Published in final edited form as:

Phys Biol. 2011 October ; 8(5): 055009. doi:10.1088/1478-3975/8/5/055009.

Efficiency of reactant site sampling in network-free simulation of rule-based models for biochemical systems

Jin Yang¹ and William S. Hlavacek^{2,3}

Jin Yang: jinyang2004@gmail.com

¹ Chinese Academy of Sciences – Max Planck Society Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Shanghai 200031, China

² Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

³ Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA

Abstract

Rule-based models, which are typically formulated to represent cell signaling systems, can now be simulated via various network-free simulation methods. In a network-free method, reaction rates are calculated for rules that characterize molecular interactions, and these rule rates, which each correspond to the cumulative rate of all reactions implied by a rule, are used to perform a stochastic simulation of reaction kinetics. Network-free methods, which can be viewed as generalizations of Gillespie's method, are so named because these methods do not require that a list of individual reactions implied by a set of rules be explicitly generated, which is a requirement of other methods for simulating rule-based models. This requirement is impractical for rule sets that imply large reaction networks (i.e., long lists of individual reactions), as reaction network generation is expensive. Here, we compare the network-free simulation methods implemented in RuleMonkey and NFsim, general-purpose software tools for simulating rule-based models encoded in the BioNetGen language. The method implemented in NFsim uses rejection sampling to correct overestimates of rule rates, which introduces null events (i.e., time steps that do not change the state of the system being simulated). The method implemented in RuleMonkey uses iterative updates to track rule rates exactly, which avoids null events. To ensure a fair comparison of the two methods, we developed implementations of the rejection and rejection-free methods specific to a particular class of kinetic models for multivalent ligand-receptor interactions. These implementations were written with the intention of making them as much alike as possible, minimizing the contribution of irrelevant coding differences to efficiency differences. Simulation results show that performance of the rejection method is equal to or better than that of the rejection-free method over wide parameter ranges. However, when parameter values are such that ligand-induced aggregation of receptors yields a large connected receptor cluster, the rejection-free method is more efficient.

1. Introduction

Protein-protein interactions in cell signaling systems involve domain-based protein interactions and site-specific post-translational modifications [1, 2]. Simulating the dynamics of cell signaling is usually a daunting task because a large (bio)chemical reaction network is typically required to capture protein-protein interactions at the level of site-specific details and submolecular domains [3, 4, 5, 6]. Even though a large-scale biochemical reaction network can be built by either manual or automated construction [7, 8, 9, 10], simulating such models is computationally inefficient because a conventional kinetic Monte Carlo simulation algorithm, for example, has a cost that depends on the size of a

network measured by the number of reactions [11] or the number of chemical species [12] in the network.

The challenge of simulating protein-protein interactions in cell signaling systems can be addressed with the rule-based modeling approach (see Ref. [4] for a review). Rule-based modeling provides a hierarchical structure to define biochemical reaction systems (Fig. 1). In a rule-based approach, molecules are modeled as structured objects composed of reactive sites, and reaction rules are used to represent interactions [4, 13, 14] (see Fig. 2 for examples of rules for ligand-receptor interactions). In general, a rule specifies local properties of individual sites (e.g., whether a site is free or occupied) in a molecule and application conditions that require checking non-local properties of sites (e.g., whether two sites are members of the same macromolecular aggregate). A rule defines a class of (unidirectional) reactions. Assuming rate laws for elementary reactions, one parameterizes the reactions implied by a rule with a single rate constant. Thus, a rule provides a compact representation of a class of reactions, which are only implicitly defined, at the cost of coarse-graining the representation of the rates of these reactions, which in principle may each be unique.

Kinetic Monte Carlo (KMC) methods have been developed for simulating the stochastic dynamics of rule-based models [15, 16, 17]. The methods of Danos et al. [15] and Yang et al. [16] avoid the requirement of specifying a chemical reaction network prior to simulation by directly sampling a rule list to generate reaction events and updating the system state in accordance with reaction rules. The computational cost is essentially independent of network size because a rule list is typically more compact and orders of magnitude smaller than the size of the corresponding reaction network.

The rule-based KMC simulation method of Yang et al. [16] may require checking application conditions for rules and therefore may involve rejection sampling that excludes non-permissible trial sites. After a rule is selected, trial sites with suitable local properties are randomly sampled to potentially undergo the chemical transformation(s) specified by the rule. Trial sites are rejected if they are not compatible with the application conditions of the rule. Random site selection in the rejection sampling steps is a constant-time operation, but efficiency might be significantly compromised in cases where null events become a dominant fraction of all sampled events. This scenario happens when the method is used to simulate multivalent ligand-receptor interactions in a sol-gel region that yields a giant ligand-receptor aggregate at equilibrium, where sampled trial sites are very likely rejected because the vast majority of sites reside in the giant aggregate and are prohibited from binding to each other by an application condition (a model assumption) that prohibits intra-aggregate binding reactions to form cyclic aggregates. The method of Yang et al. [16] has been implemented in software called NFsim [18].

A network-free simulation method that is related to the method of Yang et al. [16] but that is free of null events (i.e., rejection free) has been reported by Colvin et al. [19]. This method has been implemented in software called RuleMonkey [19]. In this method, rates associated with rules are calculated exactly using various formulas for different types of reaction classes that can be defined by rules. The rates are initially calculated as sums but the rates are thereafter iteratively updated as reaction events take place. Exact rule rates are always available and the simulation method is rejection free. Although RuleMonkey and NFsim can be used to simulate benchmark problems and the performance properties of the two software tools can be compared [18], it is not clear that the results obtained in this way represent theoretical performance limits or irrelevant coding differences.

Here, to evaluate the efficiency of rejection and rejection-free site samplings in rule-based KMC simulations, we present problem-specific implementations of the rejection method of

NFsim and the rejection-free method of RuleMonkey and we use these implementations to compare the two methods. These implementations are significantly less complicated than the RuleMonkey and NFsim codes, which are each designed to perform simulations of any model that can be encoded in the BioNetGen language (BNGL) [13]. Instead, our implementations are tailored for a single but important class of models: models for multivalent ligand-receptor interactions. In such a model, we consider a population of ligands, each with m equivalent sites, interacting with a population of cell-surface receptors, each with n equivalent sites to generate acyclic ligand-induced receptor aggregates. The class of models for m -valent ligand interaction with n -valent receptor is ideal for evaluating network-free simulation algorithms, because the size of the reaction network encompassing populated chemical species can be tuned from a handful of reactions to very large numbers of reactions [16]. This class of models is also relevant for describing phenomena that are common in cell signaling systems, including of course ligand-induced receptor clustering, which is important for signal initiation in many cell signaling systems [20], and aggregation processes mediated by intracellular signaling proteins [21].

2. Rule-based kinetic Monte Carlo with rejection or rejection-free sampling

In this section, we summarize the rejection and rejection-free KMC methods for simulating rule-based models. In a biomolecular reaction system, particularly in a system involving protein-protein interactions for signal transduction, proteins reversibly bind to one another via their binding sites (e.g., conserved domains and linear motifs). Protein sites (amino acid residues) can also be modified, by phosphorylation, methylation, etc. The state of the system (or a system configuration) is determined by states of individual sites and their connectivities. The temporal dynamics of a biochemical reaction system is the evolution of the system state in time, which can be modeled as a Markov process that is generally described by the Chapman-Kolmogorov equation (or more specifically, the chemical master equation for our case) [22]:

$$\frac{dP(Y, t)}{dt} = \sum_{Y' \neq Y} [W(Y|Y')P(Y', t) - P(Y, t)W(Y'|Y)]. \quad (1)$$

$P(Y, t)$ is the probability that the system is found in state Y , and $W(Y'|Y)$ gives the transition rate from state Y to state Y' . In a conventional chemical reaction system, a state Y is defined by the concentrations of all chemical species. In a rule-based system, a state Y is defined by the site occupancy and connectivity of individual molecules. Analytical solutions to the above master equation are only possible for very simple systems. Direct numerical integration of the above ordinary differential equations is also intractable because the size of the state space is often enormous even for a mildly-complex system that involves interactions among a few protein types. To solve this problem, Monte Carlo simulation is often applied to conduct random walks over the state space and to generate stochastic state trajectories for a system of interest.

Here we consider a biochemical system that is described by a set of reaction rules (see Fig. 1), $\mathbf{R} = \{R_1, \dots, R_M\}$. The system evolves in time with corresponding rates $\boldsymbol{\eta} = \{\eta_1, \dots, \eta_M\}$ for the rules. Such rate processes can be simulated using the standard kinetic Monte Carlo method [23]. The waiting time τ between two consecutive (reaction) events follows an

exponential distribution with a mean waiting time $\langle \tau \rangle = 1/\eta_{\text{tot}}$, where $\eta_{\text{tot}} = \sum_{i=1}^M \eta_i$ is the sum of the rule rates. The probability for a rule to be chosen is proportional to its rate. The rule-based kinetic Monte Carlo algorithm requires several additional operations per reaction event following the selections of τ and e : (1) select reactive sites admissible to the rule; (2)

excute the chemical transformation defined by rule e ; and (3) update rates in the vector η that are affected by the chemical transformation.

To choose reactive sites, we can write the rule-specific probability for a site x in a candidate set X_e to be chosen as:

$$P(x)=P_{a|s}(x)P_s(x). \quad (2)$$

We note that a “site” x denotes a set of interacting protein sites that are defined by the rule. For example, for a bimolecular interaction, x will include two interacting sites. This above equation describes a two-step procedure to find a site x : (a) choose a site x according to the sampling probability $P_s(x)$ and (b) accept the chosen x according to the acceptance probability $P_{a|s}(x)$. In the rejection method, a convenient sampling distribution $P_s(x)$ of candidate sites, usually a uniform distribution for maximum sampling efficiency, is used to generate trial sites and trial sites may be rejected according to the acceptance probability $P_{a|s}(x)$ so that the true probability $P(x)$ is recovered. Thus, the rejection method involves the sampling of nonreactive sites and introduces null events or a null process to compensate for the sampling of nonreactive sites. The null process generates a time step without a change in system state, i.e., a null event. The null process is triggered whenever a set of sampled sites is found to be nonreactive after being sampled. The rationale for rejection sampling is to gain efficiency by simplifying the procedure for sampling sites. If a rule has an application condition that is difficult to evaluate, it may be more efficient to sample sites with rejection rather than to explicitly track the sites that satisfy the application condition. We consider an application condition to be difficult to evaluate if satisfaction of the condition depends on a non-local property of a set of sites, such as the indirect connectivity of two sites [16] or the spatial configuration of sites/molecules within molecular complexes [24].

Generally, we can partition the rate of any rule into two parts:

$$\eta_e=\eta_{e,a}+\eta_{e,n} \quad e=1,\dots,M, \quad (3)$$

where $\eta_{e,a}$ and $\eta_{e,n}$ denote the actual reactive rate and the null process rate of rule e , respectively. Generally, the size of the null fraction of $\eta_{e,n}$ depends on the site sampling procedure which also affects the calculation of the rule rate. Operating with the combined rate η_e ignores any application condition by model assumption imposed on this rule in the step of site sampling. The model assumption can however be enforced by a subsequent rejection of the sampled sites that are non-permissible to react. Because the rejection method often can achieve random sampling of reactive sites, the cost per sampling step is constant and the implementation is usually straightforward. Determining whether to accept trial sites for reaction may however incur non-constant cost and complex bookkeeping, which depends on the nature of model assumptions.

The effect of rejection sampling on computational performance can be measured by the rejection ratio given as:

$$\theta=\frac{\eta_{\text{tot},n}}{\eta_{\text{tot}}}, \quad (4)$$

where $\eta_{\text{tot},n}=\sum_{i=1}^M\eta_{i,n}$ is the total rate of null processes. It should be noted that the rejection ratio is a function of the time at which rates in Eq. 4 are evaluated. Even at steady state, θ varies with time as a result of stochastic fluctuations in the steady-state rates of reactions.

For cases where θ tends to be close to 1, a majority of trial sites will be rejected and the rejection method will be inefficient. For such situations, a method without rejections (or with a substantially reduced number of rejections) is desirable. In contrast, the rejection-free method always accepts a sampled site x . In other words, $P_{a|s}(x)$ is unity for every trial site x and $P(x)$ is captured at the step of site sampling, i.e., $P(x) = P_s(x)$.

An early rejection-free Monte Carlo simulation method was developed by Bortz, Kalos and Lebowitz (BKL) [25] to simulate Ising spin systems, which overcomes the significant slowing down near the critical temperature by the classic Metropolis method [26]. The BKL algorithm is often called the “ n -fold way” method, and it has been widely used to simulate systems with high density and/or at low temperature. Gillespie later developed a stochastic simulation algorithm, equivalent to the BKL algorithm, for simulating coupled chemical reactions in a homogeneous reaction compartment [27]. Below, we describe the rejection-free procedure for simulating rule-based models.

Without loss of generality, we only consider unimolecular and bimolecular reactions. Higher-order reactions can usually be decomposed into these two types of elementary reactions. For a unimolecular reaction (e.g., a single-step site modification or a noncovalent bond dissociation between two proteins) of rule e that involves a single type of protein site A , at a given time, the candidate sites are defined in a set $X_e = \{a_i | i = 1, \dots, n_a\}$ that includes all legitimate sites designated by the rule and its application conditions. By the mean-field approximation, sites available according to a rule definition are considered to have equal probability to be chosen. The probability of a site $x = a_i$ to be chosen is

$$P(x=a_i)=P_s(x=a_i)=\frac{w_i}{Z} \quad (5)$$

where w_i is the reaction propensity for site a_i and Z is a partition function $Z = \sum_{X_e} w_i$. The quantity w_i reflects the biophysical properties that determine the reaction propensity of the site a_i . A rule defines reactions of sites from the same protein types and with identical properties. For this reason and under the law of mass action, we can interpret w_i as the specific rate (or, the rate constant) k_e for the rule. The partition function becomes the total rule rate, $Z = k_e |X_e|$, where $|X_e|$ denotes the size of the candidate set X_e (i.e., $|X_e| = n_a$). Effectively, Eq. 5 defines a uniform sampling distribution over X_e . The above analysis can be readily extended to the case of bimolecular reactions. Considering an interaction such as binding between two types of protein sites A and B , we have a candidate set $X_e = \{a_i, b_j | i = 1, \dots, n_a, j = 1, \dots, n_b\}$. The probability of a pair of sites $x = (a_i, b_j)$ to be chosen to react is

$$P(x=\{a_i, b_j\})=P_s(x=\{a_i, b_j\})=\frac{w_{ij}}{Z}. \quad (6)$$

Similarly, w_{ij} can be interpreted as a specific rate k_e that is identical for all site pairs. According to the law of mass action, the partition function $Z = \sum_{X_e} w_{ij} = k_e |X_e| = k_e n_a n_b$.

Based on the above considerations, we can summarize the rejection-free algorithm as follows.

- i. Initialization: assign copy numbers of proteins and specify initial states of individual protein sites, initialize lists of candidate sites for reaction rules, assign rate constants k for rules, and calculate the initial values of η ;

- ii. Sample a waiting time τ from an exponential distribution with a mean waiting time of $1/\eta_{\text{tot}}$. That is, the time step for the next event is calculated as $\tau = \ln(1/\rho_1)/\eta_{\text{tot}}$, where ρ_1 is a random deviate uniformly drawn from the interval (0, 1);
- iii. Sample a reaction rule e from the reaction rule list. The probability for a rule to be sampled is proportional to the rate of the rule. That is, identify the reaction rule e satisfying $\sum_{i=1}^e r_i > \rho_2 r_{\text{tot}}$, where ρ_2 is a random deviate uniformly drawn from the interval (0, 1);

Sample a set of protein sites x based on the true probability distribution $P(x)$ for site x to take part in the sampled reaction rule e . As described in Eq. 6, $P(x)$ is usually a uniform distribution over all candidates.

Update states of the sampled sites according to the rule specification by executing the chemical transformation defined by the rule;

Update lists of candidate sites for reaction rules and recalculate the rate vector η for all reaction rules;

Return to Step (ii), unless a stopping condition is satisfied.

The rejection algorithm [16] is largely the same as the rejection-free algorithm described above. The key differences relate to calculation of η and Step (iv). In the rejection algorithm, rates in η are overestimates of the true rates for those rates associated with rules involving application conditions that depend on non-local site properties (e.g., indirect connectivity or configuration within a molecular complex). Furthermore, Step (iv) is replaced with a step in which sites with local properties of genuine reactive sites are sampled and then evaluated to determine if the sites also have the non-local properties of genuine reactive sites. If the sampled sites are found to lack the non-local properties of genuine reactive sites, then the sites are rejected and a null event occurs, i.e., the procedure bypasses Steps (v) and (vi).

We note that the procedure described above does not include a step for calculating outputs (i.e., functions of the system state) for the sake of simplicity. Specified outputs can be calculated in a manner similar to the calculation of rates. In both rejection and rejection-free simulation, outputs are calculated via iterative updates, which occur upon changes in the state of a system.

In the following section, we apply the rejection-free method to simulate a multivalent ligand-receptor interaction model, and we discuss the implementation details of the rejection-free method within the context of this model.

3. Simulations of multivalent ligand-receptor binding systems

Cell-surface receptor aggregation induced by ligand binding is an important step in many signal transduction pathways [28]. In an earlier study, Goldstein and Perelson developed an equilibrium model to investigate binding interactions between trivalent extracellular ligands and bivalent cell-surface receptors [29]. The Goldstein-Perelson model predicted in theory that for certain parameter values small receptor aggregates and “superaggregates” can coexist at equilibrium. In fact, a sol-gel phase transition can happen in any multivalent ligand-receptor system in which either the ligand or the receptor has more than two binding sites and each ligand and receptor has at least two binding sites.

Recently, Yang et al. [16] developed a rule-based kinetic Monte Carlo method that was used to simulate the stochastic dynamics of a kinetic version of the Goldstein-Perelson model, which has been called the TLBR model, where TLBR represents “trivalent ligand – bivalent

receptor.” The gelation processes of ligand-receptor aggregation were observed in simulations, which was consistent with the Goldstein-Perelson theory of sol-gel phase transition. The simulation further revealed the kinetic details about formation of superaggregates. Monine et al. [24] also applied the same method to study steric effects using extended kinetic versions of the Goldstein-Perelson model. The KMC simulation method of Yang et al. [16] includes rejection sampling, which becomes inefficient when a system forms a superaggregate that contains most ligands and receptors. Below, we describe details on how to apply rejection-free sampling to simulate multivalent ligand-receptor aggregation.

3.1. Model of multivalent ligand-receptor binding

We consider a system with N_L ligands and N_R receptors. The ligand and receptor have v_l and v_r identical and independent binding sites, respectively. Three types of reactions are considered: (1) free ligand recruitment to a receptor on the cell surface, (2) crosslinking of two cell-surface receptors by a ligand that is already bound on the cell surface but has at least one free site, and (3) dissociation of a ligand-receptor bond. The rule-based specification of the TLBR model, a kinetic extension of the Goldstein-Perelson model, is illustrated in Fig. 2. The model has three reaction rules (Rules 1–3). To compare results with the Goldstein-Perelson model, we impose the same model assumptions: a) binding sites are equivalent [30], b) a ligand cannot associate with a receptor via more than one bond, and c) a ligand cannot associate with a receptor that is a member of the same aggregate (an aggregate is a ligand-receptor complex that has at least one ligand and one receptor in it). These two model assumptions prevent the formation of cyclic ligand-receptor aggregates and affect the calculation of rule rates, which can be accomplished via rejection or rejection-free sampling procedures for receptor crosslinkings. Using rejection sampling, one randomly samples a ligand-receptor site pair on the cell surface and rejects the trial pair when the ligand and receptor sites reside in the same aggregate.

3.2. Rule rates and site sampling

All rule rates are calculated according to the law of mass action. We first give equations for directly calculating η_1 and η_3 . Calculation of η_2 requires non-trivial treatment, and different treatments correspond to different implementations of the simulation algorithm (rejection or rejection-free method). We will explain these issues in detail.

The rate of free ligand recruitment to a cell-surface receptor for Rule 1, η_1 , is proportional to the product of the number of free sites on free ligands and the number of free receptor sites on the cell surface,

$$\eta_1 = k_{+1} v_l F_L (v_r N_R - N_B), \quad (7)$$

where F_L is the number of free ligands and N_B is the number of bonds which increases (decreases) by 1 upon an association (dissociation) event. We note that Rule 1 specifies a class of bimolecular reactions and the rate constant k_{+1} absorbs a volume factor (this also applies to the rate constant k_{+2}). The rate of ligand-receptor bond dissociation for Rule 3, η_3 , is proportional to the number of ligand-receptor bonds,

$$\eta_3 = k_{\text{off}} N_B. \quad (8)$$

For ligand-mediated receptor cross-linking on the cell surface (an event of Rule 2), to avoid forming cyclic ligand-receptor aggregates, one must ensure that an association event joins either two separate ligand-receptor aggregates or a ligand-receptor aggregate and a free receptor. We note that the probability for a ligand site to be chosen is proportional to the number of free receptor sites to which the ligand site may bind, i.e., the number of free receptor sites excluding ones in the same aggregate with the candidate ligand site. Once a ligand site is chosen, the probability of selecting a binding receptor site is uniform for all free receptor sites that are not in the same aggregate with the selected ligand site. At a given time, if the system has N_A ligand-receptor aggregates on the cell surface (excluding free receptors), the rate η_2 can be given by

$$\eta_2 = k_{+2} \sum_{i=1}^{N_A} l_i (v_r N_R - N_B - r_i), \quad (9)$$

where l_i and r_i are the numbers of free ligand and receptor sites in the i th aggregate, respectively. The term $(v_r N_R - N_B - r_i)$ is the total number of free receptor sites that are not in aggregate i . From the above equation, it follows that the probability that an aggregate i contributes a ligand site for a Rule 2 event is

$$P_i = \frac{k_{+2} l_i (v_r N_R - N_B - r_i)}{\eta_2}, \quad i=1, \dots, N_A. \quad (10)$$

We first sample this distribution to locate an aggregate i that provides a free ligand site and then select a free ligand site in aggregate i . All free ligand sites in aggregate i are equivalent (according to model assumptions) and therefore each of them has an equal probability to be chosen. Then, a free receptor site that is not in aggregate i is chosen to bind with the sampled ligand site. Similarly, all such free receptor sites (either in another aggregate or in the free receptor pool) are equivalent and therefore can be randomly sampled. The number of free ligand and receptor sites, l_i and r_i , can be calculated when the numbers of ligands and receptors in an aggregate are known. In an acyclic aggregate that has n_r receptors and n_l ligands, the number of bonds is given as $b = n_r + n_l - 1$, and the numbers of free receptor and ligand sites can be calculated as $r = v_r n_l - b$ and $l = v_r n_r - b$, respectively. The calculation of b is based on a common observation that a connected acyclic graph (equivalent to a tree) has a number of edges one less than the number of nodes [31]. A ligand-receptor aggregate can be represented as a bipartite graph, in which receptors and ligands are nodes and ligand-receptor bonds are edges. The numbers of receptors and ligands can be obtained by a linear-time traversal of the aggregate graph. A better approach is to maintain an aggregate list during simulation. This approach avoids routine traversal of aggregate graphs for each reaction event except for bond dissociation events, in which graph traversals are required to determine the resulting two separate aggregates. Direct calculation of η_2 using Eq. (9) has a linear-time cost scaled by the number of aggregates N_A . To achieve a constant-time cost per event, we update η_2 iteratively by accounting for the change caused by a reaction event. Eq. (9) can be rewritten as follows:

$$\eta_2 = k_{+2} \left([v_l (N_L - F_L) - N_B] (v_r N_R - N_B) - \sum_{i=1}^{N_A} l_i r_i \right). \quad (11)$$

The term $[v_l(N_L - F_L) - N_B](v_r N_R - N_B)$ accounts for all pair combinations of free ligand sites (on the cell surface) and free receptor sites under the law of mass action. We note that, in the method using rejection sampling [16], this quantity is used to calculate rate η_2 and a trial pair of ligand and receptor sites can be randomly chosen but are subject to rejection later if they are found in the same aggregate, which is the only reaction rule in the TLBR model that produces a null process because of rejection sampling. For our rejection-free

implementation, the exact rate is obtained by subtracting $u \equiv \sum_{i=1}^{N_A} l_i r_i$, where u is the number of intra-aggregate site-pair combinations, to accommodate the model assumption that prohibits intra-aggregate binding reactions. The first term in Eq. (11) can be easily recalculated with a constant cost after each reaction event. The second term can be updated iteratively by Δu (Table 1). Each reaction event changes the list of aggregates by creating a new one, or modifying or removing existing ones. For instance, a free ligand binding to a receptor may (if the receptor is also free) or may not (if the receptor is already bound to another ligand) create a new aggregate. However, most aggregates do not participate in a reaction event and thus remain unchanged. Therefore, updating Eq. (11) is straightforward by accounting for sites on one or two affected aggregates and it has a constant-time cost. For example, a merge of two aggregates i and j creates one new aggregate k (N_A decreases by 1). In this case, on the newly-formed aggregate, free ligand and receptor sites both decrease by 1 from the sums of those on the two merging aggregates, and it follows that $\Delta u = l_k r_k - l_i r_i - l_j r_j$, where $l_k = l_i + l_j - 1$ and $r_k = r_i + r_j - 1$. Formulas for all event types are given in Table 1.

The rejection method of Yang et al. [16] and the rejection-free method reported here are implemented in the C programming language and provided in the online Supplementary Material. These implementations simulate models for m -valent ligand and n -valent receptor that incorporate model assumptions about site equivalence and acyclic aggregates similar to those of the Goldstein-Perelson model. The rejection and rejection-free codes were written with the intention of making them as much alike as possible, except for the sampling strategies and rule rate calculations, to eliminate irrelevant differences. We note that our implementations of the simulation methods discussed here provide new tools for studying the equilibrium and the kinetics of multivalent ligand-receptor interactions and gelation on cell membranes, topics which have been studied intensely over the years via a variety of approaches [32, 33, 34, 35].

4. Simulation results

In this section, we present results obtained using the rejection and the rejection-free algorithms described above. Parameter values used in simulations were taken from Ref. [16].

To validate the accuracy of our problem-specific implementation of the rejection-free method, we compare simulation results with those obtained using ordinary differential equations (ODEs). The relaxation kinetics of a bivalent ligand ($v_l = 2$) – bivalent receptor ($v_r = 2$) system to equilibrium is shown in Fig. 3(a). The deterministic results were obtained by solving a small number of ODEs as described in Refs. [30, 36]. The stochastic trajectories recapitulate on average the deterministic solutions. In Fig. 3(b), the equilibrium distribution of receptor aggregates in a TLBR system agrees on average with the equilibrium results from the Goldstein-Perelson model. These results demonstrate that our problem-specific rejection-free simulation algorithm produces outcomes consistent with those obtained independently.

A phase transition in ligand-receptor clustering at equilibrium is predicted in the Goldstein-Perelson model by varying two dimensionless parameters $c_{\text{tot}} = 3k_{+1}N_L/k_{\text{off}}$ and $\beta = k_{+2}N_R/k_{\text{off}}$, which shows that finite-sized ligand-receptor aggregates can coexist with an infinite-sized polymer-like aggregate in a so-called sol-gel coexistence regime [29]. To investigate the stochastic dynamics of ligand-receptor aggregation and the algorithmic performance at different phase regimes, we use the average receptor aggregate size to measure the degree of ligand-receptor clustering, which can be calculated as

$$\Gamma = \sum_{n=1}^{N_R} n p_n = \frac{1}{N_R} \sum_{n=1}^{N_R} n^2 x_n, \quad (12)$$

where n is the size of a ligand-receptor aggregate measuring the number of receptors in the aggregate and x_n is the number of aggregates of size n . The term $p_n = nx_n/N_R$ is the fraction of receptors in aggregates of size n , which corresponds to the probability of finding an arbitrary receptor in an aggregate of size n . We note that in the above equation the number of aggregates containing single receptors, x_1 , accounts for all receptor monomers including free receptors. Therefore, at any given time the average aggregate size Γ takes a value between 1 and N_R . In practice, since the number of aggregates N_A is always much less than N_R (Fig. 4), Γ can be more efficiently calculated as follows:

$$\Gamma = \frac{F_R}{N_R} + \frac{1}{N_R} \sum_{j=1}^{N_A} n_j^2, \quad (13)$$

where F_R is the number of free receptors and n_j is the number of receptors in aggregate j . Γ can be also calculated iteratively during a simulation by accounting for the changes of aggregation caused by reaction events. To study the effect of the degree of aggregation on algorithmic efficiency by adjusting the parameters c_{tot} and β , we ran simulations of the TLBR model to equilibrium under a set of different values of the dissociation rate constant k_{off} in the range between 10^{-6} s^{-1} and 100 s^{-1} , with both N_L and N_R fixed at 5000. Figure 5(a) shows a sigmoid-like relationship between the average aggregate size $\langle \Gamma \rangle$ and the dissociation rate constant k_{off} . At smaller k_{off} (sol-gel region), $\langle \Gamma \rangle$ approaches its maximum value N_R , which indicates that a large aggregate containing most receptors exists in the system. At larger k_{off} (sol region), $\langle \Gamma \rangle$ approaches its minimum value, indicating that the majority of receptors are in the form of free receptors and ligand-bound receptor monomers. The mean aggregate number $\langle N_A \rangle$ exhibits a bell-shaped curve with a maximum value of 966 near the phase transition boundary at $k_{\text{off}} = 0.17$. Figure 5(b) shows the performance comparison between the rejection-free method described in this work and the rejection method of Yang et al. [16] in terms of CPU time per reaction event, for the TLBR model with a typical set of parameters. In most of the sol region ($k_{\text{off}} > 1 \text{ s}^{-1}$) and near the sol-gel region ($10^{-4} \text{ s}^{-1} < k_{\text{off}} < 10 \text{ s}^{-1}$), the rejection-free and rejection methods match each other in performance, and the efficiency of both methods deteriorates as $\langle \Gamma \rangle$ increases. The rejection-free method is affected by searching over a near maximum number of aggregates, which is reflected in the increases of CPU time per reaction event near the boundary of the phase transition ($10^{-1} \text{ s}^{-1} < k_{\text{off}} < 1 \text{ s}^{-1}$) compared to that of the rejection method. However, the rejection-free method is less sensitive to $\langle \Gamma \rangle$ and outperforms the rejection method in the sol-gel region ($k_{\text{off}} < 10^{-4}$), where free intra-aggregate ligand-receptor site pairs are dominant and the rejection method has a rejection ratio θ greater than 99% at equilibrium. As shown in Fig. 5(c), the rejection ratio θ has a log-linear dependence on $\langle \Gamma \rangle$ over a wide range.

To investigate the effect of valence, we varied the number of ligand binding sites v_l and kept receptor valence fixed at $v_r = 2$ (Fig. 5(b)), which can be accomplished experimentally through synthesis of multivalent ligands [37, 38, 39]. Compared to the rejection method of Yang et al. [16], the rejection-free method has better scaling of CPU time per reaction event with an increase of v_l . Except for the system with bivalent ligands, the rejection-free method is almost insensitive to changes in the number of ligand binding sites, whereas sampling by the rejection method involves large numbers of null events due to increases in intra-aggregate combinations of available ligand and receptor site pairs.

5. Discussion

We have presented problem-specific rejection and rejection-free methods for simulating multivalent ligand-receptor interactions described by reaction rules. Kinetic Monte Carlo procedures are applied to sample the rule list, identify reactant protein sites, and update protein states. In this procedure, all chemical species are formed dynamically. For this reason, the rule-based KMC methods have in general a computational cost independent of reaction network size measured by the numbers of chemical species or reactions. Although our implementations in this study are tailored to multivalent ligand-receptor interactions, general-purpose implementations of the rejection and rejection-free methods considered here are available [18, 19].

The implementation of rejection-free sampling described here in every Monte Carlo step changes the state of the system. Our implementation is closely related to the direct simulation Monte Carlo (DSMC) method developed to simulate coagulation processes where irreversible particle aggregation is considered [40]. The rejection-free method accounts for the exact rates of reaction rules and probability distributions of protein sites. In comparison to the rejection method, implementation of a rejection-free method demands more bookkeeping effort, especially for processing rules that require evaluation of non-local state information. In a more general and straightforward implementation, searching for reactant sites based on the exact probability distributions incurs a linear-time cost per reaction event scaled by the number of sites. In contrast, our rejection-free implementation uses a hierarchical search, which first searches over chemical species that contain candidate sites and after a set of chemical species is selected the algorithm searches for reactant sites in the chemical species. In the multivalent ligand-receptor interaction model, the number of chemical species during a simulation corresponds to the number of ligand-receptor aggregates, N_A . Although the rejection-free method has a higher cost per Monte Carlo step because of searching for candidate sites and the extra bookkeeping required for calculating probabilities for sites, the method outperforms the rejection method when the rejection ratio θ approaches unity.

A comparison between rejection-free and rejection methods was briefly discussed by Yang et al. [16]. The rejection-free algorithm reported by Yang et al. [16] takes a strategy of searching for a ligand-receptor site pair for cross-linking, which is less efficient for simulating formation of large ligand-receptor clusters in comparison to the algorithm presented in this report. This earlier algorithm maintains reaction probabilities of all sites for searching, which has a cost of CPU time per reaction event scaling with the number of molecules. In contrast, the current method searches over an aggregate list that is significantly smaller compared to the full list of molecule sites. For example, as shown in Fig. 5(a), the maximum number of aggregates on average (observed around $k_{\text{off}} = 0.1 \text{ s}^{-1}$) was less than one fourth of the total number of receptors ($N_R = 5000$). At other values of k_{off} , $\langle N_A \rangle$ is much less than N_R (Fig 5(a)). As shown in Fig. 4, at equilibrium $\langle N_A \rangle$ has a moderate (sublinear) dependence on the number of molecules.

The rejection and rejection-free methods have similar memory requirements (results not shown). Both require an amount of memory that depends largely on the total number of sites tracked in a simulation. However, the rejection-free method always requires more memory because a list of populated chemical species (i.e., the aggregate list) is tracked in addition to a list of individual sites, as discussed in the preceding paragraph. The extra amount of memory required may only be marginal. As can be seen in Fig. 4, the number of populated chemical species in a system can be orders of magnitude less than the number of molecules or sites.

Implementation of the rejection-free method is more complicated than implementation of the rejection method for at least two reasons. First, updates of rates associated with rules are more complicated. For example, as indicated in Table 1, in our rejection-free simulations, a number of different expressions are used to update the rate associated with Rule 2 in the TLBR model (Fig. 2). In contrast, in rejection simulations, a single expression is always used to obtain the overestimate of the rate associated with this rule [16]. Second, tracking of populated chemical species is an element of the rejection-free method but not the rejection method. As can be seen from inspection of the source codes of NFsim [18] and RuleMonkey [19], a general-purpose implementation of either method is complicated, and it is important not to rely on comparisons of such codes to assess methods. As we have shown here, in a comparison intended to control for irrelevant coding differences, the rejection and rejection-free methods appear to complement each other, with one performing better in one parameter regime and the other performing better in a different parameter regime; also, for many parameter values, the two methods appear to have comparable efficiencies (Fig. 5).

It should be noted that the rejection and rejection-free methods differ only in the handling of rules with application conditions that depend on non-local site properties. In principle, there is no barrier to treating two rules with such application conditions in different ways. Thus, the rejection and rejection-free methods could potentially be combined. A combination of the two methods might be advantageous if one rule among many is responsible for the vast majority of null events in a simulation via the rejection method. In this case, one might consider calculating the rate for this rule exactly, as in the rejection-free method, to reduce the number of null events. Rates for other rules could be determined via rejection sampling, which tends to be more efficient when the rejection ratio, θ , is small (Fig. 5). Monitoring this ratio during a simulation could allow for automatic selection of the best method for calculating each rate in η .

Both rejection and rejection-free methods require explicitly tracking connectivity between sites. This feature erodes the efficiency of simulation, in particular for simulating systems with large aggregates. In simulating multivalent ligand-receptor binding, to process an aggregate dissociation, at least one unweighted traversal of an aggregate subgraph is necessary (with no prior information about which subaggregate is smaller), which presents a major bottleneck to simulating systems in the sol-gel regime because a graph traversal has an order of cost proportional to the size of the aggregate. In the sol-gel regime, most graph traversals will happen on the giant aggregate that has a size close to that of the entire system. Experiments can provide some information about the composition of a protein aggregate but the intra-aggregate (or intra-multiprotein complex) topology usually cannot be resolved. For comparisons of model predictions against certain types of data, it may be sufficient to simulate a system without tracking the aggregate topology. In other words, this approach might be desirable when prediction of connectivity information (e.g., distribution of aggregate topologies for a particular aggregate size) is not of interest. One can refer to our recent algorithm (Q. Chang and J. Yang, arXiv:1010.0339), which improves simulation of models that account for large aggregates, such as a superaggregate [16, 29]. The algorithm avoids tracking connectivity between reaction sites and thus avoids routine graph traversal.

In summary, our results suggest that the rejection method is as efficient or more efficient than the rejection-free method for simulation of models that are not intended to capture the dynamics of large aggregates, which includes many models useful for studying cell signaling. In contrast, the rejection-free method is more efficient than the rejection method for studying the dynamics of large aggregates. Thus, the two methods are complementary.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. R. Faeder, M. I. Monine and Q. Chang for helpful discussions. We also thank the anonymous referees for their helpful critiques. This work was supported by NIH grants GM076570, GM085273 and RR18754, DOE contract DE-AC52-06NA25396, and NSFC grant 30870477. We thank the Center for Nonlinear Studies for support that allowed J.Y. to visit Los Alamos.

References

- Hunter T. Signaling—2000 and beyond. *Cell*. 2000; 100:113–128. [PubMed: 10647936]
- Pawson T, Nash P. Assembly of cell regulatory systems through protein interaction domains. *Science*. 2003; 300:445–452. [PubMed: 12702867]
- Hlavacek WS, Faeder JR, Blinov ML, Perelson AS, Goldstein B. The complexity of complexes in signal transduction. *Biotechnol Bioeng*. 2003; 84:783–794. [PubMed: 14708119]
- Hlavacek WS, Faeder JR, Blinov ML, Posner RG, Hucka M, Fontana W. Rules for modeling signal transduction systems. *Sci STKE*. 2006; 2006:re6. [PubMed: 16849649]
- Danos V, Feret J, Fontana W, Harmer R, Krivine J. Rule-based modelling of cellular signalling. *Lect Notes Comp Sci*. 2007; 4703:17.
- Mayer BJ, Blinov ML, Loew LM. Molecular machines or pleiomorphic ensembles: signaling complexes revisited. *J Biol*. 2009; 8:81. [PubMed: 19835637]
- Shapiro BE, Levchenko A, Meyerowitz EM, Wold BJ, Mjolsness ED. Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics*. 2003; 19:677. [PubMed: 12651737]
- Blinov ML, Faeder JR, Goldstein B, Hlavacek WS. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*. 2004; 20:3289–3291. [PubMed: 15217809]
- Lok L, Brent R. Automatic generation of cellular reaction networks with Molecuizer 1.0. *Nat Biotechnol*. 2005; 23:131–136. [PubMed: 15637632]
- Mallavarapu A, Thomson M, Ullian B, Gunawardena J. Programming with models: modularity and abstraction provide powerful capabilities for systems biology. *J R Soc Interface*. 2009; 6:257. [PubMed: 18647734]
- Gillespie DT. Stochastic simulation of chemical kinetics. *Ann Rev Phys Chem*. 2007; 58:35–55. [PubMed: 17037977]
- Ramaswamy R, González-Segredo N, Sbalzarini IF. A new class of highly Efficient exact stochastic simulation algorithms for chemical reaction networks. *J Chem Phys*. 2009; 130:244104. [PubMed: 19566139]
- Faeder JR, Blinov ML, Hlavacek WS. Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol Biol*. 2009; 500:113–167. [PubMed: 19399430]
- Feret J, Danos V, Krivine J, Harmer R, Fontana W. Internal coarse-graining of molecular systems. *Proc Natl Acad Sci USA*. 2009; 106:6453–6458. [PubMed: 19346467]
- Danos V, Feret J, Fontana W, Krivine J. Scalable simulation of cellular signaling networks. *Lect Notes Comput Sci*. 2007; 4807:139–157.
- Yang J, Monine MI, Faeder JR, Hlavacek WS. Kinetic monte carlo method for rule-based modeling of biochemical networks. *Phys Rev E*. 2008; 78:031910.

17. Colvin J, Monine MI, Faeder JR, Hlavacek WS, Von Hoff DD, Posner RG. Simulation of large-scale rule-based models. *Bioinformatics*. 2009; 25:910–917. [PubMed: 19213740]
18. Sneddon MW, Faeder JR, Emonet T. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat Methods*. 2011; 8:177–183. [PubMed: 21186362]
19. Colvin J, Monine MI, Gutenkunst RN, Hlavacek WS, Von Hoff DD, Posner RG. RuleMonkey: software for stochastic simulation of rule-based models. *BMC Bioinformatics*. 2010; 11:404. [PubMed: 20673321]
20. Metzger H. Transmembrane signaling: the joy of aggregation. *J Immunol*. 1992; 149:1477–87. [PubMed: 1324276]
21. Nag A, Monine MI, Faeder JR, Goldstein B. Aggregation of membrane proteins by cytosolic cross-linkers: theory and simulation of the LAT-Grb2-SOS1 system. *Biophys J*. 2009; 96:2604–2623. [PubMed: 19348745]
22. Van Kampen, NG. Stochastic processes in physics and chemistry. North-Holland: 2007.
23. Voter AF. Introduction to the kinetic Monte Carlo method. *Radiation Effects in Solids*. 2007; 235:1–23.
24. Monine MI, Posner RG, Savage PB, Faeder JR, Hlavacek WS. Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. *Biophys J*. 2010; 98:48–56. [PubMed: 20085718]
25. Bortz AB, Kalos MH, Lebowitz JL. A new algorithm for Monte Carlo simulation of Ising spin systems. *J Comput Phys*. 1975; 17:10–18.
26. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys*. 1953; 21:1087–1092.
27. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*. 1977; 81:2340–2361.
28. Goldstein B, Faeder JR, Hlavacek WS. Mathematical and computational models of immune-receptor signalling. *Nat Rev Immunol*. 2004; 4:445–456. [PubMed: 15173833]
29. Goldstein B, Perelson AS. Equilibrium theory for the clustering of bivalent cell surface receptors by trivalent ligands. Application to histamine release from basophils. *Biophys J*. 1984; 45:1109–1123. [PubMed: 6204698]
30. Perelson AS, DeLisi C. Receptor clustering on a cell-surface. I. theory of receptor cross-linking by ligands bearing two chemically identical functional groups. *Math Biosci*. 1980; 48:71–86.
31. Bollobás, B. Modern graph theory. Springer Verlag; 1998.
32. Perelson, AS. Some mathematical models of receptor clustering by multivalent ligands. In: Perelson, AS.; DeLisi, C.; Wiegel, FW., editors. *Cell surface dynamics: concepts and models*. Marcel Dekker; 1984. p. 223–276.
33. Lauffenburger, DA.; Linderman, JJ. *Receptors: models for binding, trafficking, and signaling*. Oxford University Press; 1996.
34. Guo C, Levine H. A thermodynamic model for receptor clustering. *Biophys J*. 1999; 77:2358–2365. [PubMed: 10545339]
35. Barisas BG. Aggregation and gelation of divalent cell surface receptors by rigid polyvalent ligands: examination by theoretical, kinetic and thermodynamic techniques. *Thermochim Acta*. 2003; 400:1–20.
36. Posner RG, Wofsy C, Goldstein B. The kinetics of bivalent ligand-bivalent receptor aggregation: ring formation and the breakdown of the equivalent site approximation. *Math Biosci*. 1995; 126:171–190. [PubMed: 7703593]
37. Mammen M, Choi SK, Whitesides GM. Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors. *Angew Chemie Int Edit*. 1998; 37:2754–2794.
38. Kiessling LL, Gestwicki JE, Strong LE. Synthetic multivalent ligands in the exploration of cell-surface interactions. *Curr Opin Chem Biol*. 2000; 4:696–703. [PubMed: 11102876]
39. Posner RG, Geng D, Haymore S, Bogert J, Pecht I, Licht A, Savage PB. Trivalent antigens for degranulation of mast cells. *Org Lett*. 2007; 9:3551–3554. [PubMed: 17691795]

40. Kruis FE, Maisels A, Fissan H. Direct simulation Monte Carlo method for particle coagulation and aggregation. *AIChE J.* 2000; 46:1735–1742.

Figure 1.

Diagrammatic depiction of a biochemical system described by rules and its underlying reaction network. Rules partition the entire reaction list into disjoint subsets which are consolidated by rules into rate processes denoted by $\{\eta_1, \eta_2, \dots, \eta_M\}$. The term r_{ij} denotes the j th reaction implied by rule. Note that partitions of a reaction network specified by rules are not necessarily of finite size, which is the case for many models of multivalent ligand-receptor interactions (see main text). Rule-based kinetic Monte Carlo samples the rule list (the intermediate layer) to generate reaction events and avoids simulating a system by means of a reaction network (the bottom layer).

Figure 2.

The interactions of a trivalent ligand and a bivalent cell-surface receptor (left). Graphical rules (right) that represent free ligand recruitment to the cell surface (Rule 1), receptor crosslinking by ligand (Rule 2) and ligand-receptor bond dissociation (Rule 3). Rules are parameterized by rate constants k_{+1} , k_{+2} and k_{off} , respectively. An empty (filled) circle inside a container denotes a free (bound) site. A connecting line indicates a ligand-receptor bond. An empty container indicates an arbitrary state (free or bound). An additional assumption is imposed to prohibit reactions defined by R_2 from forming cyclic ligand-receptor aggregates (see main text).

Figure 3.

Validation of the rejection-free simulation algorithm for multivalent ligand-receptor interaction systems. (a) Comparison of normalized results to the ODE solutions for the bivalent ligand and bivalent receptor interaction system. The normalization factor is N_R/n , where n is the number of receptors in an aggregate ($n = 2$ for dimers, $n = 3$ for trimers, etc.). The means and standard deviations of the simulation results obtained by the simulation method are shown on top of the continuous ODE solutions. One stochastic time trajectory is shown for the free receptor population. (b) Stochastic receptor aggregate distribution for the trivalent ligand and bivalent receptor system. The system reaches equilibrium after 350 seconds and the averages of equilibrium distributions match the results (solid curve) obtained using the model of Goldstein and Perelson [29]. Parameter values: $N_R = 300$, $N_L = 4200$, $k_{+1} = 6.67 \times 10^{-7} \text{ s}^{-1}$, $k_{+2} = 100k_{+1}$, $k_{\text{off}} = 0.01 \text{ s}^{-1}$. Initial condition: all simulations start with free ligands and free receptors without bonds.

Figure 4.

The relationship between $\langle N_A \rangle$ and N_R for the trivalent ligand–bivalent receptor system. The number of ligands N_L is set equal to the number of receptors N_R . $\langle N_A \rangle$ is obtained by calculating its mean value from 10^6 consecutive reaction events after the system reaches equilibrium. Other parameters and initial conditions are identical to the values indicated in Fig. 3.

Figure 5.

Performance comparison of the rejection-free method and the rejection method. (a) The mean average aggregate size $\langle \Gamma \rangle$ (solid line) and the average number of aggregates $\langle N_A \rangle$ (dashed line). The results are normalized by the total number of receptors, N_R . Simulation parameters: $\nu_r = 2$, $\nu_l = 3$, $N_R = N_L = 5000$. (b) Relationship between CPU time and k_{off} : rejection-free method (solid line), rejection method of Yang et al. [16] with aggregate bookkeeping (dashed line). (c) The expected rejection ratio $\langle \theta \rangle$ is plotted as a function of the average aggregate size $\langle \Gamma \rangle$. As can be seen, $\langle \theta \rangle$ appears to have a log-linear dependence on $\langle \Gamma \rangle$ over a wide range. The quantity $\langle \theta \rangle$ is calculated as the number of rejection events occurring in a simulation at equilibrium conditions divided by the total number of events required to capture 200,000 reaction events in the simulation. Other quantities reported in this figure, including CPU times per reaction event, $\langle \Gamma \rangle$ and $\langle N_A \rangle$, are calculated as averages over the 200,000 reaction events. Unless otherwise indicated, parameter values and initial conditions are the same as those given in the caption of Fig. 3.

Figure 6.

Scaling with number of binding sites on the ligand (v_l). For both the rejection-free (squares) and the rejection (circles) simulations, the receptor valence is fixed at $v_r = 2$ and the copy numbers of receptor and ligand are $N_R = N_L = 1000$. CPU times per reaction event are calculated by averaging over 200,000 reaction events after simulations reach equilibrium. Other parameters and initial conditions are the same as indicated in Fig. 3.

Table 1Formulas for calculating Δu_2

Rule	Event	Δu_2
1	A free ligand binds to a free receptor	$(v_r - 1)(v_l - 1)$
	A free ligand binds to aggregate i	$(v_l - 1)(r_i - 1) - l_i$
2	Aggregate i associates with a free receptor	$(v_r - 1)(l_i - 1) - r_i$
	Aggregates i and j associate to form a new aggregate k	$l_k r_k - l_i r_i - l_j r_j$
3	An aggregate breaks into a free ligand and a free receptor	$-(v_r - 1)(v_l - 1)$
	A ligand dissociates from aggregate i	$l_i - (v_l - 1)(r_i + 1)$
	A receptor dissociates from aggregate i	$r_i - (v_r - 1)(l_i + 1)$
	Aggregate k dissociates into two aggregates i and j	$l_i r_i + l_j r_j - l_k r_k$