



**HAL**  
open science

## Coarse-grained brownian dynamics simulation of rule-based models

Michael Klann, Loïc Paulevé, Tatjana Petrov, Heinz Koepl

► **To cite this version:**

Michael Klann, Loïc Paulevé, Tatjana Petrov, Heinz Koepl. Coarse-grained brownian dynamics simulation of rule-based models. The 11th Conference on Computational Methods in Systems Biology (CMSB'13), Sep 2013, Klosterneuburg, Austria. pp.64-77, 10.1007/978-3-642-40708-6\_6 . hal-00872963

**HAL Id: hal-00872963**

**<https://hal.science/hal-00872963>**

Submitted on 16 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Coarse-Grained Brownian Dynamics Simulation of Rule-Based Models

Michael Klann<sup>1</sup>, Loïc Paulevé<sup>1</sup>, Tatjana Petrov<sup>1</sup> and Heinz Koepl<sup>1,2</sup>

<sup>1</sup>BISON Group, Automatic Control Laboratory, ETH Zurich,  
Physikstrasse 3, 8092 Zurich, Switzerland

<sup>2</sup>IBM Research - Zurich, Rueschlikon, Switzerland

**Abstract.** Studying spatial effects in signal transduction, such as co-localization along scaffold molecules, comes at a cost of complexity. In this paper, we propose a coarse-grained, particle-based spatial simulator, suited for large signal transduction models. Our approach is to combine the particle-based reaction and diffusion method, and (non-spatial) rule-based modeling: the location of each molecular complex is abstracted by a spheric particle, while its internal structure in terms of a site-graph is maintained explicit; The particles diffuse inside the cellular compartment and the colliding complexes stochastically interact according to a rule-based scheme. Since rules operate over molecular motifs (instead of full complexes), the rule set compactly describes a combinatorial or even infinite number of reactions. The method is tested on a model of Mitogen Activated Protein Kinase (MAPK) cascade of yeast pheromone response signaling. Results demonstrate that the molecules of the MAPK cascade co-localize along scaffold molecules, while the scaffold binds to a plasma membrane bound upstream component, localizing the whole signaling complex to the plasma membrane. Especially we show, how rings stabilize the resulting molecular complexes and derive the effective dissociation rate constant for it.

## 1 Introduction

Signal transduction pathways can contain several proteins and activation steps which give rise to complex spatiotemporal dynamics. Interactions between signaling molecules do not only transmit activations, but can also localize the molecules to certain structures or compartments in the cell [20]. This work aims at a simulator that is able to handle the complexity, and includes the localization of the molecules. Current modeling techniques include ordinary differential equations (ODE), or, if space is important, the corresponding partial differential equations (PDE). If the stochasticity is significant, the chemical or reaction-diffusion master equation (CME/RDME) describes the system dynamics. More specifically, each molecule can be tracked individually in the simulation, e.g. using Brownian/Smoluchowski or Green's function dynamics (see e.g. review [13]).

Transient complex formation of proteins and their post-translational modification in signaling can lead to a combinatorial number of distinct molecular

species. Rule-based languages, such as Kappa [7] or BioNetGen [9], provide a compact representation of such combinatorial processes. Nonspatial rule-based models can be efficiently simulated [3], and are amenable to further quantitative analysis, e.g. formal model reduction [10, 11].

An extended Kappa framework was proposed in order to model the internal spatial structure of complexes that form during the process [4]. The Meredys simulator [21] in addition includes the molecule positions. Furthermore SRSim [12] provides a high resolution spatial rule-based extension, tracking the position and internal structure of all complexes. Indeed, the possibility to specify the binding angles sometimes naturally enforces a unique assembly path of a desired complex structure: for example, a polymer chain with angles between the bonds of  $\pi - 2\pi/N$  in a plane will form a ring of  $N$  monomers, and local rules are sufficient to describe the global structure [4]. However, the exact molecular geometry of signaling molecules is often not known. In such cases, the simulation including the binding angles becomes more complex, without contributing additional insights.

Here we present a framework that supports a particle-based, spatial simulation, but omits the internal geometry. Still, as the internal structure in terms of a site-graph is maintained explicit, their cooperative effect on complex stability can be investigated effectively. In particular, we derive the dissociation rate constant for rings. The simulation is applied to MAPK (Mitogen activated protein kinase) signaling in yeast, where both localization and activation is mediated by a scaffold [20].

The paper is structured as follows. In Sect. 2, the general particle-based framework and the biophysical principles of complex formation are introduced. The formal framework underlying the simulator is outlined in Sect. 3. In Sect. 4, we show the application to MAPK signaling and we discuss the results.

## 2 Coarse-grained Particle Diffusion and Reaction

The mobility of molecules in the cytoplasm is mainly governed by diffusion. Diffusion can be modeled efficiently by a random walk of the molecules of interest such that the myriad of solvent molecules can be omitted. We assume that the properties of the solvent allow us to use the Stokes-Einstein equation to obtain the (translational) diffusion coefficient  $D_i \propto r_i^{-1}$  for a given molecular radius  $r_i$  of particle  $i$ . We also assume that rotational diffusion is much faster than translational diffusion such that the actual shape of the molecules averages to a sphere with radius  $r_i$  at the temporal resolution of the method. Therefore only the position but not the orientation of the molecules has to be tracked.

### 2.1 Remark on Diffusion-Controlled Reactions

Molecules can only react with each other if they are in contact/collide, and the collision process is governed by diffusion. Accordingly the observable bulk/macrosopic reaction rate constant  $k_{ij}$  between particles  $i$  and  $j$  in solution is determined both by the rate constant of collisions  $k_D(i, j) = 4\pi(r_i + r_j)(D_i + D_j)$

(in 3D space) and the microscopic rate constant  $k'_{ij}$  that determines the reactive fraction of collisions [2]. For spherically symmetric molecules with isotropic reaction properties the microscopic and macroscopic rate constant are related in the form  $k_{ij}^{-1} = k_D(i, j)^{-1} + k'_{ij}{}^{-1}$  (Collins-Kimball model). Reactions with very high  $k_{ij} \approx k_D(i, j)$  require  $k'_{ij} \rightarrow \infty$ , i.e. every collision leads to a reaction such that they are diffusion-controlled. In contrast, reactions with  $k_{ij} \ll k_D(i, j)$  are reaction-controlled; in this case  $k'_{ij} \approx k_{ij}$ . A diffusion factor/function can be introduced  $f_D(i, j) := k_D(i, j)/(k_D(i, j) + k'_{ij})$  such that  $k = f_D k'$  [2].

More in detail, signaling molecules have specific reaction sites, i.e. non-isotropic reaction properties. Such molecules have to be in contact (by translational diffusion) and correctly aligned with their reaction sites (by rotational diffusion). We define the corresponding nanoscopic rate constant as  $k''$  and the conversion factor  $f_{DR}$  such that  $k = f_{DR} k''$ , however the derivation of  $f_{DR}$  is not straightforward and several approximations exist [2, 22]. For completeness we also introduce  $f_R = f_{DR} f_D^{-1}$  for the conversion  $k' = f_R k''$ , i.e. for integrating out the rotational diffusion effect only. In general:  $k \leq k' \leq k''$ .

Reversible reactions  $A + B \rightleftharpoons C$  with forward reaction at rate constant  $k_{AB}$  and backward  $k_C$  require to scale both rate constants with  $f(A, B)$  in order to maintain the macroscopic reaction equilibrium (dissociation constant  $K_d = k_C/k_{AB} = k'_C/k'_{AB} = k''_C/k''_{AB}$ ) [17, 14]. For consistency we define the respective conversion factors  $f = 1$  for unimolecular reactions that are not reversed by a bimolecular reaction, such that microscopic rate constants are always defined.

## 2.2 General Particle-Based Diffusion and Reaction Method

The present method implements the  $\lambda$ - $\rho$  model [8], i.e. a discrete time continuous space random walk for diffusion. Each particle position  $\mathbf{x}_i$  is updated by

$$\mathbf{x}_i(t + \Delta t) = \mathbf{x}_i(t) + \sqrt{2D_i \Delta t} \boldsymbol{\xi} \quad (1)$$

with diffusion coefficient  $D_i$  and standard normal random variable  $\boldsymbol{\xi}$ . Particles can overlap as discussed in [14], so collision testing is only needed for (static) reaction compartment boundaries. Reactions are executed with a probability that depends on their arity as follows.

**Unimolecular reactions**  $A \rightarrow \dots$  at rate constant  $k_A$  are executed in this method in every step with probability

$$P_A = 1 - \exp(-k'_A \Delta t) \approx k'_A \Delta t \text{ if } \Delta t \rightarrow 0 \quad (2)$$

for each molecule which is of type  $A$  [8]. Note that the Bernoulli-trial scheme leads to a binomial distribution, which converges to the Poisson distribution for small probabilities in each step (law of rare events). The Poissonian reaction process has exponentially distributed inter event waiting times as expected.

**Bimolecular reactions**  $A + B \rightarrow \dots$  can only occur if two molecules are closer than their collision distance ( $r_A + r_B$ ). If so, then the reaction is executed with probability

$$P_{A,B} = \frac{k'_{AB} \Delta t}{4\pi(r_A + r_B)^3/3} \quad (3)$$

as derived in [16]. The accuracy constraint  $P_{A,B} < 0.2$  gives an upper bound for  $\Delta t$ , if larger  $\Delta t$  are needed the reaction probability of [8] has to be used.

**Higher order and Hill-type reaction schemes:** for a particle-based simulation higher order reaction models have to be composed into their elementary uni- and bimolecular reaction steps (in which they also occur in nature).

### 2.3 Complexes

Complexes  $C$  that form out of  $A + B \rightarrow C$  can still be modeled as spherical particles. The complex radius  $r_C$  is obtained e.g. under the assumption that the volume/mass of  $A$  and  $B$  is redistributed into  $C$  with constant mass density, i.e.

$$r_C = \left( \frac{3}{4\pi} \left[ \frac{4\pi r_A^3}{3} + \frac{4\pi r_B^3}{3} \right] \right)^{1/3} = \left( \sum_{i \in C} r_i^3 \right)^{1/3} \quad (4)$$

as suggested in [23]. Alternatives for Eq. (4) are discussed and listed in Table 1 in [13]. The diffusion coefficient  $D_C$  is given by  $D_C = D_0/r_0 r_C^{-1}$  based on a reference  $D_0$  and  $r_0$  in the Stokes-Einstein relation. Reactions between complexes  $\alpha$  and  $\beta$  can occur when they are within their contact distance. The joint  $k'_{\alpha\beta} = \sum_{A \in \alpha} \sum_{B \in \beta} k'_{AB}$  (assume  $k'_{AB} = 0$  if no reaction between  $A$  and  $B$  is defined) could directly be used to calculate the binding probability by Eq. (3). However we decided to track each reaction individually as described in Sect. 3. Therefore also all resulting bonds in a complex are tracked individually. Such a bond in complex  $\alpha$  between  $A \in \alpha$  and  $B \in \alpha$  can break with the corresponding dissociation rate constant/probability defined for the interaction. But if the two formerly directly connected molecules  $A \in \alpha$  and  $B \in \alpha$  are still connected by other bonds, they will stay together – and **aligned with their binding sites**. Therefore  $A$  and  $B$  can rebind in every step with  $P_{AB}$ , leading to bond recovery with rate constant

$$\bar{k}_{AB} = \frac{3k''_{AB}}{4\pi(r_A + r_B)^3} \quad (5)$$

effectively in a first order reaction (cf. Eq. 3 and Eq. 2) [15]. The recovery reaction does not take place in the (relatively large) reaction volume of the whole reaction compartment but in the (relatively tiny) interaction volume of the two agents. Therefore  $\bar{k}$  is large compared to all other rates, and the lifetime of the open state of the bond  $\tau \propto \bar{k}^{-1}$  is accordingly relatively small (cf. Test Case 2).

## 3 Spatial Stochastic Simulation of Rule-based Model

Each particle of the simulation is an instantiation of a molecular species. A molecular species can be a protein, its post-translationally modified form or a protein complex that consists of proteins bound together. In order to reflect this internal structure of molecular species we represent them by *site-graphs*, in which modifications of protein residues and bonds are explicitly encoded, as introduced in Kappa [7].

*Notations.* We denote by  $\{e_1 \mapsto v_1, \dots, e_N \mapsto v_N\}$  the mapping from distinct elements  $e_i$  to values  $v_i$ . Given a mapping  $A$ ,  $\text{dom}(A)$  denotes its domain, and  $A(e)$  the value associated to  $e$  in  $A$ . We also write  $A\{e \mapsto v\}$  for the mapping  $A$  updated so that  $e$  maps to  $v$ ; and  $A \setminus e$  for the mapping  $A$  updated so that the mapping from  $e$  is removed.

### 3.1 Site-graphs

A site-graph is an undirected graph where typed nodes have sites, and edges are partial matchings on sites. Moreover, the sites which do not serve for forming edges are called *internal*, and they are assigned a value from a predefined set. The nodes of the site-graph can be interpreted as protein names, and sites of a node stand for protein binding domains. Internal states are used to encode post-translational modifications.

Let  $\mathcal{S}$  denote the set of site labels, and  $\mathcal{I}$  the set of internal values that can be assigned to sites. The function  $I : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{I})$  denotes the set of internal values that a site  $s$  can take. Let  $\mathcal{A}$  be the set of node types. Each node type is being equipped with a set of sites, defined by a signature map  $\Sigma : \mathcal{A} \rightarrow \mathcal{P}(\mathcal{S})$ . Finally, the set of admissible bindings between sites is defined by the mapping  $\mathcal{E} : \mathcal{A} \times \mathcal{S} \rightarrow \mathcal{P}(\mathcal{A} \times \mathcal{S})$  so that if  $(a', s') \in \mathcal{E}(a, s)$  then necessarily  $(a, s) \in \mathcal{E}(a', s')$ . A rule-based model is defined over a fixed *contact map* defined by the tuple  $(\mathcal{A}, \Sigma, \mathcal{E}, I)$  that we consider constant in the rest of this section.

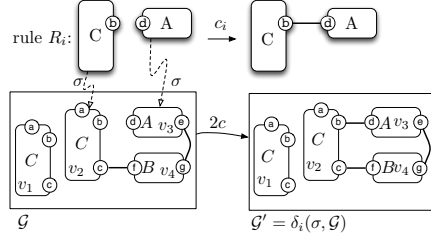
**Definition 1.** A site-graph is a tuple  $G = (V, T, F, E, \psi)$  with

1. a set of nodes  $V$ ,
2. a node type function  $T : V \rightarrow \mathcal{A}$ ,
3. a node interface function  $F : V \rightarrow \mathcal{P}(\mathcal{S})$ , such that for  $v \in V$ ,  $F(v) \subseteq \Sigma(T(v))$ ,
4. a set of edges between sites of different nodes, encoded by the function  $E : V \times \mathcal{S} \rightarrow V \times \mathcal{S}$  such that if  $E(v, s) = (v', s')$  then necessarily  $v \neq v'$ ,  $E(v', s') = (v, s)$ , and  $(T(v'), s') \in \mathcal{E}(T(v), s)$ .
5. a site evaluation function  $\psi : V \times \{s \in \mathcal{S} \mid I(s) \neq \emptyset\} \rightarrow \mathcal{I}$ , so that  $\psi(v, s) \in I(s)$ .

Site-graphs will be used in two different contexts: (i) to model physically existing complexes, also termed *concrete site-graphs* (or reaction mixtures), and (ii) to specify the local interaction patterns (rewrite rules). The concrete site-graphs must have all interfaces complete, in the sense that, for all nodes  $v \in V$ ,  $F(v) = \Sigma(T(v))$ .

**Definition 2.** Site-graph  $G = (V, T, F, E, \psi)$  is a *union* of two site-graphs  $G_1 = (V_1, T_1, F_1, E_1, \psi_1)$  and  $G_2 = (V_2, T_2, F_2, E_2, \psi_2)$ , denoted by  $G = G_1 \oplus G_2$ , if  $V_1 \cap V_2 = \emptyset$ , and  $V = V_1 \cup V_2$ ,  $F = F_1 \cup F_2$ ,  $E = E_1 \cup E_2$ ,  $\psi = \psi_1 \cup \psi_2$ .

**Definition 3.** Given a site-graph  $G = (V, T, F, E, \psi)$ , a sequence of edges  $((v_1, s_1), E(v_1, s_1)), \dots, ((v_k, s_k), E(v_k, s_k))$  such that for  $i = 1, \dots, k$ ,  $v_i \in V$ ,



**Fig. 1.** Rule application. Rule  $R_i$  can be applied to a reaction mixture  $\mathcal{G}$  via the embedding  $\sigma$  (indicated by the dotted arrows), and resulting in the reaction mixture  $\mathcal{G}'$ . The contact map is such that  $\mathcal{A} = \{A, B, C\}$ ,  $\mathcal{S} = \{a, b, c, d, e, f, g\}$ , and  $I(a) = \{u, p\}$ . More details on the formalism can be found e.g. in [19].

$s_i \in F(v_i)$ , and for  $i = 1, \dots, k-1$ ,  $E(v_i, s_i) = (v, s) \Rightarrow v_{i+1} = v$  and  $s_{i+1} \neq s$ , is called a *path* between nodes  $v_1$  and  $v_k$ . A site-graph  $G$  is connected, denoted  $\text{cc}(G)$  if there exists a path between every two nodes  $v$  and  $v'$ .

Two site-graphs can be related by an embedding function, which is important for defining the applicability of a rule to a reaction mixture (cf. Fig. 1).

**Definition 4.** The *embedding*  $\sigma$  between site-graphs  $G = (V, T, F, E, \psi)$  and  $G' = (V', T', F', E', \psi')$ , is induced by a support function  $\sigma^* : V \rightarrow V'$ , if

1.  $\sigma^*$  is injective: for all  $v, v' \in V$ ,  $[\sigma^*(v) = \sigma^*(v') \Rightarrow v = v']$ ;
2. for all  $v \in V$ ,  $T(v) = T'(\sigma^*(v))$ ;
3. for all  $v \in V$ ,  $[s \in F(v) \Rightarrow s \in F'(\sigma^*(v))]$ ;
4.  $\{(v, s) \mapsto (v', s')\} \subset E \Rightarrow \{(\sigma^*(v), s) \mapsto (\sigma^*(v'), s')\} \subset E'$ ;
5.  $\{(v, s) \mapsto i\} \subset \psi \Rightarrow \{(\sigma^*(v), s) \mapsto i\} \subset \psi'$ .

If  $\sigma^*$  is bijective, then  $\sigma$  is an *isomorphism*. The set of embeddings between the site-graph  $G$  and  $G'$  is denoted by  $\text{embed}(G, G')$ . The set of all embeddings is denoted by  $\mathbb{E}$ .

### 3.2 Rule-based models

**Definition 5.** Consider three types of elementary transformations of site-graphs, denoted by  $\delta_{\text{ae}}, \delta_{\text{de}}, \delta_{\text{ci}} : \mathbb{E} \times \mathbb{G} \rightarrow \mathbb{G}$ , with the following form:

1.  $\delta_{\text{ae}}^{v, s, v', s'}(\sigma, G) = (V, T, F, E \cup \{(\sigma^*(v), s) \mapsto (\sigma^*(v'), s'), (\sigma^*(v'), s') \mapsto (\sigma^*(v), s)\}, \psi)$  (adding an edge)
2.  $\delta_{\text{de}}^{v, s}(\sigma, G) = (V, T, F, E \setminus E((\sigma^*(v), s)) \setminus (\sigma^*(v), s), \psi)$  (deleting an edge);
3.  $\delta_{\text{ci}}^{v, s, i}(\sigma, G) = (V, T, F, E, \psi \setminus \{(\sigma^*(v), s) \mapsto i\})$  (changing the state value),

where  $G = (V, T, F, E, \psi) \in \mathbb{G}$ ,  $\sigma \in \mathbb{E}$  induced by the support function  $\sigma^* : V_i \mapsto V$ ,  $V_i$  being a set of nodes such that  $v, v' \in V_i$ ;  $v \neq v'$ ,  $(T(v'), s') \in \mathcal{E}(T(v), s)$ , and  $i \in I(s)$ .

A *rule*  $R_i$  is a triple  $(G, \delta, c)$ , where  $G$  is a site-graph,  $\delta$  is of type  $\delta_{\text{ae}}$ ,  $\delta_{\text{de}}$ , or  $\delta_{\text{ci}}$ , and  $c$  is a non-negative real number. Applying the rule to a site-graph  $\mathcal{G}$  is unique for an embedding  $\sigma \in \text{embed}(G, \mathcal{G})$ , and results in  $\mathcal{G}' = \delta(\sigma, \mathcal{G})$  (for a rigorous explanation, see [5]). In particular, for the identity support function  $\sigma^* = \mathbb{I}$ , and for  $\mathcal{G} = G$ , we get  $G' = \delta(\mathbb{I}, G)$ , which is sometimes called the right-hand-side of the rule (cf. Fig. 1).

### 3.3 Stochastic Abstract Machine

This subsection defines the syntax and semantics of the abstract formal machine for our coarse-grained spatial stochastic simulation of rule-based models.

In our machine, a *complex* is associated to a connected concrete site-graph, denoted by  $\mathcal{G}$ , and a 3-D position. The site-graph of a complex can be modified according to the rules, and may undergo a split into two site-graphs (dissociation), adding a new complex in the machine; or a merge with another site-graph (association), removing the other complex from the machine. The radius of agents  $a \in A$  is determined by the constant  $r_a$ , and the radius of the connected site-graph  $\mathcal{G}$  is given by  $\text{rad}(\mathcal{G})$ , according to Eq. (4) (Sect. 2.3):

$$\text{rad}((V, T, F, E, \psi)) = \left( \sum_{v \in V} r_{T(v)}^3 \right)^{1/3} \quad (6)$$

The syntax of the abstract machine is given in Def. 6. A machine term  $M$  is a quadruple  $(t, C, X, \mathcal{R})$  where  $t$  is the current time,  $C$  is a map from a complex  $i$  to its current site-graph  $\mathcal{G}$ ,  $X$  is a map from a complex  $i$  to its position  $\mathbf{x} \in \mathbb{R}^3$ , and  $\mathcal{R}$  the set of rules, as defined in Sect. 3.2.

The execution of the machine at time  $t$  follows Eq. (7). First, the position of each complex is updated according to a Brownian diffusion during a fixed  $\Delta t$  time (diffuse function, Def. 7). Second, the active rules at time  $t$  are applied according to the probabilities introduced in Sect. 2. The new site-graphs and positions maps give the new term of the machine at time  $t + \Delta t$ .

$$\frac{\begin{array}{l} X' = \text{diffuse}(X, C, \Delta t) \\ (C', X'') = \text{react}(C, X', \mathcal{R}, \Delta t) \end{array}}{(t, C, X, \mathcal{R}) \rightarrow (t + \Delta t, C', X'', \mathcal{R})} \quad (7)$$

The **react** function (Def. 12) applies the rules embeddings that are active at time  $t$  in a random order. Each active rule embedding (Def. 11) is specified by the set of concerned complexes (either one or two), and the quadruple  $(\sigma, G, \delta, k)$  where  $(G, \delta, k)$  is a rule in  $\mathcal{R}$ , and  $\sigma$  is the actual embedding being a map from the nodes of the rule left hand-side  $G$  to the nodes of the concrete complex site-graphs. These rule embeddings gathers the embeddings from any rule in  $\mathcal{R}$  to a single complex site-graph  $C(i)$  (**unary** function) or to the union of two complex site-graphs  $C(i_1) \oplus C(i_2)$ , assuming that complexes  $i_1$  and  $i_2$  have their distance  $|X(i_1) - X(i_2)|$  less than the sum of their radii (**neighbors** function). A rule with embedding  $\sigma$  is actually applied if (1) the embedding is still valid, i.e., previous rules application have not interfered with it; (2) the random number  $\zeta_1$  uniformly distributed in  $[0; 1]$  is less than the rule probability.

The probability of applying a rule embedding (**prob** function, Def. 8) depends on the arity of the reaction and on the modification type, as described in Sect. 2. The application of rule embedding to the site-graphs and position maps is given by the **do** function (Def. 10). In the case of agent site values changes ( $\delta_{ci}$ ) or an internal bound creation ( $\delta_{ae}$  within one complex  $i$ ), the embedded rule modification is applied to the concerned site-graphs, without any side-effect. In the case



---

**Stochastic Abstract Machine Definitions**


---

**Definition 6.** Syntax of the abstract machine.  $i_1, \dots, i_N$  are the complexes identifiers, assumed all distinct.

$$\begin{aligned}
M &::= (t, C, X, \mathcal{R}) && \text{Time } t, \text{ complex maps } C \text{ and } X, \text{ rules } \mathcal{R} \\
C &::= \{i_1 \mapsto \mathcal{G}_1, \dots, i_N \mapsto \mathcal{G}_N\} && \text{Map from a complex } i \text{ to its site-graph } \mathcal{G} \\
X &::= \{i_1 \mapsto \mathbf{x}_1, \dots, i_N \mapsto \mathbf{x}_N\} && \text{Map from a complex } i \text{ to its position } \mathbf{x} \in \mathbb{R}^3
\end{aligned}$$


---

**Definition 7.** Euler-Maruyama integration of diffusion. Returns the new position maps of complexes after a Brownian diffusion during  $\Delta t$  time, where  $\zeta$  is a standard normal random variable, and  $D_0$  and  $r_0$  are constants (cf. Eq. (1), Sect. 2.2). Note: state dependent diffusion in different compartments is introduced in Appendix A.

$$\text{diffuse}(X, C, \Delta t) \triangleq \{i \mapsto X(i) + \zeta \sqrt{2D_i \Delta t} \mid D_i = D_0 r_0 / \text{rad}(C(i)), i \in \text{dom}(X)\}$$


---

**Definition 8.** Rule probability. The rates  $k'$ ,  $k''$  and  $\bar{k}$  refers to the rule rate  $k$  modified according to Sect. 2,  $\text{agents}(\delta_{\text{ae}})$  refers to the couple of agents concerned by the bound creation, and  $r_{a_i}$  is the radius of agent  $a_i$ .

$$\begin{aligned}
\text{prob}(\{i\}, \delta_{\text{ci}}, \sigma, k, C, \Delta t) &\triangleq k' \Delta t \\
\text{prob}(\{i\}, \delta_{\text{de}}, \sigma, k, C, \Delta t) &\triangleq k'' \Delta t \text{ if } \text{cc}(\delta_{\text{de}}(\sigma, C(i))) \text{ else } k' \Delta t \\
\text{prob}(\{i\}, \delta_{\text{ae}}, \sigma, k, C, \Delta t) &\triangleq \bar{k} \Delta t \triangleq \frac{3k'' \Delta t}{4\pi(r_{a_1} + r_{a_2})^3} \text{ if } \text{agents}(\delta_{\text{ae}}) = (a_1, a_2) \\
\text{prob}(\{i_1, i_2\}, \delta, \sigma, k, C, \Delta t) &\triangleq \frac{3k' \Delta t}{4\pi(\text{rad}(C(i_1)) + \text{rad}(C(i_2)))^3} \text{ if } i_1 \neq i_2
\end{aligned}$$


---

**Definition 9.** Complex formation and dissociation. Returns the modified site-graphs and positions mappings after a merge or a split of complexes. The condition  $i_2 \notin \text{dom}(C)$  ensures that the new complex  $i_2$  is a fresh identifier.

$$\begin{aligned}
\text{merge}(i_1, i_2, \delta, \sigma, C, X) &\triangleq ((C \setminus i_2)\{i_1 \mapsto \delta(\sigma, C(i_1) \oplus C(i_2))\}, X \setminus i_2) \\
\text{split}(i_1, \delta, \sigma, C, X) &\triangleq (C\{i_1 \mapsto \mathcal{G}_1, i_2 \mapsto \mathcal{G}_2\}, X\{i_2 \mapsto X(i_1)\}) \\
&\text{if } i_2 \notin \text{dom}(C), \mathcal{G}_1 \oplus \mathcal{G}_2 = \delta(\sigma, C(i_1)), \text{cc}(\mathcal{G}_1), \text{cc}(\mathcal{G}_2)
\end{aligned}$$


---

**Definition 10.** Rule application. Returns the modified site-graphs and positions mappings. Predicate  $\text{cc}(\mathcal{G})$  is true if and only if  $\mathcal{G}$  is connected (Def. 3, Sect. 3.1).

$$\begin{aligned}
\text{do}(I, \delta_{\text{ci}}, \sigma, C, X) &\triangleq (C\{i \mapsto \delta_{\text{ci}}(\sigma, C(i)) \mid i \in I\}, X) \\
\text{do}(\{i\}, \delta_{\text{ae}}, \sigma, C, X) &\triangleq (C\{i \mapsto \delta_{\text{ae}}(\sigma, C(i))\}, X) \\
\text{do}(\{i_1, i_2\}, \delta_{\text{ae}}, \sigma, C, X) &\triangleq \text{merge}(i_1, i_2, \delta_{\text{ae}}, \sigma, C, X) \text{ if } i_1 \neq i_2 \\
\text{do}(\{i\}, \delta_{\text{de}}, \sigma, C, X) &\triangleq (C\{i \mapsto \delta_{\text{de}}(\sigma, C(i))\}, X) \text{ if } \text{cc}(\delta_{\text{de}}(\sigma, C(i))) \\
\text{do}(\{i\}, \delta_{\text{de}}, \sigma, C, X) &\triangleq \text{split}(i, \delta_{\text{de}}, \sigma, C, X) \text{ if } \text{not } \text{cc}(\delta_{\text{de}}(\sigma, C(i)))
\end{aligned}$$


---

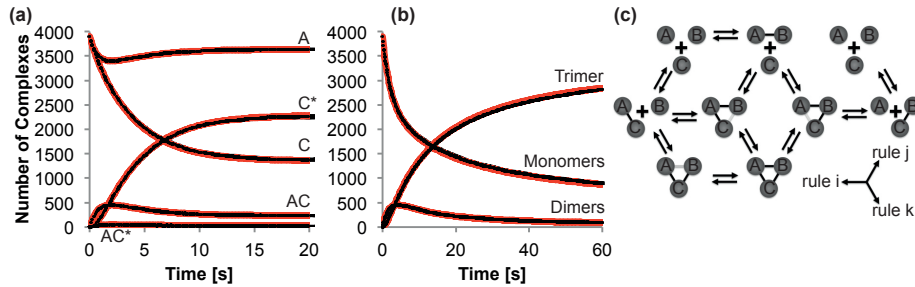
### Stochastic Abstract Machine Definitions (Continued)

**Definition 11.** Active embeddings. `unary` and `binary` returns the embedding specifications of rules  $\mathcal{R}$  to single or couples of complexes, respectively; `neighbors` returns the couples of complexes close enough to react; `act_embeds` returns the  $(I, E)$  couples where  $I$  is the set of complexes concerned by the embedding specification  $E$ .

$$\begin{aligned} \text{unary}(i, C, \mathcal{R}) &\triangleq \{(\sigma, G, \delta, k) \mid (G, \delta, k) \in \mathcal{R}, \sigma \in \text{embed}(G, C(i))\} \\ \text{binary}(i_1, i_2, C, \mathcal{R}) &\triangleq \{(\sigma, G, \delta, k) \mid (G, \delta, k) \in \mathcal{R}, \sigma \in \text{embed}(G, C(i_1) \oplus C(i_2)), \\ &\quad \sigma \notin \text{embed}(G, C(i_1)) \cup \text{embed}(G, C(i_2))\} \\ \text{neighbors}(C, X) &\triangleq \{\{i_1, i_2\} \mid \text{rad}(C(i_1)) + \text{rad}(C(i_2)) \geq |X(i_1) - X(i_2)|, \\ &\quad i_1, i_2 \in \text{dom}(C), i_1 \neq i_2\} \\ \text{act\_embeds}(C, X, \mathcal{R}) &\triangleq \{\{\{i\}, E\} \mid i \in \text{dom}(C), E \in \text{unary}(i, C, \mathcal{R})\} \\ &\quad \cup \{\{\{i_1, i_2\}, E\} \mid \{i_1, i_2\} \in \text{neighbors}(C, X), \\ &\quad E \in \text{binary}(i_1, i_2, C, \mathcal{R})\} \end{aligned}$$

**Definition 12.** Sequential application of active rule embeddings, *order is assumed to be random*; returns the modified site-graphs and position maps. The set of concerned complexes is denoted by  $I$ ,  $\oplus_{i \in I} C(i)$  denotes the union of site-graphs of complexes in  $I$ ,  $\zeta_1$  is a random variable uniformly distributed within 0 and 1, and `embed` is the set of embeddings from rule left hand-side to reaction mixture (cf. Sect. 3.1).

$$\begin{aligned} \text{apply}(\{(I, (\sigma, G, \delta, k))\} \cup Q, C, X, \Delta t) &\triangleq \text{apply}(Q, C, X, \Delta t) \\ &\quad \text{if } \sigma \notin \text{embed}(G, \oplus_{i \in I} C(i)) \\ &\quad \text{or } \zeta_1 \geq \text{prob}(I, \delta, \sigma, k, C, \Delta t) \\ &\triangleq \text{apply}(Q, C', X', \Delta t) \\ &\quad \text{if } \sigma \in \text{embed}(G, \oplus_{i \in I} C(i)), \\ &\quad \zeta_1 < \text{prob}(I, \delta, \sigma, k, C, \Delta t), \\ &\quad (C', X') = \text{do}(I, \delta, \sigma, C, X) \\ \text{apply}(\emptyset, C, X, \Delta t) &\triangleq (C, X) \\ \text{react}(C, X, \mathcal{R}, \Delta t) &\triangleq \text{apply}(\text{act\_embeds}(C, X, \mathcal{R}), C, X, \Delta t) \end{aligned}$$



**Fig. 2.** Comparison of particle-based simulation according to the presented algorithm (black symbols) and ODE model (red line, see Appendix B) for two test cases: (a) enzymatic activation, Test Case 1 and (b) trimerization, Test Case 2. (c) Overview of the reactions in the ring formation process of example (b), binding sites are omitted here.

of a complex formation ( $\delta_{ae}$  between two complexes  $i_1$  and  $i_2$ ), the two complexes are merged into one and receives the position of one of the two complexes in a non-deterministic manner (merge function, Def. 9). Finally, in the case of a bond deletion within a complex ( $\delta_{de}$ ), unless the site-graph is still connected, the complex is split in two complexes receiving the same position (split function, Def. 9).

## 4 Test Cases and Application

**Test Cases:** The accuracy of the simulation was tested on the principal reaction motifs of signaling models, where we assume the reaction-limited case for simplicity, i.e.  $k = k' = k''$ . The test cases are (in Kappa like syntax, see Fig. 1 for agents and sites):

1. Reversible enzymatic activation  $A(d) + C(b) \rightleftharpoons A(d^1), C(b^1)$ ,  
 $A(d^1), C(b^1, a_u) \rightarrow A(d^1), C(b^1, a_p)$  and  $C(a_p) \rightarrow C(a_u)$ .
2. Reversible trimerization by:  $A(d) + C(b) \rightleftharpoons A(d^1), C(b^1)$ ,  
 $A(e) + B(g) \rightleftharpoons A(e^2), B(g^2)$  and  $C(c) + B(f) \rightleftharpoons C(c^3), B(f^3)$

Fig. 2 depicts the simulation result which exactly match the ODE models derived in Appendix B, thus showing the correctness of the approach.

Test Case 2 exemplifies the formation of a ring and the cooperativity of the bonds in the ring. The ABC trimer can only dissociate if two bonds break and are open at the same time. For equal binding rate constants  $k_1$ , equal dissociation rate constants  $k_2$  and under the quasi-steady-state assumption for complexes with one broken bond the effective dissociation rate constant becomes

$$k^* = 6k_2^2 / (\bar{k}_1 + 2k_2) \quad (8)$$

as shown in Appendix B, Eq. (13). In a ring consisting of  $N$  agents the effective dissociation rate constant  $k^* = (N(N-1)k_2^2) / (\bar{k}_1 + (N-1)k_2)$ , will converge to  $Nk_2$ , i.e. with rising  $N$  the cooperativity vanishes. Especially small rings are therefore stabilized due to the high bond recovery rate constant  $\bar{k}$ . In the present example  $k_1 = 5 \times 10^5 M^{-1} s^{-1}$ ,  $k_2 = 0.2 s^{-1}$ ,  $r_A = r_B = r_C = 5nm$ ,  $D_A = D_B = D_C = 1 \mu m^2 / s$ . Then  $\bar{k}_1 = 198 s^{-1}$ , and  $k^* = 0.0012 s^{-1} \ll k_2$ , i.e. the ring structure is extremely stable. Diffusion limit  $k_D = 1.51 \times 10^8 M^{-1} s^{-1} \gg k_1$  i.e. reaction-limited regime.

**Signaling with Scaffolds in Space and Time:** In order to exemplify how the presented method can be used to model signaling in space and time, we simulate the yeast pheromone response signaling model from Thomson et al. [20] (see Fig. 3a) up to the MAPK Fus3. In the model the actual signaling molecules Ste4, Ste11, Ste7 and Fus3 can bind to the scaffold Ste5. In contrast to Thomson et al. we assume that any activation by the upstream molecule also involves a possible binding interaction between these molecules – although that interaction might be weak. The resulting additional bonds, shown in red in Fig. 3b, enable the formation of three rings around Ste5, which can stabilize



the signaling cascade is disrupted. In contrast, the additional bonds drive the equilibrium towards complete complexes.

The spatial and temporal dynamics of the activation process in the MAPK signaling cascade is depicted in Fig. 3c and d. Ste4 is a membrane bound protein, therefore all complexes containing Ste4 will likewise be membrane bound (cf. Appendix A). Fig. 3c shows how Ste5 accumulates at the plasma membrane accordingly. The 3D positions of molecules and complexes in the simulation are shown in Fig. 3e.

## 5 Discussion

The present work shows how a simulation for complex signal transduction models can be implemented, making use of a rule-based model description and including biophysical aspects as well as the spatial component. The present algorithm uses a coarse-grained description and simplified models for binding and dissociation rate constants on the macro-, micro- and nanoscopic level, thus refining the simulation method suggested in Meredys [21]. The applied rate constants and conversion factors as introduced in Sect. 2.1 could be further refined by more detailed models, which e.g. require to solve the reaction-diffusion or corresponding master equation [17, 18].

An accuracy and performance test of the particle-based simulator core in [14] shows that the presented method performs at least as good as Smoldyn [1]. Smoldyn however can handle rule-based models only via the libMoleculizer plugin. Association and dissociation of complexes require costly graph traversal. However, associations require that two complexes are in contact, which is a rare event, and dissociation does not occur more frequently (at steady state). Given the computational cost of the random walk of the molecules and neighbor finding alone, the rule-based extension does not dramatically slow down simulation. The performance of the algorithm can be improved by using a Gillespie scheme (draw exponential waiting time for reaction event) instead of the Bernoulli trials for unimolecular reactions as in [15]. In that case also the order of unimolecular reactions is given by the (ordered) waiting times instead of the random order we proposed in order to execute all reactions in an unbiased manner. For bimolecular reactions it is extremely unlikely that more than two complexes are within the reactive distance such that in most cases there is no ambiguity which reaction is to be applied. Further improvements of the performance could come from multi-scale or mixed approaches for different domains of the simulation [13].

The coarse-grained rate constants enable the calculation of emerging properties like the cooperativity between bonds, that stabilizes rings. In the MAPK signaling example also rings are formed around the scaffold Ste5. Future work can analyze the stability of signaling complexes and the dose-response curve for Ste5 now including the cooperative effect of the bonds. The simulation already includes localization sites that determine the location of the agents (cytoplasmic or membrane based as shown in Appendix A). In the future we are planning to include the transport into the nucleus such that the complete signaling path-

way can be analyzed. Furthermore formal model checking of spatial rule-based models has to be included to ensure meaningful models and simulation results.

## Additional Information

The appendix is available from [http://www.bison.ethz.ch/research/spatial\\_simulations\\_si/CMSB2013\\_Appendix.pdf](http://www.bison.ethz.ch/research/spatial_simulations_si/CMSB2013_Appendix.pdf) and the simulator as well as example files from [http://www.bison.ethz.ch/research/spatial\\_simulations](http://www.bison.ethz.ch/research/spatial_simulations).

## Acknowledgements

We thank Gerd Grünert for valuable discussions about spatial rule-based simulations and Pablo de Heras Ciechomski for the assistance with the visualization. **Funding:** M.K.: Swiss Commission for Technology and Innovation (CTI) project 12532.1 PFLS-LS in the joint ZigCell3D software development project with ScienceVisuals, Lausanne, Switzerland. L.P. and T.P.: Swiss SystemsX.ch initiative. H.K.: Swiss National Science Foundation, grant no. PP00P2\_128503.

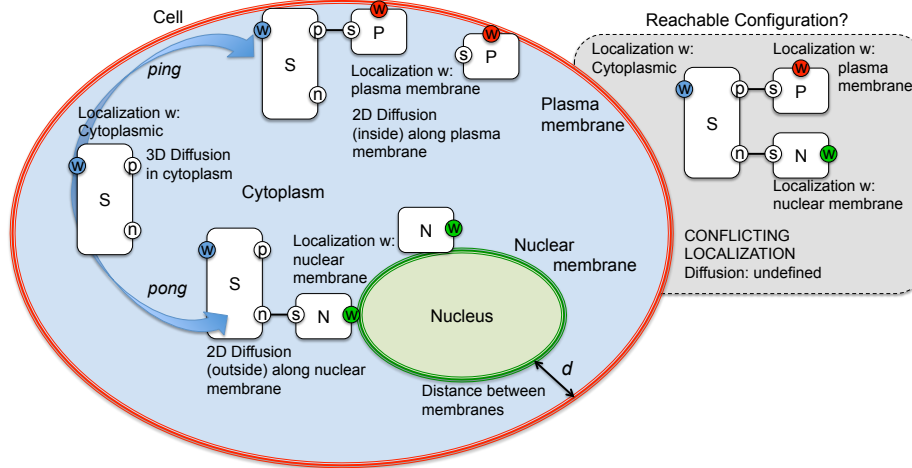
## References

1. Andrews, S.S., Addy, N.J., Brent, R., Arkin, A.P.: Detailed simulations of cell biology with smoldyn 2.1. *PLoS computational biology* 6(3), e1000705 (2010)
2. Berdnikov, V., Doktorov, A.: Steric factor in diffusion-controlled chemical reactions. *Chemical Physics* 69(1), 205–212 (1982)
3. Danos, V., Feret, J., Fontana, W., Krivine, J.: Scalable simulation of cellular signaling networks, invited paper. In: Shao, Z. (ed.) *Proceedings of the Fifth Asian Symposium on Programming Systems, APLAS'2007*, Singapore. *Lecture Notes in Computer Science*, vol. 4807, pp. 139–157. Springer, Berlin, Germany, Singapore (29 November – 1 December 2007)
4. Danos, V., Honorato-Zimmer, R., Riveri, S., Stucki, S.: Rigid geometric constraints for Kappa models. *Electronic Notes in Theoretical Computer Science* (2012)
5. Danos, V., Feret, J., Fontana, W., Harmer, R., Krivine, J.: Abstracting the differential semantics of rule-based models: exact and automated model reduction. In: *LICS 2010*. pp. 362–381 (2010)
6. Danos, V., Feret, J., Fontana, W., Krivine, J.: Abstract interpretation of cellular signalling networks. In: Logozzo, F., Peled, D.A., Zuck, L.D. (eds.) *Proceedings of the Ninth International Conference on Verification, Model Checking and Abstract Interpretation, VMCAI'2008*. *Lecture Notes in Computer Science*, vol. 4905, pp. 83–97. Springer, Berlin, Germany, San Francisco, USA (7–9 January 2008)
7. Danos, V., Laneve, C.: Formal molecular biology. *Theoretical Computer Science* 325(1), 69–110 (2004)
8. Erban, R., Chapman, S.: Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions. *Phys. Biol.* 6, 046001 (2009)
9. Faeder, J., Blinov, M., Hlavacek, W.: Rule-based modeling of biochemical systems with bionetgen. In: *Systems Biology*, pp. 113–167. Springer (2009)

10. Feret, J., Danos, V., Krivine, J., Harmer, R., Fontana, W.: Internal coarse-graining of molecular systems. *Proceedings of the National Academy of Sciences* 106(16), 6453–6458 (April 2009), <http://dx.doi.org/10.1073/pnas.0809908106>
11. Feret, J., Henzinger, T.A., Koepl, H., Petrov, T.: Lumpability abstractions of rule-based systems. In: Ciobanu, G., Koutny, M. (eds.) *MeCBIC. EPTCS*, vol. 40, pp. 142–161 (2010)
12. Gruenert, G., Ibrahim, B., Lenser, T., Lohel, M., Hinze, T., Dittrich, P.: Rule-based spatial modeling with diffusing, geometrically constrained molecules. *BMC bioinformatics* 11(1), 307 (2010)
13. Klann, M., Koepl, H.: Spatial simulations in systems biology: from molecules to cells. *International Journal of Molecular Sciences* 13, 7798–7827 (2012)
14. Klann, M., Koepl, H.: Reaction schemes, escape times and geminate recombinations in particle-based spatial simulations of biochemical reactions. *Physical Biology* accepted (2013)
15. Klann, M., Ganguly, A., Koepl, H.: Improved reaction scheme for spatial stochastic simulations with single molecule detail. *Proceedings of the International Workshop on Computational Systems Biology, Zurich, WCSB 2011* pp. 93–96 (2011)
16. Klann, M., Lapin, A., Reuss, M.: Stochastic Simulation of Reactions in the Crowded and Structured Intracellular Environment: Influence of Mobility and Location of the Reactants. *BMC Systems Biology* 5(1), 71 (2011)
17. Morelli, M., Ten Wolde, P.: Reaction Brownian dynamics and the effect of spatial fluctuations on the gain of a push-pull network. *J. Chem. Phys.* 129, 054112 (2008)
18. Mugler, A., Tostevin, F., ten Wolde, P.: Spatial partitioning improves the reliability of biochemical signaling. *Proceedings of the National Academy of Sciences* 110(15), 5927–5932 (2013)
19. Petrov, T., Feret, J., Koepl, H.: Reconstructing species-based dynamics from reduced stochastic rule-based models. In: *Proceedings of the Winter Simulation Conference*. p. 225. *Winter Simulation Conference* (2012)
20. Thomson, T., et al.: Scaffold number in yeast signaling system sets tradeoff between system output and dynamic range. *Proceedings of the National Academy of Sciences* 108(50), 20265–20270 (2011)
21. Tolle, D.P., Le Novère, N.: Meredys, a multi-compartment reaction-diffusion simulator using multistate realistic molecular complexes. *BMC systems biology* 4(1), 24 (2010)
22. Traytak, S.: Diffusion-controlled reaction rate to an active site. *Chemical physics* 192(1), 1–7 (1995)
23. Weiss, M., Elsner, M., Kartberg, F., Nilsson, T.: Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophys. J.* 87, 3518–3524 (2004)
24. ZigCell3D: from ScienceVisuals, [zigcell.sciencevisuals.com](http://zigcell.sciencevisuals.com)

## Appendix A: Spatial Rule-Based Model Checking

The following model has to be checked for undefined diffusion states:



A modifiable localization site on agents defines their location and mobility type. The value can be:

- 0 'null': no diffusion ( $D = 0$ ), i.e. static position ('bound to fixed anchor')
- 1 'pm': 2D diffusion along the plasma membrane ('bound to membrane 1').
- 2 'nm': 2D diffusion along the nuclear membrane ('bound to membrane 2').
- 3 '3D': 3D diffusion in the cytoplasmic domain ('unbound').

However, there can be reachable states with multiple location site values in one complex. In such a case, the smallest location value determines the behavior of the whole complex, and the states 'pm' and 'nm' become conflicting, because a complex can not safely be diffused along both membranes.

In order to prevent such a conflict, one has to ensure that no complex referencing both 'pm' and 'nm' location can be formed. While this condition can be enforced in the abstract machine by assigning a zero the probability to a binding between a complex with 'pm' location and a complex with 'nm' location, it is worth noticing that, in some cases, a conflict-free simulation can be warranted by static analysis ([6]).

Indeed, from `neighbors` function (Def. 11), two complexes  $i_1, i_2$  can merge only if they are close enough: for any machine term  $(t, C, X, \mathcal{R})$ ,  $\{i_1, i_2\} \in \text{neighbors}(X, C) \Leftrightarrow \text{rad}(C(i_1)) + \text{rad}(C(i_2)) \geq |X(i_1) - X(i_2)|$ . Assuming that complex  $i_1$  has location 'pm' and complex  $i_2$  has location 'nm',  $i_1$  and  $i_2$  may be merged only if  $\text{rad}(C(i_1)) + \text{rad}(C(i_2)) \geq d$ , where  $d$  is the minimal distance between nuclear and plasma membranes. It is hence sufficient that the sum of the radius of the largest possible complexes  $i_1$  and  $i_2$  is (strictly) less than  $d$  to ensure a conflict-free simulation. The largest complexes can be over-approximated from the contact map of the rule-based model, giving all the possible compositing agents. Note that in some models, it may be possible to have unbounded



complex sizes: in such a case, the sufficient condition can not be satisfied, and a dynamical disabling of rules leading to conflict has to be used.

## Appendix B: ODE Models for the Test Cases

The following ODEs describe the dynamics of the test cases in the macroscopic, well mixed limit.

**Test Case 1:** The reversible dimerization of  $A$  and  $B$  to  $AB$  is described by the concentration of the 3 species  $c_A$ ,  $c_B$  and  $c_{AB}$  and rates  $r_1 = k_1 c_A c_B$  for association and  $r_2 = k_2 c_{AB}$  for dissociation. In addition the test case case includes activation from  $B$  to  $B^*$  if bound to 'enzyme'  $A$  with rate constant  $k_3$  and spontaneous deactivation from  $B^*$  to  $B$  with rate constant  $k_4$ , independent of whether  $B^*$  is bound to  $A$ . The active form  $B^*$  has association and dissociation rate constants  $k_1^*$  and  $k_2^*$  respectively. This system has 7 rates, the 2 from above and  $r_3 = k_1^* c_A c_{B^*}$ ,  $r_4 = k_2^* c_{AB^*}$ ,  $r_5 = k_3 c_{AB}$ ,  $r_6 = k_4 c_{AB^*}$  and  $r_7 = k_4 c_{B^*}$ , which determine in the ODE system for the 5 species

$$\begin{aligned}
 \dot{c}_A &= -r_1 + r_2 - r_3 + r_4 \\
 \dot{c}_B &= -r_1 + r_2 && + r_7 \\
 \dot{c}_{AB} &= +r_1 - r_2 && - r_5 + r_6 \\
 \dot{c}_{AB^*} &= && + r_3 - r_4 + r_5 - r_6 \\
 \dot{c}_{B^*} &= && - r_3 + r_4 && - r_7
 \end{aligned}$$

**Test Case 2:** Symmetric binding of  $A + B$ ,  $B + C$  and  $C + A$  all with binding rate constant  $k_1$  and dissociation constant  $k_2$ . The full complex  $ABC$  contains three bonds in a ring structure. If a monomer (e.g  $A$ ) binds to a dimer (e.g.  $BC$ ), two bonds will have to be established accordingly, which is described by the joint rate constant  $k_3 = 2k_1$ . The bond recovery rate constant  $\bar{k}_1$  within the full complex is given by Equation (5). The rates are

$$\begin{aligned}
 r_1 &= k_1 c_A c_B & r_2 &= k_2 c_{AB} & r_3 &= k_1 c_B c_C & r_4 &= k_2 c_{BC} \\
 r_5 &= k_1 c_C c_A & r_6 &= k_2 c_{CA} & r_7 &= k_3 c_A c_B c_C & r_8 &= k_3 c_B c_C c_A \\
 r_9 &= k_3 c_C c_A c_B & r_{10} &= k_2 c_{ABC} & r_{11} &= \bar{k}_1 c_{AB.C} & r_{12} &= \bar{k}_1 c_{A.BC} \\
 r_{13} &= \bar{k}_1 c_{ABC} & r_{14} &= k_2 c_{AB.C} & r_{15} &= k_2 c_{A.BC} & r_{16} &= k_2 c_{ABC},
 \end{aligned}$$

where the dot in  $ABC$  denotes the location of the one open bond in the ring. Corresponding ODE System (where the multiplier in front of the rates has to be

chosen according to the number of bonds that are affected):

$$\begin{aligned}
\dot{c}_A &= -r_1 + r_2 - r_5 + r_6 - r_7 && + r_{14} + r_{16} \\
\dot{c}_B &= -r_1 + r_2 - r_3 + r_4 - r_8 && + r_{14} + r_{15} \\
\dot{c}_C &= +r_3 - r_4 - r_5 + r_6 - r_9 && + r_{15} + r_{16} \\
\dot{c}_{AB} &= +r_1 - r_2 && - r_9 && + r_{15} + r_{16} \\
\dot{c}_{BC} &= +r_3 - r_4 && - r_7 && + r_{14} + r_{16} \\
\dot{c}_{CA} &= +r_5 - r_6 && - r_8 && + r_{14} + r_{15} \\
\dot{c}_{ABC} &= && + r_7 + r_8 + r_9 - 3r_{10} + r_{11} + r_{12} + r_{13} \\
\dot{c}_{A.BC} &= && + 1r_{10} - r_{11} - 2r_{14} \\
\dot{c}_{AB.C} &= && + 1r_{10} - r_{12} - 2r_{15} \\
\dot{c}_{ABC.} &= && + 1r_{10} - r_{13} - 2r_{16}
\end{aligned} \tag{9}$$

The three species of the trimer with one broken bond  $A.BC$ ,  $AB.C$  and  $ABC.$  can be removed from the model using a steady state assumption:

- Lump the three states with one broken bond into the species  $XYZ.$  with balance equation in steady state

$$\dot{c}_{XYZ.} = +3k_2 c_{ABC} - \bar{k}_1 c_{XYZ.} - 2k_2 c_{XYZ.} = 0. \tag{10}$$

- Now  $c_{XYZ.}$  can be resolved as a function of  $c_{ABC}$ , namely:

$$c_{XYZ.} = \frac{3k_2}{\bar{k}_1 + 2k_2} c_{ABC}. \tag{11}$$

- The decay of the state  $ABC$  is given by the new rate  $r_{17} = -3r_{10} + \bar{k}_1 c_{XYZ.}$ . By inserting  $c_{XYZ.}$ , which was calculated above, the decay rate becomes

$$r_{17} = - \left( 3k_2 - \bar{k}_1 \frac{3k_2}{\bar{k}_1 + 2k_2} \right) c_{ABC} = \frac{2 \times 3k_2^2}{\bar{k}_1 + 2k_2} c_{ABC} \tag{12}$$

such that the effective decay rate constant of the trimer is

$$k_2^* = \frac{6k_2^2}{\bar{k}_1 + 2k_2}. \tag{13}$$

The solution of the resulting reduced ODE System

$$\begin{aligned}
\dot{c}_A &= -r_1 + r_2 - r_5 + r_6 - r_7 && + r_{17}/3 \\
\dot{c}_B &= -r_1 + r_2 - r_3 + r_4 - r_8 && + r_{17}/3 \\
\dot{c}_C &= +r_3 - r_4 - r_5 + r_6 - r_9 && + r_{17}/3 \\
\dot{c}_{AB} &= +r_1 - r_2 && - r_9 && + r_{17}/3 \\
\dot{c}_{BC} &= +r_3 - r_4 && - r_7 && + r_{17}/3 \\
\dot{c}_{CA} &= +r_5 - r_6 && - r_8 && + r_{17}/3 \\
\dot{c}_{ABC} &= && + r_7 + r_8 + r_9 - r_{17}
\end{aligned} \tag{14}$$

by Matlab shows only marginal differences if compared with the full system (9).