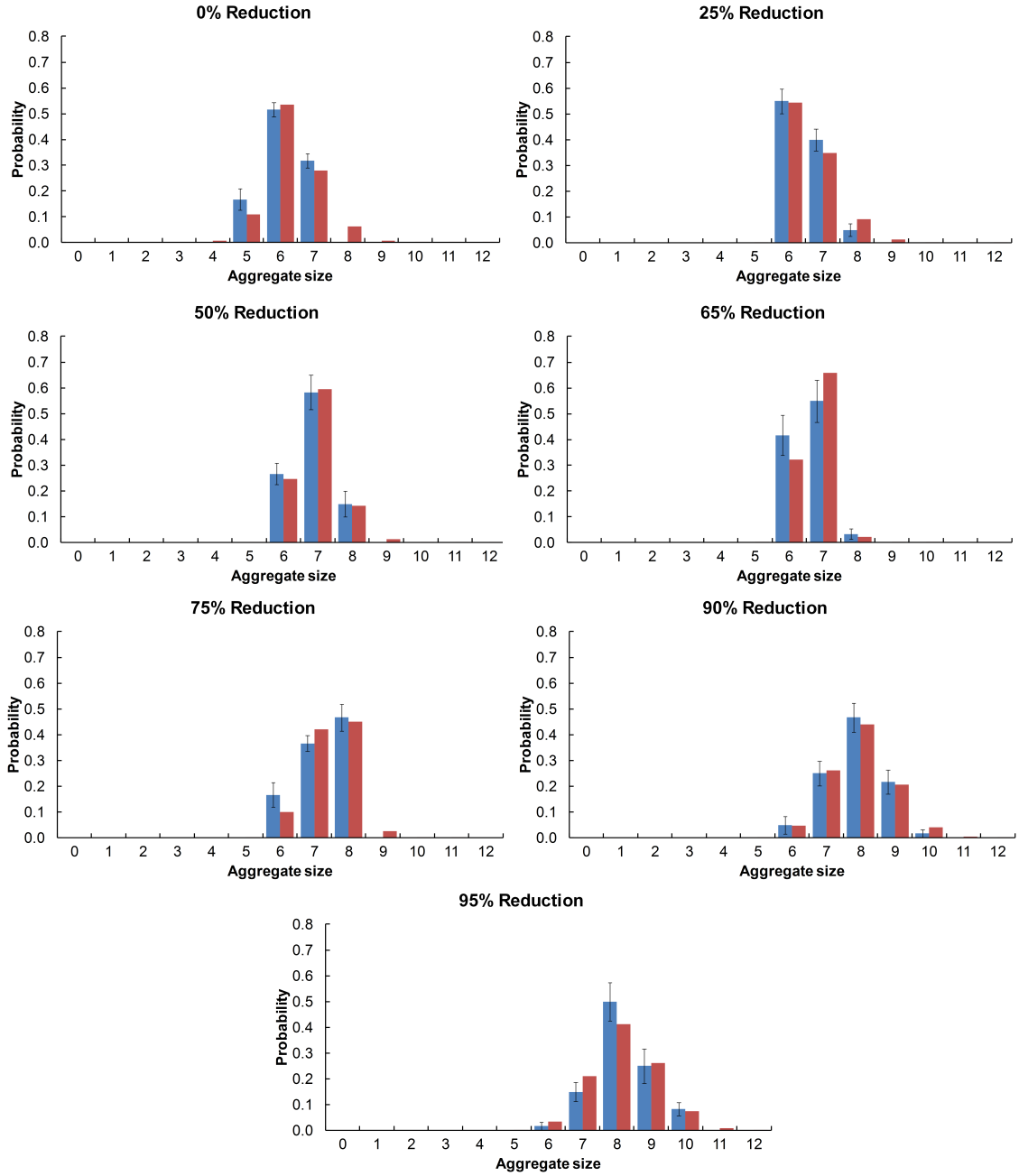

Figure 5.1: Comparison of Monte Carlo (blue) and optimized rule-based model (red) aggregate size distributions for the native Pen a 1. The error bars for the Monte Carlo data were calculated by dividing the 60 runs into 10 sets of six runs each and then calculating the standard error of the mean.

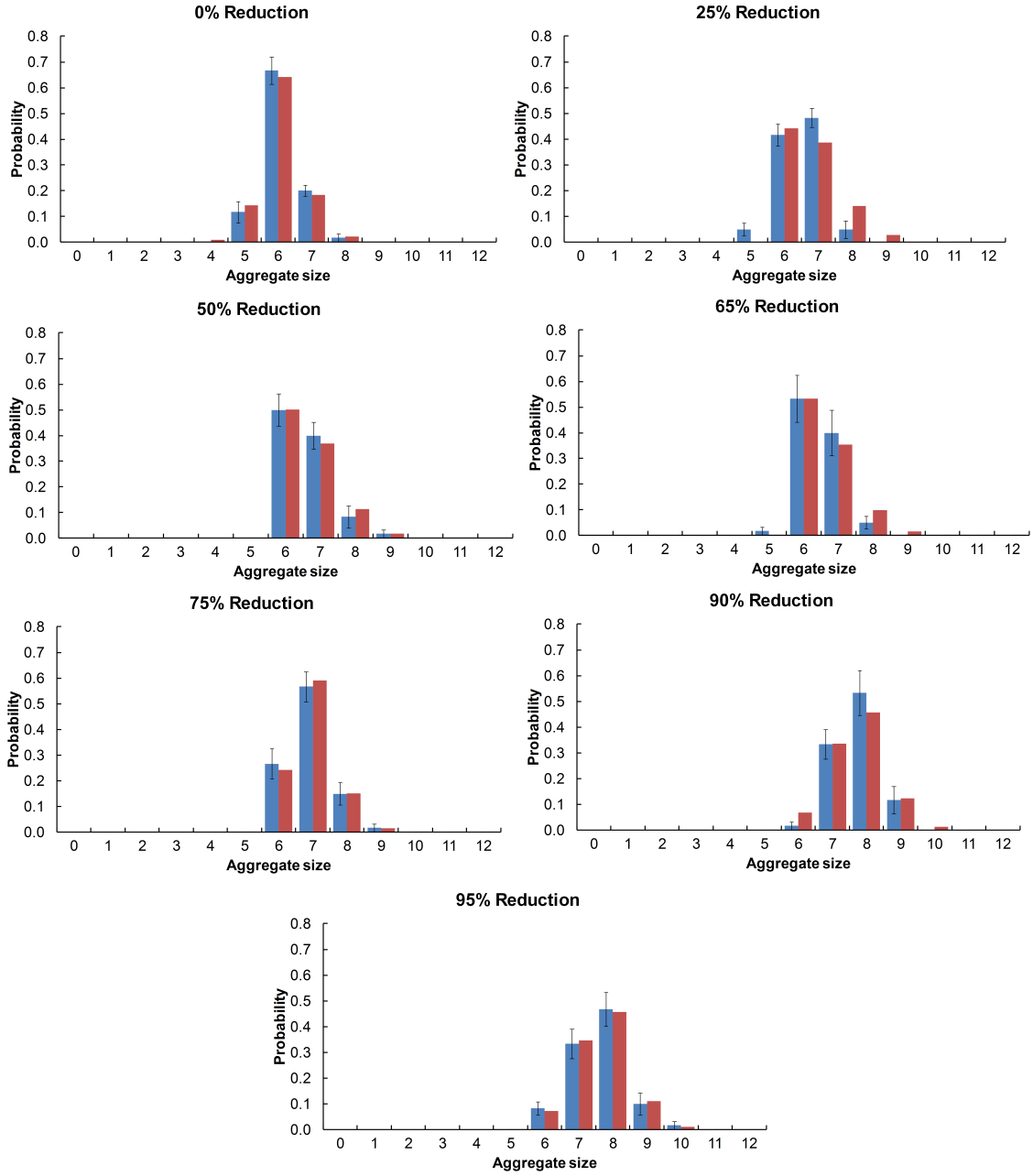

Figure 5.2: Comparison of Monte Carlo (blue) and optimized rule-based model (red) aggregate size distributions for the S-shaped Pen a 1. The error bars for the Monte Carlo data were calculated by dividing the 60 runs into 10 sets of six runs each and then calculating the standard error of the mean.

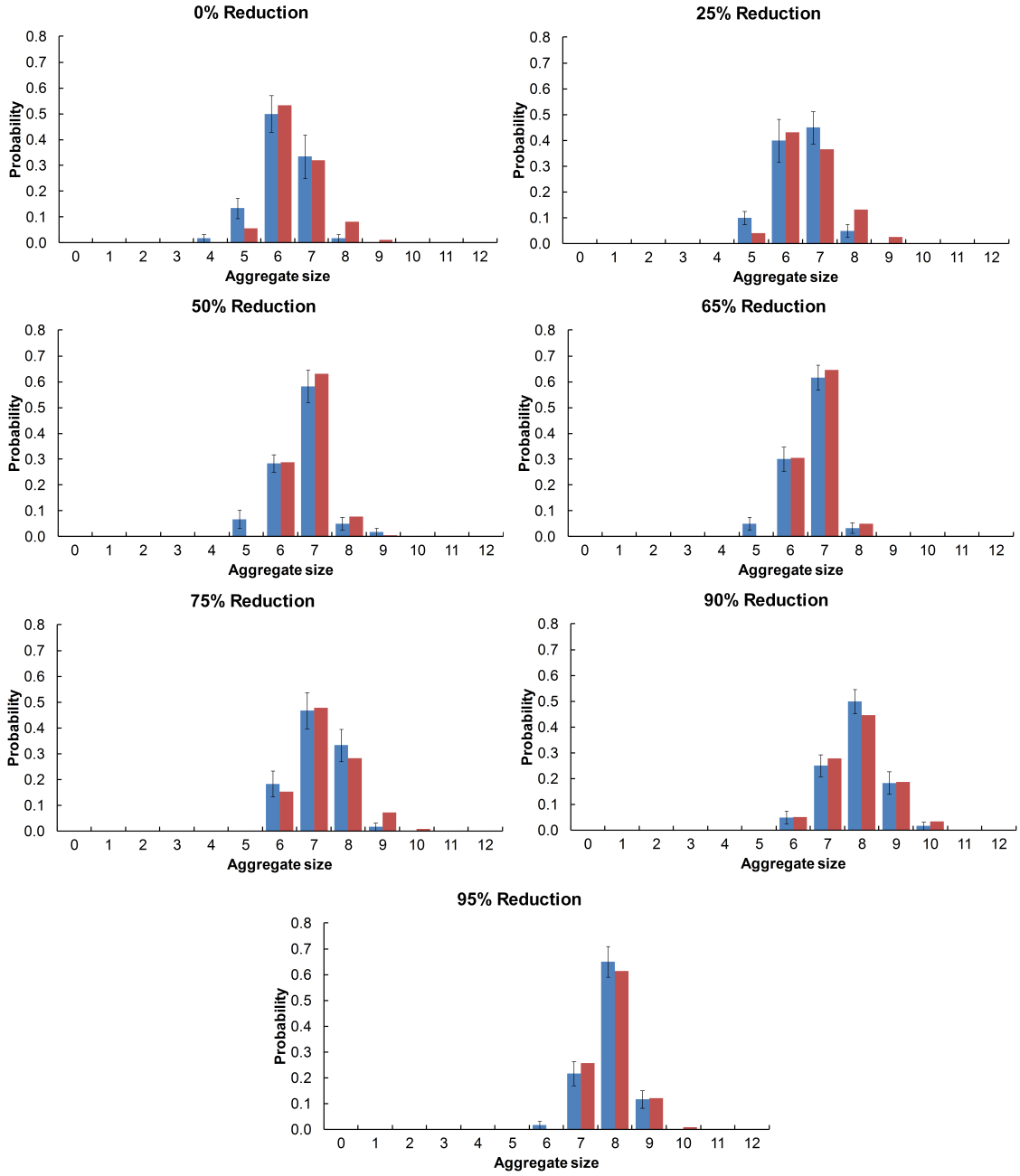

Figure 5.3: Comparison of Monte Carlo (blue) and optimized rule-based model (red) aggregate size distributions for the U-shaped Pen a 1. The error bars for the Monte Carlo data were calculated by dividing the 60 runs into 10 sets of six runs each and then calculating the standard error of the mean.

5.3 Comparison with Previous Method

In our previous work [Manavi et al., 2015], we optimized rule-based models for the native conformation Pen a 1 molecule at the 0%, 50%, 75%, and 90% reduced resolutions. We found that this older method yielded poorer fits overall than did our newer method proposed in this work. The RSS values for the fits of the rule-based model to the Monte Carlo data is shown in Table 5.7.

Table 5.7: Comparison of RSS values for the rule-based method used in this work and the rule-based method used in our previous work for various resolutions of the native Pen a 1.

Rule-based Method	Model Percent Reduction			
	0%	50%	75%	90%
New method	7.03e-04	6.30e-05	6.31e-04	1.29e-04
Old method	1.37e-03	2.28e-03	2.73e-03	1.14e-03

5.4 Cutoff Distance Models

For each Pen a 1 conformation, we plotted the optimal cutoff distance (the cutoff distance for the rule-based model that best fit the Monte Carlo data) versus the Monte Carlo resolution for the native (Figure 5.4), S-shape (Figure 5.5), and U-shape (Figure 5.6) Pen a 1 conformations. The goal of this model is to aid in understanding how the optimal cutoff distance, and hence, the steric hindrances between binding regions vary with the resolution for different conformations of Pen a 1. The error bars on some of the data points represent the range of possible cutoff distances that includes the optimal cutoff distance. Since the rule-based model is identical for every cutoff distance in this range, the optimal cutoff distance could be any of these distances. The stippled boxes around the data points group resolutions with the

same optimal cutoff distance range together; the numbers next to the boxes are the numbers of the rule sets (listed in Appendix A) that represent that cutoff distance range.

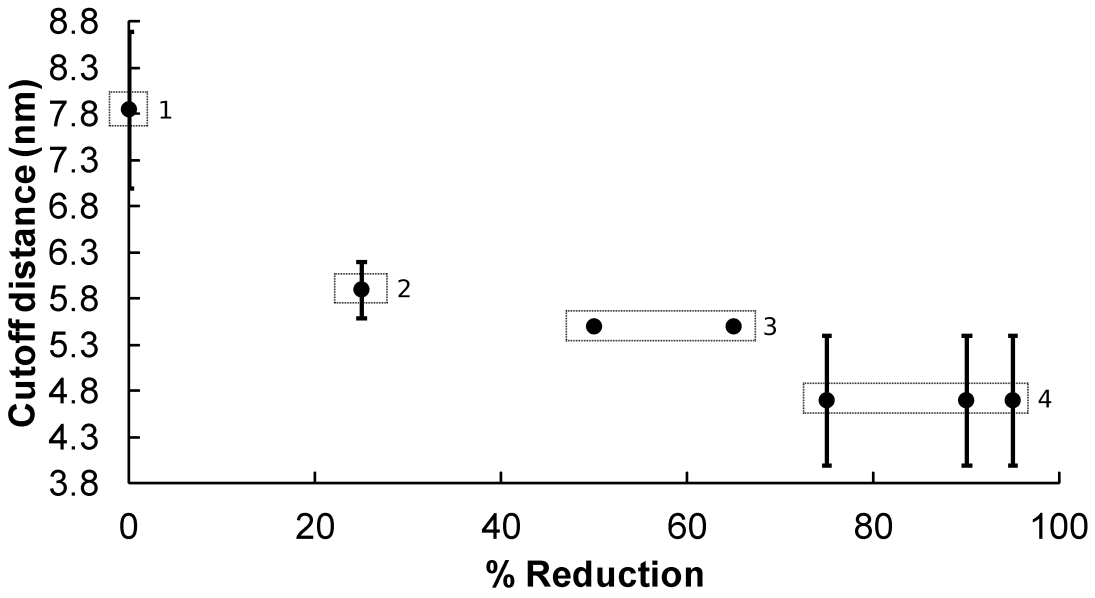

Figure 5.4: Cutoff distance versus resolution data for the native Pen a 1 conformation. The error bars on some of the data points represent the range of possible cutoff distances that includes the optimal cutoff distance. The stippled boxes around the data points group resolutions with the same optimal cutoff distance range together; the numbers next to the boxes are the numbers of the rule sets (listed in the Appendices) that represent that cutoff distance range.

5.5 Probabilities of Binding for Individual Binding Regions

The probability of binding for each individual binding region of the Pen a 1 molecule varies with resolution and with conformation. For each optimized rule-based model described in Section 5.2, the probability of binding versus model resolution is shown

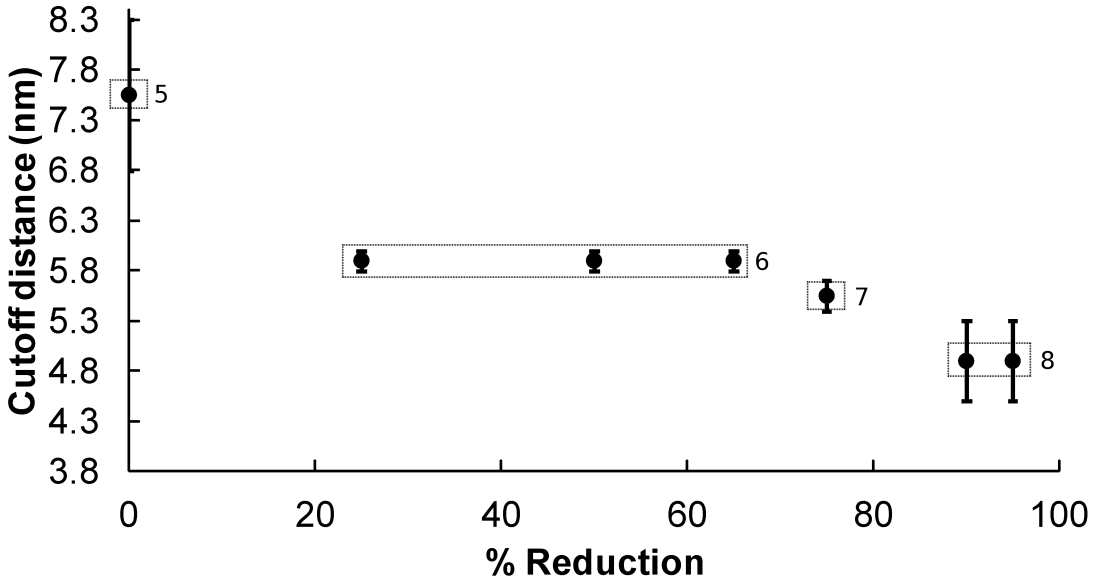

Figure 5.5: Cutoff distance versus resolution data for the S-shaped Pen a 1 conformation. The error bars on some of the data points represent the range of possible cutoff distances that includes the optimal cutoff distance. The stippled boxes around the data points group resolutions with the same optimal cutoff distance range together; the numbers next to the boxes are the numbers of the rule sets (listed in the Appendices) that represent that cutoff distance range.

for each of the six binding regions used in our model for the native (Figure 5.7), S-shaped (Figure 5.8), and U-shaped (Figure 5.9) conformations.

We observe that for each of the three conformations, region D exhibits strong variation in binding probability with resolution for both strand I and strand II. Region C exhibits strong variation for strand II of the U-shaped type. We also observe that certain pairs of binding regions display symmetry such that their binding probabilities have exactly the same value. Symmetric values correspond with overlapping data points in Figures 5.7, 5.8, and 5.9. For example, for the native type at 0% reduced resolution, the region pairs A and F, B and E, and C and D are symmetric for both strands. This symmetry occurs due to the strong similarities between

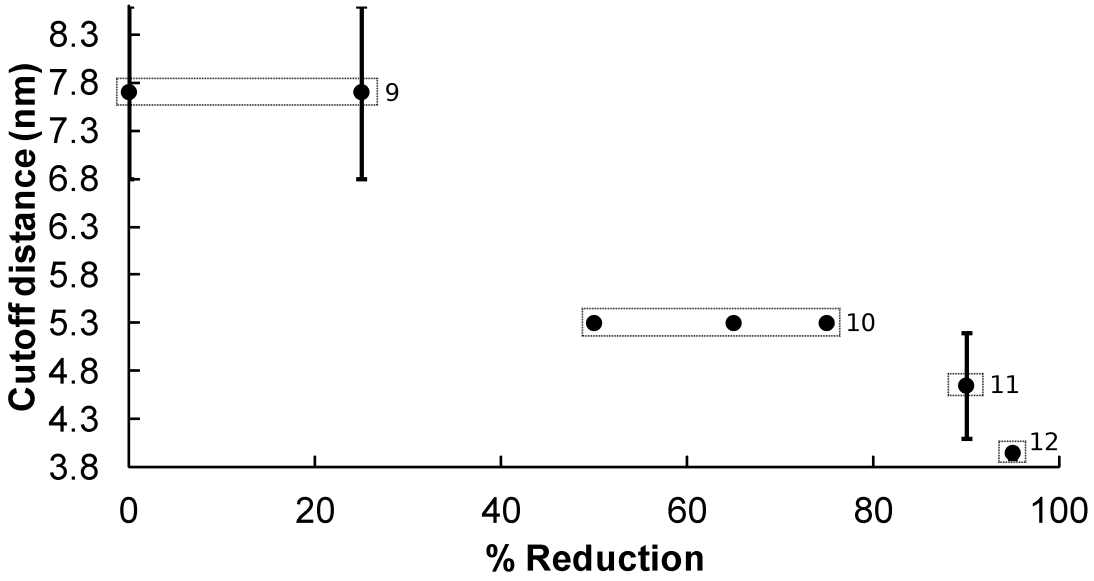

Figure 5.6: Cutoff distance versus resolution data for the U-shaped Pen a 1 conformation. The error bars on some of the data points represent the range of possible cutoff distances that includes the optimal cutoff distance. The stippled boxes around the data points group resolutions with the same optimal cutoff distance range together; the numbers next to the boxes are the numbers of the rule sets (listed in the Appendices) that represent that cutoff distance range.

the binding rules of two or more regions; for example, the binding rules of regions in symmetric groupings encode steric effects for the same number of neighboring regions.

We compare the binding probability data for our optimized rule-based model with that of the Monte Carlo simulation for the native Pen a 1 at 0% reduction in Figure 5.10. The two sets of data line up reasonably well for most of the regions in strand I, although we observe that there are large differences in the two sets of data for regions B, D, and E of strand II and for regions C and D of strand I. Our rule-based model may not be accurately capturing some details of molecular geometry, particularly differences between strands I and II. Future improvements to this model

could be made by fitting our rule-based model to the Monte Carlo binding region probability data in addition to the aggregate size data in order to improve accuracy.

5.6 Rate Constant Prediction

Our cutoff distance model has two variable parameters: the cutoff distance range, which corresponds to the rule set, and the variable rate constant k_{f2}. For the purpose of creating a model that can be used to predict the aggregate size data of higher resolutions given only lower resolution data, it is helpful to fix one parameter and only allow the other parameter to vary. Then, a trend can be more easily observed.

Due to the difficulty of fixing k_{f2} at a constant value for all resolutions and still being able to obtain rule-based models that fit the Monte Carlo data well, we have fixed the cutoff distance range for all resolutions and allowed k_{f2} to vary. The rate constant k_{f2} was optimized as described in Section 4.2.1. We have chosen to fix the cutoff distance range at the optimal range for the 25% reduced resolution. This range is 5.6-6.2 nm for the native type, 5.8-6.0 nm for the S-shaped type, and 6.8-8.6 nm for the U-shaped type. Since our goal is to be able to predict the results of the 0% reduced resolution, the rule set of the 25% reduced resolution strikes a balance between allowing good fits to the 0% reduced Monte Carlo data and still allowing decent fits to the lower resolution Monte Carlo data. The aggregate size histogram data for the 0% reduced resolution, along with the optimal rate constant versus resolution data is shown in Figure 5.11 for the native type Pen a 1, Figure 5.12 for the S-shaped type, and Figure 5.13 for the U-shaped type. This data was fitted to exponential functions.

Sources of error include error in the Monte Carlo simulation data, which comes from the somewhat small number of runs (60) used to fit our predictive exponential functions and the Monte Carlo data not always fitting the expected trend of higher

aggregate sizes with higher percent reduction in resolution (see Figure 5.12). Another source of error is the fitted function itself, which provides a predicted rate constant that is different from the actual rate constant. For comparison, Figures 5.11, 5.12, and 5.13 display the rule-based modeling histogram data corresponding to both the actual optimized rate constant for the 0% reduction Monte Carlo data and the rate constant predicted by the exponential function to fit the 0% reduction data.

Ideally, we want to develop data sets for fitting our models using only rate constant versus resolution data for lower resolutions, so that the computationally costly higher resolution Monte Carlo simulations do not need to be run. For this reason, we developed models using a method similar to that used for the models described above, with the exception that the functions are fitted to the set of data omitting the 0% reduced data, or omitting both the 0% reduced data and the 25% reduced data. The models that omit only the 0% reduced data, along with the corresponding aggregate size histogram predicted for the 0% reduced resolution, are shown in Figure 5.14 for the native type, Figure 5.15 for the S-shaped type, and Figure 5.16 for the U-shaped type. The models that omit both the 0% and 25% reduced data, along with the corresponding aggregate size histogram predicted for the 0% reduced resolution, are shown in Figure 5.17 for the native type, Figure 5.18 for the S-shaped type, and Figure 5.19 for the U-shaped type.

5.7 Discussion

5.7.1 Rate Constants versus Resolution

We expect that for models with the same cutoff distance range (and therefore, the same rule set), the optimized rate constant k_{f2} should increase as the aggregate size distribution shifts toward larger aggregates. This is because higher rate constants

correspond with increased binding, which results in larger aggregates. Looking at the histograms in Figure 5.1, Figure 5.2, and Figure 5.3, as well as the rate constant data in Table 5.1, Table 5.3, and Table 5.5, we do indeed observe this trend, which is most clearly seen by comparing the 75%, 90%, and 95% reduced resolutions of the native Pen a 1. For these resolutions, which all have the same optimized cutoff distance range, the aggregate size distribution shifts toward larger aggregates as the percent reduction in resolution increases, and there is a corresponding increase in the optimized k_{f2} value with reduction in resolution. We observe a similar feature for the 0% and 25% resolutions of the U-shaped Pen a 1.

It may be expected that this increase in k_{f2} with a reduction in resolution should also hold true for other resolutions with the same rule set, such as the 50% and 65% reduced resolutions of the native Pen a 1. However, looking at the histograms in Figure 5.1, we see that the Monte Carlo aggregate size distribution of the 65% reduced resolution is shifted toward smaller aggregate sizes than is the 50% reduced resolution. Given this data, the fact that k_{f2} is larger for the 50% reduced resolution than for the 65% reduced resolution makes sense. It should be noted that this unexpected feature of the Monte Carlo data may be due to the rather small number of runs (60) performed for each Monte Carlo resolution. We observe a similar feature for the 25%, 50%, and 65% resolutions of the S-shaped Pen a 1, for the 90% and 95% resolutions of the S-shaped Pen a 1, and for the 50% and 65% resolutions of the U-shaped Pen a 1.

5.7.2 Cutoff Distance versus Resolution

As the resolution of the Monte Carlo model decreases, the volume of the molecular models decreases, and the steric hindrance between binding regions decreases, resulting in a shift toward larger aggregates with a reduction in resolution. We ex-

pect that the cutoff distance range, which is used in this method as a measure of average steric hindrance between binding regions, should generally decrease as the resolution decreases. Looking at Figure 5.4, Figure 5.5, and Figure 5.6, this is what is observed. However, it is difficult to fit a function to this data in order to predict what the optimal cutoff distance should be at any resolution. We also observe that some of the resolutions have the same optimized cutoff distance range. For example, the 75%, 90%, and 95% resolutions of the native Pen a 1 all have a cutoff distance range of 4.0-5.4 nm.

Studying how the cutoff distance range changes with resolution provides us with useful information about how the average steric hindrance changes (or does not change) with resolution. For example, from the native Pen a 1 data, we can infer that the average steric hindrance is about the same for the 50% and 65% resolutions. It should be noted that some cutoff distance ranges are rather large, such as the range of 4.0-5.4 nm, and the exact optimized cutoff distance lies somewhere within that range. Therefore, for the 75%, 90%, and 95% resolutions, we cannot assume that the average steric hindrance is exactly the same.

Comparing this data with the distances between binding regions on the Pen a 1 molecule allows us to predict which pairs of binding regions we expect to exhibit the greatest amount of steric hindrance on each other for each resolution of the Monte Carlo model. The cutoff distance for the 50% and 65% reduced resolutions of the native type is 5.5 nm. We can predict that any pair of binding regions with this or a smaller distance between them will exhibit a significant amount of steric hindrance on each other. In this case, these pairs of binding regions on strand I would be A and B, B and C, C and D, and E and F. On strand II, however, these pairs of binding regions would be A and B, B and C, and E and F. This is due to differences in curvature between the two strands. The distance between binding regions C and D on strand I is 5.5 nm, which is the cutoff distance. On strand II, the distance between

binding regions C and D is 5.6 nm, which is outside the cutoff distance range, so we can predict that C and D do not exhibit significant steric hindrance on each other on strand II.

The cutoff distance versus resolution data may potentially be useful for identifying which resolutions are similar enough to a greater resolution that the Monte Carlo simulations with these reduced resolutions could be run instead of the greater resolution, saving computational time. If necessary, the difference in aggregate size data caused by a reduction in resolution could be corrected by constructing a rule-based model, optimizing the cutoff distance and rate constant to fit the rule-based model data to the Monte Carlo data, applying a correction to the rate constant and/or adding a couple of rules to the model, and then running the corrected rule-based model to obtain results similar to those from a high-resolution Monte Carlo model. For example, the 0% and 25% reduced resolutions of the U-shaped type have the same cutoff distance range. The 25% reduced model could be run in place of the 0% reduced model, and the rate constant of the corresponding rule-based model could be reduced a little to correct for the difference in aggregate size histograms between the two resolutions. For the native type, the 75%, 90%, and 95% resolutions have the same cutoff distance range; the 95% reduced Monte Carlo resolution could be run to save computational time. The Monte Carlo data could then be corrected to the 75% reduced resolution by reducing the rate constant, and the resulting data could be corrected even further by adding rules to the model according to the cutoff distance versus resolution data.

5.7.3 Differences Between Conformations

We observe that there are significant differences between the plots of cutoff distance versus resolution for each of the three conformations (native, S-shaped, and

U-shaped). This is not unexpected, as the differences in curvature between the three molecules means that the distances between binding regions also vary. For example, the resolutions at which the cutoff distance range changes are not the same for the three conformations. In addition, the cutoff distances themselves vary based on conformation. These differences indicate that molecular geometry plays an important role in antibody aggregation onto the Pen a 1 molecule and should be taken into consideration in future research.

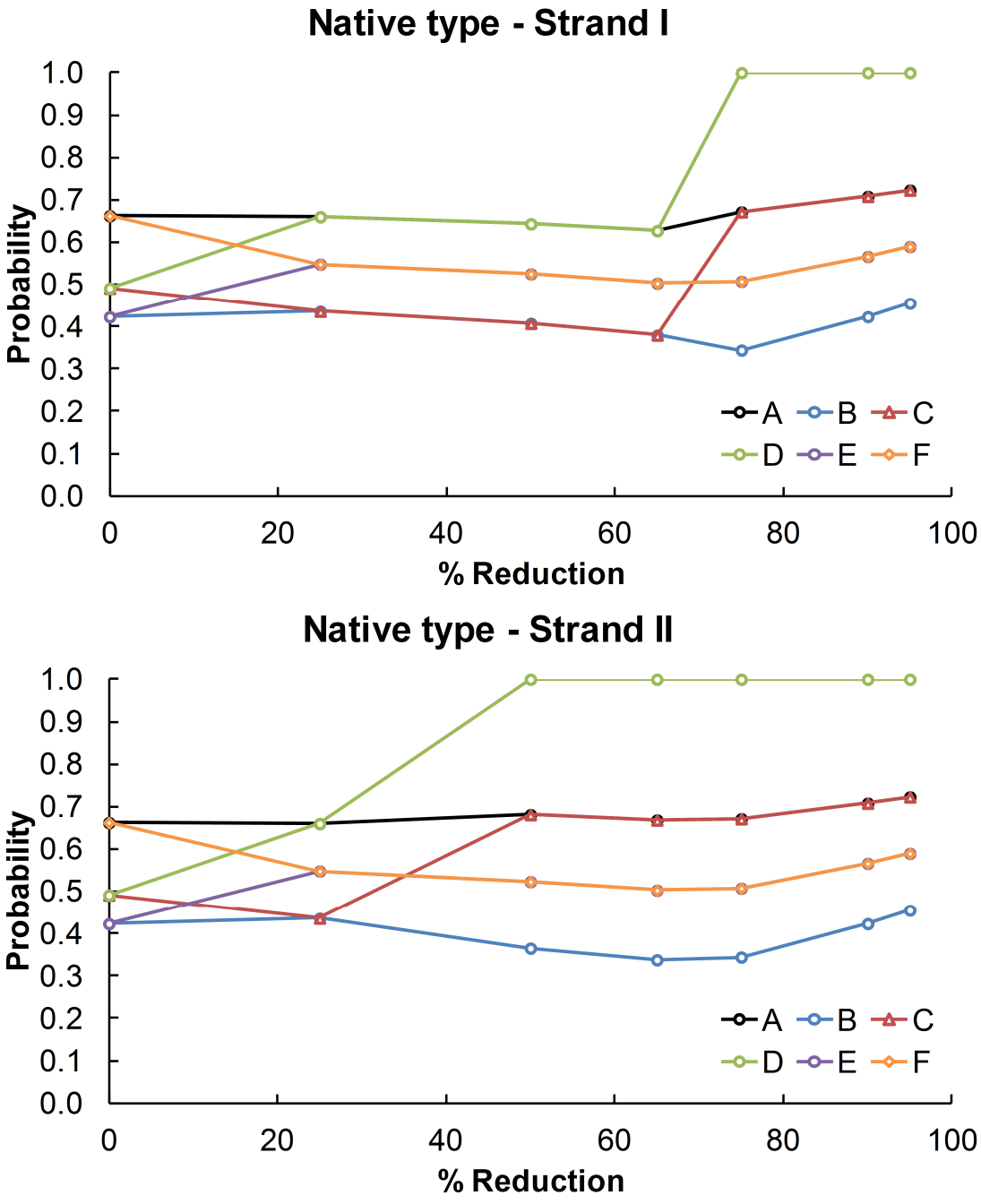

Figure 5.7: Binding probability versus resolution rule-based modeling data for strand I (top) and strand II (bottom) of the native Pen a 1 conformation. The letters A, B, C, D, E, and F represent the six binding regions used in our rule-based model.

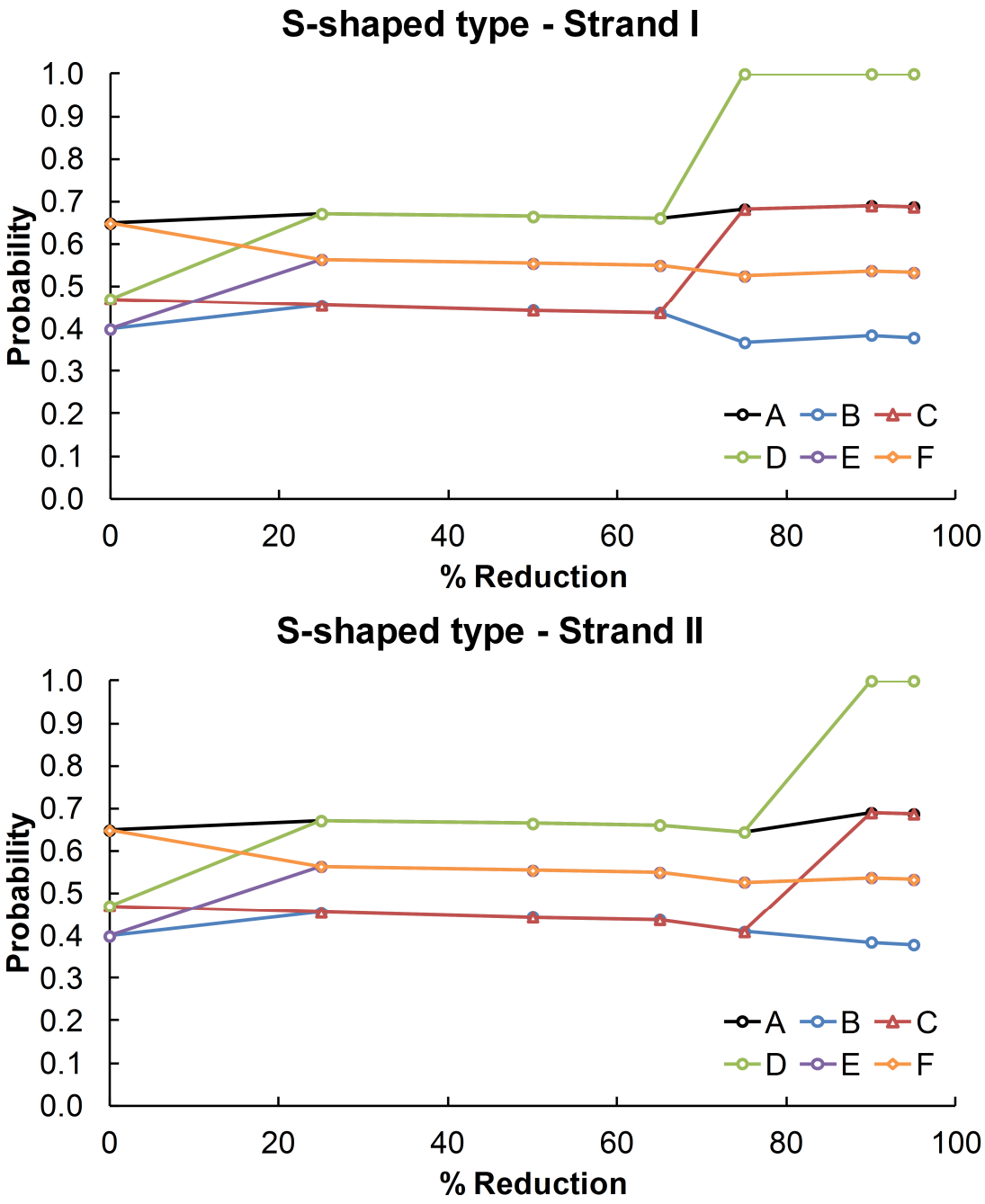

Figure 5.8: Binding probability versus resolution rule-based modeling data for strand I (top) and strand II (bottom) of the S-shaped Pen a 1 conformation. The letters A, B, C, D, E, and F represent the six binding regions used in our rule-based model.

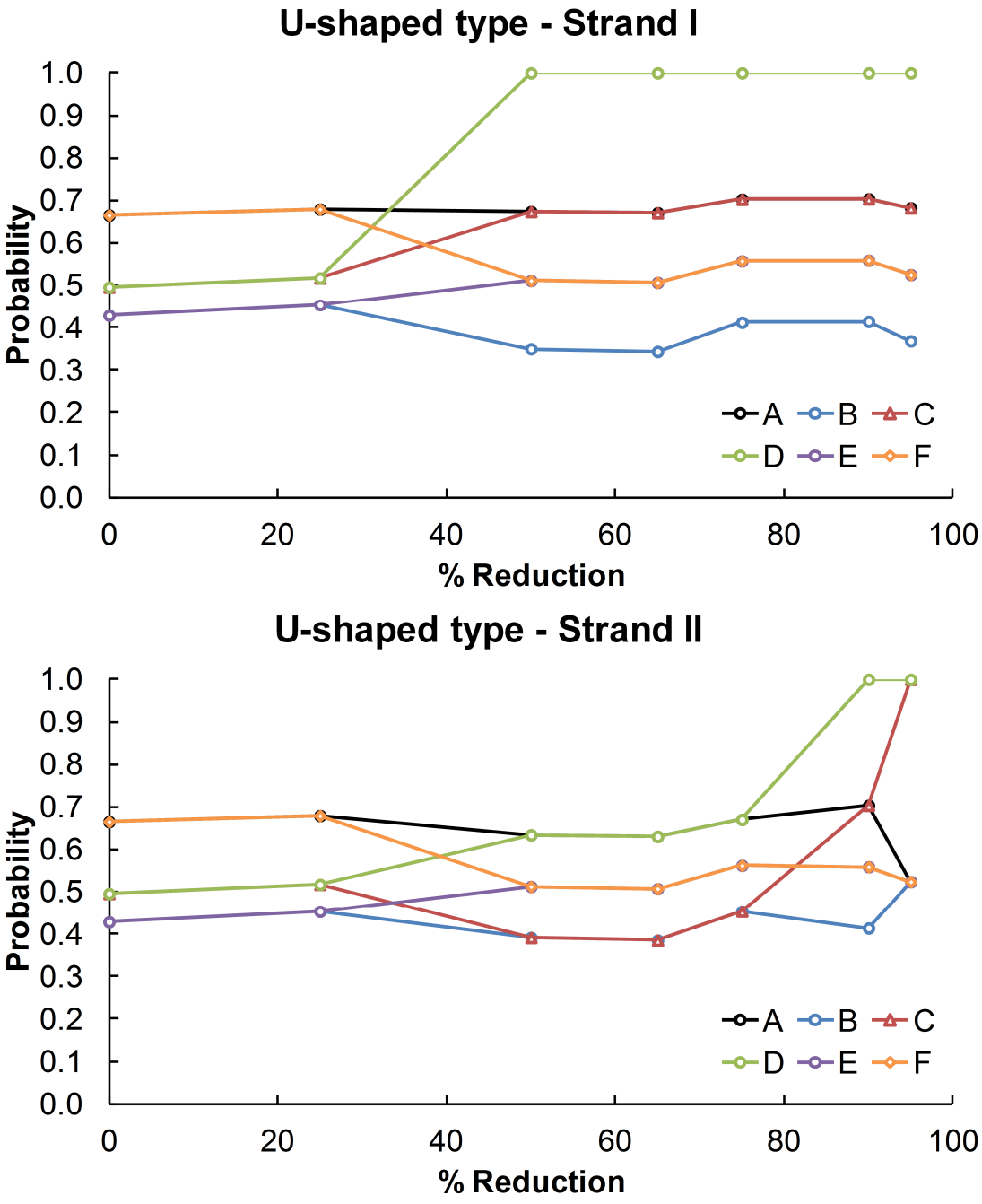

Figure 5.9: Binding probability versus resolution rule-based modeling data for strand I (top) and strand II (bottom) of the U-shaped Pen a 1 conformation. The letters A, B, C, D, E, and F represent the six binding regions used in our rule-based model.

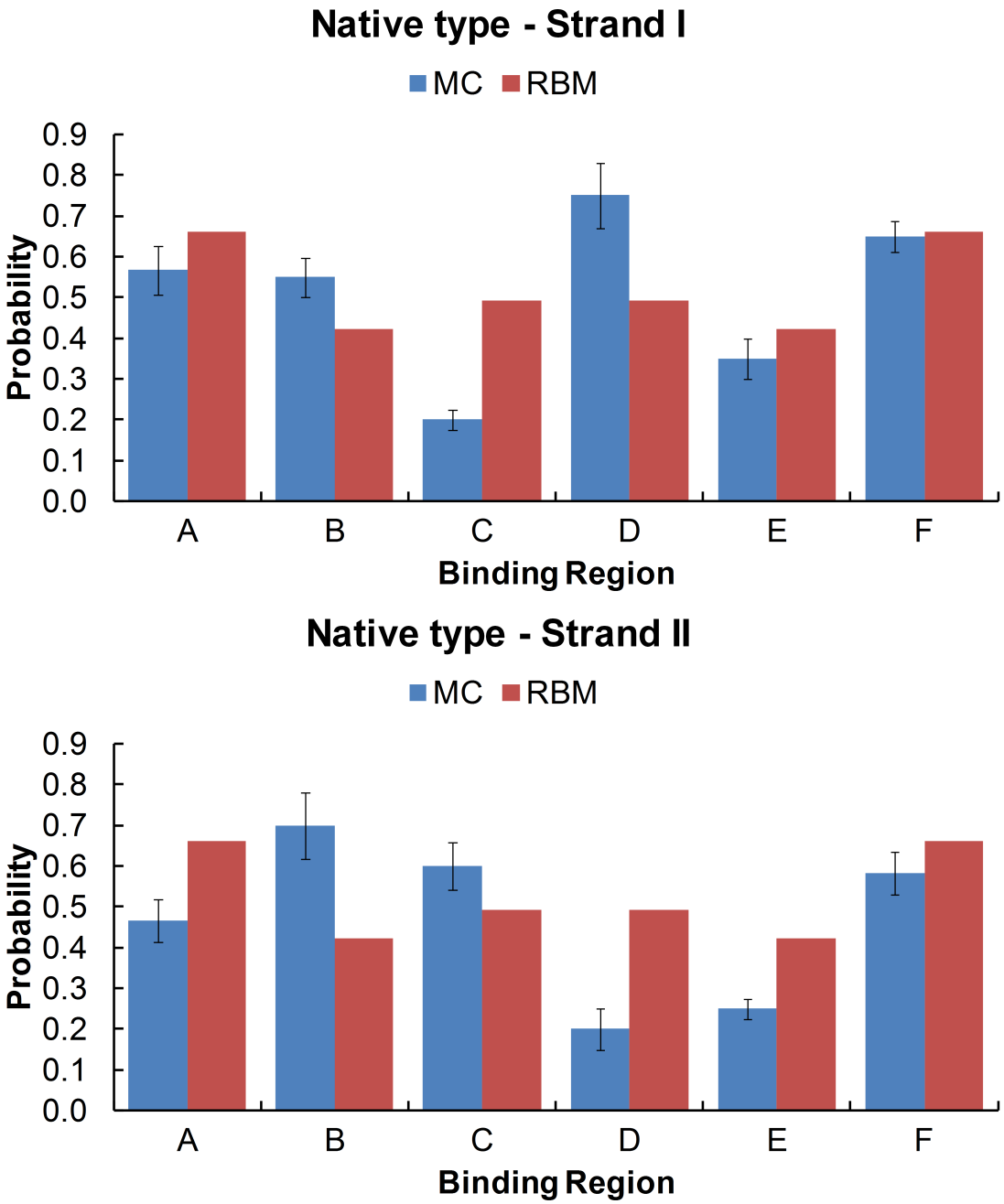

Figure 5.10: Binding probabilities for each binding region of the native Pen a 1 molecule at 0% reduced resolution for strand I (top) and strand II (bottom). Data from the Monte Carlo simulation (blue) and the optimized rule-based model (red) is shown. The letters A, B, C, D, E, and F represent the six binding regions used in our rule-based model.

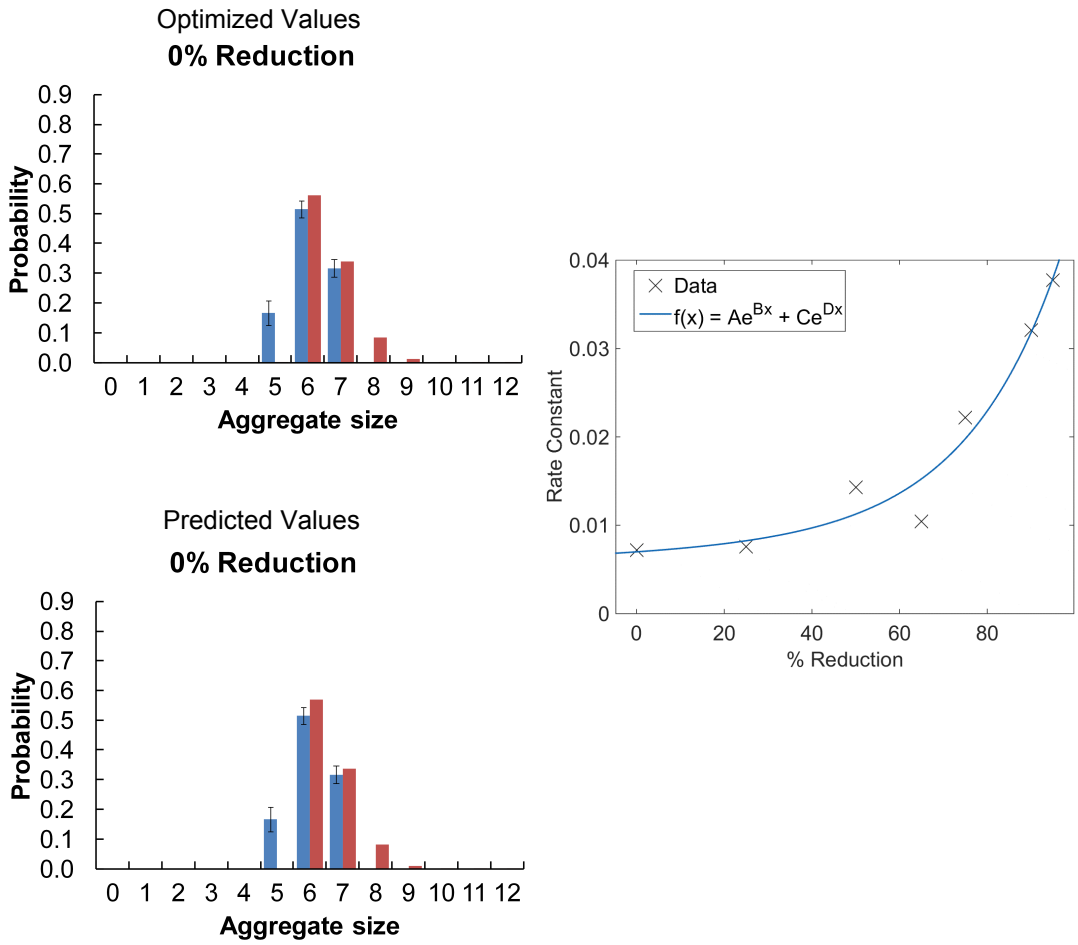

Figure 5.11: Comparison of Monte Carlo (blue) and rule-based model (red) aggregate size distributions for the 0% reduced resolution of the native Pen a 1 for the optimized rate constant (top left) and predicted rate constant (bottom left). Rate constant versus resolution data for the native Pen a 1 conformation (right). For these plots, the cutoff distance range has been fixed at 5.6-6.2 nm (see rule set 2 in the Appendices), which was found to be the optimal range for the 25% reduced resolution of the native type Pen a 1. The data points (X-shaped markers) were fitted to an exponential function (boxed equation) which is plotted as a solid line. The equation constants are A=0.0066, B=0.0027, C=0.0004, and D=0.0462.

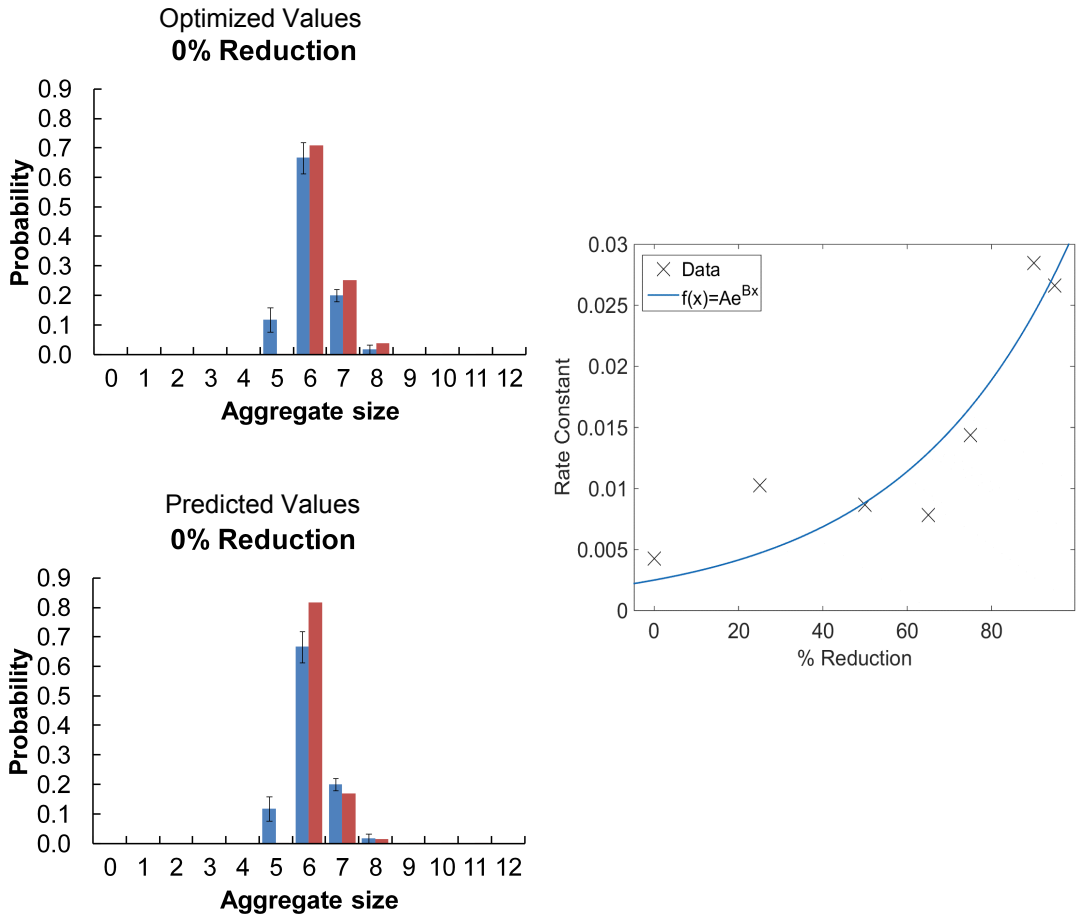

Figure 5.12: Comparison of Monte Carlo (blue) and rule-based model (red) aggregate size distributions for the 0% reduced resolution of the S-shaped Pen a 1 for the optimized rate constant (top left) and predicted rate constant (bottom left). Rate constant versus resolution data for the S-shaped Pen a 1 conformation (right). For these plots, the cutoff distance range has been fixed at 5.8-6.0 nm (see rule set 6 in the Appendices), which was found to be the optimal range for the 25% reduced resolution of the S-shaped type Pen a 1. The data points (X-shaped markers) were fitted to an exponential function (boxed equation) which is plotted as a solid line. The equation constants are A=0.0025 and B=0.0253.

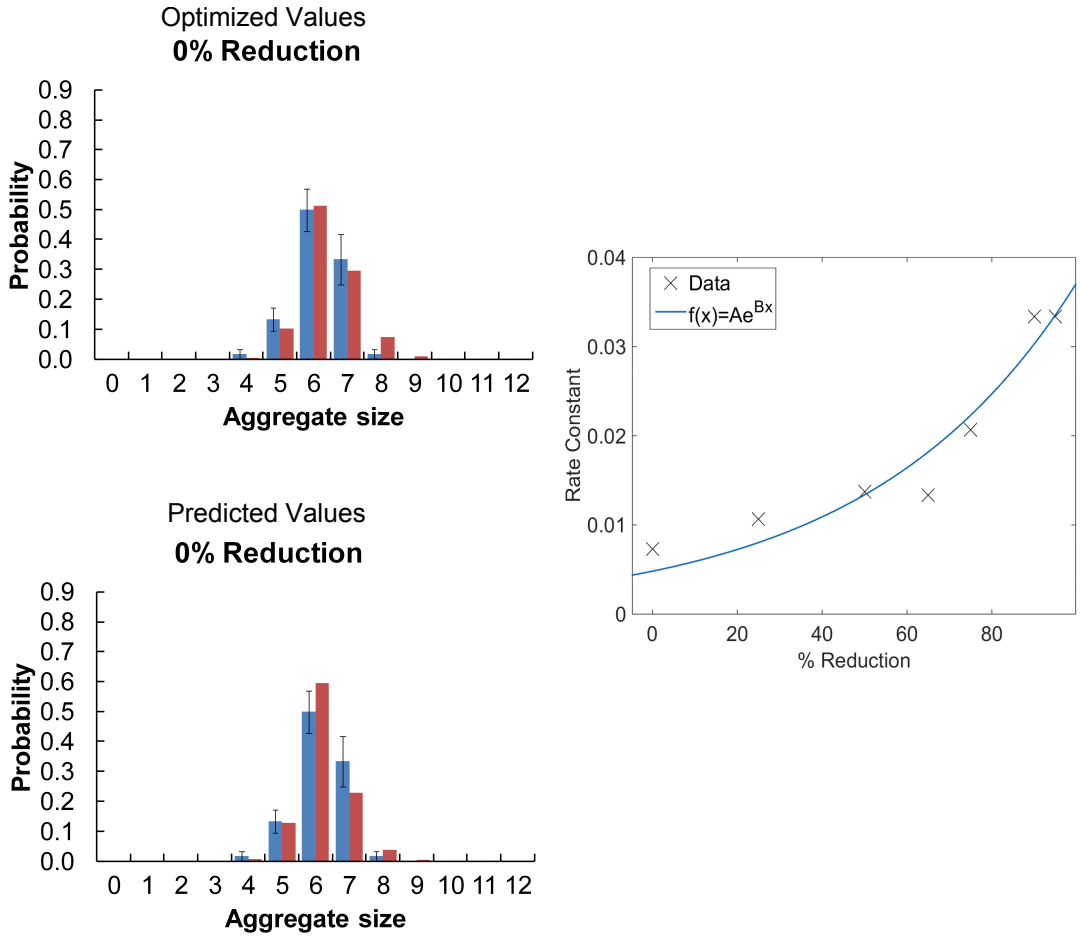

Figure 5.13: Comparison of Monte Carlo (blue) and rule-based model (red) aggregate size distributions for the 0% reduced resolution of the U-shaped Pen a 1 for the optimized rate constant (top left) and predicted rate constant (bottom left). Rate constant versus resolution data for the U-shaped Pen a 1 conformation (right). For these plots, the cutoff distance range has been fixed at 6.8-8.6 nm (see rule set 9 in the Appendices), which was found to be the optimal range for the 25% reduced resolution of the U-shaped type Pen a 1. The data points (X-shaped markers) were fitted to an exponential function (boxed equation) which is plotted as a solid line. The equation constants are A=0.0048 and B=0.0204.

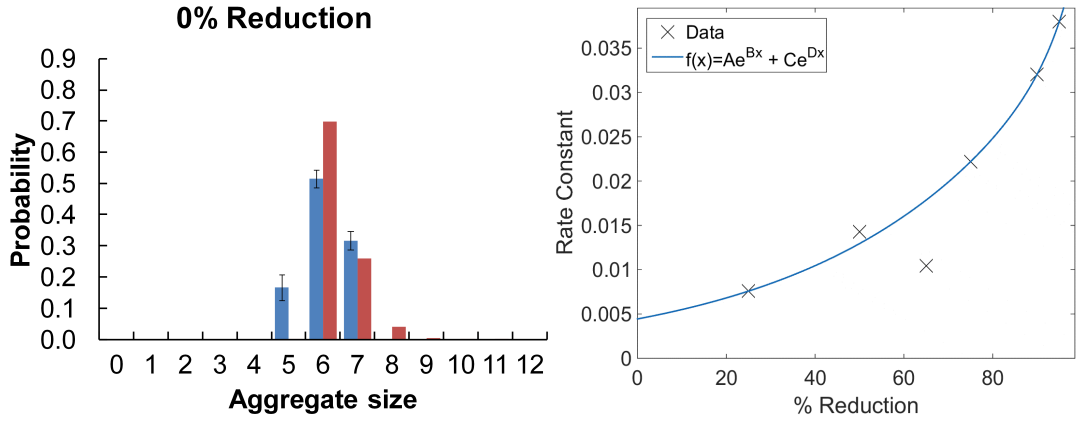

Figure 5.14: Comparison of Monte Carlo (blue) and rule-based model (red) aggregate size distributions for the 0% reduced resolution of the native Pen a 1 for the predicted rate constant (left). Rate constant versus resolution data for the native Pen a 1 conformation, omitting the 0% reduction data (right). For these plots, the cutoff distance range has been fixed at 5.6-6.2 nm (see rule set 2 in the Appendices), which was found to be the optimal range for the 25% reduced resolution of the native type Pen a 1. The data points (X-shaped markers) were fitted to an exponential function (boxed equation) which is plotted as a solid line. The equation constants are A=7.539e-11, B=0.1875, C=0.004443, and D=0.02138.

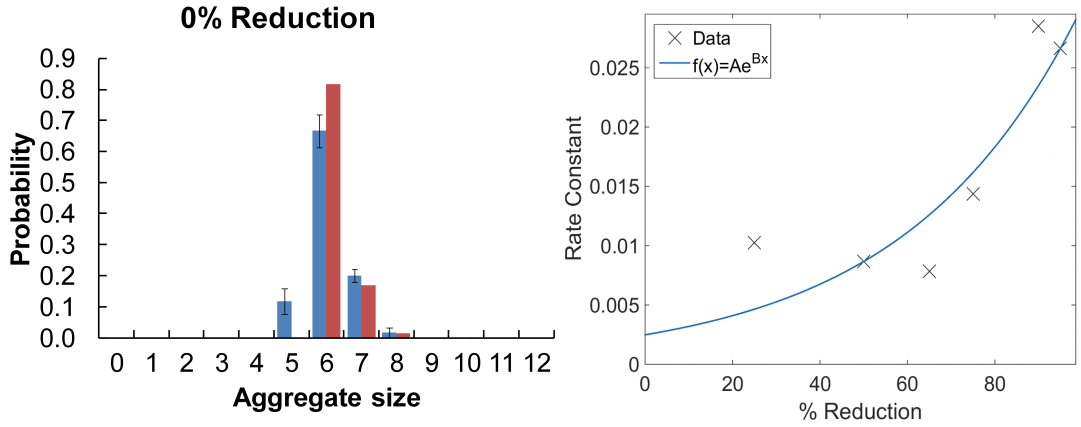

Figure 5.15: Comparison of Monte Carlo (blue) and rule-based model (red) aggregate size distributions for the 0% reduced resolution of the S-shaped Pen a 1 for the predicted rate constant (left). Rate constant versus resolution data for the S-shaped Pen a 1 conformation, omitting the 0% reduction data (right). For these plots, the cutoff distance range has been fixed at 5.8-6.0 nm (see rule set 6 in the Appendices), which was found to be the optimal range for the 25% reduced resolution of the S-shaped type Pen a 1. The data points (X-shaped markers) were fitted to an exponential function (boxed equation) which is plotted as a solid line. The equation constants are A=0.002491 and B=0.02494.

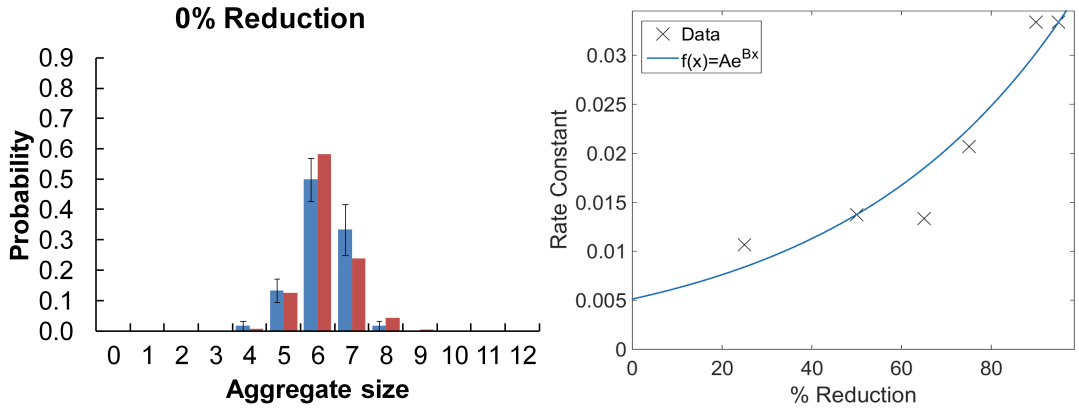

Figure 5.16: Comparison of Monte Carlo (blue) and rule-based model (red) aggregate size distributions for the 0% reduced resolution of the U-shaped Pen a 1 for the predicted rate constant (left). Rate constant versus resolution data for the U-shaped Pen a 1 conformation, omitting the 0% reduction data (right). For these plots, the cutoff distance range has been fixed at 6.8-8.6 nm (see rule set 9 in the Appendices), which was found to be the optimal range for the 25% reduced resolution of the U-shaped type Pen a 1. The data points (X-shaped markers) were fitted to an exponential function (boxed equation) which is plotted as a solid line. The equation constants are A=0.005139 and B=0.0197.

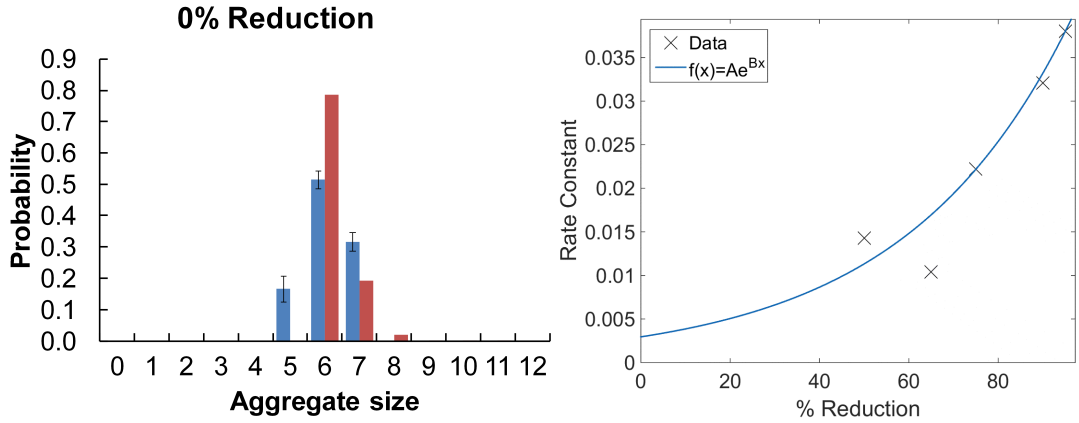

Figure 5.17: Comparison of Monte Carlo (blue) and rule-based model (red) aggregate size distributions for the 0% reduced resolution of the native Pen a 1 for the predicted rate constant (left). Rate constant versus resolution data for the native Pen a 1 conformation, omitting the 0% and 25% reduction data (right). For these plots, the cutoff distance range has been fixed at 5.6-6.2 nm (see rule set 2 in the Appendices), which was found to be the optimal range for the 25% reduced resolution of the native type Pen a 1. The data points (X-shaped markers) were fitted to an exponential function (boxed equation) which is plotted as a solid line. The equation constants are A=0.002952 and B=0.0269.

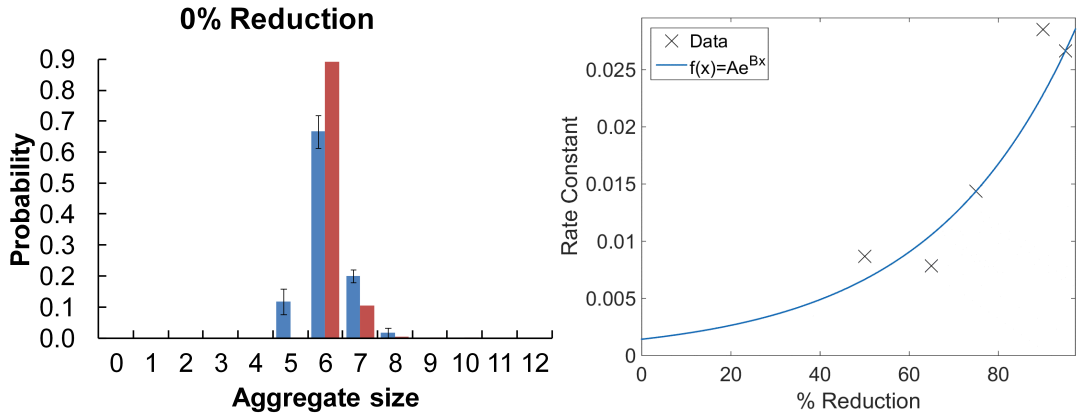

Figure 5.18: Comparison of Monte Carlo (blue) and rule-based model (red) aggregate size distributions for the 0% reduced resolution of the S-shaped Pen a 1 for the predicted rate constant (left). Rate constant versus resolution data for the S-shaped Pen a 1 conformation, omitting the 0% and 25% reduction data (right). For these plots, the cutoff distance range has been fixed at 5.8-6.0 nm (see rule set 6 in the Appendices), which was found to be the optimal range for the 25% reduced resolution of the S-shaped type Pen a 1. The data points (X-shaped markers) were fitted to an exponential function (boxed equation) which is plotted as a solid line. The equation constants are A=0.001426 and B=0.03081.

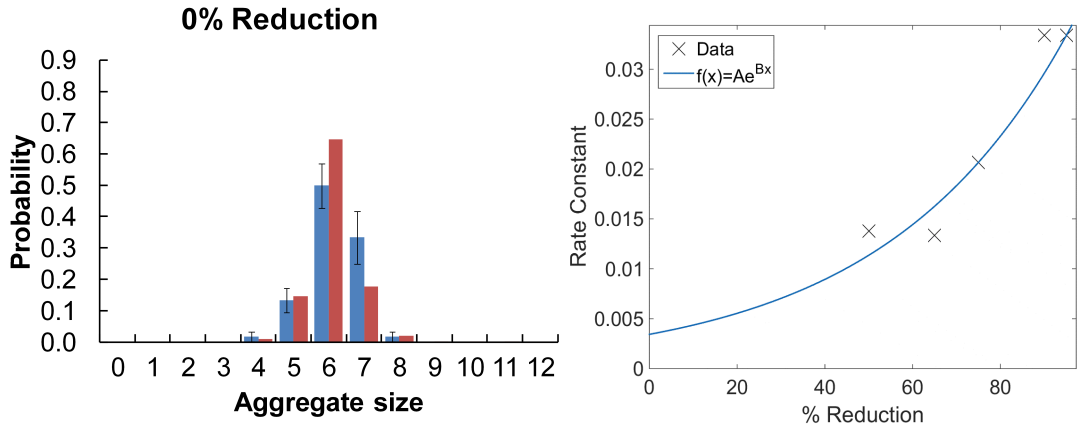

Figure 5.19: Comparison of Monte Carlo (blue) and rule-based model (red) aggregate size distributions for the 0% reduced resolution of the U-shaped Pen a 1 for the predicted rate constant (left). Rate constant versus resolution data for the U-shaped Pen a 1 conformation, omitting the 0% and 25% reduction data (right). For these plots, the cutoff distance range has been fixed at 6.8-8.6 nm (see rule set 9 in the Appendices), which was found to be the optimal range for the 25% reduced resolution of the U-shaped type Pen a 1. The data points (X-shaped markers) were fitted to an exponential function (boxed equation) which is plotted as a solid line. The equation constants are A=0.00343 and B=0.02395.

Chapter 6

Experimental Rule-Based Modeling

In this chapter, we apply a simple non-geometric rule-based model to analyze a set of data obtained from cell degranulation experiments for inclusion in this thesis. Experiment is less advanced than theory with regards to geometric rule-based modeling. We use a simple non-geometric method to model the experimental data in order to help us develop our intuition about antigen-antibody aggregate structure, which we apply to our development of a geometric rule-based method. It should also be noted that information about the conformations of the Pen a 1 molecules used in these experiments is not available, and that we use a model in which Pen a 1 has six binding sites. We justify this six-site model based on the most probable aggregate size that occurred in our 3D Monte Carlo simulations (see Section 5.2). In these experiments, the shrimp allergen Pen a 1 is used to initiate receptor cross-linking and stimulate degranulation. Receptor cross-linking refers to two or more antibody-receptor complexes binding to a single antigen molecule. Degranulation is the secretion of chemicals such as histamine from the granules (secretory vesicles) of a cell.

Currently, experimental data relating to aggregate size and structure, and the effect of the flexibility of the Pen a 1 molecule on aggregation, is not yet readily available. However, we can create a theoretical model that fits a set of cell degranulation data to gain insight into the relationship between the processes of degranulation and antigen-antibody aggregation. The specificity of the IgE antibodies (the epitopes on the Pen a 1 molecule to which the antibodies can bind), and the measure of aggregation that best corresponds to the experimental secretion data, are both unknown. In our modeling, we account for these unknowns by testing models with various Pen a 1 valencies (specifically, we tested models with six, eight, and ten epitopes per molecule) and measures of aggregation. We model the experimental dose-response data using a simple rule-based model written in the BNGL language and implemented in the stochastic simulator NFsim [Sneddon et al., 2011] in which free allergens are captured by receptors and receptors are cross-linked by Pen a 1 molecules. The "dose" in these experiments refers to the varying concentrations of Pen a 1, and the "response" refers to the percent secretion corresponding to each dose of Pen a 1. Only one cell was simulated for each experimental data point, as the computational requirements for simulating multiple cells are prohibitive.

6.1 Experimental Setup

Experiments were performed by our collaborators to determine the percent secretion of cells in response to the presence of Pen a 1. The percent secretion is defined as the amount of secretion observed divided by the total secretion possible and multiplied by 100. The total secretion possible is the total amount of secretory granule inside of the cells, which was measured by lysing the cells and measuring the total granule content. Secretion is measured by observing the number of secretory granules released by cells in response to cross-linking of cell receptors.

See Table 6.3 for the constant parameters used in the rule-based model. The experiments were conducted in wells each having 0.32 cm^2 of surface area. Initially, approximately 34,000 cells were plated the day prior to experiment. The number of cells per well at the time of experiment was measured in two cell count experiments; the average of these two values is 43,063 and is the number used in our calculations of the number of molecules per cell. The volume of media, which contained the allergen-specific IgE, per well was 100 μL.

See Table 6.1 and Table 6.2 for the quantities of Pen a 1 and IgE used in the rule-based model, respectively. For each experiment, allergen-specific IgE and Pen a 1 molecules were added at various concentrations. We look at experiments in which each well contains 15 ng/mL, 30 ng/mL, 60 ng/mL, and 120 ng/mL of tropomyosin-specific IgE, corresponding to 324, 716, 1211, and 2692 IgE per cell, respectively. The number of IgE per cell was obtained from quantifying experimental fluorescence data using flow cytometry. The specificity of the IgE antibodies used is unknown.

For each IgE concentration, the concentration of Pen a 1 molecules was varied from 0.0001 ng/mL to 10,000 ng/mL, and the percent secretion was measured at each order of magnitude of the Pen a 1 concentration. The number of copies of Pen a 1 molecules per cell was calculated for use in the rule-based model. This number ranges from two molecules for 0.0001 ng/mL Pen a 1 to 194256735 molecules for 10,000 ng/mL Pen a 1.

6.2 Rule-Based Modeling of Experimental Data

We created a simple rule-based model to fit to the experimental data. In this model, we assume that each Pen a 1 molecule is a dimer with either six, eight, or ten binding sites, with no steric effects. There are three rules: (1) a rule in which a free (unbound) Pen a 1 molecule binds to a receptor, (2) a rule in which two receptors

Table 6.1: Table showing the concentrations of Pen a 1 and the corresponding number of Pen a 1 per cell for the experimental rule-based model.

Pen a 1 concentration (ng/mL)	Number of Pen a 1 per cell
0.0001	2
0.001	19
0.01	194
0.1	1943
1	19426
10	194257
100	1942567
1000	19425673
10,000	194256735

Table 6.2: Table showing the concentrations of IgE and the corresponding number of IgE per cell for the experimental rule-based model.

IgE concentration (ng/mL)	Number of IgE per cell
15	324
30	716
60	1211
120	2692

are cross-linked by a Pen a 1 molecule, and (3) a rule in which an allergen-IgE bond dissociates (see Table 6.4).

The binding affinity k_{f1}, which is the rate associated with the first rule, and the cross-linking factor $xfactor$, which is related to the rate associated with the second rule such that $k_{x1} = xfactor \cdot k_{f1}$, are used as free parameters in fitting. It should be noted that the binding sites on a Pen a 1 molecule become more accessible once the molecule binds to a receptor, as the act of binding anchors the molecule close to the cell surface and increases the probability that other binding sites on the Pen a 1 molecule will bind to other receptors. This explains why k_{x1} must be greater than k_{f1}. The dissociation rate k_{r1} is fixed at $KD1 \cdot k_{f1}$ s^{-1}.

Table 6.3: Table showing constant parameters for the experimental rule-based model.

Number of cells per well	43063
Volume of media per well	100 μL
KD1	10^{-9} M

We used a combination of parameter scanning and GenFit [Thomas et al., 2015] to optimize the free parameters of the model in order to maximize the fit of the measure of aggregation to the experimental data. Then, we scaled $xfactor$ such that the maximum average aggregate size is consistent with that observed in experiment (no larger than 20). We first conducted a two-dimensional parameter scan, varying k_{f1} from 10^3 to 10^{12} $M^{-1}s^{-1}$, and varying $xfactor$ from 10^4 to 10^{13}. (The unit $M^{-1}s^{-1}$ is converted to $(molecule/cell)^{-1}s^{-1}$ in the rule-based model by dividing by Avogadro's number times the volume of extracellular fluid per cell.) The purpose of this scan was to identify which observable exhibited the best correlation with the experimental secretion data. We found that the average aggregate size fit the experimental data well. We define the *average aggregate size* as the total number of receptors in aggregates divided by the total number of aggregates, where an *aggregate* is defined as any cluster of bound molecules containing two or more receptors. We also found that the model in which one Pen a 1 molecule has six binding sites available to bind to IgE fits the data significantly better than the eight or ten site models (which are not shown). Since Pen a 1 has five main binding regions per strand, this implies that only three binding regions are being consistently bound. This may be due to the other two binding regions being less IgE-specific or steric effects causing blockage of other available binding sites. It should be noted that our study of Pen a 1 aggregate size distributions using a 3D rigid-body Monte Carlo method yielded six as the aggregate size with the greatest probability of occurrence at full model resolution for three energy-minimized conformations of Pen a 1 (see Section 5.2).

After parameter scanning, we optimized the binding rates further using the ge-

netic fitting program GenFit [Thomas et al., 2015]. The parameters used for GenFit are listed in Table 6.5. Fifty generations were tested, with 50 models tested per generation. Based on our initial parameter scan, k_{f1} and $xfactor$ were allowed to vary from 10^5 to 10^{10}. The average aggregate size was fitted to the experimental data curve corresponding to an IgE concentration of 120 ng/mL. Due to experimental constraints, the fitting parameter $xfactor$ was scaled down after the fitting procedure.

Table 6.4: Table showing the rules and associated rate constants for the experimental rule-based model assuming six binding sites per tropomyosin molecule. T represents a tropomyosin molecule, E represents an unbound binding site on the tropomyosin molecule, and IgE_E represents a binding site bound to an IgE antibody.

Rule Number	Reaction Rule	Binding Rate
(1)	T(E,E,E,E,E,E) + IgE $\rightarrow$ T(IgE_E,E,E,E,E,E)	k_{f1}
(2)	T(IgE_E,E) + IgE $\rightarrow$ T(IgE_E,IgE_E)	k_{x1}
(3)	T(IgE_E) $\rightarrow$ T(E) + IgE	k_{r1}

Table 6.5: Table showing the parameters for the genetic fitting program GenFit.

Generations	50
Models per generation	50
Mutation rate	0.2
Swap rate	0.5
Range of possible values	10^5-10^{10}

6.3 Results

The experimental data and the corresponding average aggregate size, defined as the number of receptors bound in aggregates divided by the total number of aggregates, produced by the optimized rule-based model are presented in Figure 6.1. The 95% confidence intervals of the optimized binding rates for this observable are $k_{f1} = 2.39303429 \times 10^9 \pm 0.05681216 \times 10^9$ $M^{-1}s^{-1}$ and $xfactor = 7.40627173 \times 10^3 \pm 0.12119755 \times 10^3$. These confidence intervals were calculated using the 20 best fits that were generated by a single run of GenFit.

The general shape of the dose-response curve, which increases as the allergen concentration increases until it peaks and then monotonically decreases for higher allergen concentrations, meets theoretical expectations [Sulzer and Perelson, 1996]. We expect that at low allergen concentrations, there will be small aggregates due to the low number of allergens per cell. As the allergen concentration rises, the average aggregate size will correspondingly increase until reaching a maximum at some concentration. At concentrations higher than this one, the large number of allergens cause enough of an increase in monovalent allergen-receptor binding that the aggregates will become smaller due to a reduced number of bound sites per allergen.

We observe that although the average aggregate size fits the experimental data well in general, there are significant differences between the two curves at Pen a 1 concentrations lower than 1 ng/mL and higher than 100 ng/mL. Possible reasons for this discrepancy include greater error in experimental measurements for the lowest and highest concentrations of Pen a 1, differences in cell signaling within a well of tens of thousands of cells versus a single cell, and the possibility of the Pen a 1 molecules being inhomogeneously distributed throughout the cells within a well.

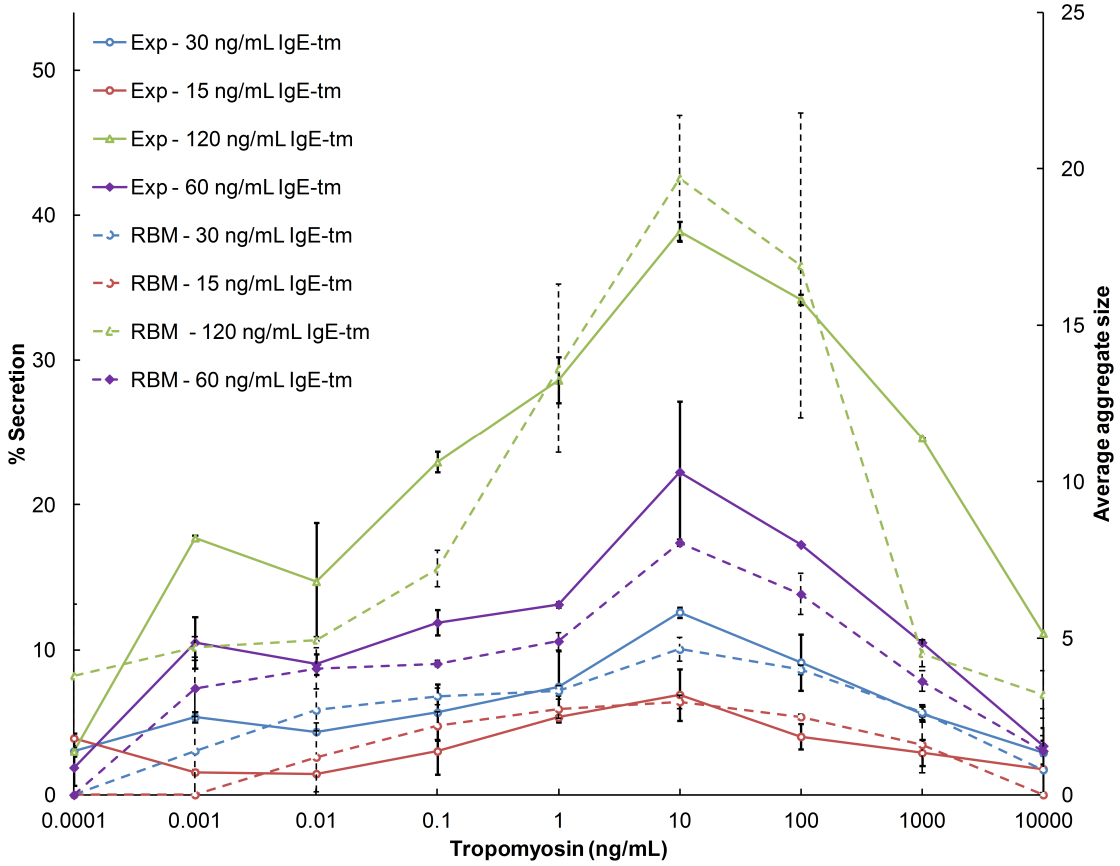

Figure 6.1: Average aggregate size generated by the rule-based model (dashed lines) plotted with the experimental percent secretion curves (solid lines) with the Pen a 1 concentration on the x-axis, for varying concentrations of allergen-specific IgE. The error bars for the model (dotted error bars) represent the standard deviation of 20 runs of NFsim. The error bars for the experimental data (solid error bars) represent the standard deviation for two experiments.

6.3.1 Related Observables

In addition to the average aggregate size, we analyze several other observables of the optimized rule-based model, using the same model (i.e. the same optimized rate constants) as that used to generate the average aggregate size data. Although these additional observables do not fit the experimental data as well as the average

aggregate size does, they improve our understanding of the model.

Two observables that are directly related to the average aggregate size are the number of aggregated receptors (see Figure 6.2) and the number of aggregates (see Figure 6.3). (The average aggregate size is defined as the number of aggregated receptors divided by the number of aggregates.) Other observables we study are the number of receptor crosslinks, defined as two receptors bound to the same Pen a 1 molecule (see Figure 6.4), and the number of bound receptor sites (see Figure 6.5).

We observe that the number of aggregated receptors experiences a sharp increase at 0.01 ng/mL of tropomyosin and increases until nearly saturating at 100 ng/mL of tropomyosin, at which point it begins to decrease. We also notice that a minimum number of aggregates occurs at 10 ng/mL of tropomyosin, which is the point at which the average aggregate size peaks. This indicates that the maximum of the average aggregate size occurring at 10 ng/mL is due to the number of aggregates reaching a minimum at this point combined with the number of aggregated receptors being near saturation.

We observe that the receptor crosslink curve undergoes a sharp increase at 0.1 ng/mL of tropomyosin, which is the same value at which the average aggregate size and the number of aggregated receptors both experience a sharp increase. This data indicates that the large increase in receptor crosslinking causes a large quantity of receptors to be bound within aggregates, allowing the formation of significantly larger aggregates starting at 0.1 ng/mL.

The number of bound receptor sites increases steadily as the concentration of tropomyosin increases. This data, combined with the decrease in receptor crosslinking and the decrease in aggregated receptors at high concentrations of tropomyosin, indicates that monovalent allergen-receptor binding becomes prevalent at high concentrations of tropomyosin.

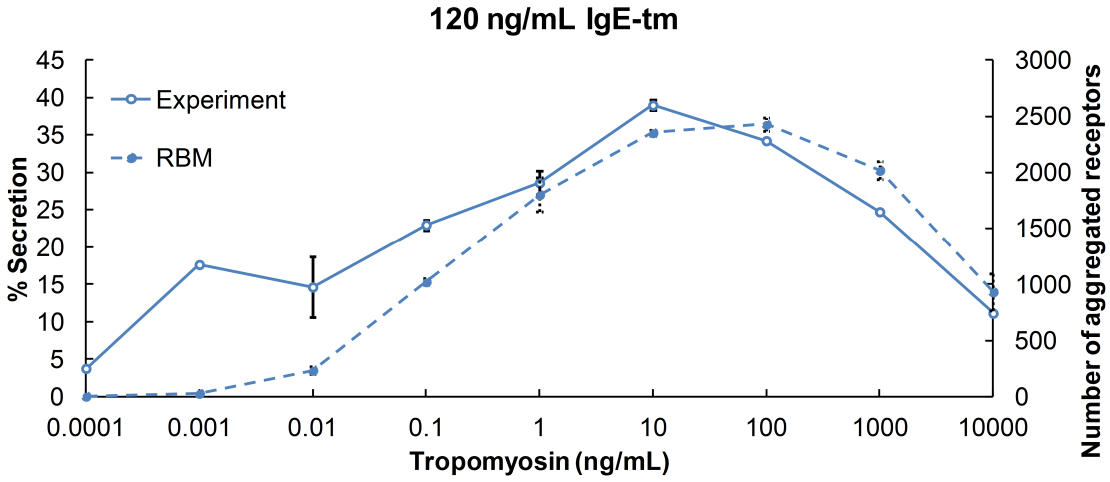

Figure 6.2: Number of aggregated receptors generated by the rule-based model (dashed line) plotted with the experimental percent secretion curve (solid line) with the Pen a 1 concentration on the x-axis, for 120 ng/mL allergen-specific IgE. The error bars for the model (dotted error bars) represent the standard deviation of 20 runs of NFsim. The error bars for the experimental data (solid error bars) represent the standard deviation for two experiments.

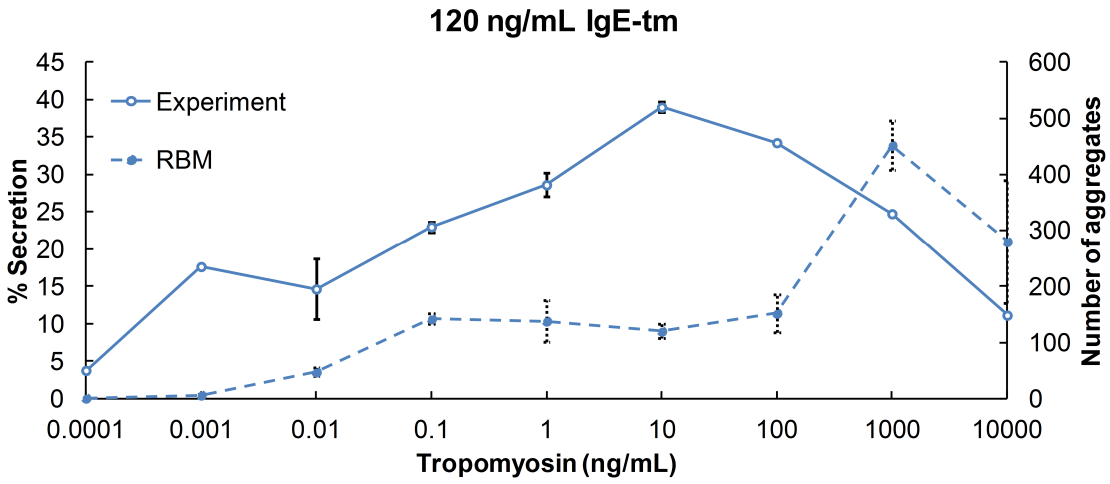

Figure 6.3: Number of aggregates generated by the rule-based model (dashed line) plotted with the experimental percent secretion curve (solid line) with the Pen a 1 concentration on the x-axis, for 120 ng/mL allergen-specific IgE. The error bars for the model (dotted error bars) represent the standard deviation of 20 runs of NFsim. The error bars for the experimental data (solid error bars) represent the standard deviation for two experiments.

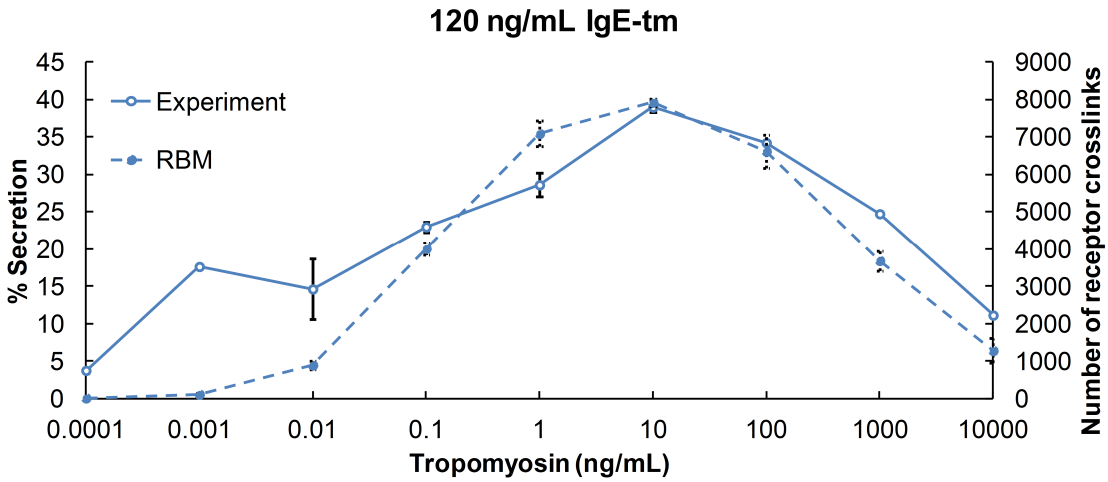

Figure 6.4: Number of receptor crosslinks generated by the rule-based model (dashed line) plotted with the experimental percent secretion curve (solid line) with the Pen a 1 concentration on the x-axis, for 120 ng/mL allergen-specific IgE. The error bars for the model (dotted error bars) represent the standard deviation of 20 runs of NFsim. The error bars for the experimental data (solid error bars) represent the standard deviation for two experiments.

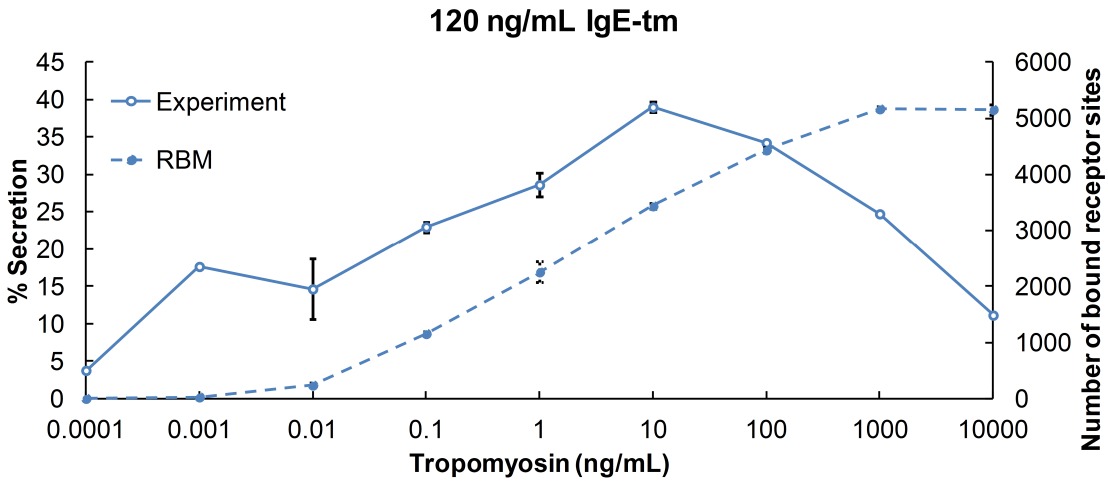

Figure 6.5: Number of bound receptor sites generated by the rule-based model (dashed line) plotted with the experimental percent secretion curve (solid line) with the Pen a 1 concentration on the x-axis, for 120 ng/mL allergen-specific IgE. The error bars for the model (dotted error bars) represent the standard deviation of 20 runs of NFsim. The error bars for the experimental data (solid error bars) represent the standard deviation for two experiments.

Chapter 7

Conclusions

We employed a 3D rigid-body Monte Carlo method inspired by robotic motions combined with a novel geometric rule-based method to simulate IgE antibody aggregation with the shrimp allergen Pen a 1. The main contributions of this work are: the development of a rule-based method that encodes steric effects between neighboring binding regions, the application of this rule-based method to the quantification of steric effects for various Monte Carlo model resolutions and various geometries of Pen a 1, and the application of this method to the prediction of results for high-resolution Monte Carlo models using only lower-resolution Monte Carlo simulations. In addition, we employed a more traditional non-geometric rule-based method to the modeling of experimental cell secretion data, using our Monte Carlo results to justify the number of binding sites in this model; we have obtained preliminary results using the average aggregate size as the main observable.

We first developed our geometric rule-based method, using three U-shaped geometries of Pen a 1 with varying degrees of curvature in order to demonstrate how our method captures the steric effects induced by curvature. We then performed extended analysis of the Monte Carlo and rule-based modeling results for three dif-

ferent energy-minimized conformations of the allergen Pen a 1: native, S-shaped, and U-shaped. For each conformation, we performed Monte Carlo simulations at seven different model resolutions, using polygon reduction of the isosurface molecular models to reduce the resolution. For each resolution, we then constructed a rule-based model of antibody aggregation using the cutoff distance and the distances between the binding regions on each strand of Pen a 1 to determine the rule set. The distances between binding regions vary with the Pen a 1 conformation. We then optimized the cutoff distance and the variable forward rate constant for each rule-based model. Additionally, in order to create a model that we can use to more easily predict the aggregate size distribution at higher resolutions, we fixed the cutoff distance as the optimal cutoff distance found for the 25% reduced Monte Carlo resolution and performed another optimization of the rate constant for each resolution. We then fitted the rate constant versus resolution data to a function.

We have shown that there are significant differences in the plots of optimal cutoff distance versus resolution for each of the three conformations. For example, the location and slope of the "dip" in the plot vary with conformation. This finding indicates that molecular geometry should perhaps play a greater role in biomolecular rule-based modeling so that important geometric effects can be captured in the model results.

We have also shown that there is a downward trend in the optimal cutoff distance with a decrease in resolution for all three of the Pen a 1 conformations. From this finding, we can conclude that the reduction in volume that results from polygon reduction of the 3D molecular models results in increased binding site accessibility. We also observed that some resolutions of the same conformation have identical optimal cutoff distances. From this data, we can potentially extract useful information about similarities between different resolutions. This may allow us to run a lower resolution model with reduced computational time rather than a higher resolution model if the

two resolutions are similar enough to each other.

In future research, we will use experimental electron microscopy (EM) data in place of the aggregation data from our Monte Carlo simulations. Cryo-EM data, which provides information about molecular geometry, is now becoming available. The locations of receptor probes in EM images can be used to estimate receptor clustering. We will also consider the effects of allergen cross-linking on our novel rule-based method.

One concern with our rate constant model is that the model may change significantly based on the number of Monte Carlo runs. We used 60 runs for this study, which is a fairly small number. We may want to run hundreds of runs, in which case some or all of the data points in the model may be significantly different than the data points obtained using 60 runs, which will change the fitted function to some degree. Future work will involve quantifying the possible error introduced into the model by differences in the number of runs.

Appendices

Appendix A

Rule Sets for Optimized Rule-Based Models

Table A.1: Rule Set (1) for Strand I (T_I) of native Pen a 1 with a cutoff distance range of 7.0-8.7 nm. 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 → 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) + 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(B,C,D) + IgE → T_I(B,IgE$_C$,D)	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(IgE$_B$,C,IgE$_D$) + IgE → T_I(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}

Table A.2: Rule Set (1) for Strand II (T_{II}) of native Pen a 1 with a cutoff distance range of 7.0-8.7 nm. 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 A.3: Rule Set (2) for Strand I (T_I) of native Pen a 1 with a cutoff distance range of 5.6-6.2 nm. 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 → 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) + 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(B,C,D) + IgE → T_I(B,IgE$_C$,D)	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(IgE$_B$,C,IgE$_D$) + IgE → T_I(IgE$_B$,IgE$_C$,IgE$_D$)	k_{f2}
D	T_I(C,D) + IgE → T_I(C,IgE$_D$)	k_{f1}
	T_I(IgE$_C$,D) + IgE → T_I(IgE$_C$,IgE$_D$)	k_{f2}
E	T_I(E,F) + IgE → T_I(IgE$_E$,F)	k_{f1}
	T_I(E,IgE$_F$) + IgE → T_I(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}

Table A.4: Rule Set (2) for Strand II (T_{II}) of native Pen a 1 with a cutoff distance range of 5.6-6.2 nm. 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 $\rightarrow$ T_{II}(IgE$_A$,B)	k_{f1}
	T_{II}(A,IgE$_B$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$)	k_{f2}
B	T_{II}(A,B,C) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,C)	k_{f1}
	T_{II}(IgE$_A$,B,C) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,C)	k_{f2}
	T_{II}(A,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,IgE$_C$)	k_{f2}
	T_{II}(IgE$_A$,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,IgE$_C$)	k_{f2}
C	T_{II}(B,C,D) + IgE $\rightarrow$ T_{II}(B,IgE$_C$,D)	k_{f1}
	T_{II}(IgE$_B$,C,D) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$,D)	k_{f2}
	T_{II}(B,C,IgE$_D$) + IgE $\rightarrow$ T_{II}(B,IgE$_C$,IgE$_D$)	k_{f2}
	T_{II}(IgE$_B$,C,IgE$_D$) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$,IgE$_D$)	k_{f2}
D	T_{II}(C,D) + IgE $\rightarrow$ T_{II}(C,IgE$_D$)	k_{f1}
	T_{II}(IgE$_C$,D) + IgE $\rightarrow$ T_{II}(IgE$_C$,IgE$_D$)	k_{f2}
E	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,F)	k_{f1}
	T_{II}(E,IgE$_F$) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}
F	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(E,IgE$_F$)	k_{f1}
	T_{II}(IgE$_E$,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}

Table A.5: Rule Set (3) for Strand I (T_I) of native Pen a 1 with a cutoff distance range of 5.5 nm. 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 → 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) + 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(B,C,D) + IgE → T_I(B,IgE$_C$,D)	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(IgE$_B$,C,IgE$_D$) + IgE → T_I(IgE$_B$,IgE$_C$,IgE$_D$)	k_{f2}
D	T_I(C,D) + IgE → T_I(C,IgE$_D$)	k_{f1}
	T_I(IgE$_C$,D) + IgE → T_I(IgE$_C$,IgE$_D$)	k_{f2}
E	T_I(E,F) + IgE → T_I(IgE$_E$,F)	k_{f1}
	T_I(E,IgE$_F$) + IgE → T_I(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}

Table A.6: Rule Set (3) for Strand II (T_{II}) of native Pen a 1 with a cutoff distance range of 5.5 nm. 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 $\rightarrow$ T_{II}(IgE$_A$,B)	k_{f1}
	T_{II}(A,IgE$_B$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$)	k_{f2}
B	T_{II}(A,B,C) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,C)	k_{f1}
	T_{II}(IgE$_A$,B,C) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,C)	k_{f2}
	T_{II}(A,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,IgE$_C$)	k_{f2}
	T_{II}(IgE$_A$,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,IgE$_C$)	k_{f2}
C	T_{II}(B,C) + IgE $\rightarrow$ T_{II}(B,IgE$_C$)	k_{f1}
	T_{II}(IgE$_B$,C) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$)	k_{f2}
D	T_{II}(D) + IgE $\rightarrow$ T_{II}(IgE$_D$)	k_{f1}
E	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,F)	k_{f1}
	T_{II}(E,IgE$_F$) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}
F	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(E,IgE$_F$)	k_{f1}
	T_{II}(IgE$_E$,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}

Table A.7: Rule Set (4) for Strand I (T_I) of native Pen a 1 with a cutoff distance range of 4.0-5.4 nm. 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) + IgE $\rightarrow$ T_I(B,IgE$_C$)	k_{f1}
	T_I(IgE$_B$,C) + IgE $\rightarrow$ T_I(IgE$_B$,IgE$_C$)	k_{f2}
D	T_I(D) + IgE $\rightarrow$ T_I(IgE$_D$)	k_{f1}
E	T_I(E,F) + IgE $\rightarrow$ T_I(IgE$_E$,F)	k_{f1}
	T_I(E,IgE$_F$) + IgE $\rightarrow$ T_I(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 A.8: Rule Set (4) for Strand II (T_{II}) of native Pen a 1 with a cutoff distance range of 4.0-5.4 nm. 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 $\rightarrow$ T_{II}(IgE$_A$,B)	k_{f1}
	T_{II}(A,IgE$_B$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$)	k_{f2}
B	T_{II}(A,B,C) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,C)	k_{f1}
	T_{II}(IgE$_A$,B,C) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,C)	k_{f2}
	T_{II}(A,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,IgE$_C$)	k_{f2}
	T_{II}(IgE$_A$,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,IgE$_C$)	k_{f2}
C	T_{II}(B,C) + IgE $\rightarrow$ T_{II}(B,IgE$_C$)	k_{f1}
	T_{II}(IgE$_B$,C) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$)	k_{f2}
D	T_{II}(D) + IgE $\rightarrow$ T_{II}(IgE$_D$)	k_{f1}
E	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,F)	k_{f1}
	T_{II}(E,IgE$_F$) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}
F	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(E,IgE$_F$)	k_{f1}
	T_{II}(IgE$_E$,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}

Table A.9: Rule Set (5) for Strand I (T_I) of S-shaped Pen a 1 with a cutoff distance range of 6.8-8.3 nm. 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 → 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) + 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(B,C,D) + IgE → T_I(B,IgE$_C$,D)	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(IgE$_B$,C,IgE$_D$) + IgE → T_I(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}

Table A.10: Rule Set (5) for Strand II (T_{II}) of S-shaped Pen a 1 with a cutoff distance range of 6.8-8.3 nm. 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 A.11: Rule Set (6) for Strand I (T_I) of S-shaped Pen a 1 with a cutoff distance range of 5.8-6.0 nm. 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 → 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) + 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(B,C,D) + IgE → T_I(B,IgE$_C$,D)	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(IgE$_B$,C,IgE$_D$) + IgE → T_I(IgE$_B$,IgE$_C$,IgE$_D$)	k_{f2}
D	T_I(C,D) + IgE → T_I(C,IgE$_D$)	k_{f1}
	T_I(IgE$_C$,D) + IgE → T_I(IgE$_C$,IgE$_D$)	k_{f2}
E	T_I(E,F) + IgE → T_I(IgE$_E$,F)	k_{f1}
	T_I(E,IgE$_F$) + IgE → T_I(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}

Table A.12: Rule Set (6) for Strand II (T_{II}) of S-shaped Pen a 1 with a cutoff distance range of 5.8-6.0 nm. 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 $\rightarrow$ T_{II}(IgE$_A$,B)	k_{f1}
	T_{II}(A,IgE$_B$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$)	k_{f2}
B	T_{II}(A,B,C) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,C)	k_{f1}
	T_{II}(IgE$_A$,B,C) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,C)	k_{f2}
	T_{II}(A,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,IgE$_C$)	k_{f2}
	T_{II}(IgE$_A$,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,IgE$_C$)	k_{f2}
C	T_{II}(B,C,D) + IgE $\rightarrow$ T_{II}(B,IgE$_C$,D)	k_{f1}
	T_{II}(IgE$_B$,C,D) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$,D)	k_{f2}
	T_{II}(B,C,IgE$_D$) + IgE $\rightarrow$ T_{II}(B,IgE$_C$,IgE$_D$)	k_{f2}
	T_{II}(IgE$_B$,C,IgE$_D$) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$,IgE$_D$)	k_{f2}
D	T_{II}(C,D) + IgE $\rightarrow$ T_{II}(C,IgE$_D$)	k_{f1}
	T_{II}(IgE$_C$,D) + IgE $\rightarrow$ T_{II}(IgE$_C$,IgE$_D$)	k_{f2}
E	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,F)	k_{f1}
	T_{II}(E,IgE$_F$) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}
F	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(E,IgE$_F$)	k_{f1}
	T_{II}(IgE$_E$,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}

Table A.13: Rule Set (7) for Strand I (T_I) of S-shaped Pen a 1 with a cutoff distance range of 5.4-5.7 nm. 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) + IgE $\rightarrow$ T_I(B,IgE$_C$)	k_{f1}
	T_I(IgE$_B$,C) + IgE $\rightarrow$ T_I(IgE$_B$,IgE$_C$)	k_{f2}
D	T_I(D) + IgE $\rightarrow$ T_I(IgE$_D$)	k_{f1}
E	T_I(E,F) + IgE $\rightarrow$ T_I(IgE$_E$,F)	k_{f1}
	T_I(E,IgE$_F$) + IgE $\rightarrow$ T_I(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 A.14: Rule Set (7) for Strand II (T_{II}) of S-shaped Pen a 1 with a cutoff distance range of 5.4-5.7 nm. 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 $\rightarrow$ T_{II}(IgE$_A$,B)	k_{f1}
	T_{II}(A,IgE$_B$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$)	k_{f2}
B	T_{II}(A,B,C) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,C)	k_{f1}
	T_{II}(IgE$_A$,B,C) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,C)	k_{f2}
	T_{II}(A,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,IgE$_C$)	k_{f2}
	T_{II}(IgE$_A$,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,IgE$_C$)	k_{f2}
C	T_{II}(B,C,D) + IgE $\rightarrow$ T_{II}(B,IgE$_C$,D)	k_{f1}
	T_{II}(IgE$_B$,C,D) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$,D)	k_{f2}
	T_{II}(B,C,IgE$_D$) + IgE $\rightarrow$ T_{II}(B,IgE$_C$,IgE$_D$)	k_{f2}
	T_{II}(IgE$_B$,C,IgE$_D$) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$,IgE$_D$)	k_{f2}
D	T_{II}(C,D) + IgE $\rightarrow$ T_{II}(C,IgE$_D$)	k_{f1}
	T_{II}(IgE$_C$,D) + IgE $\rightarrow$ T_{II}(IgE$_C$,IgE$_D$)	k_{f2}
E	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,F)	k_{f1}
	T_{II}(E,IgE$_F$) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}
F	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(E,IgE$_F$)	k_{f1}
	T_{II}(IgE$_E$,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}

Table A.15: Rule Set (8) for Strand I (T_I) of S-shaped Pen a 1 with a cutoff distance range of 4.5-5.3 nm. 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) + IgE $\rightarrow$ T_I(B,IgE$_C$)	k_{f1}
	T_I(IgE$_B$,C) + IgE $\rightarrow$ T_I(IgE$_B$,IgE$_C$)	k_{f2}
D	T_I(D) + IgE $\rightarrow$ T_I(IgE$_D$)	k_{f1}
E	T_I(E,F) + IgE $\rightarrow$ T_I(IgE$_E$,F)	k_{f1}
	T_I(E,IgE$_F$) + IgE $\rightarrow$ T_I(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 A.16: Rule Set (8) for Strand II (T_{II}) of S-shaped Pen a 1 with a cutoff distance range of 4.5-5.3 nm. 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 $\rightarrow$ T_{II}(IgE$_A$,B)	k_{f1}
	T_{II}(A,IgE$_B$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$)	k_{f2}
B	T_{II}(A,B,C) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,C)	k_{f1}
	T_{II}(IgE$_A$,B,C) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,C)	k_{f2}
	T_{II}(A,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,IgE$_C$)	k_{f2}
	T_{II}(IgE$_A$,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,IgE$_C$)	k_{f2}
C	T_{II}(B,C) + IgE $\rightarrow$ T_{II}(B,IgE$_C$)	k_{f1}
	T_{II}(IgE$_B$,C) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$)	k_{f2}
D	T_{II}(D) + IgE $\rightarrow$ T_{II}(IgE$_D$)	k_{f1}
E	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,F)	k_{f1}
	T_{II}(E,IgE$_F$) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}
F	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(E,IgE$_F$)	k_{f1}
	T_{II}(IgE$_E$,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}

Table A.17: Rule Set (9) for Strand I (T_I) of U-shaped Pen a 1 with a cutoff distance range of 6.8-8.6 nm. 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 → 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) + 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(B,C,D) + IgE → T_I(B,IgE$_C$,D)	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(IgE$_B$,C,IgE$_D$) + IgE → T_I(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}

Table A.18: Rule Set (9) for Strand II (T_{II}) of U-shaped Pen a 1 with a cutoff distance range of 6.8-8.6 nm. 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 A.19: Rule Set (10) for Strand I (T_I) of U-shaped Pen a 1 with a cutoff distance range of 5.3 nm. 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) + IgE $\rightarrow$ T_I(B,IgE$_C$)	k_{f1}
	T_I(IgE$_B$,C) + IgE $\rightarrow$ T_I(IgE$_B$,IgE$_C$)	k_{f2}
D	T_I(D) + IgE $\rightarrow$ T_I(IgE$_D$)	k_{f1}
E	T_I(E,F) + IgE $\rightarrow$ T_I(IgE$_E$,F)	k_{f1}
	T_I(E,IgE$_F$) + IgE $\rightarrow$ T_I(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 A.20: Rule Set (10) for Strand II (T_{II}) of U-shaped Pen a 1 with a cutoff distance range of 5.3 nm. 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 $\rightarrow$ T_{II}(IgE$_A$,B)	k$_{f1}$
	T_{II}(A,IgE$_B$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$)	k$_{f2}$
B	T_{II}(A,B,C) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,C)	k$_{f1}$
	T_{II}(IgE$_A$,B,C) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,C)	k$_{f2}$
	T_{II}(A,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,IgE$_C$)	k$_{f2}$
	T_{II}(IgE$_A$,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,IgE$_C$)	k$_{f2}$
C	T_{II}(B,C,D) + IgE $\rightarrow$ T_{II}(B,IgE$_C$,D)	k$_{f1}$
	T_{II}(IgE$_B$,C,D) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$,D)	k$_{f2}$
	T_{II}(B,C,IgE$_D$) + IgE $\rightarrow$ T_{II}(B,IgE$_C$,IgE$_D$)	k$_{f2}$
	T_{II}(IgE$_B$,C,IgE$_D$) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$,IgE$_D$)	k$_{f2}$
D	T_{II}(C,D) + IgE $\rightarrow$ T_{II}(C,IgE$_D$)	k$_{f1}$
	T_{II}(IgE$_C$,D) + IgE $\rightarrow$ T_{II}(IgE$_C$,IgE$_D$)	k$_{f2}$
E	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,F)	k$_{f1}$
	T_{II}(E,IgE$_F$) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k$_{f2}$
F	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(E,IgE$_F$)	k$_{f1}$
	T_{II}(IgE$_E$,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k$_{f2}$

Table A.21: Rule Set (11) for Strand I (T_I) of U-shaped Pen a 1 with a cutoff distance range of 4.1-5.2 nm. 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) + IgE $\rightarrow$ T_I(B,IgE$_C$)	k_{f1}
	T_I(IgE$_B$,C) + IgE $\rightarrow$ T_I(IgE$_B$,IgE$_C$)	k_{f2}
D	T_I(D) + IgE $\rightarrow$ T_I(IgE$_D$)	k_{f1}
E	T_I(E,F) + IgE $\rightarrow$ T_I(IgE$_E$,F)	k_{f1}
	T_I(E,IgE$_F$) + IgE $\rightarrow$ T_I(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 A.22: Rule Set (11) for Strand II (T_{II}) of U-shaped Pen a 1 with a cutoff distance range of 4.1-5.2 nm. 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 $\rightarrow$ T_{II}(IgE$_A$,B)	k_{f1}
	T_{II}(A,IgE$_B$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$)	k_{f2}
B	T_{II}(A,B,C) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,C)	k_{f1}
	T_{II}(IgE$_A$,B,C) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,C)	k_{f2}
	T_{II}(A,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(A,IgE$_B$,IgE$_C$)	k_{f2}
	T_{II}(IgE$_A$,B,IgE$_C$) + IgE $\rightarrow$ T_{II}(IgE$_A$,IgE$_B$,IgE$_C$)	k_{f2}
C	T_{II}(B,C) + IgE $\rightarrow$ T_{II}(B,IgE$_C$)	k_{f1}
	T_{II}(IgE$_B$,C) + IgE $\rightarrow$ T_{II}(IgE$_B$,IgE$_C$)	k_{f2}
D	T_{II}(D) + IgE $\rightarrow$ T_{II}(IgE$_D$)	k_{f1}
E	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,F)	k_{f1}
	T_{II}(E,IgE$_F$) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}
F	T_{II}(E,F) + IgE $\rightarrow$ T_{II}(E,IgE$_F$)	k_{f1}
	T_{II}(IgE$_E$,F) + IgE $\rightarrow$ T_{II}(IgE$_E$,IgE$_F$)	k_{f2}

Table A.23: Rule Set (12) for Strand I (T_I) of U-shaped Pen a 1 with a cutoff distance range of 3.9-4.0 nm. 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 → 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) + 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(B,C) + IgE → T_I(B,IgE$_C$)	k_{f1}
	T_I(IgE$_B$,C) + IgE → T_I(IgE$_B$,IgE$_C$)	k_{f2}
D	T_I(D) + IgE → T_I(IgE$_D$)	k_{f1}
E	T_I(E,F) + IgE → T_I(IgE$_E$,F)	k_{f1}
	T_I(E,IgE$_F$) + IgE → T_I(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}

Table A.24: Rule Set (12) for Strand II (T_{II}) of U-shaped Pen a 1 with a cutoff distance range of 3.9-4.0 nm. 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) + IgE → T_{II}(A,IgE$_B$)	k_{f1}
	T_{II}(IgE$_A$,B) + IgE → T_{II}(IgE$_A$,IgE$_B$)	k_{f2}
C	T_{II}(C) + IgE → T_{II}(IgE$_C$)	k_{f1}
D	T_{II}(D) + IgE → T_{II}(IgE$_D$)	k_{f1}
E	T_{II}(E,F) + IgE → T_{II}(IgE$_E$,F)	k_{f1}
	T_{II}(E,IgE$_F$) + IgE → T_{II}(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}

Appendix B

Additional Rule Sets for Curvature Demonstration

Table B.1: 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}

Table B.2: 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}

Binding Site	Reaction Rule	Binding Rate
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}

References

[Andrews et al., 2009] Andrews, N. L., Pfeiffer, J. R., Martinez, A. M., Haaland, D. M., Davis, R. W., Kawakami, T., Oliver, J. M., Wilson, B. S., and Lidke, D. S. (2009). Small, mobile FcϵRI receptor aggregates are signaling competent. *Immunity*, 31(3):469–479.

[Ayuso et al., 2002] Ayuso, R., Lehrer, S., and Reese, G. (2002). Identification of continuous, allergenic regions of the major shrimp allergen Pen a 1 (tropomyosin). *International Archives of Allergy and Immunology*, 127(1):27–37.

[Blinov et al., 2004] Blinov, M. L., Faeder, J. R., Goldstein, B., and Hlavacek, W. S. (2004). BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20(17):3289–3291.

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

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

[Danos and Laneve, 2004] Danos, V. and Laneve, C. (2004). Formal molecular biology. *Theoretical Computer Science*, 325(1):69–110.

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

[Goddard et al., 2005] Goddard, T. D., Huang, C. C., and Ferrin, T. E. (2005). Software extensions to UCSF Chimera for interactive visualization of large molecular assemblies. *Structure*, 13(3):473–482.

[Gruenert et al., 2010] Gruenert, G., Ibrahim, B., Lenser, T., Lohel, M., Hinze, T., and Dittrich, P. (2010). Rule-based spatial modeling with diffusing, geometrically constrained molecules. *BMC Bioinformatics*, 11:307.

[Hoard et al., 2015] Hoard, B., Jacobson, B., Manavi, K., and Tapia, L. (2015). Extending rule-based methods to model molecular geometry. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*.

[Ivanciuc et al., 2002] Ivanciuc, O., Schein, C. H., and Braun, W. (2002). Data mining of sequences and 3D structures of allergenic proteins. *Bioinformatics*, 18(10):1358–1364.

[Ivanciuc et al., 2003] Ivanciuc, O., Schein, C. H., and Braun, W. (2003). SDAP: Database and computational tools for allergenic proteins. *Nucleic Acids Research*, 31(1):359–362.

[Kerr et al., 2009] Kerr, R., Bartol, T., Kaminsky, B., Dittrich, M., Chang, J., Baden, S., Sejnowski, T., and Stiles, J. (2009). Fast Monte Carlo simulation methods for biological reaction-diffusion systems in solution and on surfaces. *SIAM J. Sci. Comput.*, 30(6):3126–3149.

[Kuntz, 2014] Kuntz, A. (2014). Statistical and graph based analysis of protein aggregation in simulation. Undergraduate Honors thesis, CS Dept., Univ. of New Mexico, Albuquerque, NM.

[Mahajan et al., 2014] Mahajan, A., Barua, D., Cutler, P., Lidke, D. S., Espinoza, F. A., Pehlke, C., Grattan, R., Kawakami, Y., Tung, C.-S., Bradbury, A. R. M., Hlavacek, W. S., and Wilson, B. S. (2014). Optimal aggregation of FcεRI with a structurally defined trivalent ligand overrides negative regulation driven by phosphatases. *ACS Chemical Biology*, 9(7):1508–1519.

[Manavi et al., 2015] Manavi, K., Jacobson, B., Hoard, B., and Tapia, L. (2015). Influence of model resolution on geometric simulations of antibody aggregation. Submitted to *Robotica*.

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

[Manavi et al., 2012] Manavi, K., Wilson, B. S., and Tapia, L. (2012). 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)*, pages 458–465.

[Maus et al., 2011] Maus, C., Rybacki, S., and Uhrmacher, A. M. (2011). Rule-based multi-level modeling of cell biological systems. *BMC Systems Biology*, 5:166.

[Ollivier et al., 2010] Ollivier, J. F., Shahrezaei, V., and Swain, P. S. (2010). Scalable rule-based modelling of allosteric proteins and biochemical networks. *PLoS Comput. Biol.*, 6(11).

[Reese et al., 2006] Reese, G., Schicktanz, S., Lauer, I., Randow, S., Luttkopf, D., Vogel, L., Lehrer, S., and Vieths, S. (2006). Structural, immunological and functional properties of natural recombinant Pen a 1, the major allergen of Brown Shrimp, Penaeus aztecus. *Clinical & Experimental Allergy*, 36(4):517–524.

[Reese et al., 2005] Reese, G., Viebranz, J., Leong-Kee, S. M., Plante, M., Lauer, I., Randow, S., Moncin, M. S.-M., Ayuso, R., Lehrer, S. B., and Vieths, S. (2005). Reduced allergenic potency of VR9-1, a mutant of the major shrimp allergen Pen a 1 (tropomyosin). *The Journal of Immunology*, 175(12):8354–8364.

[Smith et al., 2012] Smith, A. M., Xu, W., Sun, Y., Faeder, J. R., and Marai, G. E. (2012). RuleBender: Integrated modeling, simulation and visualization for rule-based intracellular biochemistry. *BMC Bioinformatics*, 13(8).

[Sneddon et al., 2011] Sneddon, M., Faeder, J., and Emonet, T. (2011). Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nature Methods*, 8(2):177–183.

[Stefan et al., 2014] Stefan, M., Bartol, T., Sejnowski, T., and Kennedy, M. (2014). Multi-state modeling of biomolecules. *PLoS Computational Biology*, 10(9):e1003844.

[Sulzer and Perelson, 1996] Sulzer, B. and Perelson, A. S. (1996). Equilibrium binding of multivalent ligands to cells: effects of cell and receptor density. *Mathematical Biosciences*, 135:147–185.

[Thomas et al., 2015] Thomas, B., Chylek, L., Colvin, J., Sirimulla, S., Clayton, A., Hlavacek, W., and Posner, R. (2015). GenFit: a fitting tool compatible with BioNetGen, NFsim, and distributed computing environments. *Los Alamos Technical Report LA-UR-15-25157*.

[Tolle and Novère, 2010] Tolle, D. P. and Novère, N. L. (2010). Meredys, a multi-compartment reaction-diffusion simulator using multistate realistic molecular complexes. *BMC Syst. Biol.*, 4(24).

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

[Xu et al., 1998] Xu, K., Goldstein, B., Holowka, D., and Baird, B. (1998). Kinetics of multivalent antigen DNP-BSA binding to IgE-FcϵRI in relationship to the stimulated tyrosine phosphorylation of FcϵRI. *J. Immunol.*, 160(7):3225–3235.

[Yang et al., 2008] Yang, J., Monine, M. I., Faeder, J. R., and Hlavacek, W. S. (2008). Kinetic Monte Carlo method for rule-based modeling of biochemical networks. *Phys. Rev. E*, 78(3):031910.

[Zhang et al., 2006] Zhang, J., Leiderman, K., Pfeiffer, J. R., Wilson, B. S., Oliver, J. M., and Steinberg, S. L. (2006). Characterizing the topography of membrane receptors and signaling molecules from spatial patterns obtained using nanometer-scale electron-dense probes and electron microscopy. *Micron*, 37(1):14–34.