Biophysical Journal, Volume 96

Supporting Material

Aggregation of Membrane Proteins by Cytosolic Cross-Linkers: Theory and Simulation of the LAT-Grb2-SOS1 System

Ambarish Nag, Michael I. Monine, James R. Faeder, and Byron Goldstein

SUPPLEMENTAL MATERIAL: Stochastic model of LAT-Grb2-SOS1 interactions

In this section, we present a rule-based kinetic Monte Carlo (KMC) algorithm developed to model interactions in the LAT-Grb2-SOS1 system. The algorithm is described elsewhere (1), and it is based on the well-known direct Gillespie method (2, 3). We assume that all reactants in the system are well-mixed. Due to the multivalent nature of interacting molecules the number of types of all possible molecular configurations can be very large. All of these configurations, however, need to be taken into account, and this makes the standard Gillespie algorithm inefficient. The main feature of our approach is that it avoids explicit generation of the set of all possible molecular configurations (commonly referred to as "chemical species") by tracking instead the states of a finite set of molecules to which transformations are applied. Transformations are selected based on sampling the set of possible reaction events determined by the reaction rules.

Data structures

We define LAT, Grb2 and SOS1 molecules as objects with multiple binding units. In general, an ith molecule of type A is represented by a set of addresses $A_i = \{p^B_{i1}, p^B_{i1}, ...\}$, where $p^B_{ij} = \{m^B_{ij}, s^B_{ij}\}$, m^B_{ij} is an index of a molecule of type B bound to jth binding unit on A_i through a unit on B with index s^B_{ij} . If there is no bond with another molecule, $p^B_{ij} = \{-1, -1\}$, i.e. empty. We define the following types of binding units: Grb2 binding sites on LAT and SOS1, and SH2 and SH3 domains on Grb2. For simplicity, we refer to all binding units as sites; we also use the following notations: L≡LAT, G≡Grb2 and S≡SOS1. Thus, taking into account the number and type of sites on molecules, we can define $L_i = \{p^{G-SH2}_{i1}, p^{G-SH2}_{i2}, p^{G-SH2}_{i3}\}$, $G_j = \{p^L_j, p^S_j\}$ and $S_k = \{p^{G-SH3}_{k1}, p^{G-SH3}_{k2}\}$ for every molecule of LAT, Grb2 and SOS1, respectively (i, j, k are indices of molecules).

To retain the connectivity between particular molecules and their binding sites, we use arrays of molecules. Note that LAT-Grb2-SOS1 interactions are structurally constrained: only linear chains and branched structures are possible. Thus, information about molecular connectivity is required to prevent formation of ring structures in the model. To determine the structure of multimolecular complexes, we apply the breadth-first traversal method (4). If two reactive sites selected for an association reaction belong to molecules of the same complex, this reaction must be rejected.

We assume that the reaction kinetics depend on the type of reactive

sites, occupancy of other sites at the same Grb2 and SOS1 molecules, and whether the molecules are in solution or tethered to a membrane. (See the reaction scheme in Fig. S1). We also assume that the binding kinetics at free LAT sites does not depend on the occupancy of other sites on the same LAT molecule. This approach enables the reduction of the reaction network of all possible configurations (single molecules and multimolecular chains) to a much smaller network of reactants representing sites on single molecules (L, G, S), complexes in solution (G-S, G-S-G), short chains attached to the membrane at one end (-L-G, -L-G-S, -L-G-S-G), and fragments of long chains (-L-G-S-G-L-) attached to the membrane at both ends. Note that in this scheme, long chains containing more than two LAT molecules are included naturally.

Based on the assumptions given above, we partition free reactive sites and bonds (pairs of bound sites) into lists and interaction classes. Each interaction class comprises reactive sites that can participate in a reaction of only one type. In accordance with the reaction scheme presented in Fig. S1, we define 10 lists of free reactive sites available for association reactions and 10 lists of bonds specified for dissociation reactions. Lists of reactive sites are specified in Tab. 1. Each class of association reactions contains two lists of free reactive sites, and each class of dissociation reactions contains one list of bound sites. The structure of interaction classes is defined in Tab. 2. The ith list of free reactive sites is defined as $Y_i = \{p^A_{i1}, p^A_{i2}, ..., p^A_{ij}, ...\}$, where $i = 1, ..., 10$, $p^A_{ij} = \{m^A_{ij}, s^A_{ij}\}$ is an address of a molecule of type A (A=G, L or S) with index m^A_{ij} and a site with index s^A_{ij}, j denotes a position in list Y_i. Similarly, each list of bound sites contains pairs of addresses, $Y_i = \{..., \{p^A_{ij}, p^B_{ij}\}, ...\}$, where $i = 11, ..., 20$, p^A_{ij} and p^B_{ij} are site addresses of two bound molecules A and B, respectively, index j denotes a position in the list. Note that p^A_{ij} and p^B_{ij} from the lists of bound sites correspond to the non-empty addresses in the arrays of molecules.

Reaction classes and rules

In the rule-based KMC method, molecules interact according to a set of reaction rules. Each rule is associated with a particular reaction type and gives a map of transformations of reactant lists. The rules are defined in such a way that reactive sites are sampled from classes $X_1, ..., X_{20}$ (see Tab. 2), and any reaction causes exchange of site addresses between the reactant lists. Updates associated with each rule are schematically shown in Fig. S2. Note that every reaction rule updates only those lists, which include molecular configurations affected by a reaction. To illustrate how the reactant lists are

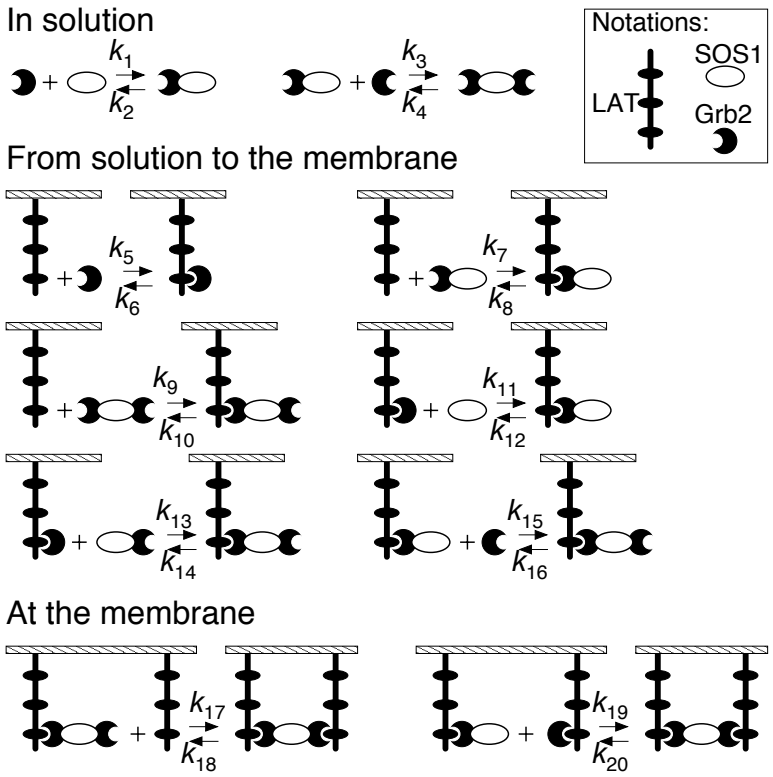


Figure S1: Scheme of the reduced reaction network in the LAT/Grb2/SOS1 system.

updated, we consider binding of free molecules of G and L. This reaction is associated with a shift of randomly selected p_{1i}^{G-SH2} from list Y_1 and p_{4j}^{L} from list Y_4 to list of bonds Y_{12}, see Rule 5 in Fig. S2. This reaction also modifies the probability of subsequent G + S interactions: when G is bound to the membrane through L, then its free SH3 domain can bind a membrane-bound S molecule with reaction rate constant k_{+GS}^{surf} (Rule 19), which is different from the rate constants in solution, k_{+GS} and σk_{+GS} (Rules 11, 13 and 15). Therefore, the surface configuration of G should be distinguished from its configuration in solution. Thus, an additional update for SH3 domain of the same G molecule is required: the address of this reactive site, p_{2l}^{G-SH3}, is shifted from Y_2 to Y_7.

Table 1: Lists of reactive sites in the rule-based KMC model.

Notation	Description of the site type and its context
Y_1	free SH2 domains on single G molecules
Y_2	free SH3 domains on single G molecules
Y_3	free sites on single S molecules
Y_4	free sites on single or bound L molecules
Y_5	free sites on S in complexes G-S
Y_6	free SH2 domains on G in complexes G-S
Y_7	free SH3 domains on G in complexes -L-G
Y_8	free SH2 domains on G in complexes G-S-G
Y_9	free sites on S in complexes -L-G-S
Y_{10}	free SH2 domains on G in complexes -L-G-S-G
Y_{11}	G-S bonds in bimolecular complexes G-S
Y_{12}	G-L bonds in complexes -L-G
Y_{13}	G-S bonds in complexes G-S-G
Y_{14}	G-S bonds in complexes -L-G-S
Y_{15}	G-L bonds in complexes -L-G-S
Y_{16}	LG-SG bonds in complexes -L-G-S-G
Y_{17}	LGS-G bonds in complexes -L-G-S-G
Y_{18}	G-L bonds in complexes -L-G-S-G
Y_{19}	G-S bonds in complexes -L-G-S-G-L-
Y_{20}	G-L bonds in complexes -L-G-S-G-L-

Algorithm of the rule-based KMC model

1. Compute rates for each reaction class i,

$$r_i = k_i \prod_{n=1}^{n_i} |X_{in}| \quad \text{and} \quad r_{\text{tot}} = k_i \sum_{i=1}^{20} r_i,$$

where n denotes a position of reactant list Y_j in class X_i, i.e., $X_{in} = Y_j$, and k_i is the rate constant for class i, see Tab.2.

2. Choose the time step according to

$$\tau = -\ln(\rho_1)/r_{\text{tot}},$$

where ρ_1 is a uniform random number on (0,1).

Table 2: Reaction classes in the rule-based KMC model.

Index	Molecularity, n_i	Reaction class	Rate constant
1	2	$X_1 = \{Y_2, Y_3\}$	k_{+GS}
2	1	$X_2 = \{Y_{11}\}$	k_{-GS}
3	2	$X_3 = \{Y_2, Y_5\}$	σk_{+GS}
4	1	$X_4 = \{Y_{13}\}$	k_{-GS}
5	2	$X_5 = \{Y_1, Y_4\}$	k_{+GL}
6	1	$X_6 = \{Y_{12}\}$	k_{-GL}
7	2	$X_7 = \{Y_4, Y_6\}$	k_{+GL}
8	1	$X_8 = \{Y_{15}\}$	k_{-GL}
9	2	$X_9 = \{Y_4, Y_8\}$	k_{+GL}
10	1	$X_{10} = \{Y_{18}\}$	k_{-GL}
11	2	$X_{11} = \{Y_3, Y_7\}$	k_{+GS}
12	1	$X_{12} = \{Y_{14}\}$	k_{-GS}
13	2	$X_{13} = \{Y_2, Y_9\}$	σk_{+GS}
14	1	$X_{14} = \{Y_{17}\}$	k_{-GS}
15	2	$X_{15} = \{Y_5, Y_7\}$	σk_{+GS}
16	1	$X_{16} = \{Y_{16}\}$	k_{-GS}
17	2	$X_{17} = \{Y_4, Y_{10}\}$	$\bar{k}_{+GL}$
18	1	$X_{18} = \{Y_{20}\}$	k_{-GL}
19	2	$X_{19} = \{Y_7, Y_9\}$	k_{+GS}^{surf}
20	1	$X_{20} = \{Y_{19}\}$	k_{-GS}

3. Choose a reaction rule by finding the smallest J such that

$$\rho_2 r_{\text{tot}} \leq \sum_{i=1}^{J} r_i,$$

where ρ_2 is a uniform random number on $(0, 1)$.

4. Choose a site for each reactant type M of the n_J types of sites in class J by picking $p^M \in X_{JM}$ at random.

5. Check if the two reactive sites selected for the association reaction are members of the same complex. If this is the case, reject the reaction. Otherwise, accept the reaction and update the lists of sites (as shown in Fig. S2), arrays of molecules and reaction rates.

6. Update time.

Steps 1-6 are repeated till a desired end time, t_{end}, is reached.

Modification of the model to account for the inhibitor

The inhibitor (the C-terminal fragment of SOS1) is introduced as a new monovalent species. Interaction of the inhibitor with Grb2 is characterized by constants k_{+GS}, σk_{+GS} and k_{-GS}. Taking advantage of the fact that both the inhibitor and SOS1 have the same kinetic properties, we do not need to specify additional reactant classes and reactant rules. The new species is defined as an extent to the array of SOS1 molecules. The total number of elements in this array is given by $S_T + I_T$. Thus, to distinguish between the two types of the reactant while making updates for sites of the selected SOS1 molecule with index m, the following rule is used: if $m \leq S_T$, the molecule is considered as a bivalent species and updates for its second site are required, otherwise, no updates should be done for the second site.

The numerical algorithm is implemented in the C language.

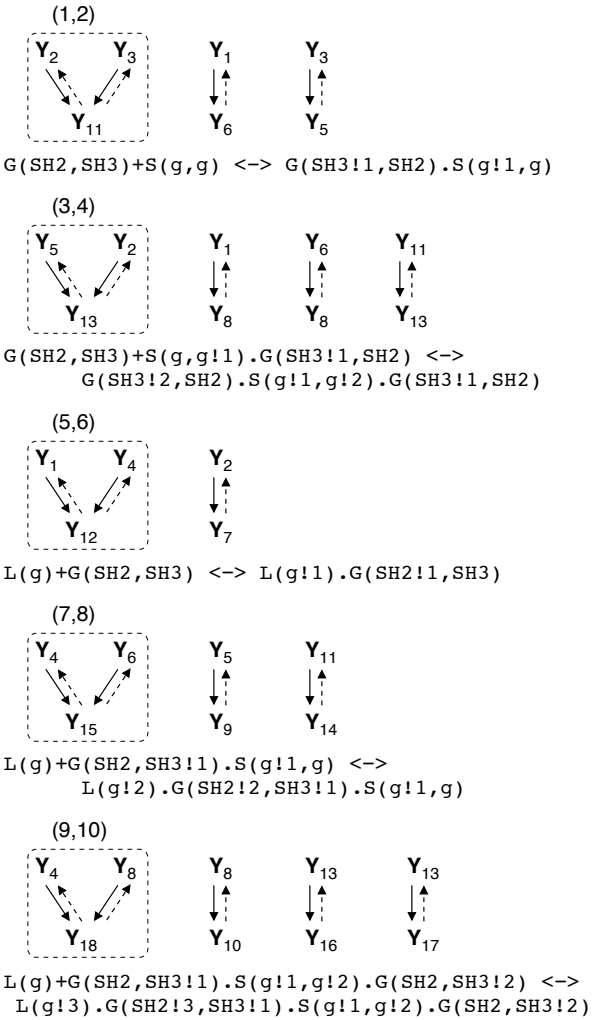
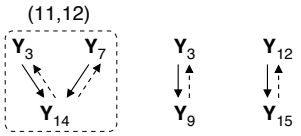
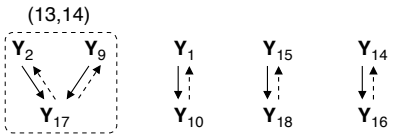


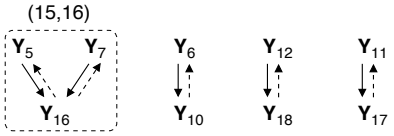
Figure S2: Rules for transformation of reactant classes. Rules for transformation of reactant classes. Numbers in brackets denote indices of association and dissociation reactions, respectively, in accordance with notations used in Tab. 2 and Fig. S1. See also Tab. 1 for definitions of reactant lists. For each association reaction, addresses of two sites are selected from the corresponding lists of free sites at random and moved to a list of bonds (in the directions shown by continued arrows). Updates associated with dissociation reactions are shown by dashed arrows. Frames outline lists of reactants and products of reactions. Transitions shown next to each frame denote updates for lists that include adjacent sites, i.e., the sites that do not participate in the reaction, but belong to the same interacting molecules. Reaction rules are also expressed in terms of the BioNetGen language (5, 6).



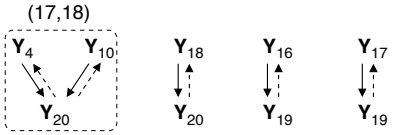

```
G(SH3,SH2!+)+S(g,g) <-> G(SH3!1,SH2!+).S(g!1,g)
```



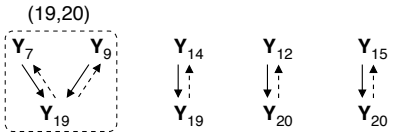

```
G(SH3,SH2!+)+S(g,g!1).G(SH3!1,SH2) <->
    G(SH3!2,SH2!+).S(g!2,g!1).G(SH3!1,SH2)
```




```
G(SH3,SH2)+S(g,g!1).G(SH3!1,SH2!+) <->
    G(SH3!2,SH2).S(g!2,g!1).G(SH3!1,SH2!+)
```




```
L(g)+G(SH2,SH3!1).S(g!1,g!2).G(SH2!3,SH3!2).L(g!3) <->
  L(g!4).G(SH2!4,SH3!1).S(g!1,g!2).G(SH2!3,SH3!2).L(g!3)
```




```
G(SH3,SH2!+)+S(g,g!1).G(SH3!1,SH2!+) <->
    G(SH3!2,SH2!+).S(g!2,g!1).G(SH3!1,SH2!+)
```

Figure S2 (continued): Rules for transformation of reactant classes.

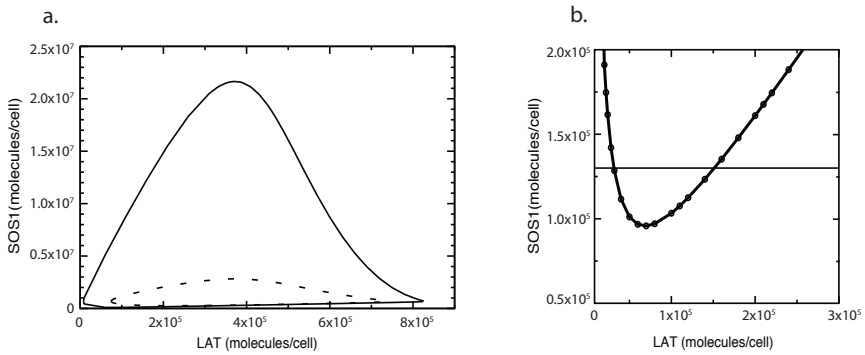


Figure S3 : Effect of changing the equilibrium crosslinking constant $\bar{K}_{GL}$ on the size of the sol-gel coexistence region. a. $\bar{K}_{GL} = 1.7 \times 10^{16}$ mole^{-1} cm^2 (solid curve) and $\bar{K}_{GL} = 1.7 \times 10^{15}$ mole^{-1} cm^2 (dashed curve). The Grb2 concentrations used in calculating the curves, $G_T = 1.3 \times 10^6$ molecules/cell, is the value estimated for Jurkat E6.1 cells. b. Re-plot of a portion of the solid curve in (a). The horizontal line corresponds to the concentration of SOS1 in Jurkat E6.1 cells, $S_T = 1.3 \times 10^5$ molecules/cell.

References

1. Yang J., Monine, M. I., Faeder, J. R., and Hlavacek, W. S. Kinetic Monte Carlo method for rule-based modeling of biochemical networks. *Phys. Rev. E.* 78:31910,2008.

2. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340, 1977.

3. Cao, Y., Li, H., and Petzold, L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J. Chem. Phys.* 121:4059, 2004.

4. Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C., Introduction to Algorithms. MIT Press, Cam-bridge, 2001, 2nd ed.

5. Hlavacek, W. S., Faeder, J. R., Blinov, M. L., Posner, R. G., Hucka, M., and Fontana, W. Rules for modeling signal-transduction systems. *Sci. STKE.* re6, 2006.

6. Blinov, M. L., Faeder, J. R., Goldstein, B., Hlavacek, W. S. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics.* 20:3289, 2004