Hybrid Calculus of Wrapped Compartments*

Mario Coppo[1], Ferruccio Damiani[1], Maurizio Drocco[1], Elena Grassi[1,2], Eva Sciacca[1], Salvatore Spinella[1], Angelo Troina[1]

[1]Dipartimento di Informatica, Università di Torino

[2]Molecular Biotechnology Center, Dipartimento di Genetica, Biologia e Biochimica, Università di Torino

The modelling and analysis of biological systems has deep roots in Mathematics, specifically in the field of ordinary differential equations (ODEs). Alternative approaches based on formal calculi, often derived from process algebras or term rewriting systems, provide a quite complementary way to analyze the behaviour of biological systems. These calculi allow to cope in a natural way with notions like compartments and membranes, which are not easy (sometimes impossible) to handle with purely numerical approaches, and are often based on stochastic simulation methods. Recently, it has also become evident that stochastic effects in regulatory networks play a crucial role in the analysis of such systems. Actually, in many situations it is necessary to use stochastic models. For example when the system to be described is based on the interaction of few molecules, when we are at the presence of a chemical instability, or when we want to simulate the functioning of a pool of entities whose compartmentalised structure evolves dynamically. In contrast, stable metabolic networks, involving a large number of reagents, for which the computational cost of a stochastic simulation becomes an insurmountable obstacle, are efficiently modelled with ODEs. In this paper we define a hybrid simulation method, combining the stochastic approach with ODEs, for systems described in CWC, a calculus on which we can express the compartmentalisation of a biological system whose evolution is defined by a set of rewrite rules.

1 Introduction

The most common approach of biologists to describe biological systems is based on the use of deterministic mathematical means like, e.g., ordinary differential equations (ODEs for short). ODEs make it possible to abstractly reason on the behaviour of biological systems and to perform a quantitative *in silico* investigation. This kind of modelling, however, becomes more and more difficult, both in the specification phase and in the analysis processes, when the complexity of the biological systems taken into consideration increases. This has probably been one of the main motivations for investigating the description of biological systems by means of formalisms developed in Computer Science for the description of computational entities [21]. Different formalisms have either been applied to (or have been inspired from) biological systems. Automata-based models [2, 16] have the advantage of allowing the direct use of many verification tools such as model checkers. Rewrite systems [10, 20, 4] usually allow describing biological systems with a notation that can be easily understood by biologists. Both automata-like models and rewrite systems present, in general, problems from the point of view of compositionality, which allows studying the behaviour of a system componentwise. Compostionality, instead, is in general ensured by Process calculi, included those commonly used to describe biological systems [21, 19, 6]. Quantitative simulations of biological models represented with these kind of frameworks (see, e.g. [19, 11, 15, 3, 12, 8]) are usually developed via a stochastic method derived by Gillespie's algorithm [13].

*This research is founded by the BioBITs Project (*Converging Technologies* 2007, area: Biotechnology-ICT), Regione Piemonte.

To appear in EPTCS.

The ODE description of biological systems determines *continuous*, *deterministic* models in which variables describe the concentrations of the species involved in the system as functions of the time. These models are based on average reaction rates, measured from real experiments which relate to the change of concentrations over time, taking into account the known properties of the involved chemicals, but possibly abstracting away some unknown mechanisms. Given the reaction equations (together with their rates) and the initial amount for each species, an ODEs model can be constructed by writing a differential equation for each biochemical specie whose concentration changes over time.

In contrast to the deterministic model, *discrete*, *stochastic* simulations involve random variables. Therefore, the behaviour of a reaction is not determined a priori but characterized statistically. Since biological reactions fall in the category of stochastic systems (the very basic steps of every molecular reaction can be described only in terms of its probability of occurrence), stochastic kinetic models are increasingly accepted as the best way to represent and simulate genetic and biochemical networks. Moreover, when the system to be described is based on the interaction of few molecules, or we want to simulate the functioning of a little pool of cells it is necessary to use stochastic models.

The stochastic approach is always valid when the deterministic one is, and it may be valid when the ordinary deterministic is not (i.e. in a nonlinear system in the neighborhood of a chemical instability). Actually, in the last years it has become evident that stochastic effects in regulatory networks play a crucial role in the analysis of such systems (for example in case of multi-stable systems). In contrast, metabolic networks involving large numbers of molecules are most often modelled deterministically. Thus, because of the bimodal nature of biological systems, it may happen that a purely deterministic model does not accurately capture the dynamics of the considered system, and a stochastic description is needed. However, the computational cost of a discrete simulation often becomes an insurmountable obstacle. Computationally, the ODEs method is extremely more efficient. Thus, when the deterministic approach is applicable, it might be profitable to take advantage of its efficiency, and move to the stochastic approach when it is not. In a hybrid model, some reactions are modelled in a discrete way (i.e. computed, probabilistically, according to a stochastic method) and others in a continuous way (i.e. computed, in a deterministic way, by a set of ODEs).

Hybrid models for the simulation of biological systems have been presented in the last few years for purely mathematical models [22, 14, 9]. In this paper we adapt the hybrid simulation technique within the programming language approach to describe and analyse the dynamics of biological systems.

In [8] we proposed the *Calculus of Wrapped Compartments* (CWC for short), a simplification of the Calculus of Looping Sequences (CLS for short) [4, 3]. Starting from an alphabet of atomic elements, CWC terms are defined as multisets of elements and compartments. Elements can be localized by compartmentalisation and the structure of a compartment can be specified by detailing the elements of interest on its membrane. The evolution of the system is driven by a set of rewrite rules modelling the reactions of interest. We provided CWC with a stochastic operational semantics from which a continuous time Markov chain can be build following the standard Gillespie's approach [13].

In this paper we define a hybrid simulation method for systems described in CWC; thus: (1) we are able to simulate systems with compartments, (2) we use the stochastic simulation method when the deterministic one is not valid, and (3) we exploit the efficiency of the deterministic approach whenever it is applicable.

Summary. Section 2 introduces the CWC formalism. Section 3 recalls the stochastic and the deterministic simulation methods. Section 4 introduces the hybrid simulation technique and Section 5 applies it to the analysis of the HIV-1 transactivation mechanism. Finally, in Section 6, we draw our conclusions.

Simple terms syntax

$t \quad ::= \quad a \;\mid\; (\overline{a} \rfloor \overline{t})^{\ell}$

Structural congruence

$\overline{t}\ u\ w\ \overline{v} \equiv \overline{t}\ w\ u\ \overline{v}$

if $\overline{a} \equiv \overline{b}$ and $\overline{t} \equiv \overline{u}$ then $(\overline{a} \rfloor \overline{t})^{\ell} \equiv (\overline{b} \rfloor \overline{u})^{\ell}$

Figure 1: CWC term syntax and structural congruence rules

2 The Calculus of Wrapped Compartments

Like most modelling languages based on term rewriting (notably CLS), a CWC (biological) model consists of a term, representing the system and a set of rewrite rules which model the transformations determining the system's evolution. The calculus presented here is a slight variant of the one introduced in [8]. Namely, compartments are enriched with a nominal type which identifies the set of rewrite rules that can be applied on that compartment.

Terms and Structural Congruence A *term* of the CWC calculus is intended to represent a biological system. A *term* is a multiset of *simple terms*. Simple terms, ranged over by t, u, v, w, are built by means of the *compartment* constructor, $(-\rfloor -)^{-}$, from a set $\mathscr{A}$ of *atomic elements* (*atoms* for short), ranged over by a, b, c, d, and from a set $\mathscr{L}$ of *compartments types* (represented as *labels* attached to compartments and rules), ranged over by $\ell, \ell', \ell_1, \ldots$. The syntax of simple terms is given at the top of Figure 1. We write $\overline{t}$ to denote a (possibly empty) multiset of simple terms $t_1 \cdots t_n$. Similarly, with $\overline{a}$ we denote a (possibly empty) multiset of atoms. The set of simple terms will be denoted by $\mathscr{T}$. The set of terms (multiset of simple terms) and the set of multisets of atoms will be denoted by $\overline{\mathscr{T}}$ and $\overline{\mathscr{A}}$, respectively. Note that $\overline{\mathscr{A}} \subseteq \overline{\mathscr{T}}$.

A term $\overline{t} = t_1 \cdots t_n$ should be understood as the multiset containing the simple terms $t_1, \ldots, t_n$. Therefore, we introduce a relation of structural congruence, following a standard approach in process algebra. The CWC *structural congruence* is the least equivalence relation on terms satisfying the rules given at the bottom of Figure 1. From now on we will always consider terms modulo structural congruence. Then a simple term is either an atom or a compartment $(\overline{a} \rfloor \overline{t})^{\ell}$ consisting of a *wrap* (represented by the multiset of atoms $\overline{a}$), a *content* (represented by the term $\overline{t}$) and a *type* (represented by the label ℓ). We write the empty multiset as $\bullet$ and denote the union of two multisets $\overline{u}$ and $\overline{v}$ as $\overline{u}\ \overline{v}$. Let's extend the notion of subset (denoted as usual as $\subseteq$) between terms interpreted as multisets.

An example of term is $\overline{t} = a\ b\ (c\ d \rfloor e\ f)^{\ell}$ representing a multiset consisting of two atoms a and b (for instance two molecules) and an ℓ-type compartment $(c\ d \rfloor e\ f)^{\ell}$ which, in turn, consists of a wrap (a membrane) with two atoms c and d (for instance, two proteins) on its surface, and containing the atoms e (for instance, a molecule) and f (for instance a DNA strand). See Figure 2 for some graphical representations.

Rewrite Rules, Variables, Open Terms and Patterns A rewrite rule is defined as a pair of terms (possibly containing variables), which represent the patterns defining the system transformations, together with a label ℓ representing the compartment type on which the rule can be applied. Compartments are identified by the notion of (labelled) reduction context introduced below. A rule is applicable in a compartment if its content matches the left-hand side of the rule via a proper instantiation of its variables

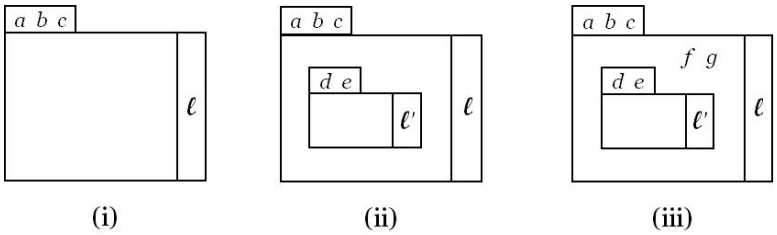

Figure 2: **(i)** represents $(a\ b\ c \rfloor \bullet)^{\ell}$; **(ii)** represents $(a\ b\ c \rfloor (d\ e \rfloor \bullet)^{\ell'})^{\ell}$; **(iii)** represents $(a\ b\ c \rfloor (d\ e \rfloor \bullet)^{\ell'}\ f\ g)^{\ell}$

(note this instantiation is in general not unique). A system transformation is obtained by replacing the reduced subterm by the corresponding instance of the right-hand side of the rule.

In order to formally define the rewriting semantics, we introduce the notion of open term (a term containing variables) and pattern (an open term that may be used as left part of a rewrite rule). In order to respect the syntax of terms, we distinguish between "wrap variables" which may occur only in compartment wraps (and can be replaced only by multisets of atoms) and "term variables" which may only occur in compartment contents or at top level (and can be replaced by arbitrary terms). Therefore, we assume a set of *term variables*, $\mathscr{V}_{\overline{\mathscr{T}}}$, ranged over by X,Y,Z, and a set of *wrap variables*, $\mathscr{V}_{\overline{\mathscr{A}}}$, ranged over by x,y,z. These two sets are disjoint. We denote by $\mathscr{V}$ the set of all variables $\mathscr{V}_{\overline{\mathscr{T}}} \cup \mathscr{V}_{\overline{\mathscr{A}}}$, and with ρ any variable in $\mathscr{V}$.

- *Open terms* are terms which may contain occurrences of wrap variables in compartment wraps and term variables in compartment contents or at top level. They can be seen as multisets of *simple open terms*. More formally, open terms, ranged over by O and simple open terms, ranged over by o, are defined in the following way:

$$\begin{array}{lcl} O & ::= & \overline{o} \\ o & ::= & a \ \mid\ X \ \mid\ (\overline{a}\ \overline{x} \rfloor \overline{o})^{\ell} \end{array}$$

  We denote with $\mathscr{O}$ the set of open terms. An open term is *linear* if each variable occurs at most once.

- *Patterns*, ranged over by P, and *simple patterns*, ranged over by p, are the linear open terms defined in the following way:

$$\begin{array}{lcl} P & ::= & \overline{p} \\ p & ::= & t \ \mid\ (\overline{a}\ x \rfloor \overline{p}\ X)^{\ell} \end{array}$$

  We denote with $\mathscr{P}$ the set of patterns.

An *instantiation* is a partial function $\sigma : \mathscr{V} \to \overline{\mathscr{T}}$. An instantiation must preserve the type of variables, thus for $X \in \mathscr{V}_{\overline{\mathscr{T}}}$ and $x \in \mathscr{V}_{\overline{\mathscr{A}}}$ we have $\sigma(X) \in \overline{\mathscr{T}}$ and $\sigma(x) \in \overline{\mathscr{A}}$, respectively. Given $O \in \mathscr{O}$, with $O\sigma$ we denote the term obtained by replacing each occurrence of each variable $\rho \in \mathscr{V}$ appearing in O with the corresponding term $\sigma(\rho)$.

Let Σ denote the set of all the possible instantiations and $Var(O)$ denote the set of variables appearing in $O \in \mathscr{O}$.

A *rewrite rule* is a triple (ℓ, P, O), also denoted by $\ell : P \longmapsto O$, where $P \in \mathscr{P}$ and $O \in \mathscr{O}$ are such that $Var(O) \subseteq Var(P)$. The label ℓ denotes the type of the compartments where the rule can be applied. A

rewrite rule $\ell : P \longmapsto O$ then states that a subterm $P\sigma$, obtained by instantiating variables in P by some instantiation function σ, can be transformed into the subterm $O\sigma$ within any compartment of type ℓ. We use the special label $\top \in \mathscr{L}$ to denote the type of the top level of a term.

Contexts The definition of reduction for CWC systems is completed by resorting to the notion of reduction context. To this aim, the syntax of terms is enriched with a new element $\Box$ representing a hole. *Reduction context* (ranged over by C) are defined by:

$$C \;::=\; \Box \;\;\big|\;\; C\;\bar{t} \;\;\big|\;\; (\bar{a}\rfloor C)^{\ell}$$

where $\bar{a} \in \overline{\mathscr{A}}$, $\bar{t} \in \overline{\mathscr{T}}$ and $\ell \in \mathscr{L}$. We denote with $\mathscr{C}$ the infinite set of contexts.

By definition, every context contains a single hole $\Box$. Let us assume $C, C' \in \mathscr{C}$. With $C[\bar{t}]$ we denote the term obtained by replacing $\Box$ with $\bar{t}$ in C; with $C[C']$ we denote context composition, whose result is the context obtained by replacing $\Box$ with C' in C. For example, given $C = (a\ b\rfloor \Box)^{\ell}\ i$, $C' = (c\ d\rfloor \Box)^{\ell'}\ g\ h$ and $\bar{t} = e\ f$, we get $C[C'[\bar{t}]] = (a\ b\rfloor (c\ d\rfloor e\ f)^{\ell'}\ g\ h)^{\ell}\ i$.

In order to apply a rule within a compartment of the correct type we define a function that, given a context, returns the label of the innermost compartment containing the hole. If the hole appears at top level, the distinguished label $\top$ is returned. The function LAB is defined as follows:

$$\textsc{Lab}(C) = \begin{cases} \top & \text{if } C = \Box\ \bar{t} \\ \ell & \text{if } C = C'[(\bar{a}\rfloor \Box\ \bar{t})^{\ell}] \end{cases}$$

Qualitative Reduction Semantics A CWC *system* over a set $\mathscr{A}$ of atoms and a set $\mathscr{L}$ of labels is represented by a set $\mathscr{Q}_{\mathscr{A},\mathscr{L}}$ ($\mathscr{Q}$ for short when $\mathscr{A}$ and $\mathscr{L}$ are understood) of rewrite rules over $\mathscr{A}$ and $\mathscr{L}$.

The *qualitative reduction semantics* of a CWC system $\mathscr{Q}$ is the least transition relation satisfying the following rule:

$$\frac{\ell : P \longmapsto O \in \mathscr{Q} \quad \sigma \in \Sigma \quad C \in \mathscr{C} \quad \textsc{Lab}(C) = \ell}{C[P\sigma] \rightarrow C[O\sigma]}$$

Modelling Guidelines In this section we give some explanations and general hints about how CWC could be used to represent the behaviour of various biological systems. Here, entities are represented by terms of the rewrite system, and events by rewrite rules.

First of all, we should select the biomolecular entities of interest. Since we want to describe cells, we consider molecular populations and membranes. Molecular populations are groups of molecules that are in the same compartment of the cells and inside them. As we have said before, molecules can be of many types: we classify them as proteins, chemical moieties and other molecules.

Membranes are considered as elementary objects: we do not describe them at the level of the phospholipids they are made of. The only interesting properties of a membrane are that it may have a content (hence, create a compartment) and that in its phospholipid bilayer various proteins are embedded, which act for example as transporters and receptors. Since membranes are represented as multisets of the embedded structures, we are modeling a fluid mosaic in which the membranes become similar to a two-dimensional liquid where molecules can diffuse more or less freely [23].

Compartment labels are useful to identify the kind of a compartment. For example, we may use compartment labels to denote a nucleus within a cell, the different organelles, etc..

Biomolecular Event	CWC Rewrite Rules
State change	$a \longmapsto b$
Complexation	$a\ b \longmapsto c$
Decomplexation	$c \longmapsto a\ b$
State change on membrane	$(a\ x \rfloor X) \longmapsto (b\ x \rfloor X)$
Complexation on membrane	$a\ (b\ x \rfloor X) \longmapsto (c\ x \rfloor X)$ $(b\ x \rfloor a\ X) \longmapsto (c\ x \rfloor X)$
Decomplexation on membrane	$(c\ x \rfloor X) \longmapsto a\ (b\ x \rfloor X)$ $(c\ x \rfloor X) \longmapsto (b\ x \rfloor a\ X)$
Membrane crossing	$a\ (x \rfloor X) \longmapsto (x \rfloor a\ X)$ $(x \rfloor a\ X) \longmapsto a\ (x \rfloor X)$
Catalyzed membrane crossing	$a\ (b\ x \rfloor X) \longmapsto (b\ x \rfloor a\ X)$ $(b\ x \rfloor a\ X) \longmapsto a\ (b\ x \rfloor X)$
Membrane joining	$a\ (x \rfloor X) \longmapsto (a\ x \rfloor X)$ $(x \rfloor a\ X) \longmapsto (a\ x \rfloor X)$
Catalyzed membrane joining	$a\ (b\ x \rfloor X) \longmapsto (a\ b\ x \rfloor X)$ $(b\ x \rfloor a\ X) \longmapsto (a\ b\ x \rfloor X)$ $(x \rfloor a\ b\ X) \longmapsto (a\ x \rfloor b\ X)$

Table 1: Guidelines for modelling biomolecular events in CWC

Table 1 lists the guidelines (taken from [8]) for the abstraction into CWC rules of some basic biomolecular events, some of which will be used in our applications.[1] Entities are associated with CWC terms: elementary objects (genes, domains, etc...) are modelled as atoms, molecular populations as CWC terms, and membranes as atom multisets. Biomolecular events are associated with CWC rewrite rules.

The simplest kind of event is the change of state of an elementary object. Then, there are interactions between molecules: in particular complexation, decomplexation and catalysis. Interactions could take place between simple molecules, depicted as single symbols, or between membranes and molecules: for example a molecule may cross or join a membrane. Finally, there are also interactions between membranes: in this case there may be many kinds of interactions (fusion, vesicle dynamics, etc...).

3 Quantitative Simulation Models for CWC

A quantitative operational semantics for CWC can be defined by associating to the rewriting rules of CWC the kinetic constant k of the modeled chemical reaction. A *quantitative rewrite rule* is then a quadruple (ℓ, P, O, k), denoted with $\ell : P \overset{k}{\longmapsto} O$, where ℓ, P and O are in Section 2, and $k \in \mathbb{R}^{\geq 0}$.

In this section we introduce two (standard) quantitative simulation methods for CWC based respectively on Gillespie's stochastic simulation algorithm [13] and on the (deterministic) solution of ordinary differential equations. These two approaches, that will be presented separately in this section, will be integrated in the next section defining the hybrid semantics of CWC.

A prototype implementation of the hybrid CWC calculus (which encompasses both the pure stochastic and the deterministic versions of the calculus) is available [1].

Remark 3.1 *Notice that the stochastic, deterministic and hybrid approaches introduced in this section simulate well-stirred system of molecules, confined to a constant volume and in thermal equilibrium at some constant temperature. In these conditions we can describe the systems state by specifying only*

[1] Compartment labels are omitted for simplicity, just notice that the rules shown in the table can be specified to apply only within a given type of compartment.

μ	**A**	**B**	**C**
A	0.15	0.2	0.15
B	0.2	0.15	0.2
C	0.15	0.2	0.15

Table 2: Interaction Matrix

the molecular populations, ignoring the positions and velocities of the individual molecules. Different approaches such as Molecular Dynamics, Partial Differential Equations or Lattice-based methods are required in case of molecular crowding, anisotropy of the medium or canalization.

Running Example In order to illustrate the quantitative semantics of CWC we consider, as a running example, a toy case study based on a compartmentalized variant of a system studied in ecology to describe a generalized competitive Lotka-Volterra dynamics [24]. The variant consists in an schema of ecoregions bounded by geographical frontiers rendered by compartments. The system has 3 competitive species in 2 environmental compartments. The supposed population dynamics depend on the following parameters:

- *the interaction matrix* $\mu_{i,j} \geq 0$, where the element $\mu_{i,j}$ represents the relative strength that species i has on the population of species j, i.e. the competition between species in the same environment due to incompatibility;
- *the carrying capacity* $K_i \geq 0$ is the population size of the species i that the environment can sustain indefinitely assuming no interaction between the species;
- *the migration rate* d_i associated to each species which migrates between the compartments.

In our tests, we set the interaction matrix in accordance to Table 2. The carrying capacities are $K_i = 100$ and the migration rates between the compartments are: $d_A = 0.01$, $d_B = 0.01$, $d_C = -0.01$. The migration rates are positive for species A and B moving from the compartment to the outside, and negative for species C which moves in the reverse path. This system has a wide set of possible behaviour, particularly the compartmentalization can be interpreted as the case of an ecological frontier like a river or a mountain which partially separates the different populations.

The set of CWC rules adopted in our toy case study is given in Figure 3, where the rates of competition of the species i against the species j were calculated as $k_{i,j} = \frac{\mu_{i,j}}{K_i}$. The two environmental compartments are represented by the implicit top level compartment (of type $\top$) and by an explicit compartment of type IN. Rules $(N_1 - N_3)$ model the migration of the three species between the two compartments. The other rules model the competition between species (rules $B_4 - B_9$) and their reproduction capacity (rules $B_1 - B_3$). Note that rule (B_1), is indeed a compact representation for the rules:

$$(B_1') \quad \top : A \xmapsto{1} A\ A \qquad\qquad (B_1'') \quad IN : A \xmapsto{1} A\ A$$

Similarly for the rules $(B_2), \ldots, (B_9)$. The simulations will be performed for 35 time units, with the starting term:[2]

$$2 \times C\ (\bullet \rfloor 2 \times A\ 2 \times B)^{IN}.$$

[2]The notation $n \times a$, where n is a natural number and a is an atom, denotes the multiset containing n occurrences of a.

(N_1) $\top : (x \rfloor A\ X)^{IN} \overset{d_A}{\longmapsto} (x \rfloor X)^{IN}\ A$

(N_2) $\top : (x \rfloor B\ X)^{IN} \overset{d_B}{\longmapsto} (x \rfloor X)^{IN}\ B$

(N_3) $\top : (x \rfloor X)^{IN}\ C \overset{-d_C}{\longmapsto} (x \rfloor C\ X)^{IN}$

(B_1) $\top, IN : A \overset{1}{\longmapsto} A\ A$

(B_2) $\top, IN : B \overset{1}{\longmapsto} B\ B$

(B_3) $\top, IN : C \overset{1}{\longmapsto} C\ C$

(B_4) $\top, IN : A\ A \overset{k_{A,A}}{\longmapsto} \bullet$

(B_5) $\top, IN : B\ B \overset{k_{B,B}}{\longmapsto} \bullet$

(B_6) $\top, IN : C\ C \overset{k_{C,C}}{\longmapsto} \bullet$

(B_7) $\top, IN : A\ B \overset{k_{A,B}}{\longmapsto} \bullet$

(B_8) $\top, IN : A\ C \overset{k_{A,C}}{\longmapsto} \bullet$

(B_9) $\top, IN : B\ C \overset{k_{B,C}}{\longmapsto} \bullet$

Figure 3: CWC rules for the test case

3.1 Stochastic Evolution

A stochastic simulation model for biological systems can be defined by incorporating a collision-based stochastic framework along the line of the one presented by Gillespie in [13], which is, *de facto*, the standard way to model quantitative aspects of biological systems. The idea of Gillespie's algorithm is that a rate constant is associated with each considered chemical reaction. Such a constant is obtained by multiplying the kinetic constant of the reaction by the number of possible combinations of reactants that may occur in the system. The resulting rate is then used as the parameter of an exponential distribution modelling the time spent between two occurrences of the considered chemical reaction. Following the law of mass action, it is necessary to count the number of reactants that are present in a system in order to compute the exact rate of a reaction. The same approach has been applied, for instance, to define the quantitative semantics of the stochastic π-calculus [18, 19].

The use of exponential distributions to represent the (stochastic) time spent between two occurrences of chemical reactions allows describing the system as a Continuous Time Markov Chain (CTMC), and consequently allows verifying properties of the described system analytically and by means of stochastic model checkers.

The number of reactants in a reaction represented by a rewrite rule is evaluated considering the number of distinct occurrences, in the same context, of subterms to which the rule can be applied producing the same term. For instance in evaluating the application rate of the stochastic rewrite rule $R = \top : a\ b \overset{k}{\longmapsto} c$ to the term $\bar{t} = a\ a\ b\ b$ we must consider the number of the possible combinations of reactants of the form $a\ b$ in $\bar{t}$. Since each occurrence of a can react with each occurrence of b, this number is 4. So the application rate of R is $k \cdot 4$.

The evaluation of the application rate of a reduction rule containing variables is more complicate since there can be many different ways in which variables can be instantiated to match the subterm to be reduced, and this must be considered to correctly evaluate the application rate. Given two terms $\bar{t}, \bar{u}$ and a reduction rule R we can compute the number of possible applications of the rule R to the term $\bar{t}$ in a context $C[\,]$ of type ℓ, resulting in the term $C[\bar{u}]$. We denote this number by $\textsc{Occ}(R, C[\bar{t}], C[\bar{u}])$, where the

function OCC is analogous to the one defined for SCLS in [3]. The *stochastic reduction semantics* of a CWC system $\mathscr{Q}$ is the least labelled transition relation satisfying the following rule:

$$\frac{R = \ell : P \xmapsto{k} O \in \mathscr{Q} \quad \sigma \in \Sigma \quad C \in \mathscr{C} \quad \text{LAB}(C) = \ell}{C[P\sigma] \xrightarrow{k \cdot \text{OCC}(R, C[P\sigma], C[O\sigma])} C[O\sigma]}$$

Reductions determined by a rule R are labelled with their rates. The rate of a reduction is obtained as the product of the rewrite rate constant and the number of occurrences of the rule within the starting term (thus counting the exact number of reactants to which the rule can be applied and which produce the same result). The rate associated with each transition in the stochastic reduction semantics is the parameter of an exponential distribution that characterizes the stochastic behaviour of the activity corresponding to the applied rewrite rule. The stochastic semantics is essentially a *Continuous Time Markov Chain* (CTMC). Given a term $\bar{t}$, a global time δ and all the reductions $e_1, \ldots, e_M$ that can be applied to $\bar{t}$, with rates $r_1, \ldots, r_M$ such that $r = \sum_{i=1}^{M} r_i$, the standard simulation procedure that corresponds to Gillespie's simulation algorithm [13] consists of the following two steps:

1. The time $\delta + \tau$ at which the next stochastic reduction will occur is randomly chosen with τ exponentially distributed with parameter r;
2. The reduction e_i that will occur at time $\delta + \tau$ is randomly chosen with probability $\frac{r_i}{r}$.

Running Example: Stochastic Simulations We performed several stochastic simulations of the toy case study, showing many, different, possible evolutions. Two of these runs are shown in Figure 4.

A characteristic of this example is that in the initial phase, mainly due to the small number of individuals involved, the evolution of the system is strongly determined by random events that can change dramatically the destiny of the species. In the first experiment, on the top of Figure 4, the populations A and C overtake population B both inside and outside compartment *IN*. The second experiment, on the bottom of Figure 4, shows a completely different fate for the populations: namely, population B overtakes populations A and C inside the compartment *IN* while population C overtakes populations A and B outside the compartment. Note that the cases shown here are just two possible examples of the many different destinies for the three populations. The second one, in particular, differs sensibly from the average behaviour appearing in the deterministic simulation which is shown in Figure 5.

3.2 Deterministic Evolution

The standard way to express the evolution of a biochemical system is via ODEs. We define the deterministic reduction semantics for a subset of CWC quantitative rewrite rules, that we call biochemical rewrite rules, expressing biochemical reactions.

Biochemical rewrite rules are the quantitative rewrite rules of the form $\ell : \overline{a} \xmapsto{k} \overline{b}$, where $\overline{a}$ and $\overline{b}$ are multisets of atomic elements.

All the reactants of a biochemical reaction are completely specified, since both sides of the rules do not contain variables. Moreover, biochemical reactions are local to a single compartment. Reactions that invoke and/or change the structure of compartments cannot be expressed with biochemical rewrite rules. Actually, referring to Table 1, we notice that biochemical rewrite rules can be used to model state change, complexation and decomplexation: these are exactly the kinds of reactions naturally eligible to be simulated with ODEs.

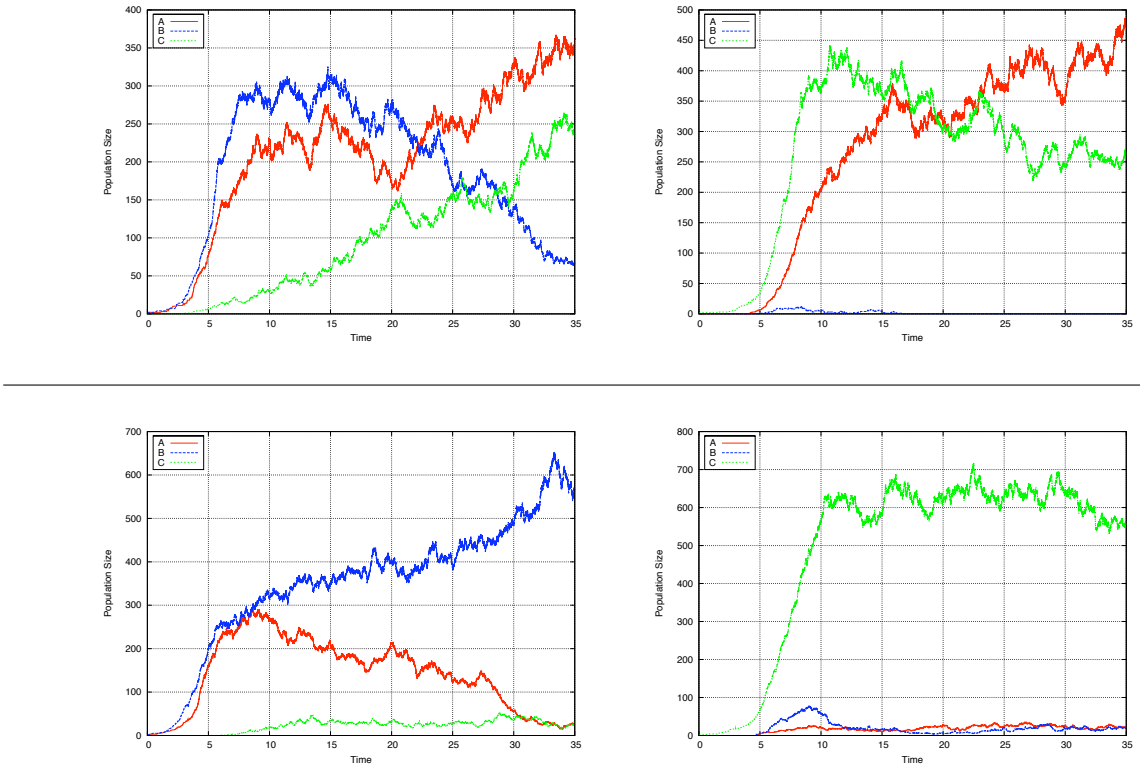

Figure 4: Two different runs of the stochastic simulations showing the different behaviour of the dynamics of the competitive species inside the compartment *IN* (on the left side) and outside the compartment (on the right side)

A CWC system $\mathscr{Q}$ consisting of r biochemical rewrite rules represents a system of r biochemical reactions. Its deterministic semantics is defined by extracting from $\mathscr{Q}$ a system of ODEs to be used for simulating the evolution of the involved multisets of atoms. For every label ℓ, let

- $a_1,\ldots,a_{n_\ell}$ ($n_\ell \geq 1$) denote the n_ℓ species of atoms that may occur at top level within a compartment of type ℓ, and
- $\mathscr{Q}_\ell$ denote the set of rules with label ℓ.

The i-th rule in the set $\mathscr{Q}_\ell$ is denoted by

$$\ell : \bar{a}_i \overset{k_i}{\longmapsto} \bar{b}_i \quad i = 1,2,\ldots,|\mathscr{Q}_\ell|$$

For all species a_j ($j = 1,2,\ldots,n_\ell$) let $\alpha^-_{i,j}$ be the number of atoms of species a_j consumed by the i-th rule and $\alpha^+_{i,j}$ the number of atoms of species a_j produced by the i-th rule. The $n_\ell \times |\mathscr{Q}_\ell|$ stoichiometric matrix Λ_ℓ is defined by $\nu_{i,j} = \alpha^+_{i,j} - \alpha^-_{i,j}$.[3]

Let $[a]$ denote the concentrations of the atoms of specie a occurring at top level in a given compartment of type ℓ. If $\bar{a}_i = a_{i_1}\ \ldots\ a_{i_{r_i}}$ ($r_i \geq 1$), let $[\bar{a}_i]$ denote the product $[a_{i_1}]\ \ldots\ [a_{i_{r_i}}]$ of the concentrations

[3] Many of the $\alpha^-_{i,j}$, $\alpha^+_{i,j}$ are usually 0.

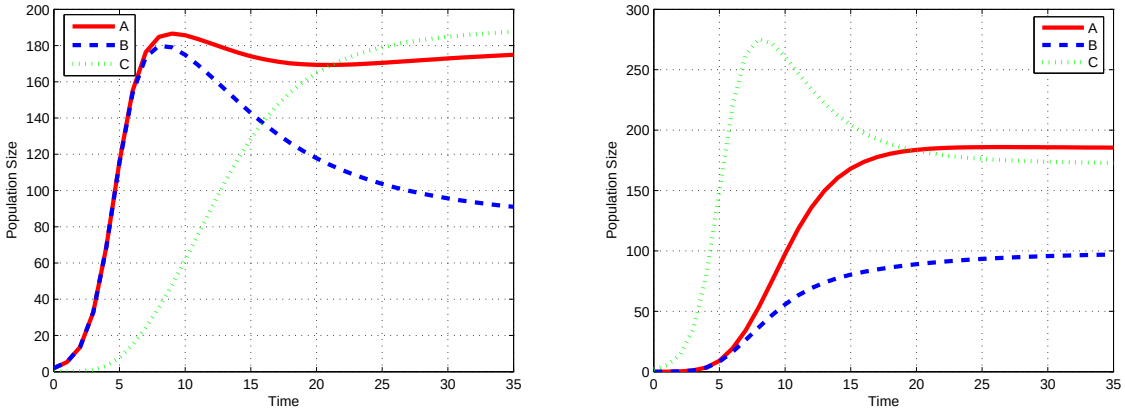

Figure 5: Deterministic simulation of the dynamics of the competitive species inside the compartment *IN* (left figure) and outside the compartment (right figure)

of the species occurring in $\bar{a}_i$ in the considered compartment.[4] The evolution of the given compartment of type ℓ is modelled by the following system of ODEs:

$$\ell : \frac{d[s_j]}{dt} = \sum_{i=1}^{|\mathscr{D}_\ell|} \nu_{i,j} \cdot k_i \cdot [\bar{a}_i]$$

Computationally, ODEs are well studied and understood. They can be solved using a variety of numerical methods, from the Euler method to higher-order Runge-Kutta methods or stiff methods, many of which are readily available in software packages that can be easily incorporated into existing simulation code.

Running Example: Deterministic Simulations To perform a deterministic simulation of the toy case study we have to remodel it by using only biochemical rewrite rules. This can be done by phrasing the compartmentalisation by using a different name for the species occurring in the compartment *IN*, namely: A^{IN}, B^{IN} and C^{IN}. Then, the three non-biochemical rewrite rules (N_1), (N_2) and (N_3) can be converted into the following biochemical rewrite rules:

$$(B_{10})\ \top : A^{IN} \overset{d_A}{\longmapsto} A \qquad (B_{11})\ \top : B^{IN} \overset{d_B}{\longmapsto} B \qquad (B_{12})\ \top : C \overset{-d_C}{\longmapsto} C^{IN}$$

The conversion of the biochemical rules $(B_1), \ldots, (B_9)$ is straightforward. For instance rule (B_1) is converted into the two rules:

$$(B_1')\ \top : A \overset{1}{\longmapsto} A\ A \qquad\qquad (B_1'')\ \top : A^{IN} \overset{1}{\longmapsto} A^{IN}\ A^{IN}$$

The converted starting term is $2 \times C\ 2 \times A^{IN}\ 2 \times B^{IN}$. The results of the deterministic simulation are shown in Figure 5.

[4] Being $\bar{a}_i$ a multiset, if an element a_j occurs h times in $\bar{a}_i$ (for instance, for $h = 2, i_p = i_q$ for some $(1 \leq p, q \leq i_{r_i},\ p \neq q)$ its concentration is considered h times in $[\bar{a}_i]$, i.e. we take the h-th power of $[a_j]$.

Remark 3.2 *The conversion of the original* CWC *system for the toy case studies into a* CWC *system using only the top level compartment and biochemical rewriting rules has been straightforward since the original system has a fixed compartment structure. Such a conversion, as well as the direct representation of the modelled biological system in terms of ODEs, might be quite complicate (or even impossible) in case of biological systems that during their evolution may change the structure of the compartments, maybe by creating a possibly unbounded number of new compartments.*

4 Hybrid Evolution

The stochastic approach is based on a probabilistic simulation method that manages the evolution of exact integer quantities and often requires a huge computational time to complete a simulation. The ODEs numerical approach computes a unique deterministic and fractional evolution of the species involved in the system and achieves very efficient computations. In this section we combine both methods within CWC, defining a hybrid simulation technique.

Given a CWC system $\mathscr{Q}$ we partition it into a set of biochemical rewrite rules $\mathscr{B}$ and a set of non-biochemical rewrite rules $\mathscr{N}$. Rules in $\mathscr{N}$ are always applied by using the stochastic method. Rules in $\mathscr{B}$ might be applied with the ODEs approach. In general $\mathscr{B}$ might contain both rules that model evolution of large numbers of molecules according to very fast reactions (whose execution is suitable to be correctly computed with ODE) and rules that model very slow reactions or reactions that involve a very small number of reagents. In the latter case it is convenient to compute the execution of the associated rule according to the stochastic approach.

According to the state of the system, a rule might be dynamically interpreted either as stochastic or deterministic. For instance, during a simulation, it might happen that a given biochemical rewrite rule $\ell : \bar{a}_i \stackrel{k_i}{\longmapsto} \bar{b}_i \in \mathscr{B}$ is applied initially according to the stochastic semantics, since the associated compartment contains a very small number of reagents. After the system has evolved for some time, however, the concentration of the reagents involved in the rule can be substantially increased and it becomes convenient to model the corresponding reaction according to the deterministic approach.

Actually, at the beginning of each simulation step we build, for each compartment in the term, a system of ODEs for the simulation of the biochemical rules in that compartment which (1) are sufficiently fast and (2) involve reagents with a sufficient concentration. For the remaining rules the evolution is determined by the stochastic simulation algorithm.

In order to describe the hybrid semantics we assume that, given a CWC term $\bar{t}$, each compartment of $\bar{t}$ is univocally identified by an index ι. The index of the (implicit) compartment at the top level will be denoted by ι_0. The *biochemical reagents* of a compartment $(\bar{a} \rfloor \bar{t})^\ell$ with index ι, written $\mathrm{BR}(\iota)$, are expressed by the multiset of the atomic elements appearing in the top level of $\bar{t}$. For example, given the term

$$\bar{t} = a\ a\ b\ (c \rfloor (d\ e \rfloor \bullet)^{\ell'}\ f)^\ell\ (c \rfloor\ f\ g)^\ell$$

and assuming that the compartment $(c \rfloor (d\ e \rfloor \bullet)^{\ell'}\ f)^\ell$ has index ι_1, the compartment $(d\ e \rfloor \bullet)^{\ell'}$ has index ι_2 and the compartment $(c \rfloor\ f\ g)^\ell$ has index ι_3, we have that $\mathrm{BR}(\iota_0) = a\ a\ b$, $\mathrm{BR}(\iota_1) = f$, $\mathrm{BR}(\iota_2) = \bullet$ and $\mathrm{BR}(\iota_3) = f\ g$.

A basic point of our hybrid approach is the criterium to determine, at each computation stage, the reductions to compute in stochastic or in a deterministic way. In this paper we have chosen simply to put a threshold on the number of possible reagents and on the speed of the reaction, but other more sophisticated criteria should be investigated.

Let $\bar{t}$ denote the whole term and let I denote the set of compartment indexes occurring in $\bar{t}$.

1. For each compartment $\iota \in I$:
   - Let ℓ be the label of ι, let $\mathscr{D}_\iota = \mathscr{B}_\ell$ and let $\mathscr{S}_\iota = \emptyset$.
   - For each biochemical rule $B_i = \ell : \bar{a}_i \overset{k_i}{\longmapsto} \bar{b}_i \in \mathscr{D}_\iota$ let $\bar{a}_i = a_{i_1} \ \ldots \ a_{i_{r_i}}$ ($r_i \geq 1$) and let $[a_{i_1}]^\iota, \ldots, [a_{i_{r_i}}]^\iota$ denote the concentrations of the species occurring in $\bar{a}_i$ within the multiset $\mathrm{BR}(\iota)$. Let K_i^ι be the rate of the application of rule B_i in ι. Namely, $K_i^\iota = k_i \cdot [a_{i_1}]^\iota \cdot \ldots \cdot [a_{i_{r_i}}]^\iota$. If $K_i^\iota < \phi$ or $min\{[a_{i_1}]^\iota, \ldots, [a_{i_{r_i}}]^\iota\} < \psi$ remove B_i from $\mathscr{D}_\iota$ and put it into $\mathscr{S}_\iota$.
2. Considering the rules in $\bigcup_{\iota \in I} \mathscr{S}_\iota \cup \mathscr{N}$ select according to Gillespie's method and to the semantics given in Section 3.1 a stochastic transition step $C[P\sigma] \xrightarrow{k \cdot \mathrm{Occ}(R, C[P\sigma], C[O\sigma])} C[O\sigma]$), where $R = \ell : P \overset{k}{\longmapsto} O \in \mathscr{S}_{\iota'} \cup \mathscr{N}_\ell$
   Let τ be the corresponding time interval.[†]
3. For each compartment ι in I:
   - Let $\mathscr{E}_\iota$ denote the system of ODEs for the rules in $\mathscr{D}_\iota$ in the compartment ι as explained in Section 3.2 without considering, in the compartment ι' where the stochastic transition step takes place, the active reagents appearing in the left part P of the stochastically applied rule. (If $\mathscr{D}_\iota = \emptyset$ then $\mathscr{E}_\iota = \emptyset$.)
   - Apply the system of ODEs $\mathscr{E}_\iota$ to the biochemical reagents $\mathrm{BR}(\iota)$ of the compartment for a time duration τ.
4. Update the term $\bar{t}$ according to the right part O of the chosen stochastic rule and to the applications of the systems of ODEs.

[†] In some rare case, it may happen that no rule in $(\bigcup_{\iota \in I} \mathscr{S}_\iota) \cup \mathscr{N}$ is applicable. In such cases the evolution of the system must be determined for some time τ according to the deterministic semantics only. In our implementation we choose as τ the maximum time calculated by Gillespie's algorithm for each of the applicable biochemical rules in $\bigcup_{\iota \in I} \mathscr{D}_\iota$.

Figure 6: Steps performed by an hybrid simulation iteration

Given a term $\bar{t}$ to reduce, a rate thresholds ϕ and a concentration threshold ψ, each iteration of the hybrid reduction semantics performs the four steps listed in Figure 6, where, for every label ℓ, the subsets of $\mathscr{B}$ and $\mathscr{N}$ containing the rules with label ℓ are denoted by $\mathscr{B}_\ell$ and $\mathscr{N}_\ell$, respectively. The first step identifies, for each compartment $\iota \in I$ (where I is the set of all compartment indexes occurring in $\bar{t}$), two disjoint sets of biochemical rules, namely $\mathscr{D}_\iota$ (to be applied deterministically) and $\mathscr{S}_\iota$ (to be applied, together with the rules in $\mathscr{N}$, according to the stochastic method). The second step selects, considering only the rules in $\bigcup_{\iota \in I} \mathscr{S}_\iota \cup \mathscr{N}$ the next rule to be applied stochastically. The third step computes a system of ODEs $\mathscr{E}_\iota$ for each compartment $\iota \in I$ and applies the ODEs for the time duration selected by the stochastic step. The fourth step updates the terms according to the results of the simulation.[5]

In general, if reactions are fast enough, the deterministic ODEs simulation approximate better the exact stochastic simulations. This is the idea behind the use of the threshold ϕ. The use of ψ, instead, allows to prevent the rounding approximation error that may derive when we are dealing with species at low concentrations. Combined together, the thresholds ϕ and ψ affect the level of approximation we

[5] Note that since ODEs deal with fractional quantities, a rounding operation will be needed before computing the next stochastic step.

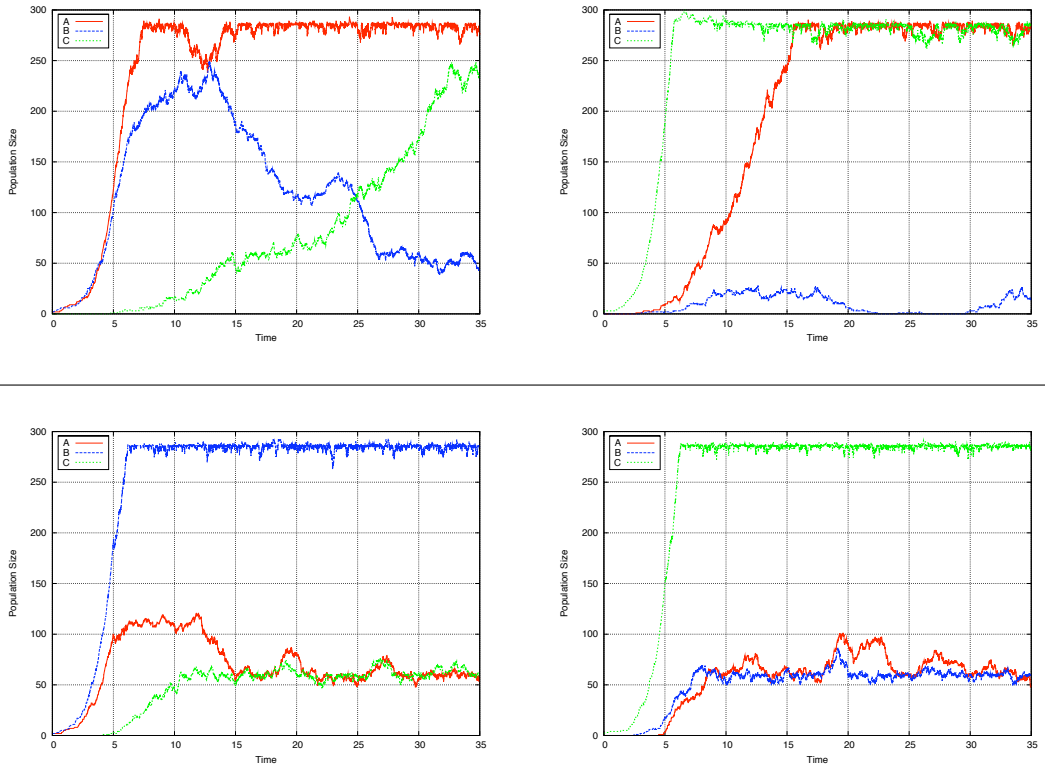

Figure 7: Two different runs of the hybrid simulations showing the different behaviour of the dynamics of the competitive species inside the compartment *IN* (on the left side) and outside the compartment (on the right side)

want to use in our simulations. Notice that with $\phi = +\infty$ all reactions will be considered *too slow* and the simulation will be computed with the purely stochastic method.

Running Example: Hybrid Simulations The hybrid simulations were performed by using thresholds $\phi = 60$ and $\psi = 60$, which were determined as feasible thresholds to catch the initial stochastic effects for the multistable behaviour of the dynamic system.

In Figure 7 we report two runs of the hybrid simulations showing two different evolutions of the competitive species inside the compartment *IN* (on the left side) and outside the compartment (on the right side). The hybrid method allows the synthesis of these evolutions of the system (among the many possible ones) by identifying, through the choice of the thresholds ϕ and ψ, the system state in which the populations are considered large enough to allow differential equations to take the government of the system evolution. In this example in fact, as it is easy to verify, while the destiny of the species is determined by the stochastic bootstrap, after the populations have reached the chosen threshold their evolution becomes regulated only by ODE transitions. The deterministic computation of the evolution of the system then tracks an average behaviour reducing the computational load. At this point, in fact, stochastic fluctuations, can be considered irrelevant for the description of the overall behaviour of the system. The moment in which the simulations move from the stochastic method to the deterministic one

can be easily captured in the graphs shown in Figure 7.

The hybrid method provides a faster final solution with respect to the pure stochastic approach by a factor of 10 giving a very satisfiable qualitative analysis of its behaviour.

5 A Real Model of Different Cellular Fate

To assess the soundness and efficiency of our hybrid approach on a real biological problem we decided to apply it to a well known system where stochastic effects play a fundamental role in determining its development: the HIV-1 transactivation mechanism.

After a cell has been infected, the retrotransposed DNA of the virus is integrated in the host genome and it begins its transcription in *mRNA* and then the translation to yield viral proteins; the initial speed of this mechanism, however, is fairly slow. The speedup of the viral production process is determined by a regulation system driven by the viral protein *TAT*: this protein is capable of binding cellular factors of the host to produce the *pTEFb* complex which in its acetylated form is able to bind to the integrated viral genome and speed up the transcription machinery, thus ending in more viral proteins and, therefore, more *TAT*, determining a positive loop.

The time scale during which this loop is triggered is affected by many factors: the initial low *TAT* production and the rate of its degradation, the equilibrium between the active (acetylated) and inactive form of *pTEFb* and so on. As a consequence, the stochastic oscillations in this events are considered pivotal in determining when viral proteins are produced in a sufficient quantity to determine cellular lysis and viral spreading. Since HIV is known to stay dormant and inactive in some types of cells and since the time between the infection and the high viral production rate related to the active phase of AIDS is variable, this transactivation mechanism is of great interest.

We decided to follow the direction taken in a previous study about this system (see [25]), in which an experimental setting is developed where a fluorescent protein, *GFP*, is the only one encoded by an engineered viral genome, along with TAT. In [25] they were able to identify different evolutions in the *GFP* level over time: cellular clones with exactly the same genome showed two different behaviour, one produced a high quantity of *GFP* (they called it "bright") and the other one with very little *GFP* ("off"). This work also reported that a purely stochastic simulation was able to individuate this bifurcation; a later work (see [14]) confirmed these results performing purely stochastic and mixed deterministic-stochastic simulations.

Since CWC systems are able to represent compartments, we slightly modified the original set of rules used in these works to explicitly represent the cytoplasm and the nucleus of an infected cell; all the kinetic rates were maintained, the one for *TAT* nuclear import has been determined from the literature (see [17]). The set of rules we adopted is given in Figure 8, where we refer to the cytoplasm as the $\top$ compartment while η is the label used for the nucleus. As regards the rules: (B_1) represents the slow basal rate of viral *mRNA* transcription; (N_1) describes the *mRNA* export from the nucleus to the cytoplasm; (B_2) and (B_3) express the translations of this *mRNA* into *GFP* and *TAT* proteins, respectively; (N_2) and (N_3) represent the nuclear import and export of *TAT*; (B_4) and (B_5)models the binding and unbinding of *TAT* with (not represented here) host cellular factors and the viral genome portion *LTR* that forms *pTEFb* which, when acetylated (by rule (B_6)) determines an higher transcriptional activity, which is represented in (B_8) by the unbinding that releases *LTR* and *TAT* and creates an *mRNA* molecule (note the higher rate with respect to (B_1)); (B_7) represents the *pTEFb* deacetylation and (B_9), (B_{10}) and (B_{11}) model the degradation processes of the proteins and the *mRNA* (note that *mRNA* degrades both in the nucleus and in the cytoplasm, the other proteins only degrade in the cytoplasm; also note how the compartment

$$(B_1) \quad \eta : LTR \xmapsto{1\times10^{-9}} LTR\ mRNA$$
$$(N_1) \quad \top : (x \rfloor mRNA\ X)^{\eta} \xmapsto{7.2\times10^{-4}} (x \rfloor X)^{\eta}\ mRNA$$
$$(B_2) \quad \top : mRNA \xmapsto{0.5} mRNA\ GFP$$
$$(B_3) \quad \top : mRNA \xmapsto{1.32\times10^{-3}} mRNA\ TAT$$
$$(N_2) \quad \top : (x \rfloor X)^{\eta}\ TAT \xmapsto{8.5\times10^{-3}} (x \rfloor TAT\ X)^{\eta}$$
$$(N_3) \quad \top : (x \rfloor TAT\ X)^{\eta} \xmapsto{7.2\times10^{-4}} (x \rfloor X)^{\eta}\ TAT$$
$$(B_4) \quad \eta : TAT\ LTR \xmapsto{1.5\times10^{-4}} pTEFb$$
$$(B_5) \quad \eta : pTEFb \xmapsto{17\times10^{-3}} TAT\ LTR$$
$$(B_6) \quad \eta : pTEFb \xmapsto{1\times10^{-3}} pTEFb\_ac$$
$$(B_7) \quad \eta : pTEFb\_ac \xmapsto{0.13} pTEFb$$
$$(B_8) \quad \eta : pTEFb\_ac \xmapsto{0.1} LTR\ TAT\ mRNA$$
$$(B_9) \quad \top : GFP \xmapsto{3.01\times10^{-6}} \bullet$$
$$(B_{10}) \quad \top : TAT \xmapsto{4.3\times10^{-5}} \bullet$$
$$(B_{11}) \quad \top, \eta : mRNA \xmapsto{4.8\times10^{-5}} \bullet$$

Figure 8: CWC rules for the TAT transactivation system

labelling mechanism allows to express this fact in a simple and elegant way).

We performed 100 purely stochastic simulations (i.e. setting $\phi = +\infty$) and 100 hybrid simulations (using $\phi = 0.5$ and $\psi = 10$). The values of ϕ and ψ for the hybrid simulation were determined, after a few experiments, as the right values to be able to grasp the stochastic effects for this system. The initial term of our simulations is represented by the CWC term

$$75000 \times GFP\ 5 \times TAT\ (\bullet \rfloor LTR)^{\eta},$$

while the time interval of our simulations has been fixed to 10^6 seconds (the same parameters are used in [25, 14]). Both our stochastic and hybrid simulations clearly showed the two possible evolutions of the system which correspond to the "bright" and the "off" cellular populations (in order to display the double destiny, almost all the biochemical rewrite rules have to be simulated with the stochastic approach). As could be seen in Figure 9, the hybrid simulations are comparable to the purely stochastic ones and, even with the relatively high thresholds used in this particular case, the hybrid simulations were computationally more efficient (almost 40% faster).[6]

6 Conclusions

As we have seen, CWC allows to model cellular interaction, localisation and membrane structures. Other formalisms were developed to describe membrane systems. Among them we cite Brane Calculi [6] and P-Systems [20].

CWC can describe situations that cannot be easily captured by the previously mentioned formalisms, which consider membranes as atomic objects (extensions of P-Systems with objects on membranes can

[6]Comparisons are made using the same stochastic engine, in both cases with no particular optimization.

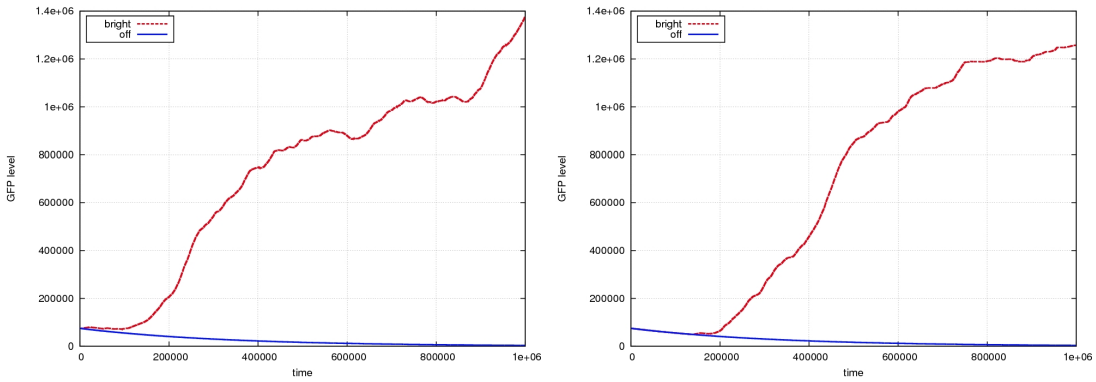

Figure 9: Two different simulations pure stochastic (on the left) and hybrid (on the right) started with the same parameters: "bright" and "off" behaviour

be found in [5, 7]). Representing the membrane structure as a multiset of the elements of interest allows the definition of different functionalities depending on the type and the number of elements on the membrane itself.

In this paper we have defined a hybrid simulation technique for systems described in CWC, which combines the stochastic approach with the deterministic one obtained through ODEs. The method alternates discrete transitions, computed probabilistically according to the stochastic method, and continuous transitions, computed in a deterministic way by a set of ODEs. Our technique turns out to accurately capture the dynamics of systems that exhibit stochastic effects and takes advantage, whenever the deterministic approach is applicable, of the efficiency of the ODEs simulation method.

The running examples used to make comparisons between the different quantitative simulation methods, and the HIV-1 transactivation mechanism are challenging tests for our hybrid methodology. On the one hand, the running example shows that, from the methodological point of view, several runs of hybrid simulations on a model of dynamic competitive populations allow to "synthesize" a set of stochastic experiments avoiding the statistical assumptions about the initial distributions of the parameters that are needed in a purely deterministic or purely stochastic analysis. On the other hand, the simulation of the HIV-1 transactivation mechanism follows a simulation which is *almost* purely stochastic: only a few rules pass the threshold condition, thus the computational gain of the deterministic approach is, in this particular case, very limited (even if still sensible).

Compartment labels introduced in this paper are a novelty with respect to the original CWC calculus presented in [8]. As we have seen, these labels are necessary when building a system of ODEs for a compartment of type ℓ. However, we might exploit these labels as an intrinsical information about the properties of a compartment. For example, assuming that compartments of the same type have approximatively the same volume, we might use the compartment type to define a set of biochemical rules whose kinetics incorporate the information about the volume of the compartment on which the rule could be applied. Suppose, in practice, to analyse a system in which two different kind of cells may interact. Let's call ℓ_1 and ℓ_2 the compartment types of the two kinds of cells. Suppose, then, that particles a and b are free to float between these cells and the top level interspace hosting all the cells. Finally, particles a and b may interact by complexation and produce the particle c. If it holds that the top level interspace on which the different cells float has around 100x the volume of a cell of type ℓ_1 and if a cell of type ℓ_1 has around 3x the volume of a cell of type ℓ_2, we can express the different speeds of the $a-b$ complexation in the different compartments (according to their volumes) with the three following

rules:

$$\top : a\ b \overset{k}{\longmapsto} c, \qquad \ell_1 : a\ b \overset{k\cdot 100}{\longmapsto} c, \qquad \ell_2 : a\ b \overset{k\cdot 300}{\longmapsto} c.$$

Actually, it is crucial to consider in detail the volumes of the involved compartments and to consider adequate kinetics for the biochemical rules used to simulate the system behaviour. We notice, in particular, that the approach based on ODEs directly translates chemical reactions into mathematical equations and computes the concentrations over time of the involved species (usually the *molar concentration*, which denotes the number of moles of a given substance per liter). Models based on the stochastic approach, instead, simulate the activity of each single individual involved in the evolution of the system. Such a delicate difference between the two methods should be carefully taken into account when developing the set of rules to be simulated with the hybrid approach. The version of CWC with labelled compartments presented in this paper simplifies this kind of analysis and allows for more accurate simulations.

Acknowledgements

We gratefully acknowledge the helpful comments and suggestions received from the anonymous reviewers of MeCBIC 2010.

The authors also wish to thank Sergio Rabellino and the ICT staff of the Computer Science Department of the University of Turin for providing technical support and assistance in running the simulations. Finally, we thank Prof. Nello Balossino (University of Turin) who made us available the computing resources of the laboratory Segnali e Immagini "G. Tamburelli".

References

[1] M. Aldinucci, M. Coppo, F. Damiani, M. Drocco, E. Giovannetti, E. Grassi, E. Sciacca, S. Spinella & A. Troina (2010): *CWC Simulator*. Dipartimento di Informatica, Università di Torino. http://sourceforge.net/projects/cwcsimulator/.

[2] R. Alur, C. Belta & F. Ivancic (2001): *Hybrid Modeling and Simulation of Biomolecular Networks*. In: *HSCC*, *LNCS* 2034, Springer, pp. 19–32.

[3] R. Barbuti, A. Maggiolo-Schettini, P. Milazzo, P. Tiberi & A. Troina (2008): *Stochastic Calculus of Looping Sequences for the Modelling and Simulation of Cellular Pathways*. Transactions on Computational Systems Biology IX, pp. 86–113.

[4] R. Barbuti, A. Maggiolo-Schettini, P. Milazzo & A. Troina (2006): *A Calculus of Looping Sequences for Modelling Microbiological Systems*. Fundam. Inform. 72(1-3), pp. 21–35.

[5] R. Brijder, M. Cavaliere, A. Riscos-Núñez, G. Rozenberg & D. Sburlan (2008): *Membrane systems with proteins embedded in membranes*. Theor. Comput. Sci. 404(1-2), pp. 26–39. Available at `http://dx.doi.org/10.1016/j.tcs.2008.04.002`.

[6] L. Cardelli (2005): *Brane Calculi*. In: *CMSB*, *LNCS* 3082, Springer, pp. 257–278.

[7] M. Cavaliere & S. Sedwards (2008): *Decision problems in membrane systems with peripheral proteins, transport and evolution*. Theor. Comput. Sci. 404(1-2), pp. 40–51. Available at `http://dx.doi.org/10.1016/j.tcs.2008.04.003`.

[8] M. Coppo, F. Damiani, M. Drocco, E. Grassi & A. Troina (2010): *Stochastic Calculus of Wrapped Compatnents*. In: *QAPL'10*, 28, EPTCS, pp. 82–98.

[9] A. Crudu, A. Debussche & O. Radulescu (2009): *Hybrid stochastic simplifications for multiscale gene networks*. BMC Syst Biol 3:89.

[10] V. Danos & C. Laneve (2004): *Formal molecular biology*. Theor. Comput. Sci. 325(1), pp. 69–110.

[11] P. Degano, D. Prandi, C. Priami & P. Quaglia (2006): *Beta-binders for Biological Quantitative Experiments*. *Electr. Notes Theor. Comput. Sci.* 164(3), pp. 101–117. Available at `http://dx.doi.org/10.1016/j.entcs.2006.07.014`.

[12] L. Dematté, C. Priami & A. Romanel (2008): *The Beta Workbench: a computational tool to study the dynamics of biological systems*. *Briefings in Bioinformatics* 9(5), pp. 437–449. Available at `http://dx.doi.org/10.1093/bib/bbn023`.

[13] D. Gillespie (1977): *Exact stochastic simulation of coupled chemical reactions*. *J. Phys. Chem.* 81, pp. 2340–2361.

[14] M. Griffith, T. Courtney, J. Peccoud & WH. Sanders (2006): *Dynamic partitioning for hybrid simulation of the bistable HIV-1 transactivation network*. *Bioinformatics* 22, pp. 2782–2789.

[15] J. Krivine, R. Milner & A. Troina (2008): *Stochastic Bigraphs*. *Electron. Notes Theor. Comput. Sci.* 218, pp. 73–96.

[16] H. Matsuno, A. Doi, M. Nagasaki & S. Miyano (2000): *Hybrid Petri net representation of gene regulatory network*. In: *Prooceedings of Pacific Symposium on Biocomputing*, World Scientific Press, pp. 341–352.

[17] N. Nitin, L. LaConte, WJ. Rhee & G. Bao (2009): *Tat peptide is capable of importing large nanoparticles across nuclear membrane in digitonin permeabilized cells*. *Ann Biomed Eng* 37, pp. 2018–2027.

[18] C. Priami (1995): *Stochastic pi-Calculus*. *Comput. J.* 38(7), pp. 578–589.

[19] C. Priami, A. Regev, E. Y. Shapiro & W. Silverman (2001): *Application of a stochastic name-passing calculus to representation and simulation of molecular processes*. *Inf. Process. Lett.* 80(1), pp. 25–31.

[20] G. Păun (2002): *Membrane computing. An introduction*. Springer.

[21] A. Regev & E. Shapiro (2002): *Cells as computation*. *Nature* 419, p. 343.

[22] H. Salis & Y. Kaznessis (2005): *Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions*. *J Chem Phys* 122(5):54103.

[23] SJ. Singer & GL. Nicolson (1972): *The fluid mosaic model of the structure of cell membranes*. *Science* 175, pp. 720–731.

[24] S. Smale (1975): *On the differential equations of species in competition*. *Journal of Mathematical Biology* 3, pp. 5–7.

[25] LS. Weinberger, JC. Burnett, JE. Toettcher, AP. Arkin & DV. Schaffer (2005): *Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity*. *Cell* 122, pp. 169–182.