Simulation of Kohn's Molecular Interaction Maps Through Translation into Stochastic CLS+

Roberto Barbuti[1], Daniela Lepri[2], Andrea Maggiolo-Schettini[1],
Paolo Milazzo[1], Giovanni Pardini[1], and Aureliano Rama[1]

[1] Dipartimento di Informatica, Università di Pisa
{*barbuti,maggiolo,milazzo,pardinig,rama*}*@di.unipi.it*
[2] Institutt for Informatikk, Universitetet i Oslo
leprid@ifi.uio.no

Abstract. Kohn's Molecular Interaction Maps (MIMs) are a graphical notation for describing bioregulatory networks at the molecular level. Even if the meaning of Kohn's diagrams can be often easily understood, in many cases, due to the lack of a precise mathematical semantics, the notation can be ambiguous. By this paper we achieve two goals. Firstly, we give a precise meaning to MIMs by their translation into a formalism, the Stochastic Calculus of Looping Sequences (SCLS+), with a mathematical semantics. Further, by this translation we provide MIMs with all the tools developed for SCLS+, namely analysers and simulators. The ability of SCLS+ to specify compartments allows us to easily translate MIMs descriptions also when membranes are involved in the interactions.

1 Introduction

The definition of a diagrammatic graphical language able to describe biochemical networks in a clearly visible and unambiguous way is an important step towards the understanding of cell regulatory mechanisms. One of the most well designed and rigidly defined proposals of graphical language are Kohn's Molecular Interaction Maps (MIMs) [1, 10, 11]. In these maps, biochemical components of bioregulatory networks are depicted using a notation similar to the "wiring diagrams" used in electronics, and various types of interactions that may occur between the components can be represented. Interactions includes complex formations, phosphorylations, enzyme catalysis, stimulation and inhibition of biochemical reactions, DNA transcription, etc.

The use of a single MIM diagram to describe all the many interactions in a biochemical network allows the tracing of pathways within the network, for instance with the aid of computer simulation. However, even if the meaning of MIM symbols is clear and easy to understand, there is a lack of univocal interpretation when more symbols are combined, hence some diagrams cannot be used directly as an input for a simulation tool. This is confirmed by the distinction made by Kohn in [10] between *heuristic maps*, which may have more

than one interpretation, and *explicit maps*, that are unambiguous. The conclusion of Kohn is that only the latter should be used to perform simulations, by translating them into a list of chemical reactions.

In this paper we face the problem of allowing the simulation of a set of diagrams larger than the explicit ones. In particular, we consider the Stochastic CLS formalism [2], that is a formal language based on term rewriting for the description of biochemical systems. We recall its definition by adopting the syntax of CLS+ [15, 16] (we call this variant Stochastic CLS+). We show that, by translating MIMs into Stochastic CLS+ terms and rewrite rules, we can simulate more diagrams than the set of explicit ones.

As regards related work, in [4] a simple example of MIM has been modeled using the Beta–binders formalism [19]. Another example of translation of MIMs in a formal syntax (concurrent constraint programming) can be found in [3]. Moreover, other graphical languages for biochemical networks have been defined [12, 18, 20]. Among these, the notation introduced in [12] (which has been compared with MIMs in [13]) seems to be another promising proposal, as it has been used to model a real complex example of signalling pathway [17] and it is supported by useful software tools [8]. A different approach to the graphical description of biochemical networks based on graph rewriting is proposed in [6, 7]. Finally, present work is based on the approach proposed in [14].

2   Stochastic CLS+

In this section we recall the definition of the Stochastic Calculus of Looping Sequences (SCLS for short) and the syntax of an extention called CLS+, to define a variant we call Stochastic CLS+. For the sake of brevity we omit some technical details. Missing details of the calculus can be found in [2, 15, 16].

The Stochastic CLS+ formalism is basically a term rewriting framework including some typical features of process calculi for concurrency. A model in the Stochastic CLS+ is composed by a term describing the inital state of the modeled system, and a finite set of rewrite rules to be applied to terms, describing the events that may occur in the system. In the definition of Stochastic CLS+ terms, that follows, we assume an alphabet of elements $\mathcal{E}$ ranged over by $a, b, c, \ldots$.

Terms T, branes B and sequences S of the Stochastic CLS+ are given by the following grammar:

$$T ::= S \mid (B)^L \rfloor T \mid T \mid T \qquad B ::= S \mid B \mid B \qquad S ::= \epsilon \mid a \mid S \cdot S$$

We denote with $\mathcal{T}$, $\mathcal{B}$ and $\mathcal{S}$ the sets of all terms and all sequences, respectively. Note that $\mathcal{E} \subset \mathcal{S} \subset \mathcal{B} \subset \mathcal{T}$.

A term can be either (i) a sequence of symbols in $\mathcal{E}$ with ϵ as the empty sequence and $\cdot$ as the sequencing operator, or (ii) a looping sequence $(B)^L$ (that is a parallel composition of sequences, with $\mid$ as the parallel composition operator) containing another term T, with $\rfloor$ as the containment operator, or (iii) the parallel composition (the juxtaposition) of two terms.

The calculus comes with structural congruence relations on sequences $\equiv_S$ and terms $\equiv$. The former, $\equiv_S$, is defined as the least congruence on $\mathcal{S}$ for which $(S, \cdot, \epsilon)$ is a monoid, while the latter, $\equiv$, is the least congruence, including $\equiv_S$, closed under $(\cdot)^L \rfloor \cdot$ and such that $(\mathcal{T}, |, \epsilon)$ is a commutative monoid.

A rewrite rule is a pair of terms with variables, called *patterns* P_1, P_2 representing a portion of the described system before and after the occurrence of the modeled event. A rewrite rule can be applied to a term T if there exists a subterm of T which is structurally equivalent to an instantiation of P_1, by replacing the subterm with the corresponding instantiation of P_2. Variables $\mathcal{V}$ are of different kinds. *Element variables* $x, y, \ldots \in \mathcal{X}$, which can be instantiated by elements in $\mathcal{E}$, *sequence variables* $\widetilde{x}, \widetilde{y}, \ldots \in \mathcal{SV}$, for sequences in $\mathcal{S}$, *brane variables* $\overline{X}, \overline{Y}, \ldots \in \mathcal{BV}$ for branes in $\mathcal{B}$, and *term variables* $X, Y, \ldots \in \mathcal{TV}$ for terms in $\mathcal{T}$. Formally, patterns are defined as follows.

$$P ::= SP \quad | \quad (BP)^L \rfloor P \quad | \quad P \,|\, P \quad | \quad X \qquad BP ::= SP \quad | \quad BP \,|\, BP \quad | \quad \overline{X}$$
$$SP ::= \epsilon \quad | \quad a \quad | \quad SP \cdot SP \quad | \quad \widetilde{x} \quad | \quad x$$

We denote by $\mathcal{P}$ the set of all Patterns and by Σ the set of all instantiation functions $\sigma : \mathcal{V} \mapsto \mathcal{T}$. Instantiation functions are required to respect the types of the variables.

Finally, a *rewrite rule* is a triple (P', P'', f), where $P', P'' \in \mathcal{P}$ and $f : \Sigma \mapsto \mathbb{R}^+$, such that $P' \not\equiv \epsilon$ and $Var(P'') \subseteq Var(P')$.

In what follows, we will use the notation $P' \overset{f}{\mapsto} P''$ for a rewrite rule (P', P'', f), and we will call f the *rate function* of the rewrite rule. Such a function gives the rate value of the rule, depending on the actual term used for the instantiation of the variables in P'. We say that a rewrite rule (P', P'', f) is ground if both P' and P'' are ground.

Now, a Stochastic CLS+ model consisting of a set of rewrite rules $\mathcal{R}$ and of an initial ground term T_0, evolves by means of a sequence of application of rewrite rules in $\mathcal{R}$ to term T_0. After each application, a new term is obtained which describes the state of the system after the occurrence of the event modeled by the applied rule. At each step, the rule to be applied is randomly chosen with a probability which depends on an *actual application rate*. Such an actual rate is the value obtained by the rate function multiplied by the number of possible positions in the term where the rule can be applied. The actual application rate is used also as the parameter of an exponential distribution to determine the quantity of time spent by the occurrence of the described event.

More precisely, at each step a set of *applicable ground rewrite rules* $AR(\mathcal{R}, T)$ is computed which contains all the ground rules that can be applied to T and that are obtained by instantiating variables in the rules in $\mathcal{R}$. In each of these ground rules we have $r = f(\sigma)$, where f is the rate function of the rewrite rule from which it was istantiated by means of an instantiation function σ. By the finiteness of $\mathcal{R}$ and of T we have that $AR(\mathcal{R}, T)$ is a finite set of ground rewrite rules. For each ground rule R in $AR(\mathcal{R}, T)$ and for each possible term T' that can be obtained by the application R, the number of different application positions in T where R can be applied producing T' is computed. Such a number, called the *application cardinality* of R leading from T to T', is denoted as $AC(R, T, T')$,

and is the number that must be multiplied by the rate constant of R to obtain the actual application rate.

The semantics of the Stochastic CLS+ can now be given as a labeled transition system, in which a transition corresponds to the application of a rule and its label contains a reference to the applied rule and the actual application rate. In the definition of the semantics we use a notion of *context* to identify the position in the term where a rewrite rule is applied.

Formally, *Term Contexts C_T* and *Brane Context C_B* are given by the following grammar:

$$C_T \quad ::= \quad \square \quad | \quad C_T \,|\, T \quad | \quad T \,|\, C_T \quad | \quad (B)^L \,\rfloor\, C_T$$

$$C_B \quad ::= \quad \square \quad | \quad C_B \,|\, T \quad | \quad T \,|\, C_B \quad | \quad (C'_B)^L \,\rfloor\, T \quad | \quad (B)^L \,\rfloor\, C_B$$

$$C'_B \quad ::= \quad \square \quad | \quad C'_B \,|\, T \quad | \quad T \,|\, C'_B$$

where $T \in \mathcal{T}$, $B \in \mathcal{B}$ and $S \in \mathcal{S}$. Context $\square$ is called the *empty context*. $\mathcal{C_T}$ is the set of all the Term Contexts and $\mathcal{C_B}$ is the set of all Brane Contexts.

With $C[T]$ we denote the term obtained by replacing $\square$ with T in C. The structural congruence relation can be easily extended to contexts, namely $C \equiv C'$ if and only if $C[\epsilon] \equiv C'[\epsilon]$.

Definition 1 (Semantics). *Given a finite set of rewrite rules $\mathcal{R}$, the semantics of the Stochastic CLS+ is the least labeled transition relation satisfying the following inference rules:*

$$\frac{R = T' \overset{r}{\mapsto} T'' \in AR(\mathcal{R}, C[T']) \quad C \in \mathcal{C_T}}{C[T'] \xrightarrow{R, r \cdot AC(R, C[T'], C[T''])} C[T'']} \qquad \frac{R = B' \overset{r}{\mapsto} B'' \in AR(\mathcal{R}, C[B']) \quad C \in \mathcal{C_B}}{C[B'] \xrightarrow{R, r \cdot AC(R, C[B'], C_B[T''])} C[T'']}$$

The semantics of the calculus is a transition relation describing all the possible evolutions of the modeled system. From such a relation a Continuous Time Markov Chain (CTMC) can be easily derived, allowing the verification of properties of the system. However, since the whole CTMC describing the system has often a huge number of states, hence its construction is often unfeasible, we can follow a standard simulation procedure that corresponds to Gillespie's simulation algorithm [9]. A complete simulator for Stochastic CLS has been already implemented in F#, based on this simulation strategy [5].

3 Molecular Interaction Maps

In this section we recall the definition of Kohn's Molecular Interaction Maps (MIMs) and we show how they can be translated into Stochastic CLS+. We refer to the definition of MIMs that can be found in [11]. We present both MIMs and their translation incrementally, by showing first the diagrams for basic molecular interactions and then their extension with contingency symbols.

A species in a MIM is depicted as a box containing the species name (Fig. 1.a). In the case of a DNA site, the box is placed over a thick line representing a DNA strand, and more than one site can be placed over the same line (Fig. 1.c). Also a bullet (Fig. 1.b) is used to denote a species when it is the result of a reaction, and to denote different instances of the same species (see the dimer $AB{:}AB$ in Fig. 4).
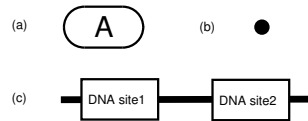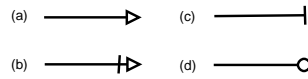
Fig. 1. Species in MIMs.
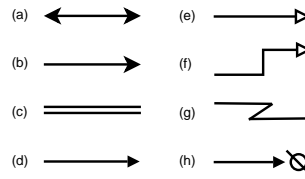

Fig. 2. Contingency arrows.

Fig. 3. Reaction symbols.

Basic Diagrams Basic MIM diagrams are composed by species and DNA fragments related each other by some reaction symbols. Reaction symbols are arrows, and they are listed in Fig. 3. In the figure, arrow (a) connects two species and denotes the reversible binding of them; (b) points to one species and denotes a covalent modification (phosphorylation, acetylation, etc. . .), the type of the modification is usually written at the tail of the arrow; (c) connects two species and denotes a covalent binding; (d) connects two species and denotes a stoichiometric conversion, namely the species at the tail of the arrow disappears while the pointed one appears; (e) is like (d) without the loss of the species at the tail of the arrow; (f) connects a DNA strand and a species and denotes DNA transcription; (g) represents the cleavage of a covalent bond; finally, (h) is connected to a single species and represents its degradation.

Extended Diagrams MIM diagrams can be extended to allow the use of modifiers to reactions: these modifiers are called "Contingencies". Contingency arrows start from a species (or a compound of species) and point to reaction, meaning that the presence (or absence) of the species influences the reaction. This influence is expressed as a change in the rate at which the reaction can happen. Contingency arrows are listed in Fig. 2. In the figure, arrow (a) stimulates the reaction pointed (increase its reaction rate) if the species is present in the environment; (b) is the necessity contingency and signals that the reaction can happen only if the species is present (if not present, the reaction rate drops to zero); inhibition contingency (c) is dual to the necessity one in that the reaction is allowed only if the species at hand is not present in the environment; finally, (d) is the catalysis arrow which signal that the reaction have a much higher reaction rate if the species is present than if it is not.

Example Here we will show a small example of a real life MIM. In Fig.4 one can see a MIM depicting a process of RNA synthesis inside the cell nucleus, activated by the presence of a species (B) outside the outer membrane of the cell through interaction with a membrane channel (group A). Species B can bind to group A that is present onto the plasma membrane forming group $A{:}B$. Two such groups
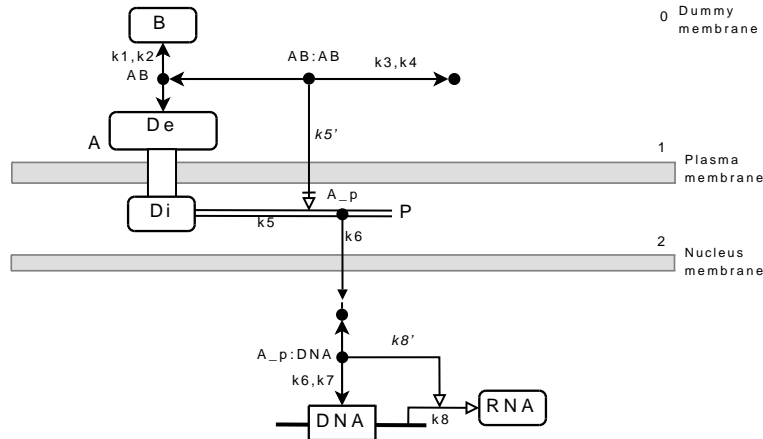
Fig. 4. Example of an explicit MIM with rate constants

can bind together, forming compound *AB:AB*, which allows the phosphorilation of a third *A* group which can then migrate inside the nucleus membrane. There, the group *A_p* can bind to a DNA fragment and together they stimulate the synthesis of a fragment of RNA.

4   Translating Maps into Stochastic CLS+

In this section we provide a precise semantics for MIMs by its translation into Stochastic CLS+. Since there is no formal description of MIMs, we introduce an *intermediate encoding* of MIMs, that allows us to formally specify the source of our translation. This intermediate encoding is sufficiently high level to allow a straightforward construction of the encoding from a diagram. Since we are interested in the possibility of simulating the MIMs, we need additional information about the initial number and position of the molecules of each species. This information, which is not provided by MIM diagrams, has to be included in the intermediate encoding.

4.1   Intermediate encoding of MIM

An intermediate encoding is composed of a *Membrane Structure*, a *Set of Species*, and a *Set of Reactions*. The first specifies the (static) membrane structure, the second describes position and initial number of the molecules of each species, the third captures the interaction capabilities of the species and their contingencies. We assume that membranes are uniquely identified by natural numebers in $\mathbb{N}$. We represent each different species (simple or compound) that appears in a MIM diagram as a simple CLS+ sequence S.

Definition 2 (Membrane Structure of a MIM). *A* Membrane Structure *MS is a set of tuples of the form* $\langle i, C \rangle \in \mathbb{N} \times \mathscr{P}(\mathbb{N})$, *where i is the* membrane

number*, that uniquely identifies it, and C contains the identification numbers of the membranes which are immediately contained in membrane i.*
We denote with $\mathcal{MS}$ the set of all possible Membrane Structures.

Definition 3 (Set of Species). *A* Set of Species *SS is a set of tuples of the form $\langle S, (i,j), q \rangle \in \mathcal{S} \times (\mathbb{N} \times \{0,1\}) \times \mathbb{N}$ where each tuple corresponds to a CLS+ sequence S (representing a species), whose position in the membrane hierarchy is described by (i,j), and whose quantity is q. In particular, i is the membrane id and j specifies if it is either on the surface of the membrane, if $j = 0$, or inside the membrane, if $j = 1$. We denote by $\mathcal{SS}$ the set of all possible Sets of Species.*

Definition 4 (Set of Contingencies, CS). *A* Set of Contingencies *CS is a set of pairs of the form $\langle C, k \rangle \in \mathcal{SS} \times \mathbb{R}$, where each pair corresponds to a contingency symbol, C is the Set of Species that must exist for the contingency to be verified and k is the reaction rate constant.*
We denote with $\mathcal{CS}$ the set of all possible Sets of Contingencies.

Definition 5 (Set of Reactions, RS). *Given a MIM, its* Set of Reactions *RS is a set of tuples of the form $\langle CS, R, P \rangle \in \mathcal{CS} \times \mathcal{SS} \times \mathcal{SS}$ where each tuple corresponds to a reaction symbol in the MIM. In particular, CS are the contingencies, and R and P specify reactants and products of the reaction, respectively. We denote with $\mathcal{RS}$ the set of all possible Sets of Reactions.*

The construction of an intermediate encoding of a MIM diagram is almost straightforward. First of all, assuming to identify each membrane with a natural number, and each species with a simple CLS sequence S, we can construct a Membrane Structure and a Set of Species by inspecting the MIM diagram. Please note that we have to specify, in the Set of Species, the initial quantity of each molecular species in each position where they appear. To simplify the translation, we assume the existance of a dummy outer membrane with $id = 0$ that contains the whole system.

For each reaction symbol of the MIM diagram, we create a tuple in the Set of Reactions specifying the contigency species, the reactants and products. Reactans specify the species that are needed for the reaction to happen, that are replaced by the product species once the reaction occurs. This representation is enough to describe the behaviour of any MIM reaction symbol (fig. 3).

As regards contingecies, the intermediate representation treats the different MIM symbols available (fig. 2) uniformly, by modeling each contigency as a set of species whose presence in the environment affects the (constant) basal rate of the reaction. First of all, the basal reaction rate to use when no contingeny are involved is also represented in the Set of Contingecies, by a tuple of the form $\langle \emptyset, k_0 \rangle$. We call it *neutral contingency*, and assume that each reaction has one. To avoid ambiguity in determining the basal rate of a reaction for which more than one contingecy is verified, we impose some constraints on the Set of Contingencies. In particular, we require that, for any two contingecies $\langle C_1, k_1 \rangle, \langle C_2, k_2 \rangle \in CS$ there exists a contingency $\langle C_1 \cup C_2, k_3 \rangle$ specifying the rate to use (k_3) when all the species of C_1 and C_2 are present. In this way, we can determine the basal

reaction rate as the one given by the *most specific* contingency, i.e. the one for which biggest set of contingency species are present.

4.2 From the intermediate encoding to Stochastic CLS+

The translation is composed of two parts. The first concerns translation of the initial state of the MIM system, described by its Membrane Structure and a Set of Species, into a CLS+ term. The second deals with the interaction symbols, by translating the Set of Reactions of the MIM into a set of rewrite rules.

Translation of the initial state In the following, we need to extract, from a Set of Species SS, its subset of species appearing on a precise position, i.e. on the surface or inside a given membrane. We denote this subset as $SS^{ij} \in \mathcal{SS}$, where i is the membrane id and j specifies if we need the elements on the surface of the membrane, if $j = 0$, or those inside, if $j = 1$. Formally, $SS^{ij} = \{\langle S, (i, j), q\rangle \in SS\}$. Moreover, we denote by $membrane(SS) \subset \mathbb{N}$ the ids of the membranes appearing in a given SS, i.e. $membrane(SS) = \{i \in \mathbb{N} | \langle S, (i, j), q\rangle \in SS\}$.

The following definitions formally describe the construction of the initial CLS+ term corresponding to a MIM.

Definition 6 (Set of species Translation Function, ϕ). *Given a set of species SS, the function $\phi : \mathcal{SS} \to \mathcal{T}$ constructs the corresponding CLS+ term formed by the parallel composition of all species with their multiplicities, i.e. $\phi\langle\{\langle S_1, p_1, q_1\rangle, \ldots, \langle S_n, p_n, q_n\rangle\}\rangle = q_1 \oplus S_1 | \cdots | q_n \oplus S_n$, where $n \oplus T$ stands for a parallel composition of n times T, that is $T | \ldots | T$ of length n $(0 \oplus T \equiv \epsilon)$.*

Note that the term produced by the function ϕ is always a brane $\mathcal{B}$ since it never contains a looping sequence operator.

Definition 7 (Term Translation Function, $[\![\cdot]\!]$). *Given a Membrane Structure $MS \in \mathcal{MS}$ and a Set of Species $SS \in \mathcal{SS}$, the Term Translation Function $[\![\cdot]\!]$ is defined by the following rule schema:*

$$\frac{\langle j_k, C_k\rangle \in MS \qquad [\![\langle j_k, C_k\rangle]\!] \mapsto T_k \qquad k = 1, \ldots, n}{[\![\langle i, \{j_1, \ldots, j_n\}\rangle]\!] \mapsto (i \,|\, \phi(SS^{i0}))^L \,\rfloor\, (\phi(SS^{i1}) \,|\, T_1 \ldots \,|\, T_n)}$$

Note that SS^{i0} is the set of species present on the surface of the membrane i while SS^{i1} contains the species inside membrane i.

Definition 8 (Initial Term of a MIM). *Given a Molecular Interaction Map, described by a Membrane Structure $MS \in \mathcal{MS}$ and a Set of Species SS, the initial CLS+ term is $T_0 = [\![\langle 0, C\rangle]\!]$, where $\langle 0, C\rangle \in MS$.*

The initial term is constructed by Term Translation Function applied to the outer membrane with $id = 0$. This translation computes a CLS+ term, where each membrane of the MIM is represented by a looping sequence, and their containment hierarchy is preserved. The species appearing on the surface and inside the membranes are put in the correct position in the corresponding looping sequences. Finally, a special symbol i is present on each looping sequence, denoting the membrane id from which it has been constructed.

Translation of the interactions Given a membrane structure MS and a set of nodes N, we denote by $subtree(MS, N) \in \mathcal{MS}$ the smallest subtree of MS induced by the nodes N, that is the minimal tree containing all the nodes in N. We also assume the function $root : \mathcal{MS} \to (\mathbb{N} \times \mathscr{P}(\mathbb{N}))$ that returns the root node of the membrane structure (where a node is the pair of its id and the set of ids of the nodes immediately contained in it).

Each reaction of a MIM, represented by an element of a Set of Reactions RS, is translated into a stochastic CLS+ rewrite rule. This translation is performed by the reaction translation function, which uses an auxiliary function, the parallel pattern builder, for constructing the patterns which compose the rewrite rule.

Definition 9 (Parallel Pattern Builder Function, $[\![\cdot]\!]_{pp}$). *Given a membrane structure MS, the* parallel pattern builder function $[\![\cdot]\!]_{pp} : (\mathbb{N} \times \mathscr{P}(\mathbb{N})) \times (\mathcal{SS} \times \mathcal{SS}) \to \mathcal{P} \times \mathcal{P}$ *is defined as follows:*

$$
\begin{array}{c}
X_0 = new(\mathcal{BV}) \qquad X_1 = new(\mathcal{TV}) \\[4pt]
\langle c_j, C_j \rangle \in MS \qquad [\![\langle c_j, C_j \rangle, (SS_1, SS_2)]\!]_{pp} \mapsto (P1_j, P2_j) \qquad j = 1, 2, \ldots, k \\[4pt]
\phi(SS_1^{i0}) = T1_{i0} \quad \phi(SS_1^{i1}) = T1_{i1} \qquad \phi(SS_2^{i0}) = T2_{i0} \quad \phi(SS_2^{i1}) = T2_{i1} \\
\hline
[\![\langle i, \{c_1, c_2, \ldots c_k\}\rangle, (SS_1, SS_2)]\!]_{pp} \mapsto (\\[4pt]
(i \,|\, T1_{i0} \,|\, X_0)^L \,\rfloor\, (T1_{i1} \,|\, P1_1 \,|\, P1_2 \,|\, \ldots \,|\, P1_k \,|\, X_1), \\[4pt]
(i \,|\, T2_{i0} \,|\, X_0)^L \,\rfloor\, (T2_{i1} \,|\, P2_1 \,|\, P2_2 \,|\, \ldots \,|\, P2_k \,|\, X_1))
\end{array}
$$

where $X_i = new(\ldots)$ means that the term variable X_i has not previously used in the current application of $[\![\cdot]\!]_{pp}$, $P_i \in \mathcal{P}$ and $T_i \in \mathcal{T}$.

The parallel pattern builder function takes as arguments a root membrane (and its children), two sets of species (the reactants and the products of a reaction) and it gives as a result the pair of reactants-products expressed as CLS+ terms. The $[\![\cdot]\!]_{pp}$ function is used in the definition of the function for translating a reaction into a stochastic CLS+ rewrite rule.

The reaction translation function also uses the following function xc to deal with contingencies. Function xc is applied to a pattern, obtained by the parallel pattern builder, and a set of species SS, representing the contingencies of the reaction. It associates, with each term variable X appearing in P, a term containing the species from SS that can be instantianted in the variable X. In this way, we can keep track of the variables in which the contigency species can appear. Formally, function $xc : \mathcal{P} \times \mathcal{SS} \to \mathscr{P}(\mathcal{BV} \cup \mathcal{TV} \times \mathcal{T})$ is defined as:

$$
\begin{aligned}
xc((i \,|\, B_1 \,|\, \overline{X_1})^L \,\rfloor\, (P_2 \,|\, X_2),\ SS) &= \{\langle \overline{X_1}, \phi(SS^{i0})\rangle, \langle X_2, \phi(SS^{i1})\rangle\} \cup xc(P_2, SS) \\
xc(P1 \,|\, P2, SS) &= xc(P_1, SS) \cup xc(P_2, SS) \\
xc(T, SS) &= \emptyset
\end{aligned}
$$

Definition 10 (Reaction Translation Function, $[\![\cdot]\!]_r$). *Let MS be the Membrane Structure of a MIM. Given a reaction $\langle CS, RS, PS \rangle$, the* Reaction Trans-

lation Function $[\![\cdot]\!]_r : (\mathcal{CS} \times \mathcal{SS} \times \mathcal{SS}) \to \Re$ *is defined as follows:*

$$(P_1, P_2) = [\![root(MS'), (CS, RS, PS)]\!]_{pp} \qquad MS' = subtree(MS, N)$$

$$N = membrane((\bigcup_{j=1}^{n} C_j) \cup RS \cup PS) \qquad CS = \{(C_1, k_1), (C_2, k_2), \ldots, (C_n, k_n)\}$$

$$\frac{xc(P_1, C_i) \mapsto \{\langle X_1, T_1^i\rangle, \langle X_2, T_2^i\rangle, \ldots, \langle X_m, T_m^i\rangle\} \quad i = 1, \ldots, n}{[\![\langle CS, RS, PS\rangle]\!]_r = (P_1, P_2, f)}$$

where P_1, P_2 *are CLS+ patterns, and the rate function* f *is defined as follows:*

$$f(\sigma) = basalRate(\sigma) \cdot \prod_{\langle S,(i,j),q\rangle \in RS} \binom{q}{occ(P_1\sigma, (i,j), S)}$$

$$basalRate(\sigma) = \begin{cases} k_1 & if\ \exists U_1, \ldots, U_m.\quad \sigma(X_1) \equiv T_1^1\,|\,U_1 \wedge \ldots \wedge \sigma(X_m) \equiv T_m^1\,|\,U_m \\ k_2 & if\ \exists U_1, \ldots, U_m.\quad \sigma(X_1) \equiv T_1^2\,|\,U_1 \wedge \ldots \wedge \sigma(X_m) \equiv T_m^2\,|\,U_m \\ \vdots & \qquad\vdots \\ k_n & if\ \exists U_1, \ldots, U_m.\quad \sigma(X_1) \equiv T_1^n\,|\,U_1 \wedge \ldots \wedge \sigma(X_m) \equiv T_m^n\,|\,U_m \end{cases}$$

The function f computes the rate of the application of the rewrite rule as the product of a basal rate and of the number of different reactions represented by this application. The basal rate is a constant rate that depends on the contingencies. In the definition of $basalRate$ we are assuming that, if more than one condition is satisfied (i.e. more than one contingency is applicable), then the most specific one is used. The constraints we imposed on the set of contincencies (see Sec.4.1) ensure that the definition of $basalRate$ is unambiguous.

The number of different reactions represented by the application of a rewrite rule takes into account the reactions involving molecules of a same species S. Considering a single compartment i, j, the reaction can happen among any subset of molecules of that species S, having the right cardinality. Therefore, for each reactant species S in each compartment i, j, described by a tuple in RS, the number of reactions corresponds to the binomial coefficient of q, i.e. the number of molecules required, and $occ(P_1\sigma, (i, j), S)$, i.e. the number of molecules present in compartment (i, j) in the actual term $P_1\sigma$.

Definition 11 (Set of Rewrite Rules of a MIM, RR). *Given a Molecular Interaction Map, described by a Membrane Structure MS and whose reactions are encoded in the Set of Reactions $RS = \{R_1, \ldots, R_n\}$, the Set of Rewrite Rules $RR \subset \Re$ of the corresponding Stochastic CLS+ model is $RR = \{[\![R_1]\!]_r, [\![R_2]\!]_r, \ldots, [\![R_n]\!]_r\}$.*

5   Applications

Now we show a small example of the translation process we defined by translating part of MIM of Fig.4 which showed a process of RNA synthesis. In the first phase of the process, protein A binds to protein B thus forming the complex $A{:}B$ onto the plasma membrane. Then two complexes can form a dimer which can be phosphorilated. In the following, we show these two reactions translated

as Stochastic CLS+ terms and rewrite rules.

Firstly, we define the membrane structure: $MS = \{\langle 0, \{1\}\rangle, \langle 1, \{2\}\rangle, \langle 2, \varnothing\rangle\}$.

Then we want to formalize the reaction where B binds to A ($A \mid B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AB$).

To do so, we need to describe two reaction tuples for this reaction, one for the complexation and another one for the de-complexation:

$$R_1 = \langle\{\langle\varnothing, k_1\rangle\}, \{\langle A, (1,0), 1\rangle, \langle B, (0,1), 1\rangle\}, \{\langle AB, (1,0), 1\rangle\}\rangle$$
$$R_2 = \langle\{\langle\varnothing, k_2\rangle\}, \{\langle AB, (1,0), 1\rangle\}, \{\langle A, (1,0), 1\rangle, \langle B, (0,1), 1\rangle\}\rangle$$

As one can see, no contingency influences these reactions, hence the only one present is the neutral one. We place the product on membrane 1, since the complex should be still connected to the membrane. The resulting rewrite rules for these reactions are:

(R_1) $\quad (0 \mid X_0)^L \rfloor (B \mid (1 \mid A \mid X_2)^L \rfloor X_3 \mid X_1) \overset{f_1}{\mapsto} (0 \mid X_0)^L \rfloor ((1 \mid AB \mid X_2)^L \rfloor X_3 \mid X_1)$

(R_2) $\quad (0 \mid X_0)^L \rfloor ((1 \mid AB \mid X_2)^L \rfloor X_3 \mid X_1) \overset{f_2}{\mapsto} (0 \mid X_0)^L \rfloor (B \mid (1 \mid A \mid X_2)^L \rfloor X_3 \mid X_1)$

where the basalRate of both f_1 and f_2 have constant results, respectively k_1 and k_2.

Now we describe the phosphorylation of the A component allowed by the presence of the dimer $AB{:}AB$ onto the plasma membrane. This modification will be described by a single reaction tuple since it is not reversible. The presence of a "necessity" contingency pointing to the reaction will be a tuple in the Set of Contingencies of the Reaction Set. Thus, we will formalize the following (mutually exclusive) reactions $A \overset{k_5}{\rightleftharpoons} pA$ and $AB{:}AB \mid A \overset{k_5'}{\rightleftharpoons} AB{:}AB \mid pA$.

A single reaction tuple will model both reactions:

$$R_3 = \langle\{\langle\varnothing, k_5\rangle, \langle\{\langle AB{:}AB, (0,1), 1\rangle\}, k_5'\rangle\}, \{\langle A, (1,0), 1\rangle\}, \{\langle pA, (1,0), 1\rangle\}\rangle$$

where k_5 is equal to zero in our example (since the phosphorylation needs the presence of $AB{:}AB$). The resulting rewrite rule for this reaction tuple is:

(R_3) $\quad (0 \mid X_0)^L \rfloor ((1 \mid A \mid X_2)^L \rfloor X_3 \mid X_1) \overset{f}{\mapsto} (0 \mid X_0)^L \rfloor ((1 \mid pA \mid X_2)^L \rfloor X_3 \mid X_1)$

$$basalRate(\sigma) = \begin{cases} k_5' & \text{if } \sigma(X_1) \equiv AB{:}AB \mid T \\ k_5 & \text{otherwise} \end{cases}$$

6 Conclusions

In this paper we have given a formal definition of the semantics of Kohn's Molecular Interaction Maps by providing a translation to a variant of the Stochastic Calculus of Looping Sequences (called Stochastic CLS+). Such a definition allows understanding and reasoning on Kohn's maps unambiguosly. Moreover, having CLS+ executables, this translation provides MIMs with simulators and other tools and methodologies developed for CLS+. Future works will involve the contruction of graphical interfaces allowing the input of Molecular Interaction Maps and the automatic translation of them in CLS+ terms and rules. This will allow biologists to use the tools for CLS+ with MIMs as interfaces.

References

1. M. I. Aladjem, S. Pasa, S. Parodi, J. N. Weinstein, Y. Pommier, and K. W. Kohn. Molecular interaction maps–a diagrammatic graphical language for bioregulatory networks. *Science's STKE*, 2004(222):pe8, 2004.
2. R. Barbuti, A. Maggiolo-Schettini, P. Milazzo, P. Tiberi, and A. Troina. Stochastic CLS for the modeling and simulation of biological systems. *Transactions on Computational Systems Biology IX, Springer LNCS journal*, 5121, 2008.
3. L. Bortolussi, S. Fonda, and A. Policriti. Constraint-based simulation of biological systems described by molecular interaction maps. *Bioinformatics and Biomedicine, IEEE International Conference on*, 0:288–293, 2007.
4. F. Ciocchetta, C. Priami, and P. Quaglia. Modeling kohn interaction maps with beta-binders: An example. *Transactions on Computational Systems Biology III, LNCS Subline*, 3737:33–48, 2005.
5. CLSm: simulation tool `http://www.di.unipi.it/~milazzo/biosims`.
6. V. Danos and C. Laneve. Formal molecular biology. *Theoretical Computer Science*, 325(1):69–110, 2004.
7. J. R. Faeder, M. L. Blinov, and W. S. Hlavacek. Graphical rule-based representation of signal-transduction networks. In *Symposium on Applied Computing (SAC)*, pages 133–140. ACM, 2005.
8. A. Funahashi, M. Morohashi, and H. Kitano. CellDesigner: a process diagram editor for gene–regulatory and biochemical networks. *BIOSILICO*, 1(5), 2003.
9. D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
10. K. W. Khon. Molecular interaction maps as information organizers and simulation guides. *CHAOS*, 11(1):84–97, 2001.
11. K. W. Khon, M. I. Aladjem, J. N. Weinstein, and Y. Pommier. Molecular interaction maps of bioregulatory networks: A general rubric for systems biology. *Molecular Biology of the Cell*, 17:1–13, 2006.
12. H. Kitano. A graphical notation fo biochemical networks. *BIOSILICO*, 1(5), 2003.
13. K. W. Kohn and M. I. Aladjem. Circuit diagrams for biological networks. *Molecular Systems Biology*, 2006.
14. D. Lepri. A formal semantics for molecular interaction maps. Master Thesis in Computer Science, University of Pisa, 2008.
15. P. Milazzo. *Qualitative and Quantitative Formal Modeling of Biological Systems*. PhD thesis, Computer Science Department - University of Pisa, 2007.
16. P. Milazzo. *Formal Modeling in Systems Biology. An approach from Theoretical Computer Science*. VDM - Verlag Dr. Muller, Saarbrucken, Germany, 2008.
17. K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano. A comprensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology*, 2005.
18. I. Pirson, N. Fortemaison, C. Jacobs, S. Dremier, J. E. Dumont, and C. Maenhaut. The visual display of regulatory information and networks. *Trends in Cell Biology*, 10:404–408, 2000.
19. C. Priami and P. Quaglia. Beta binders for biological interactions. In *CMSB'04*, volume 3082 of *LNCS*, pages 20–33. Springer, 2004.
20. Systems biology graphical notation, `http://sbgn.org`.