Computers in Biology and Medicine 107 (2019) 153–160

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Application of adaptive-network-based fuzzy inference systems to the parameter optimization of a biochemical rule-based model

Brittany R. Hoard[1]

University of New Mexico, 1 University of New Mexico, Albuquerque, NM, 87131, United States

ARTICLE INFO

Keywords:
Optimization
Fuzzy inference systems
Adaptive algorithms
Biological system modeling
Molecular geometry

ABSTRACT

In this study, the binding of allergens to antibody-receptor complexes was investigated. This process is important in understanding the allergic response. A BioNetGen model that simulates this process, combined with a novel method for encoding steric effects via the optimization of the cutoff distance and the rule binding rate, was previously developed. These parameters were optimized by fitting the model output to the output of a 3D simulation that explicitly represents molecular geometry.

In this work, the parameters for the BioNetGen model were optimized using an adaptive-network-based fuzzy inference system in order to predict the rule rate and cutoff distance given a residual-sum-of-squares value or a probability distribution. The fuzzy systems were constructed using fuzzy c-means clustering with existing data from BioNetGen model parameter scans used as the training data. Fuzzy systems with various input data and number of clusters were created and tested. Their performance was analyzed with regard to the effective optimization of the rule-based model.

The study found that the fuzzy system that uses a residual-sum-of-squares value as the input value performs acceptably well. However, the performance of the fuzzy systems that use probabilities as their input values performed inconsistently in the tests and need further development. This methodology could potentially be modified for use in fitting other biological models.

1. Background

1.1. Introduction

Fuzzy inference systems (FISs) are models that consist of a set of IF-THEN rules in which the antecedents and/or consequents of the rules are fuzzy rather than crisp [26]. The rules can be provided by a human expert, but it can be highly useful for the rules to be constructed automatically using only a set of training data and an appropriate learning algorithm. Various FIS learning algorithms have been developed. FISs can be applied to regression and classification problems in various fields including robotics, data mining, prediction, estimation, control, and computational biology.

In this paper, a novel application of FISs in the field of biology is examined; namely, the optimization of the parameters of a biochemical rule-based model implemented in BioNetGen [2]. Ideally, this method would require only one parameter scan to generate the set of data used to train the FIS. In general, BioNetGen models consist of a set of “rules” that control interactions between molecules. In a previous line of work, a BioNetGen model was developed that captures the binding interactions between molecules of the shrimp tropomyosin Pen a 1 and IgE antibody-receptor complexes that lead to the formation of aggregates. This biological process is of interest because the size and structure of these aggregates is important in understanding the allergic response in shrimp-allergic human subjects; aggregate size and structure is theorized to be linked to the strength of the allergic response. Such knowledge could be useful in developing treatments for people with allergies, such as the administration of recombinant hypoallergens [25]. The purpose of using this BioNetGen model is to determine the probability that an aggregate of a particular size will be formed. The term *aggregate size* is used to refer to the number of IgE antibody-receptor complexes bound to a single Pen a 1 molecule. The BioNetGen model output was used to calculate these probabilities.

The Pen a 1 allergen is a shrimp tropomyosin. It is a dimer; it has a double-stranded coiled structure. In previous experimental work, ten binding regions of the double-stranded Pen a 1 molecule have been identified (five per strand) [1,13,23]. However, in previous work performed by colleagues, one large binding region was effectively split into two regions based on a study of conditional binding probabilities of these two regions [17]. Consequently, in this work, each Pen a 1

E-mail address: bhoard@unm.edu.

[1] Present address: 600 N Pantano Rd, Unit 529, Tucson, AZ, 85710, United States.

https://doi.org/10.1016/j.compbiomed.2019.01.021
Received 4 September 2018; Received in revised form 22 January 2019; Accepted 22 January 2019

molecule is treated as having a total of 12 binding regions (six per strand). This means that there are 13 possible aggregate sizes (0–12 regions can be bound).

Two parameters of the biochemical rule-based model must be optimized: the cutoff distance and the rule binding rate, which is referred to throughout this paper as the "rule rate". The *cutoff distance* is the distance between tropomyosin binding sites at or below which steric effects become strong enough to significantly reduce the probability of a binding event taking place at a neighboring tropomyosin binding site. Steric effects are encoded between the binding sites by changing the set of rules according to the cutoff distance. The *rule rate* is the probability that an event encoded by that rule will occur. For this simple model, it was assumed that the same rate is associated with each rule. Previously, the determination of the rule rates for the rule-based model was achieved via parameter scanning, which can be time-consuming and risks skipping over the best fit if the step size is too large. An algorithm that uses the Metropolis method has also been employed, though it can also be time-consuming. FISs only take a maximum of a few minutes to train for this model, and they can produce output instantaneously given a set of input values. The greater efficiency and possibly greater accuracy of the FIS method of parameter optimization are the motivations for applying this method to the optimization of this BioNetGen model.

The method presented in this paper could potentially be modified for implementation in other biological models. For example, the method could be applied to any other rule-based model in order to optimize the rule rates. In addition, fuzzy inference systems could be used to optimize various parameters, such as reaction rates, decay coefficients, and production coefficients, for a wide variety of biological models, as long as experimental data or another type of training data is available. A common reason to optimize the parameters of a biological model is to fit the model output to a set of experimental data. Because experimental data for this particular biological process is not currently available, aggregate size data from a three-dimensional rigid-body Monte Carlo simulation previously developed by collaborators was used instead [18]. The rule rate and cutoff distance for this biological model are optimized to reduce the difference between the aggregate size distribution of the Monte Carlo data and that of the rule-based model. This difference is quantified by calculating the residual sum-of-squares (RSS) value between the aggregate size distribution generated by the rule-based model and that generated by the Monte Carlo simulation.

There are multiple options for designing an FIS for this optimization problem since there are 13 possible aggregate sizes, a cutoff distance, a rule rate, and an RSS value associated with each training data point. All of these parameters could be used as input parameters for the FIS. A single-input FIS with only one input parameter or a multiple-input FIS with more than one input parameter could potentially be constructed. The main contributions of this work are the creation and testing of different FISs, including single-input and multiple-input systems, and an analysis of their performance with regards to the effective optimization of the rule-based biological model implemented using BioNetGen. The results suggest that the single-input RSS system performs acceptably well, but that the performance of the multiple-input systems is inconsistent, with one system performing well and the other systems performing poorly.

1.2. Related work

1.2.1. Fuzzy inference systems

There have been numerous learning algorithms developed for FISs. One of the earliest and most widely used algorithms is the Wang-Mendel technique [26], in which the input and output spaces are divided into fuzzy regions that form the basis of the fuzzy rules, and each rule is assigned a degree of usefulness. The adaptive-network-based fuzzy inference system (ANFIS) [14,15] is a two-stage model. The forward stage consists of multiple layers, including fuzzification, inference, and other calculations; parameter learning using the least-squares method takes place in the backward stage. In the subtractive clustering and fuzzy c-means method [3,27], rule cluster centers are determined by calculating the distance of each data point from every other data point, and then optimizing the clusters using fuzzy c-means. The MOGUL method [5,9] uses iterative rule learning to generate chromosomes that consist of the rule membership function parameters. In the fuzzy inference rules by descent method [21], the antecedent membership function is an isosceles triangle, and the consequent part of the rule is a real number obtained using a descent method.

1.2.2. Biological rule-based modeling

Biological signaling systems can be composed of macromolecules, which can exist in a large number of functionally distinct states. The number of possible states scales exponentially with the number of possibilities for modification [28]. When modeling these systems, the problem of how to specify such large systems is an important issue.

One solution is implicit specification. With this solution, sets of reactions are coarse-grained into rules. Rules define the conditions needed for molecular transformation reactions and interactions between molecules [4]. All rules have rate laws associated with them [4]. All reactions specified by a single rule are associated with the same rate law. The rules can usually be specified independently of each other. Rule-based specification methods include BioNetGen [2], which was used in this project, Kappa-calculus [6], ANC [22], and ML-Rules [20].

Biological rule-based modeling methods include population-based, particle-based, and hybrid methods. Population-based methods include numerical integration of ordinary differential equations and partial differential equations as well as the stochastic Gillespie algorithm. With population-based methods, when a rule is applied, the size of one or more of the populations is changed. A population is made up of all molecules of the same state and same species. Methods to reduce the size of the state space, which can be quite large, have been proposed [28].

On the other hand, particle-based rule evaluation, which is network-free, involves the tracking of individual particles (molecules and molecular complexes) throughout the simulation [4]. At any point in time, only the existing particles, their states, and the possible reactions for the existing particles are needed to continue the simulation. Spatial particle-based methods include SRSim [7] and MCell [16].

With the biological rule-based modeling approach that was used in this work, the biological rules are constrained by the geometry of the molecules involved in the model. This differs from traditional approaches to rule-based modeling.

1.2.3. Geometric molecular modeling

The spatial simulation software SRSim [7] combines rule-based modeling, molecular dynamics, and a stochastic simulator. It allows molecular geometry to be provided by the user via data files. The geometric modeling method that was used in this work is different in that it is a purely rule-based ordinary differential equation model that does not need additional data files to run because the molecular geometry is encoded into the rules.

The Meredys [24] software is stochastic and particle-based. It uses Brownian dynamics to simulate reaction-diffusion systems. Meredys requires the specification of details such as molecule positions, molecular geometry, reaction site positions, and reaction types. The geometric rule-based method that was used in this work is population-based. Furthermore, it only requires the distances between binding regions on a single allergen molecule in order to create the model.

The 3D Monte Carlo method used in this project was previously developed by collaborators [17]. It simulates molecular systems on a larger scale than other computational methods for modeling two-molecule ligand-receptor docking. In addition, the method that was used for this project employs more realistic geometric molecular models than do existing methods for self-assembly of molecular structures, such as those employing simple bead models [8].

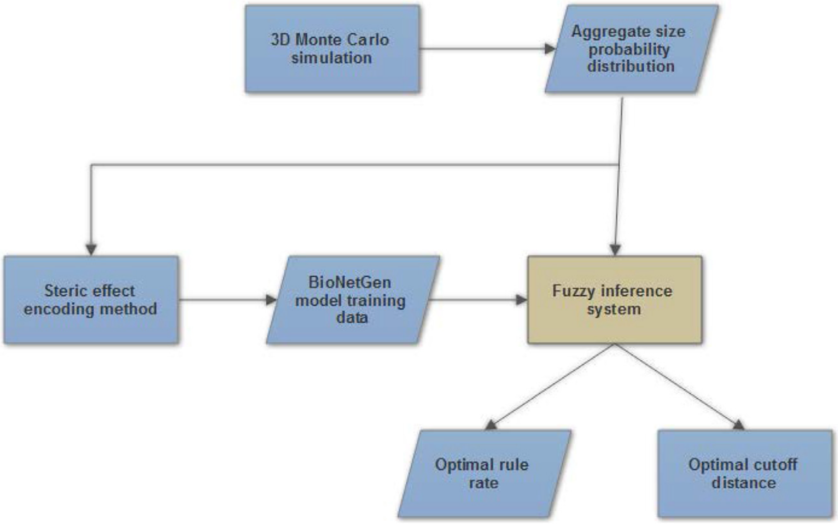
3D Monte Carlo simulation
Aggregate size probability distribution
Steric effect encoding method
BioNetGen model training data
Fuzzy inference system
Optimal rule rate
Optimal cutoff distance

Fig. 1. Flowchart showing the interdependence between the various components of the overall method.

1.3. Contributions

The contributions of this work are as follows:

1. The proposal of an effective method for fitting the parameters of a biological rule-based model using fuzzy inference systems.
2. The development of fuzzy inference systems that achieve accuracy greater than or equal to that of the previously employed method with regards to finding optimal values for the parameters of the biological rule-based model.
3. The improvement of the efficiency of this fitting method compared to the previously used fitting method.

2. Methods

For the application of a FIS to the BioNetGen model, the ANFIS method was chosen based on its easy-to-use implementation in MATLAB and its customization options. This method employs an adaptive network that is composed of nodes and directional links that connect the nodes, and at least some of the nodes are adaptive (the nodes are associated with parameters that are adapted to minimize an error of measure according to the learning rules of the network) [14]. ANFIS uses hybrid learning rules that combine the gradient method and the least squares estimate [15].

An initial fuzzy inference system (FIS), which includes the fuzzy rule base and membership functions, was generated using fuzzy c-means clustering. The parameters of this initial system were then trained further using ANFIS. Further detail regarding the construction of the FIS can be found in the online MathWorks documentation [19]. One parameter of particular interest is the number of clusters used in fuzzy c-means clustering. During this process, clusters are identified within the training data, and these clusters are then employed in the generation of the FIS. The number of clusters may be specified by the user and can have a significant effect on the results, so this study includes systems with various numbers of clusters used.

Ideally, two FISs are constructed: one that can accurately predict the rule rate, and another that can accurately predict the cutoff distance. Having such FISs would be helpful for optimization of the BioNetGen model, as ideally, only a small set of training data would be needed to train the FISs. These FISs could then be provided with the Monte Carlo aggregate size probability distribution as the input. The FISs would then accurately predict the rule rate and cutoff distance that best corresponds to that particular distribution. This method would be far less time-consuming than multiple-parameter scans and even Metropolis-based algorithms.

The application of FISs to the BioNetGen model is based on previous work using a 3D geometric Monte Carlo simulation to model the aggregation of IgE antibodies onto the shrimp allergen tropomyosin [17]. For this simulation, isosurface models of Pen a 1 were created from all-atom structures of shrimp tropomyosin. These structures were obtained from the Protein Data Bank (PDB:1CG1) and the Structural Database of Allergenic Proteins (SDAP Model #284) [12,13]. The tropomyosin molecule used contains 568 amino acids and 4577 atoms. It is a dimer and has a double-stranded coiled structure.

In previous work, three different conformations of the Pen a 1 molecule were studied in terms of the results of the Monte Carlo model, namely the aggregate size probability distributions. In addition, various resolutions of the Monte Carlo model were used to determine how the model resolution affects the results. In a previous study, a rule-based model implemented in BioNetGen that implicitly represents antigen geometry was used to quantify the differences in the Monte Carlo results that arise as a result of different allergen conformations and resolutions of the Monte Carlo model [10,11]. This work employed a method to construct sets of rules based directly on the distances between the IgE binding regions of the tropomyosin. With this method, once the cutoff distance is specified, the set of rules is constructed according to the cutoff distance and the distances between the binding regions. Hence, the number of rules in the model varies according to the cutoff distance.

In this work, the ANFIS method, implemented in MATLAB, was used to train the FISs. The FISs were constructed using fuzzy c-means clustering with existing aggregate size data from BioNetGen model parameter scans as the training data. Either an RSS value or some of the 13 possible aggregate size probabilities were used as input variables, and a rule rate or cutoff distance was the model output. For the final test, all of the Monte Carlo aggregate size probabilities were used as input variables for the FIS. A flowchart illustrating the relationships between the components of the method is shown in Fig. 1.

The performance of the FISs was evaluated by calculating the percent error between the output values generated by the FIS and the actual values, and by simply observing the output values and seeing if they are reasonable for my expectations of the particular model. This point is further discussed in the Results section. For the final test, the FIS was evaluated by comparing the RSS values to the minimum RSS value predicted by the Metropolis algorithm.

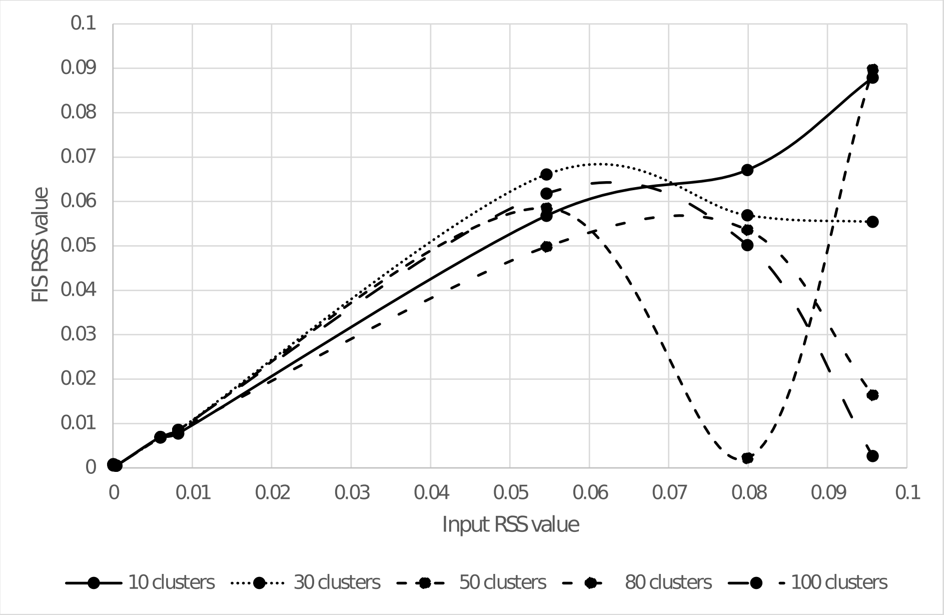

Fig. 2. RSS values for the BioNetGen models with the rule rates predicted by the FIS with various numbers of clusters using the selected RSS values as input.

2.1. Metropolis algorithm

Previously, an algorithm based on the Metropolis algorithm was employed to optimize the rule rate. This algorithm searches for the minimum RSS value between the Monte Carlo data and the BioNetGen rule-based modeling data. The algorithm is described in detail in Ref. [11] and will be summarized here as follows.

If the current RSS value is σ_c, the current rule rate is k_c, the RSS value for a new rate constant k_n is σ_n, and the difference between σ_n and σ_c is $\Delta\sigma$, then the rule rate is determined according to the following:

If $\sigma_n > \sigma_c$, then $\sigma_c = \sigma_n$ & $k_c = k_n$ w/prob. $e^{-\Delta\sigma/T}$
If $\sigma_n > \sigma_c$, then $\sigma_c = \sigma_c$ & $k_c = k_c$ w/prob. $1 - e^{-\Delta\sigma/T}$
If $\sigma_n \leq \sigma_c$, then $\sigma_c = \sigma_n$ & $k_c = k_n$ w/prob. 1.

If σ_n is greater than σ_c, then σ_n is accepted with a probability that is dependent on $\Delta\sigma$ and the simulated annealing temperature T. If σ_n is accepted, then the rule rate is incremented by the specified step size. If σ is decreasing, then a step size of 1x10^{-5} molecule$^{-1}s^{-1}$ is used to determine the next σ_n; otherwise, a step size of 1x10^{-4} molecule$^{-1}s^{-1}$ is used.

On the other hand, if σ_n is rejected, then a new rule rate will be selected randomly from the entire range of possible rule rate values, from 0.00 to 0.40 molecule$^{-1}s^{-1}$.

2.2. Computational experiments

In order to construct a fuzzy system using ANFIS, the genfis3 function in MATLAB was first run to create a Sugeno-type FIS structure using fuzzy c-means clustering to extract a set of rules and membership functions that model the training data. This function allows the specification of the number of clusters used to model the data. This parameter was varied throughout this study. The other parameters were set to their default values; the number of training epochs was set to 10, the initial step size was set to 0.01, the step size increase rate was set to 1.1, and the step size decrease rate was set to 0.9.

The data used to train the FIS consisted of the rule rates, the aggregate size probability distributions, the cutoff distances, and the RSS values. This data was obtained by running a parameter scan of the rule rate from 0.000 $molecule^{-1}s^{-1}$ to 0.020 $molecule^{-1}s^{-1}$ with a step size of 0.001 $molecule^{-1}s^{-1}$ over a range of cutoff distances from 3.5 nm to 10.0 nm with a step size of 0.1 nm. This training set consists of 1365 data points and is included in the supplementary information files for this article.

Before developing the more complex multiple-input FIS, there is another type of FIS that was developed that could be applied to optimization: a single-input FIS that uses an RSS value as an input variable and the rule rate or cutoff distance as the output. The disadvantage of this FIS is that it is not known what the minimum possible RSS value is; however, the author has an idea from previous experience with the biological models of what a “good” RSS value should be. Different “good” RSS values can be used to find an accurate rule rate or cutoff distance. The advantage of this type of FIS is that it is far less computationally expensive than a multiple-input FIS.

In order to determine how well the FIS can predict the rule rate and cutoff distance, five RSS values within the range of the training data are randomly selected (but with none matching that of any of the training values). Since the lowest RSS value is likely to be outside of the range of the training data, it is also important to know how well the FIS performs with this type of input. Two small RSS values below the lowest RSS value in the range of the training data are chosen for this purpose. After training the FIS, all of the seven RSS values are used as the input, and the FIS is then used to predict the rule rate and cutoff distance for each of these seven values. A simulation for the BioNetGen model is then run using the rule rate and cutoff distance predicted by the FIS. The actual RSS value calculated for the BioNetGen simulation is noted. Ideally, the inputted and predicted RSS values should be similar.

3. Results

It should be noted that for all inputs, the number of fuzzy sets used in the FIS was set to the default value set by MATLAB, which is two. Furthermore, the form of the two membership functions was set to the MATLAB default, which was a generalized bell-shaped membership function.

3.1. Single-input FIS

Figs. 2 and 3 show the results of this test. Fig. 2 displays the predicted RSS values for the five randomly selected inputted RSS values and the two non-randomly selected inputted RSS values. In this figure, the inputted RSS values are plotted on the x-axis, and the RSS values predicted by the FIS are plotted on the y-axis. There are five curves, one for each of the number of clusters used in the initial construction of the

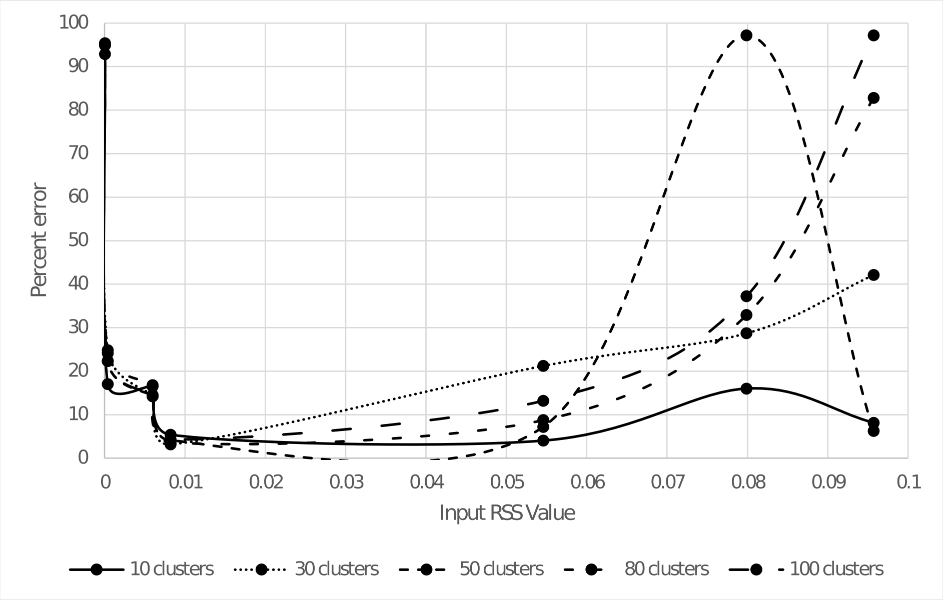

Fig. 3. Percent error between the inputted and predicted RSS values.

Table 1
Rule rates predicted by the FIS compared with the actual rule rates.

Selected rate	Clusters				
	10	30	50	80	100
3.26×10^{-3}	3.01×10^{-3}	3.26×10^{-3}	3.26×10^{-3}	3.27×10^{-3}	3.29×10^{-3}
7.43×10^{-3}	7.02×10^{-3}	7.47×10^{-3}	7.42×10^{-3}	7.42×10^{-3}	7.43×10^{-3}
7.91×10^{-3}	7.43×10^{-3}	8.07×10^{-3}	7.92×10^{-3}	7.91×10^{-3}	7.91×10^{-3}
9.76×10^{-3}	9.88×10^{-3}	9.69×10^{-3}	9.75×10^{-3}	9.76×10^{-3}	9.77×10^{-3}
1.78×10^{-2}	1.78×10^{-2}	1.78×10^{-2}	1.78×10^{-2}	1.78×10^{-2}	1.78×10^{-2}

FIS. Fig. 3 displays the percent error between the inputted and predicted RSS values for each number of clusters.

From Fig. 2, it can be observed that the FIS performed the best for this test with only 10 clusters. Furthermore, it can be observed from Fig. 3 that the system performed especially poorly for the higher RSS values, with errors of over 90% for two tests. This could be due to the fact that a high RSS value may correspond to a large range of poorly fitting data sets, while a lower RSS value corresponds with a much smaller number of well-fitting data sets.

The lowest RSS value, 0.00004, also performed poorly, probably because this RSS is too small to be achievable. It is well outside the range of the training data. The system performed reasonably well for the next lowest RSS input value, 0.0004, which is also outside the range of the training data, but close to the minimum value found by the Metropolis algorithm, 0.000481558. Many of the predicted RSS values are close to this Metropolis value, which is an indication that the performance of this FIS is similar to that of the Metropolis fitting algorithm. Since the predicted RSS values for these two input values are all less than 0.001, and since it is ultimately desired to use this system to find reasonably good fits (generally corresponding to RSS values less than 0.001), this performance is acceptable.

3.2. Random rule rates

The purpose of this test is to determine how accurately the FIS can predict a rule rate given a set of aggregate size probabilities corresponding to that rule rate. Firstly, five rule rates were selected from a uniform distribution in the interval [0.00,0.02] $molecule^{-1}s^{-1}$, and were each specified as the variable rule rate for the same BioNetGen model. The rule base varies with the cutoff distance, so for this initial test, a cutoff distance of 4.7 nm was specified for all five BioNetGen models. This distance was chosen because the Metropolis fitting algorithm found that the best fitting BioNetGen model has this cutoff distance. The BioNetGen aggregate size probability distributions were generated for each of the five rule rates. The probability values for aggregate sizes 5 through 10 were then provided to the FIS, which was used to predict the rule rates. (Not all 13 of the size values were used as input variables because 13 input variables makes the model computationally intractable. Aggregate sizes 5 though 10 were selected because these sizes are most likely to correspond to values that are large enough to be significant.) These predicted rule rates should ideally be similar to the actual rule rates used to generate the BioNetGen model.

Table 1 and Fig. 4 show that this FIS performed well for predicting the rule rates given a set of aggregate size probabilities generated by BioNetGen, with most of the percent error values being less than one. It can also be observed that the system performed better with a higher number of clusters for this test.

3.3. Random cutoff distances

The purpose of this test is to determine how accurately the FIS can predict a cutoff distance given a set of aggregate size probabilities corresponding to that cutoff distance. Firstly, five cutoff distances were selected from a uniform distribution in the interval [3.5,10.0] nm, and were each specified as the cutoff distance for a BioNetGen model with the same rule rates. For this test, the variable rule rate was specified as 0.00983 $molecule^{-1}s^{-1}$ for all models. This rule rate was chosen because the Metropolis fitting algorithm found that the best fitting BioNetGen model has this rate. The aggregate size probability distributions were generated for each of the five cutoff distances. These probability values were then fed into the FIS, which predicted the cutoff distances. These predicted cutoff distances should ideally be similar to the actual cutoff distances used to generate the BioNetGen model.

Table 2 and Fig. 5 show that this FIS performed reasonably well at predicting cutoff distances given a set of aggregate size probabilities generated by BioNetGen, although some of the error values are rather high, with a few greater than five percent and one greater than nine percent. This could potentially pose a problem for this system, a cutoff distance accurate to within 0.1 nm is desired according to the cutoff distance step size. As it stands, this FIS is better suited to finding an

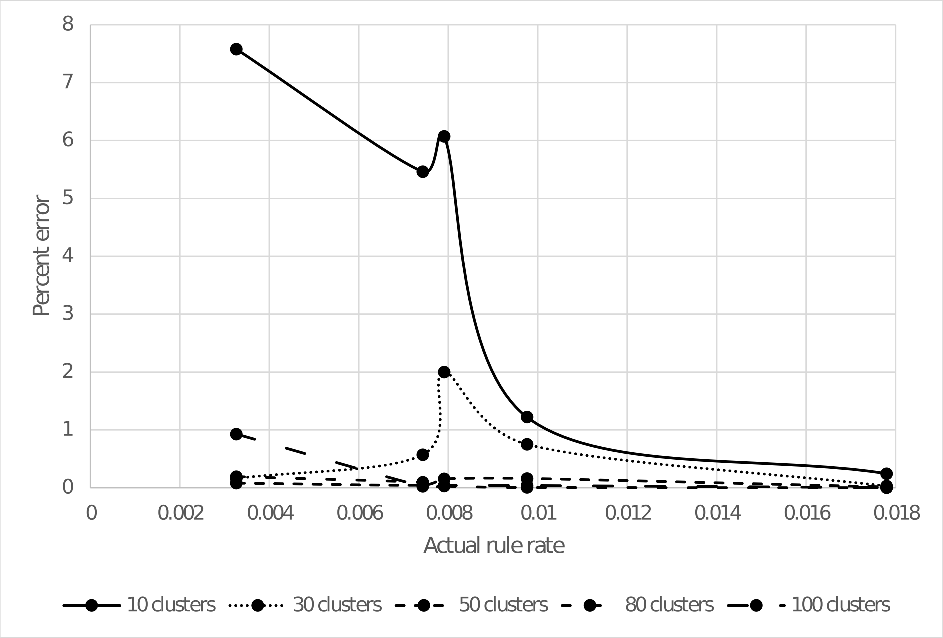

Fig. 4. Percent error between the actual and predicted rule rates.

Table 2
Cutoff distances predicted by the FIS compared with the actual cutoff distances.

Selected distance	Clusters				
	10	30	50	80	100
4.31	4.71	4.66	4.54	4.55	4.55
5.86	5.72	5.84	5.81	5.80	5.80
6.90	6.55	6.61	6.84	6.81	6.79
7.84	8.03	8.13	8.13	8.13	8.11
8.83	9.51	9.40	9.40	9.44	9.42

approximate "best" cutoff distance that can then be used to select a small range of cutoff distances that can be further optimized to find the true best cutoff distance (and rule rate). It can also be observed that the system tends to perform better with a higher number of clusters for this test.

3.4. Monte Carlo model prediction

The final test of the FIS is whether the Monte Carlo data can be used as input to predict a rule rate and a cutoff distance that correspond to a good fit. In this test, a set of Monte Carlo aggregate size probabilities was provided to the FIS, and the rule rate and cutoff distance were then predicted. These predicted values are shown in Table 3 and Table 4. The predicted rule rate and cutoff distance were tested by using these values as the rule rate and cutoff distance in the BioNetGen model, generating the aggregate size distribution from BioNetGen, and comparing this distribution to the Monte Carlo data. Since it is not known what the optimal RSS value is, the results are compared to that of the Metropolis-

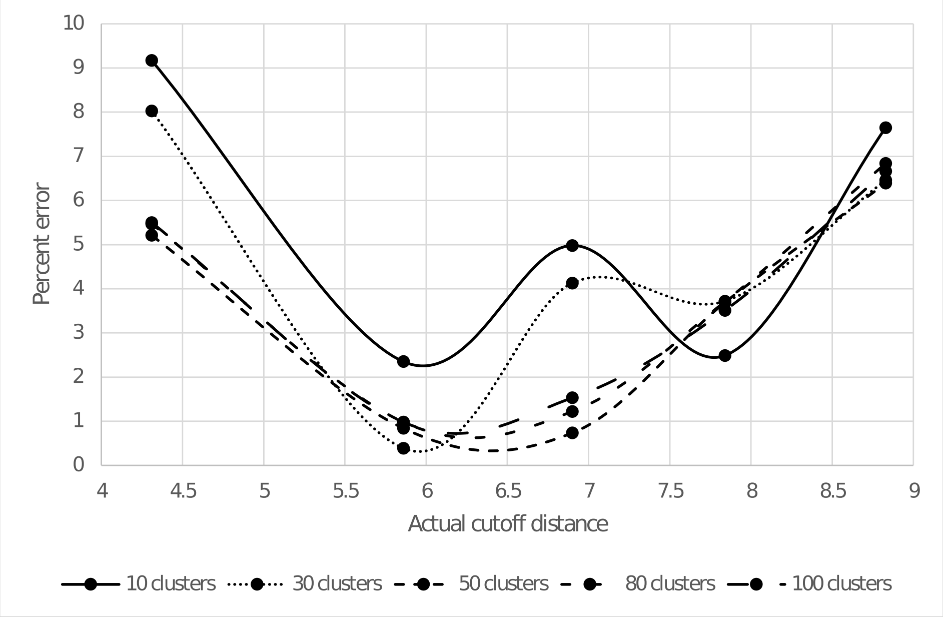

Fig. 5. Percent error between the actual and predicted cutoff distances.

Table 3
FIS-predicted rule rates for the Monte Carlo model prediction. The leftmost column represents the Monte Carlo aggregate sizes used as input variables.

Input sizes	Clusters				
	10	30	50	80	100
5,6,7,8,9,10	2.53×10^{-3}	4.17×10^{-2}	4.01×10^{-3}	4.74×10^{-3}	3.73×10^{-3}
6,7,8,9	5.82×10^{-3}	5.86×10^{-3}	4.82×10^{-3}	4.94×10^{-3}	5.08×10^{-3}
7,8,9	1.34×10^{-2}	4.52×10^{-3}	5.68×10^{-3}	5.54×10^{-3}	6.11×10^{-3}

Table 4
FIS-predicted cutoff distances for the Monte Carlo model prediction. The leftmost column represents the Monte Carlo aggregate sizes used as input variables.

Input sizes	Clusters				
	10	30	50	80	100
5,6,7,8,9,10	3.78	5.34	5.27	2.96	5.06
6,7,8,9	4.58	4.78	1.61	3.52	4.52
7,8,9	5.13	0.10	0.53	3.89	3.90

Table 5
RSS values for the BioNetGen model using the FIS-predicted rule rate and cutoff distance. The leftmost column represents the Monte Carlo aggregate sizes used as input variables.

Input sizes	Clusters				
	10	30	50	80	100
5,6,7,8,9,10	2.34×10^{-2}	1.73×10^{-2}	2.70×10^{-3}	8.96×10^{-2}	2.95×10^{-3}
6,7,8,9	1.44×10^{-3}	1.42×10^{-3}	8.94×10^{-2}	1.60×10^{-2}	1.88×10^{-3}
7,8,9	1.02×10^{-3}	9.01×10^{-2}	8.75×10^{-2}	1.46×10^{-2}	1.35×10^{-2}

based optimization algorithm to see how well the FIS compares.

The minimum RSS value found for this BioNetGen model using a Metropolis-based algorithm is 0.000481558. Comparing this value with the results in Table 5, it can be observed that all of the FIS-predicted values are at least one order of magnitude higher than this value. It can also be noted that the performance of this system is inconsistent, and many RSS values are unacceptably high (greater than 0.01). This FIS needs improvement before it can be used as a tool for finding best fits.

4. Conclusions

The purpose of this study was the creation of a FIS that can accurately predict best-fit rule rates and cutoff distances for a BioNetGen rule-based model. Different FISs were tested using three different FIS input variables: (a) an RSS value, (b) a set of aggregate size probabilities, and (c) a set of aggregate size probabilities directly from the Monte Carlo simulation data. (System (a) corresponds to Section 3.1 of the article, System (b) corresponds to Sections 3.2 and 3.3, and System (c) corresponds to Section 3.4). For System (a), the output of interest is the RSS value predicted by the FIS. For System (b), the outputs of interest are the predicted rule rate and cutoff distance, which are separated into two different tests. For System (c), the previous two systems are combined such that the output of interest is the RSS value calculated using the rule rate and cutoff distance predicted by the FIS.

A FIS corresponding to (a) or (c) would be especially useful for this optimization problem. The FIS based on (a) performed well for low RSS values and could potentially be used to predict rule rates and cutoff distances that result in a good fit for the BioNetGen model (although it will not necessarily find the best possible fit, it could come very close). The FIS based on (b) consistently performed well at predicting rule rates, but its prediction of cutoff distances was rather inaccurate. However, this system could still be used to narrow down the range of cutoff distances. Finally, the FIS based on (c) performed poorly overall. It can also be noted that the optimal number of clusters varies depending on the type of input. For most tests, a higher number of clusters was linked to better performance of the FIS, although this was not always the case.

The results of this study suggest that the use of a FIS for fitting parameters of a biological model has the potential to be effective and efficient. Future work on this problem could involve testing fuzzy training algorithms other than ANFIS and clustering or rule construction algorithms other than fuzzy c-means clustering. Furthermore, ANFIS and clustering parameters other than the number of clusters, such as change in step size rate, training epoch number, and initial step size, could be tested to see if any of these parameters have a significant effect on the FIS performance.

Conflicts of interest

The author declares that there are no competing interests.

Authorship statement

The sole author of this paper conceived and designed the study, acquired the data, analyzed and interpreted the data, drafted and revised the article, and approved the final version for submission.

Data statement

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Conflicts of interest

None declared.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

Thanks to the peer reviewers, faculty, and fellow students who contributed to the improvement of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2019.01.021.

References

[1] R. Ayuso, S. Lehrer, G. Reese, Identification of continuous, allergenic regions of the major shrimp allergen Pen a 1 (tropomyosin), Int. Arch. Allergy Immunol. 127 (1) (2002) 27–37.
[2] M.L. Blinov, J.R. Faeder, B. Goldstein, W.S. Hlavacek, BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains, Bioinformatics 20 (17) (2004) 3289–3291.
[3] S. Chiu, Method and Software for Extracting Fuzzy Classification Rules by Subtractive Clustering, Fuzzy Information Processing Society, NAFIPS, 1996, pp. 461–465.
[4] L.A. Chylek, L.A. Harris, C.S. Tung, J.R. Faeder, C.F. Lopez, W.S. Hlavacek, Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems, WIRESBM 6 (2014) 13–36.
[5] O. Cordon, M.J. del Jesus, F. Herrera, M. Lozan, MOGUL: a methodology to obtain genetic fuzzy rule-based systems under the iterative rule learning approach, Int. J. Intell. Syst. 14 (1999) 1123–1153.
[6] V. Danos, C. Laneve, Formal molecular biology, Theor. Comput. Sci. 325 (1) (2004) 69–110.
[7] G. Gruenert, B. Ibrahim, T. Lenser, M. Lohel, T. Hinze, P. Dittrich, Rule-based

spatial modeling with diffusing, geometrically constrained molecules, BMC Bioinf. 11 (2010) 307.
[8] H.W. Hatch, J. Mittal, V.K. Shen, Computational study of trimer self-assembly and fluid phase behavior, J. Chem. Phys. 142 (16) (2015).
[9] F. Herrera, M. Lozano, J. Verdegay, A learning process for fuzzy control rules using genetic algorithms, Fuzzy Sets Syst. 100 (1998) 143–158.
[10] B. Hoard, B. Jacobson, K. Manavi, L. Tapia, Extending rule-based methods to model molecular geometry, IEEE Int. Conf. Bioinf. Biomed. (BIBM) (2015) 587–594.
[11] B. Hoard, B. Jacobson, K. Manavi, L. Tapia, Extending rule-based methods to model molecular geometry and 3D model resolution, BMC Syst. Biol. 10 (48) (2016).
[12] O. Ivanciuc, C. Schein, W. Braun, Data mining of sequences and 3d structures of allergenic proteins, Bioinformatics 18 (10) (2002) 1358–1364.
[13] O. Ivanciuc, C. Schein, W. Braun, SDAP: Database and computational tools for allergenic proteins, Nucleic Acids Res. 31 (1) (2003) 359–362.
[14] J. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man Cybern. 23 (3) (1993) 665–685.
[15] J. Jang, C. Sun, E. Mizutani, Neuro-fuzzy and Soft Computing: a Computational Approach to Learning and Machine Intelligence, Prentice-Hall, Inc, 1997.
[16] R. Kerr, T. Bartol, B. Kaminsky, M. Dittrich, J. Chang, S. Baden, T. Sejnowski, J. Stiles, Fast Monte Carlo simulation methods for biological reaction-diffusion systems in solution and on surfaces, SIAM J. Sci. Comput. 30 (6) (2009) 3126–3149.
[17] K. Manavi, B. Jacobson, B. Hoard, L. Tapia, Influence of model resolution on geometric simulations of antibody aggregation, Robotica 34 (8) (2016) 1754–1776.
[18] K. Manavi, B. Wilson, L. Tapia, Simulation and analysis of antibody aggregation on cell surfaces using motion planning and graph analysis, Proc. the ACM Conf. Bioinf. Comput. Biol. Biomed. (ACM-BCB) (2012) 458–465.
[19] MathWorks, Fuzzy inference system modeling, URL, 2018. https://www.mathworks.com/help/fuzzy/mamdani-fuzzy-inference-systems.html.
[20] C. Maus, S. Rybacki, A.M. Uhrmacher, Rule-based multi-level modeling of cell biological systems, BMC Syst. Biol. 5 (2011) 166.
[21] H. Nomura, I. Hayashi, N. Wakami, A learning method of fuzzy inference rules by descent method, IEEE Int. Conf. Fuzzy Syst. (1992) 203–210.
[22] J.F. Ollivier, V. Shahrezaei, P.S. Swain, Scalable rule-based modelling of allosteric proteins and biochemical networks, PLoS Comput. Biol. 6 (11) (2010).
[23] G. Reese, J. Viebranz, S. Leong-Kee, M. Plante, I. Lauer, S. Randow, M.M. Moncin, R. Ayuso, S. Lehrer, S. Vieths, Reduced allergenic potency of VR9-1, a mutant of the major shrimp allergen Pen a 1 (tropomyosin), J. Immunol. 175 (12) (2005) 8354–8364.
[24] D.P. Tolle, N.L. Novère, Meredys, a multi-compartment reaction-diffusion simulator using multistate realistic molecular complexes, BMC Syst. Biol. 4 (24) (2010).
[25] A.M. Vargas, A. Mahajan, K.S. Tille, B.S. Wilson, C.P. Mattison, Cross-reaction of recombinant termite (coptotermes formosanus) tropomyosin with ige from cockroach and shrimp allergic individuals, Ann. Allergy Asthma Immunol. 120 (3) (2018) 335–337.
[26] L.X. Wang, J.M. Mendel, Generating fuzzy rules by learning from examples, IEEE Transactions on Systems, Man, and Cybernetics 22 (6) (1992) 1414–1427.
[27] R. Yager, D. Filev, Generation of fuzzy rules by mountain clustering, J. Intell. Fuzzy Syst. 2 (3) (1994) 209–219.
[28] J. Yang, M.I. Monine, J.R. Faeder, W.S. Hlavacek, Kinetic Monte Carlo method for rule-based modeling of biochemical networks, Phys. Rev. E 78 (3) (2008) 031910.

Brittany R. Hoard obtained a Master of Science degree in the Nanoscience and Microsystems program at the University of New Mexico. She also holds a Bachelor of Science in Physics with a minor in Computer Science from the Pennsylvania State University. For her Master's thesis, Brittany conducted research involving modeling antigen-antibody aggregation using rule-based methods and using these methods to help quantify differences in aggregate size results for a 3D rigid-body Monte Carlo simulation caused by model resolution. Since then, Brittany has been working in industry, doing primarily data analysis and software development with scientific and engineering applications.