Math. Struct. in Comp. Science (2013), *vol.* 23, *pp.* 471–503.
doi:10.1017/S0960129512000199

Multi-level modelling via stochastic multi-level multiset rewriting[†]

NICOLAS OURY and GORDON PLOTKIN

University of Edinburgh, Edinburgh, United Kingdom
Email: {Gordon.Plotkin;Nicolas.Oury}@ed.ac.uk

Received 11 November 2010; revised 14 July 2011

We present a simple stochastic rule-based approach to multi-level modelling for computational systems biology. Populations are modelled using multi-level multisets; these contain both species and agents, with the latter possibly containing further such multisets. Rules are pairs of such multisets, but they may now also include variables (as well as species and agents), together with an associated stochastic rate.
We give two illustrative examples. The first is an extracellular model of virus infection, coupled with an intracellular model of viral reproduction; this model can demonstrate successive waves of infection. The second is a model of cell division in which a repressor protein is diluted in successive generations, so eventually repression no longer occurs.
The multi-level multiset approach can also be seen in terms of stochastic term rewriting for the theory of a commutative monoid equipped with extra constants (for the species) and unary operations (for the agents). We further discuss the relationship of this approach with two others: Krivine *et al.*'s stochastic bigraphs, restricted to Milner's place graphs, and Coppo *et al.*'s Stochastic Calculus of Wrapped Compartments. These various relationships provide evidence for the fundamental nature of the approach.

1. Introduction

We present a simple rule-based formalism for multi-level modelling of biological processes. We are interested particularly in modelling both inter- and extra-cellular events, for example signalling and cell division. To that end, we present a stochastic rule-based formalism of multi-level multisets. These are, essentially, nested multisets; more exactly, they are finite multisets whose elements are either species names or pairs of an agent name, for example, Cell, and a multi-level multiset, with this multiset nesting carried on only to finite depth. The agents serve to indicate a lower level and its kind. The rules use similar multisets, but also allow variables.

Multi-level modelling involves multi-scale modelling, and much work has been done on both. For reviews, see Meier-Schellersheim *et al.* (2009), Bauer *et al.* (2009), Grieneisen and Scheres (2009), Chickarmane *et al.* (2010) and Noble (2002); some specific systems are BioCharts (Kugler *et al.* 2010; Harel and Kugler 2010), Simmune (Meier-Schellersheim *et al.* 2006) and CompuCell3D (Cickovski *et al.* 2007). Most of these systems have specific

[†] This work was supported by BBSRC/EPSRC Grant BB/D019621/1, and by a Royal Society–Wolfson Award.

modelling scenarios in mind. The computer science community has provided general purpose formalisms, often taking ideas from process calculus. The first of these was Regev *et al.*'s BioAmbients (Regev *et al.* 2004), which has features of the stochastic pi-calculus (Priami *et al.* 2001) and constructs that enable the movement of agents into and out of other agents. Another example is Cardelli's Brane Calculi (Cardelli 2005), which has, essentially, our multi-level multisets in algebraic form together with other structure, including actions, which distinguishes it from rule-based approaches.

There are also rule-based formalisms, such as: Dynamical Grammars (Mjolsness and Yosiphon 2006); Milner's bigraphs (Milner 2009), which were adapted to biological ends in Krivine *et al.* (2008); and Coppo *et al.*'s Stochastic Calculus of Wrapped Compartments (Coppo *et al.* 2010a; Coppo *et al.* 2010b), which is a descendent of the Stochastic Calculus of Looping Sequences (Barbuti *et al.* 2008a; Barbuti *et al.* 2008b). Păun introduced dynamic compartments into membrane computing, which can also be considered a rule-based formalism: see, for example, Păun (2001), Păun (2008) and Frisco (2009); see also Spicher *et al.* (2008) for work on stochastic dynamic compartments. The work presented in the current paper is closely related, though, in its current formulation, the range of rules in membrane computing seems less varied. Bitonal membrane systems (Cardelli 2008) are in the brane calculi family, but stripped of actions and equipped with rules instead, so they can be viewed as particular rule-based systems, more or less of the kind considered here.

A standard formalism for reactions is multiset rewriting over a set of constants standing for various species. If each rule is given a rate, we obtain stochastic multiset rewriting, which is equivalent to stochastic Petri nets. All this is well known, as are the uses of multiset rewriting and Petri nets for modelling biological systems. Here, as indicated above, we extend these ideas in just one way by adding unary function symbols to the rewriting formalism, and this single extension enables us to do multi-level modelling.

An example stochastic multiset rewriting rule is

$$\text{Glucose}, \text{ATP} \xrightarrow{k_1} \text{Gluc-6-P}, \text{ADP}.$$

This models a reaction from the glycolysis pathway with stochastic rate k_1. Suppose we also wish to indicate that the reaction only takes place inside cells. Then we introduce a unary function symbol Cell and write the same rule 'one level down':

$$\text{Cell}(\text{Glucose}, \text{ATP}, x) \xrightarrow{k_1} \text{Cell}(\text{Gluc-6-P}, \text{ADP}, x).$$

The variable x is used to indicate that, other than the molecule of glucose and the molecule of ATP, the contents of the cell remain unchanged. We can also write rules that mix levels. For example, a rule of the form

$$\text{H}_2\text{O}, \text{Cell}(x) \xrightarrow{k_2} \text{Cell}(\text{H}_2\text{O}, x)$$

might be used for the passive transport of water into cells.

Multiset rewriting with species can be viewed algebraically as rewriting modulo an associative–commutative (AC) operation with a zero and additional constants (Bezem *et al.* 2003). Similarly, our multi-level multiset rewriting can be seen as rewriting modulo

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

an AC operation with a zero, additional constants and additional unary function symbols. As stochasticity is also present, we are led to consider *stochastic* term rewriting. Curiously, although very natural, the idea of stochastic term rewriting seems novel; however, there has been some work on probabilistic term rewriting (Bournez and Kirchner 2002; Bournez and Hoyrup 2003; Agha *et al.* 2006).

We give a more detailed comparison of our formalism with that of Krivine *et al.*'s and Coppo *et al.*'s in Section 5; in particular, it corresponds to 'half' of Milner's bigraphs – his place graphs. There is no doubt that the three formalisms are very closely related. We would argue that ours is particularly simple to understand, and easy and natural to use when modelling multi-level biological systems. Furthermore, by working within term rewriting, with its algebraic setting, one employs a very standard approach; this contrasts with other rule-based approaches which, while perfectly sound, are, perhaps, somewhat *ad hoc*. We expect the term rewriting approach to provide a sound basis for further extension; this point is discussed further with our conclusions in Section 6. Overall, the fact that all three approaches are very much the same encourages us to think that multi-level multisets provide a fundamental structure for multi-level rule-based modelling.

1.1. *Structure of the paper*

We begin, in Section 2, with an illustrative example of viral infection and reproduction. The model demonstrates a second wave of infection; it seems to be the first multi-level model to do so. Spicher et al (Spicher *et al.* 2008) have also given a stochastic viral model in the form of a stochastic dynamical P systems model of the Semliki viral life cycle. We have implemented our rule-based formalism using a version of the standard Gillespie direct method (Gillespie 1977), and we give illustrative runs of our examples. We present our stochastic multi-level multiset rewriting formalism in Section 3. Populations are, as we have seen, modelled as multi-level multisets. Each rule has a given base rate, and the rate at which it applies in a given population is the product of the base rate and its activity, that is, the number of ways in which it can match the population.

We present another example in Section 4, again giving illustrative runs; this example was inspired by Rosenfeld *et al.* (2005), and features cell division. Next, we compare our work with the previous work of Krivine *et al.* and Coppo *et al.* in Section 5, and also discuss the algebraic formulation of stochastic multi-level multiset rewriting. Finally, we discuss some possibilities for future work in Section 6.

2. An example: viral infection

We consider a simple model of viral infection taken from Srivastava (2002) and Haseltine *et al.* (2005). There are two levels: intracellular and extracellular. A simple intracellular model of infected cells is given in Srivastava (2002), and this is combined with an extracellular model of viral infection given in Haseltine *et al.* (2005). Intracellularly, infected cells can incorporate viral protein into their genome, and then produce viral structural protein and genomic viral nucleic acid, both of which may degrade. Extracellularly, viruses can

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

invade uninfected cells, and infected cells can produce viruses from viral structural protein and genomic viral nucleic acid, or die.

The intracellular part of the model involves the following species:

gen	genomic viral nucleic acid
str	viral structural protein
tem	template viral nucleic acid
V_1, V_2	viral enzymes.

These take part in the following reactions, which are modelled in the standard way as stochastic multiset rewriting rules:

$$\begin{aligned} \mathrm{gen}, \mathrm{V}_1 &\xrightarrow{k_1} \mathrm{tem}, \mathrm{V}_1 \\ \mathrm{tem}, \mathrm{V}_2 &\xrightarrow{k_2} \mathrm{str}, \mathrm{tem}, \mathrm{V}_2 \\ \mathrm{tem} &\xrightarrow{k_3} \mathrm{gen}, \mathrm{tem} \\ \mathrm{str} &\xrightarrow{k_4} \\ \mathrm{gen} &\xrightarrow{k_{10}} . \end{aligned}$$

Turning to the extracellular part of the model, we introduce two new species, Virus and Cell, which are again modelled by constants, together with an 'infected cell agent' ICell, which is modelled by a unary function symbol. (We may consider uninfected cells as species since the only internal activity we are modelling is that of infected cells, which are modelled using the agent ICell.)

The rules are as follows:

$$\begin{aligned} \mathrm{Virus}, \mathrm{Cell} &\xrightarrow{k_6} \mathrm{ICell}(\mathrm{gen}, 80\mathrm{V}_1, 40\mathrm{V}_2) \\ \mathrm{ICell}(\mathrm{gen}, \mathrm{str}, x) &\xrightarrow{k_5} \mathrm{Virus}, \mathrm{ICell}(x) \\ \mathrm{Virus} &\xrightarrow{k_7} \\ \mathrm{ICell}(x) &\xrightarrow{k_8} \\ \mathrm{Cell} &\xrightarrow{k_9} . \end{aligned}$$

The first rule describes the infection of previously uninfected cells; note that there is no reinfection of infected cells in this model. If we wished to allow infected cells to be further infected, then, instead of working with a constant for uninfected cells and a unary function symbol for infected ones, we would simply work with a unary function symbol for cells – uninfected or not.

The second rule describes viral production; it can be read as saying that, given an infected cell whose population contains one molecule each of gen and str, and the rest of whose population is x, then a virus is produced and the population of the infected cell becomes x. The last three rules concern viral and cell death.

The intracellular reactions can only happen inside (multi-level multisets representing) infected cells, as it is only there that the relevant (sub-)populations will occur. Of course, this can only be seen from the model as a whole. If we wished to make the matter explicit, we could replace the intracellular part of the model by rules involving ICell. For example,

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

Table 1. *The stochastic rates used for the simulation*

k_1	$3.125\ 10^{-4}$	k_6	5
k_2	25	k_7	0.08
k_3	1	k_8	0.005
k_4	1.99	k_9	0
k_5	$7.5\ 10^{-6}$	k_{10}	0.25

the first rule would be replaced by the rule

$$\mathrm{ICell}(\mathrm{gen}, \mathrm{V}_1, x) \xrightarrow{k_1} \mathrm{ICell}(\mathrm{tem}, \mathrm{V}_1, x).$$

We will illustrate the model by showing the results of a few example simulations of this system. Each run shows the number of cells and the numbers of the different molecules plotted against the time in days. Table 1 gives the stochastic rates, in day^{-1}; the figures are taken from the original papers.

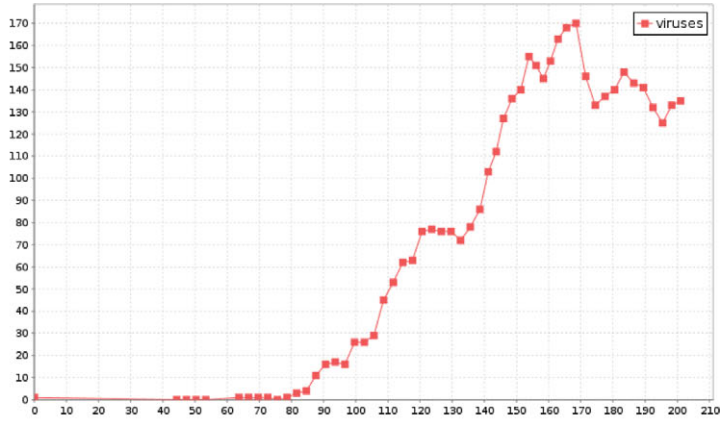
We first simulate the model with an initial population of one cell and one virus. The virus infects the cell and the mechanism to produce more virus is started inside the cell. The result of one such run is shown in Figure 1. In this simulation, we set k_9 to 0 in order to prevent the infected cell from dying before the end of the simulation.

Since we are simulating a stochastic system, different runs will give different results. For example, the delay before the first new virus is produced usually varies from 30 to 100 days. This simulation is similar to those obtained in Srivastava (2002).

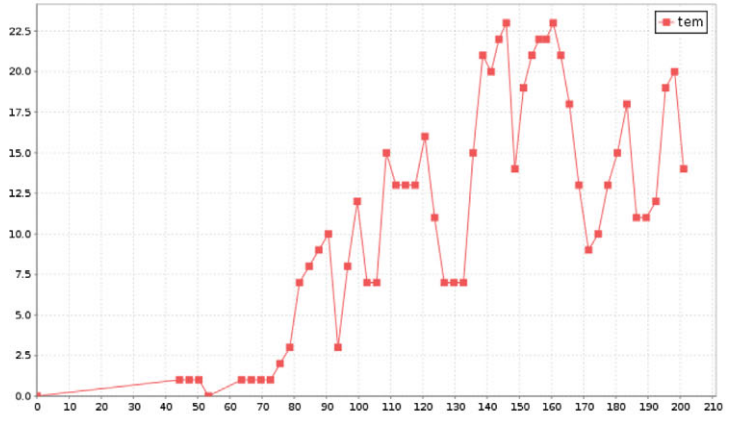
The next simulation is of an infection of many cells by many viruses. Figure 2 shows the result of the infection of 100 cells by 200 viruses. This simulation is similar to the problem studied in Haseltine *et al.* (2005). However, in that paper, the species in the different cells were averaged to enable the construction of differential equations. Here, we benefit from multi-level multiset rewriting by being able to compute an exact stochastic simulation of every cell.

But the real benefit from this framework comes with the simulation of the infection of many cells with few viruses. In this situation, the averaging used in Haseltine *et al.* (2005) is no longer valid: 10% of cells being infected is different from each cell being 10% infected, which does not make much sense. In particular, it is not clear how one would adapt the framework to simulate two or more rounds of viral infection. Figure 3 shows the result of the infection of 100 cells by 5 viruses. In particular, this simulation shows a second wave of infection starting after 125 days (this delay varies stochastically from simulation to simulation), which occurs when enough viral proteins have been made inside the cells infected during the first wave.

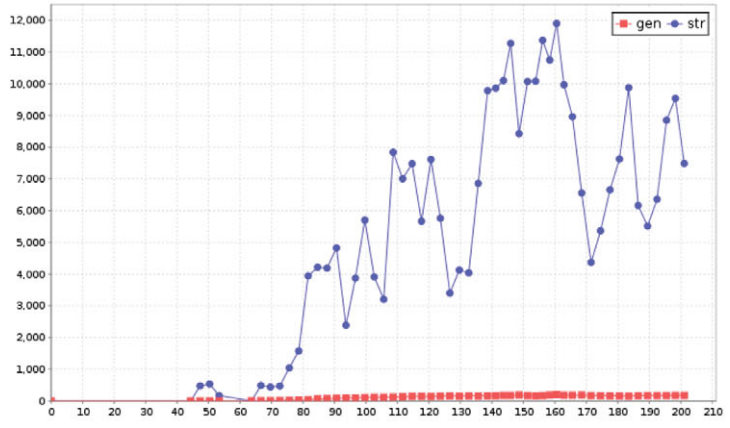
https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press



(a) Virus



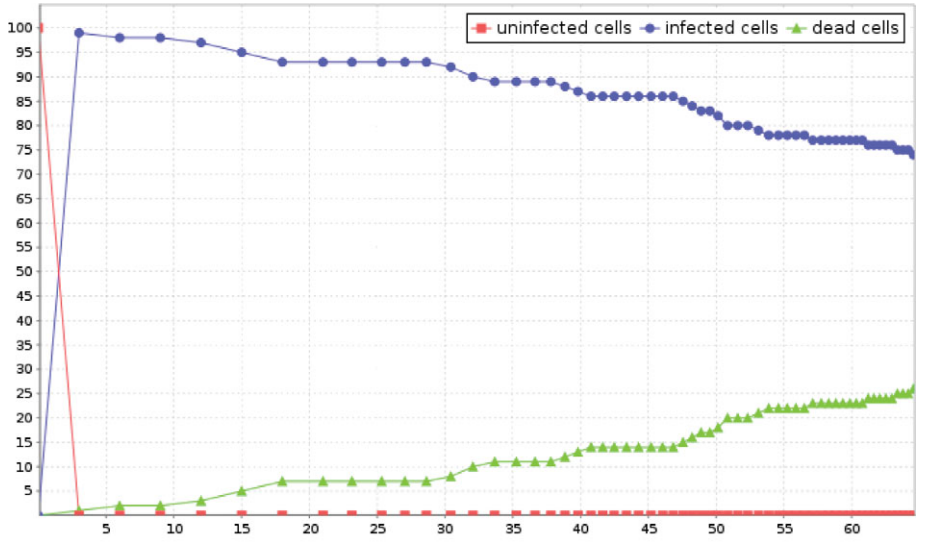
(b) tem



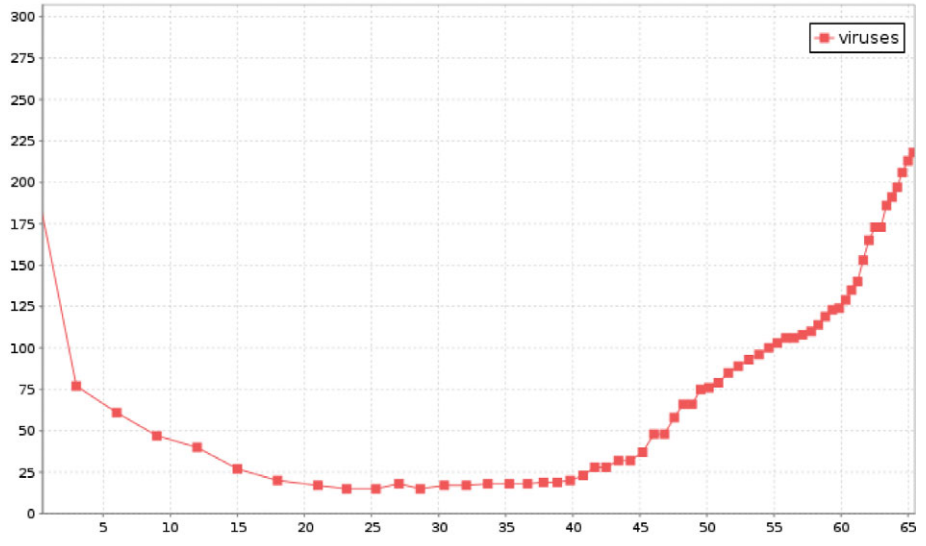
(c) str and a low level of gen

Fig. 1. (Colour online) Simulation of the infection of one cell by one virus

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press



(a) uninfected, infected and dead cells



(b) Virus

Fig. 2. (Colour online) Simulation of the infection of 100 cells by 200 viruses

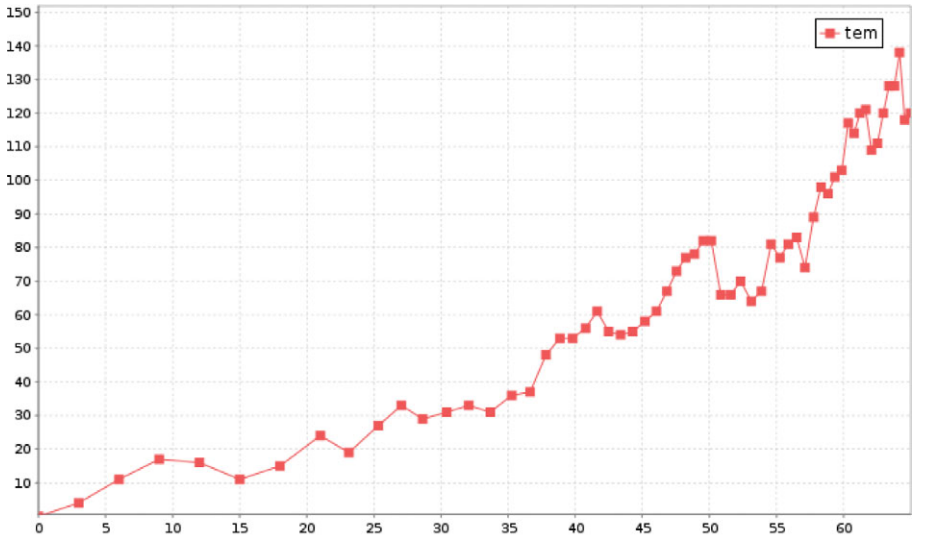
3. Stochastic multi-level multiset rewriting

In this section, we explain:

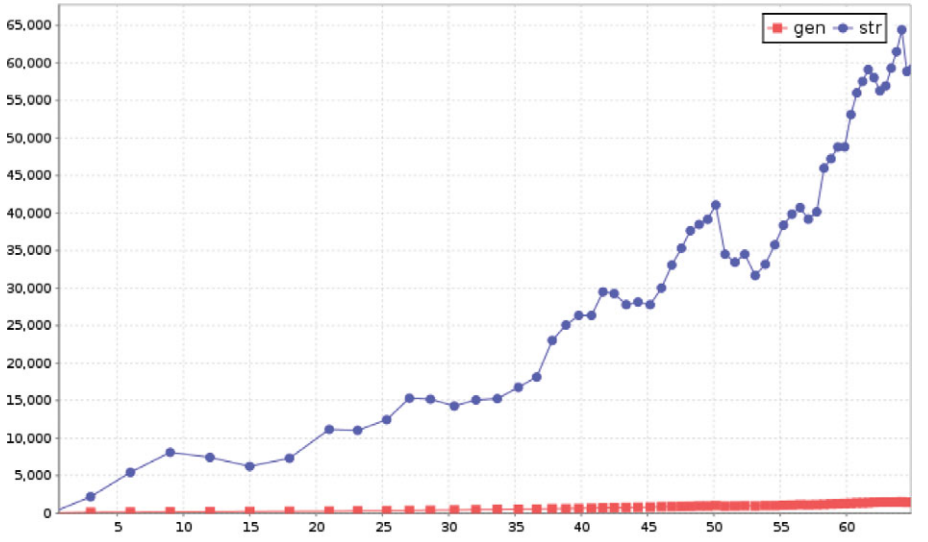
— multi-level multiset terms, which we use to model population states;

— rules and their application, which we use to model system transformations such as reactions, transport and cell creation and division; and

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press



(a) tem



(b) str and a low level of gen

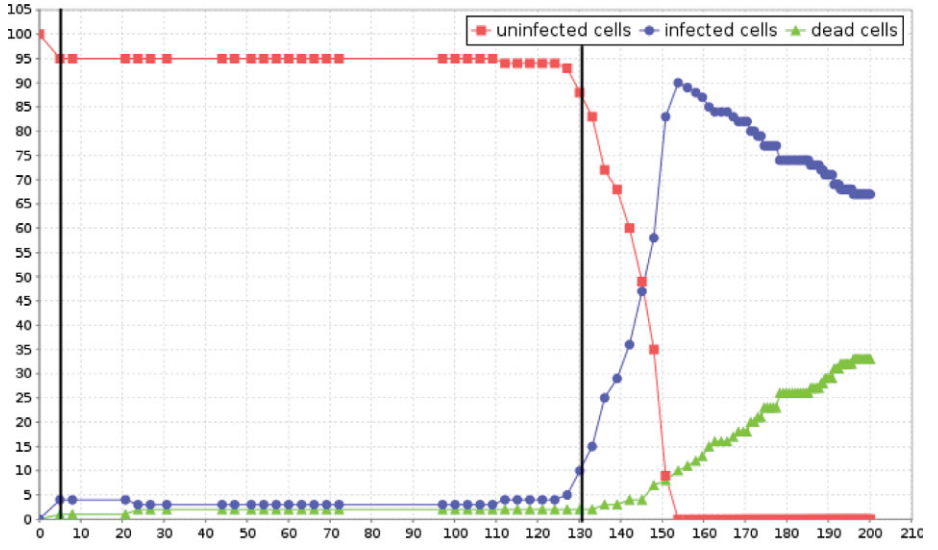
Fig. 2. (continued) Simulation of the infection of 100 cells by 200 viruses

— the activity of rules, which together with their application, determines a stochastic rate matrix.

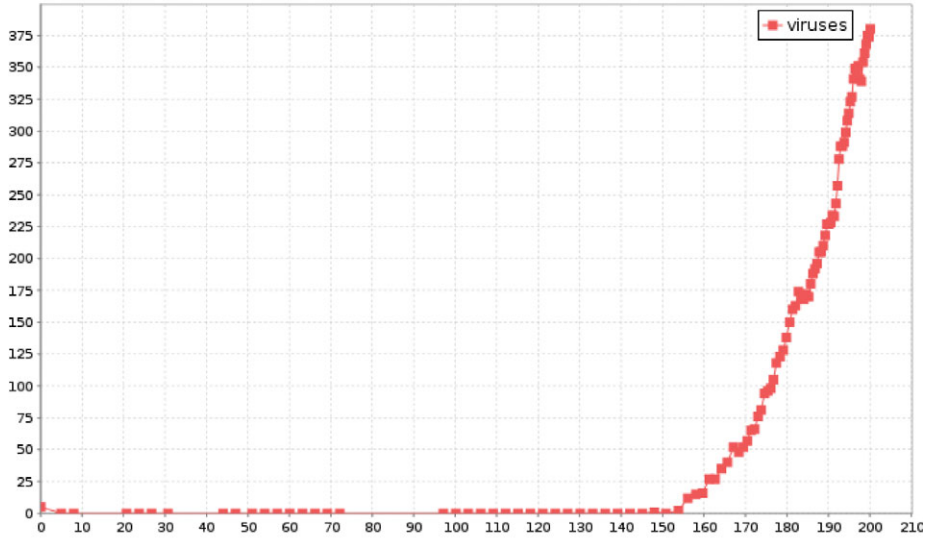
Our implementation samples a run from the corresponding stochastic process.

We need some notation and conventions for multisets. We may identify objects with their corresponding singleton multisets; we use, possibly empty, lists $M_0, \ldots, M_{n-1}$ of multisets to denote their multiset sum; and we write nM (where $n \in \mathbb{N}$) for the n-fold

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press



(a) uninfected, infected and dead cells. The vertical lines show the two waves of infection



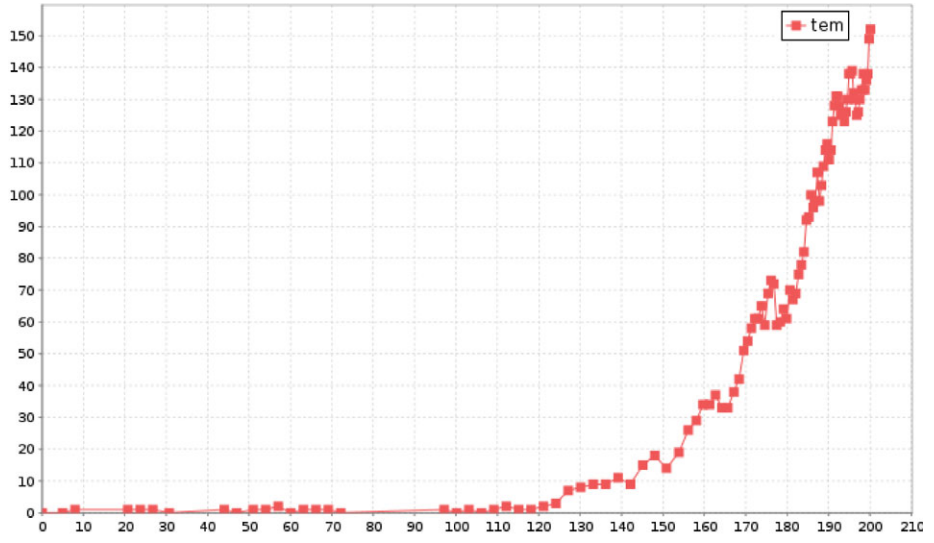
(b) Virus

Fig. 3. (Colour online) Simulation of the infection of 100 cells by 5 viruses

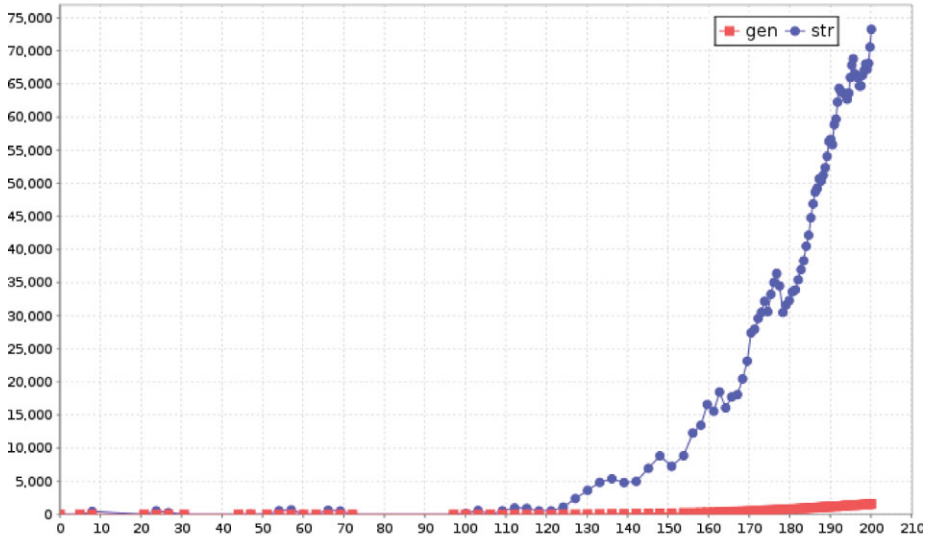
multiset sum of a multiset M with itself. We also use standard notation, such as 0, $M+N$, or $\sum_{i=0}^{n-1} M_i$ for various sums of multisets.

Beginning with states, we assume we are given two disjoint sets Spec of *species* and Agent of *agents* (a more neutral terminology, emphasising the algebraic point of view, would be *constants* and *unary function symbols*). We then define *multi-level multisets* and

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press



(a) tem



(b) str and a low level of gen

Fig. 3. (continued) Simulation of the infection of 100 cells by 5 viruses

atomic multi-level multisets as follows, where we further assume we have a countably infinite set of variables (which are disjoint from Spec and Agent):

— Every finite multiset $a_0, \ldots, a_{n-1}$ ($n \geqslant 0$) of atomic multi-level multisets is a multi-level multiset.
— Every S in Spec is an atomic multi-level multiset.

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

— Every variable x is an atomic multi-level multiset.
— If t is a multi-level multiset and $A \in \text{Agent}$, then $A(t)$ is an atomic multi-level multiset.

It is convenient to refer to multi-level multisets (respectively, atomic multi-level multisets) simply as *terms* (respectively, *atomic terms*). A term, or atomic term, is *ground* if it contains no variables. Ground terms are used to model populations. For example,

$$\text{Virus}, 20\text{Cell}, \text{ICell}(\text{gen}, 80\text{V}_1, 40\text{V}_2), \text{ICell}(\text{gen}, \text{tem}, \text{str}, 80\text{V}_1, 40\text{V}_2)$$

models a population consisting of one virus, twenty uninfected cells, a cell that has (perhaps) just been infected and an infected cell that is ready to produce a virus.

A term of the form $A(t)$ is said to be an *agent atomic term (with agent A)*. We define the *height* of terms and atomic terms by setting

$$|a_0, \ldots, a_{n-1}| = \max_{i=0}^{n-1} |a_i|$$

$$|S| = |x| = 0$$

and

$$|A(t)| = 1 + |t|.$$

The *subterm* relation between terms is defined to be the least reflexive relation between them such that if t is a subterm of t', then it is a subterm of both $A(t')$ and $t' + t''$. The *wide subterm* relation between terms is defined to be the least reflexive relation between them such that if t is a wide subterm of t', then it is a wide subterm of $A(t')$. Both relations are easily seen to be partial orders. As an example, S, x is a wide subterm of $A(S, x)$, but S is only a subterm of it.

Rules, ranged over by R, are pairs of terms l, r, together with an associated stochastic rate $k \in \mathbb{R}$, and are written as follows:

$$l \xrightarrow{k} r.$$

The height of such a rule is the maximum of the heights of l and r. For example, the kind of rules used above for modelling reactions between species have height 0, with both sides ground.

It is interesting to consider the various forms such rules may take when used to model cellular behaviour. Here are some possible rules of height 1:

Transport

$$M, A(N, x) \xrightarrow{k} M', A(N', x)$$

Creation

$$M \xrightarrow{k} M', A(N', x)$$

Death

$$M, A(N, x) \xrightarrow{k} M', N'$$

where M, N, and so on, are multisets of species. The names indicate their possible uses for modelling. One may ask if these provide sufficient possibilities for modelling extra- and

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

intra-cellular behaviour, at least to a first approximation. In this respect, the following, which is taken from Amir-Kroll *et al.* (2008) and slightly rewritten is germane:

> We have constructed the generic base of GemCell by appreciating the fact that any cell, in response to its surroundings, carries out only five types of behaviour:
>
> — Export (secretion of molecules, electricity, and so on);
> — Import (receiving signals, metabolites, phagocytosis, and so on);
> — Death;
> — Movement (including shape change and adherence); and
> — Replication.

The transport rule accounts for export and import, and the death rule for death. Movement and replication raise important issues, and point to the need for, and possibilities of, a further development of our formalism. For movement we could try rules of the following form:

$$M, A(N, x) \xrightarrow{k} M', B(N', x)$$

where A and B are cells in two given different places, or of given different shapes. However, this is an impoverished notion of place and shape. For adherence, we might try

$$M, A(N_1, x), B(N_2, y) \xrightarrow{k} M', C(N', x, y)$$

where C models A and B adhering to each other. Adherence is analogous to complex formation or binding, but at a higher level; as before, this is an impoverished notion.

For replication, we might consider the following kind of rule:

$$M, A(N, x, y) \xrightarrow{k} A(N', x), A(N'', y). \tag{1}$$

But there is a puzzle here as the left-hand side is non-linear, so there is a natural question as to how to interpret the stochastic rate: does each match of the left-hand side have the same rate, or is the rate somehow parcelled out among the possible matches? In the example below, we follow Krivine *et al.* (2008) and avoid the problem, preferring instead to explicitly model the division in two of the contents of the replicating cell.

There are several natural conditions one can place on rules, and, as alluded to in Krivine *et al.* (2008), it is interesting to discuss which are natural when modelling biological systems, and which are not. The first such condition is:

— *No creation*: $\mathrm{Var}(r) \subseteq \mathrm{Var}(l)$

This is standard in term rewriting; it is also required from a biological point-of-view, for where would the value of the 'new' variable come from? For the others we first define some possible conditions on a term t.

— *Uniqueness*: Any variable occurs at most once in t.
— *Unicity*: t has no subterm containing two variable occurrences.
— *Generality*: Every wide subterm of t contains a variable.

The last two conditions are equivalent to the condition that every wide subterm of t contains exactly one variable. The unicity condition means that not only cannot two

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

distinct variables occur in a subterm, but also that any variable occurring in it can have multiplicity at most 1.

We say that a rule $l \xrightarrow{k} r$ satisfies the *no equality*, *no splitting* or *no emptiness* condition according to whether l satisfies the first, second or third of the above conditions, respectively. All three have a linearity flavour, and are all natural from a biological point of view. For the first, surely no step of a modelled biological process can depend on two sub-populations being identical? The third states that no step of a modelled biological process can depend on a population being exactly specified (for example, that there are *exactly* so many molecules rather than at *least* so many molecules). The 'no splitting' requirement is less clearly unnatural biologically; the questions it presents in the case of replication were discussed above.

There are corresponding 'dual' rule conditions. We say that a rule $l \xrightarrow{k} r$ satisfies the *no vanishing*, *no duplication*, *no merging* or *no complete prescription* condition according to whether $\mathrm{Var}(l) \subseteq \mathrm{Var}(r)$ or r satisfies the first, second or third of the above conditions on a term, respectively.

The 'no duplication' condition seems reasonable since a biological process that exactly duplicates a population seems unlikely. However, in contrast to the corresponding possible conditions on the left-hand side of a rule, there does not seem to be any strong reason from a biological point of view to impose any of the other conditions. For the rest, the first, cell death certainly does result in vanishing: though it is at least natural not to be forced to model every detail of degradation. As for the 'no merging' condition, it is surely common for the contents of two agents to merge; equally, one may wish to prescribe completely the initial modelled part of the population of a cell, as was done, for example, in the above example of infected cells.

Given this discussion, from now on we will impose all of the first set of conditions on our rules but only the 'no duplication' condition of the second set. The rules in our examples and discussions obey all of these conditions, apart from the fact that they do not have a top-level variable. However, this is only for the sake of presentation, and every such rule written as $l \xrightarrow{k} r$ should be regarded as being in the form $l + x \xrightarrow{k} r + x$, where x is some canonically chosen variable not occurring in l or r.

We next turn to matching multiset terms against each other, more precisely, to finding the multiset of matches and their multiplicities. This is needed in order to define the stochastic process associated with a given finite set of rules. First, a *substitution* σ is a finitely based function from variables to terms, that is, a function that acts as the identity on all but finitely many variables. Such a function σ can be denoted by $[t_0/x_0, \ldots, t_{n-1}/x_{n-1}]$, where the variables $x_0, \ldots, x_{n-1}$ are all distinct, $x_i\sigma = t_i$, for $i = 0, \ldots, n-1$, and σ acts as the identity on any other variable (it is common to use the postfix form of application for substitutions).

Substitutions are extended to act on all terms and atomic terms as follows:

$$(a_0, \ldots, a_{n-1})\sigma = a_0\sigma, \ldots, a_{n-1}\sigma$$
$$S\sigma = S$$
$$A(t)\sigma = A(t\sigma).$$

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

A *match* of a term l against another term t is a substitution σ such that $l\sigma = t$. For example, the substitution $\sigma = [(n-m)S/x]$ is a match of the term $l = mS + x$ against the term $t = nS$, assuming that $n \geqslant m$ (there is otherwise no match). This match can be thought of as occurring in several ways according to which of the m S's of l is matched against which of the n S's of t. So we say that the *multiplicity* of a match is the m-fold falling product of n,

$$n^{\underline{m}} =_{\text{def}} n(n-1)\dots(n-(m-1)).$$

We also define a *symmetry* of l to be a permutation θ of the variables of l leaving it invariant, that is, such that $l\theta = l$, where we identify θ with the substitution that acts as the identity on all variables not in l.

We now define finite multisets $m(l;t)$ and $m(a;a')$ of substitutions of (atomic) terms against (atomic) ground terms; $m(l;t)$ is intended to be the multiset of matches of the term l against the term t, where the multiplicity of a substitution is its multiplicity as a match of l against t (and similarly for $m(a;a')$). The definition is only for terms l (respectively, atomic terms al) satisfying the above three conditions, and ground terms t (respectively, atomic ground terms at):

$$m(a_0,\dots,a_{m-1},x;a'_0,\dots,a'_{n-1}) = \sum_{f:[m]\hookrightarrow[n]} m(a_0,a'_{f(0)})\circ\cdots\circ m(a_{m-1},a_{f(m-1)})\circ \left[\sum_{j\notin f([m])} a_j/x\right]$$

and

$$m(S;a') = \begin{cases} [\,] & (a' = S) \\ 0 & (\text{otherwise}) \end{cases}$$

$$m(A(l);a') = \begin{cases} m(l;t) & (a' = A(t)) \\ 0 & (\text{otherwise}) \end{cases}$$

where we take the composition of multisets of substitutions to be the natural extension of the usual composition of substitutions to multisets. Note that if $m(l;t)(\sigma) \neq 0$, then σ acts as the identity on variables not in l.

One can show that σ is a match of l and t if and only if it has non-zero multiplicity in $m(l;t)$. This justifies its definition as far as as its elements are concerned. Later in the paper we will reformulate this in terms of counting tree embeddings.

We can separate $m(l;t)$ into a species and an agent part. We write l in the form $X + l' + x$, and t in the form $Y + t'$ where X and Y are finite multisets of species, and l' and t' are multisets of agent atomic terms. We define the *X-fold falling multiset sequential product of Y* to be

$$Y^{\underline{X}} =_{\text{def}} \prod_{S\in\text{Spec}} Y(S)^{\underline{X(S)}},$$

noting that this is essentially a finite product. Then we have

$$m(l;t) = Y^{\underline{X}}(m(l'+x;t') + [(Y \dot{-} X)/x]),$$

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

which allows us to add substitutions pointwise. One might have expected to see the binomial of multisets here:

$$\binom{Y}{X} = \prod_{S \in \mathrm{Spec}} \binom{Y(S)}{X(S)}.$$

This would be the case if we had counted the number of matches up to symmetry, that is, by dividing $m(l;t)$ by the number of symmetries of l.

We now turn to defining the application of rules to terms. First we need to define *contexts*, which are terms $C[\,]$ with a (single) hole $[\,]$ in them. They are defined inductively, taking $[\,]$ to be a context, and $A(C[\,])$ and the multiset $C[\,], t$ to be contexts if $C[\,]$ is. Ground contexts are those containing no variables. Given a context $C[\,]$, we can obtain a term $C[u]$ (context $C[D[\,]]$) by filling-in the hole $[\,]$ with a term u (respectively, a context $D[\,]$); we omit the definition.

We can now define the transition relation $\longrightarrow_R$ between ground terms of the application of the rule $R = l \xrightarrow{k} r$ by setting

$$t \longrightarrow_R t'$$

to hold when t has the form $C[u]$, for a context $C[\,]$ and there is a substitution σ that is a match of l against u, and is such that $t' = C[r\sigma]$; note that $C[\,]$ will necessarily be ground. A transition relation can then also be defined for a finite set $\mathscr{R}$ of rules by putting

$$t \longrightarrow_{\mathscr{R}} t' \iff \exists R \in \mathscr{R}.\, t \longrightarrow_R t'.$$

These are qualitative relations, by which we mean that no account is taken of the rates of the rules. To do so, we first need to define a narrower class of contexts, *viz.* the *wide contexts* $\mathrm{W}[\,]$. They are defined inductively, taking $[\,]$ to be a wide context, and the multiset $A(\mathrm{W}[\,]), t$ to be a wide context if $\mathrm{W}[\,]$ is. Note that every context can be written in the form $W[[\,], t]$. Using this, it is not hard to see that we get the same relation if we allow all contexts here since l obeys the generality condition. Note too that a context $\mathrm{C}[\,]$ is wide if and only if every term t is a wide subterm of $\mathrm{C}[t]$.

Wide contexts are needed to avoid a possibility of double-counting when defining stochastic rates. For example, consider the rule

$$R = S, x \xrightarrow{k} S', x.$$

We have $S, S' \to_R 2S'$, but that can be shown in two ways, using either of the contexts $[\,]$ or $[\,], S'$, and only the first of these is wide.

We need a count $\mathrm{occ}_t(\mathrm{W}[\,], u)$ of the number of ways in which a ground term t can have the form $\mathrm{W}[u]$, for a given wide context $W[\,]$ and term u:

$$\mathrm{occ}_t([\,], u) = \begin{cases} 1 & (t = u) \\ 0 & (t \neq u) \end{cases}$$

$$\mathrm{occ}_{a_0,\ldots,a_{n-1}}((A(\mathrm{W}[\,]), l), u) = \sum_{i=0}^{n-1} \sum \{\mathrm{occ}_{t'}(\mathrm{W}[\,], u) \mid a_i = A(t'), \; a_0, \ldots, a_{i-1}, a_{i+1}, \ldots, a_{n-1} = l\}.$$

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

The corresponding quantitative notion for a rule $R = l \xrightarrow{k} r$ is a stochastic matrix Q_R. This is a function from pairs of ground terms to non-negative reals, where, for distinct ground terms t, t',

$$Q_R(t,t') = k \sum_{t=\mathrm{W}[u]} \mathrm{occ}_t(\mathrm{W}[\,],u) \sum_{\substack{u = l\sigma \\ t' = \mathrm{W}[r\sigma]}} m(l;u)(\sigma),$$

and, on the diagonal, $Q_R(t,t)$ is, as usual, one minus the sum of the off-diagonal entries $Q_R(t,t')$ (it is easy to see that almost all the off-diagonal entries are 0). Note the use of the representation of multisets as functions to $\mathbb{N}$ in this definition. This can be extended, analogously to the earlier case, to a stochastic matrix for a finite set $\mathscr{R}$ of rules by defining

$$Q_{\mathscr{R}}(t,t') = \sum_{R \in \mathscr{R}} Q_R(t,t').$$

We will now describe how to simulate the CTMC given by this stochastic rate matrix in terms of choosing and applying rules from $\mathscr{R}$. The *activity* of a rule $R =_{\mathrm{def}} l \xrightarrow{k} r$ on a term t is defined by

$$\mathrm{Act}(R,t) = \sum_{t' \neq t} Q_R(t,t'),$$

which is equal to

$$k \sum_{t=\mathrm{W}[u]} \mathrm{occ}_t(\mathrm{W}[\,],u) \sum_{u=l\sigma} m(l;u)(\sigma).$$

The simulation has a current time, initialised to 0, and a current state t, and proceeds by cycling through the following sequence, for as many times as are required:

— If $\lambda =_{\mathrm{def}} \sum_{R \in \mathscr{R}} \mathrm{Act}(R,t)$ is zero, stop the simulation.
— Choose τ from the exponential distribution $1 - e^{-\lambda\tau}$ and add it to the current time.
— Choose rule $R =_{\mathrm{def}} l \xrightarrow{k} r$ from $\mathscr{R}$ with probability $\lambda^{-1}\mathrm{Act}(R,t)$.
— Choose a wide context $\mathrm{W}[\,]$, a u such that $t = \mathrm{W}[u]$ and a σ such that $l\sigma = u$ with probability

$$\frac{k\,\mathrm{occ}_t(\mathrm{W}[\,],u)(m(l;u)\sigma)}{\mathrm{Act}(R,t)}.$$

— Update t to $\mathrm{W}[r\sigma]$.

In a simulation, the species populations are normally plotted against time. This is fine for species, but what about agents? For example, we might wish to graph the number of agents A containing 3 molecules of species S. We can achieve this by plotting the activity of suitable patterns given by terms l obeying the three conditions. For example, for a species S, we can plot the term $S + x$, and for the agent example we can plot the pattern $A(3S + x) + y$. The activity of a pattern l in a ground term t is defined to be

$$\mathrm{Act}(l,t) = \sum_{l\sigma = t} m(l;t)(\sigma).$$

Returning to the question of symmetries, the activity of a rule with left-hand side $mS + x$ in the population nS will be $n^{\underline{m}}$ rather than the more usual $\binom{n}{m}$. However, we cannot just

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

simply divide the activity by the number of symmetries of the left-hand side. For example, consider the rule

$$A(x), A(y) \xrightarrow{k} B(x), C(y)$$

where we would not wish to divide by 2 since the right-hand side distinguishes between x and y. Adapting a suggestion[†] of Russ Harmer made in the context of Danos and Laneve's κ (Danos and Laneve 2003), we may prefer to divide the activity of a rule by the number of symmetries of its left-hand side that extend to a symmetry of its right-hand-side. In a sense, this is only a matter of convention, since the division can always be absorbed into the rate constant. However, we will in this paper adopt the simpler position of not doing any symmetry division, though we do not argue that it is the superior choice. In practice, the symmetry issue comes up for species, but, in our (very limited) experience, not for agents.

4. Another example

In this section, we give a model of an experiment presented in Rosenfeld *et al.* (2005). We do not so much aim to model the experiment exactly as to illustrate how one can use multi-level multiset rewriting to model a biological process in which cell division plays an important role.

Rosenfeld *et al.* introduced a method of measuring the gene regulation function of a given promoter. They created a high concentration of a repressor protein yfp in a single cell. This protein targets the chromosomally integrated promoter Pro of the gene for a protein cfp. At cell division, each daughter cell receives approximately half the population of the repressor protein. So, after a few divisions, the concentration is low enough to allow the production of cfp to take place.

For our model we need to represent different cells with different content evolving independently. We begin by giving a set of rules for the reactions occurring inside cells. (For practical reasons, our model actually uses a modified version of these rules, as explained below.)

We first introduce two different species, yfp and cfp, for modelling the repressor protein and the gene product. We next introduce a species mRNA for the messenger RNA; this is used to give a (much-simplified) model of the transcription–translation of cfp,

Modelling the promoter is less straightforward. To approximate repression cooperativity, which has a measured value of around 2 in the paper, we model the promoter as having two binding sites, each of which can bind to a yfp. Transcription can only occur when both binding sites are empty. This could be modelled by using different species, one for each of the three possible promoter states: free; one binding site occupied; both binding sites occupied. This approach is very intuitive, but would make it harder to model cell division.

Instead, we use an agent Pro to model the promoter and two species FreeBS and OccBS to represent the state (respectively, free or occupied) of each of the two binding sites. For

[†] Personal communication.

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

example, a promoter with two free binding sites is modelled by Pro(2FreeBS), whereas a promoter with one binding site occupied is modelled by Pro(OccBS, FreeBS). (The model is symmetric in both binding sites; non-symmetric models are also possible.)

The following rules then model repression and transcription (we do not model the nucleus and transport between it and the cytoplasm):

$$\begin{aligned} \mathrm{Pro}(\mathrm{FreeBS}, x), \mathrm{yfp} &\xrightarrow{r_{\mathrm{free-occ}}} \mathrm{Pro}(\mathrm{OccBS}, x) \\ \mathrm{Pro}(\mathrm{OccBS}, x) &\xrightarrow{r_{\mathrm{occ-free}}} \mathrm{Pro}(\mathrm{FreeBS}, x), \mathrm{yfp} \\ \mathrm{Pro}(\mathrm{2FreeBS}, x) &\xrightarrow{r_{\mathrm{transcr}}} \mathrm{Pro}(\mathrm{2FreeBS}, x), \mathrm{mRNA}. \end{aligned}$$

Note that whenever the last rule applies, x will be (matched to) 0.

Translation is modelled by the rule

$$\mathrm{mRNA} \xrightarrow{r_{\mathrm{transl}}} \mathrm{cfp}.$$

Both mRNA and cfp can degrade, and so we have the following two rules:

$$\begin{aligned} \mathrm{mRNA} &\xrightarrow{r_{\mathrm{m-degrad}}} \\ \mathrm{cfp} &\xrightarrow{r_{\mathrm{c-degrad}}}. \end{aligned}$$

This completes our representation of the rules for modelling the intracellular part of the process. However, there is a difficulty with modelling the extracellular part: the experiment produces an exponential number of cells, making faithful simulation impractical for our current implementation. There were 20 cell divisions in the experiment, so we would have to simulate around 2^{20} cells.

To solve this problem, we only model one cell of each generation: the first simulated cell is given a marker; only one of the two descendants of a marked cell receives a marker; and the system simulates only cells that have been marked.

To this end, instead of a single agent Cell, we use a family of agents Cell_u, where $u \subseteq \{\mathrm{mark}, \mathrm{dup}\}$. The marker 'mark' is used to determine the cells whose evolution is being modelled; the marker 'dup' is used to control DNA replication.

To make sure we only simulate the marked cell, we adapt the above rules by embedding them in a marked agent. For example, the first three rules are replaced by the following six:

$$\begin{aligned} \mathrm{Cell}_u(\mathrm{Pro}(\mathrm{FreeBS}, x), \mathrm{yfp}, y) &\xrightarrow{r_{\mathrm{free-occ}}} \mathrm{Cell}_u(\mathrm{Pro}(\mathrm{OccBS}, x), y) && (\mathrm{mark} \in u) \\ \mathrm{Cell}_u(\mathrm{Pro}(\mathrm{OccBS}, x), y) &\xrightarrow{r_{\mathrm{occ-free}}} \mathrm{Cell}_u(\mathrm{Pro}(\mathrm{FreeBS}, x), \mathrm{yfp}, y) && (\mathrm{mark} \in u) \\ \mathrm{Cell}_u(\mathrm{Pro}(\mathrm{2FreeBS}, x), y) &\xrightarrow{r_{\mathrm{transcr}}} \mathrm{Cell}_u(\mathrm{Pro}(\mathrm{2FreeBS}, x), \mathrm{mRNA}, y) && (\mathrm{mark} \in u). \end{aligned}$$

(Clearly, it would be preferable to allow parametric agents directly in the formalism, thereby permitting parametric rules. For example, the above six rules would become three parametric ones. This point is discussed further in Section 6.)

Our remaining rules model cell division. The first rule models DNA replication. However, we must ensure that this happens only once in the cell cycle. To this end, we make use of the marker dup, which acts as a token allowing DNA to be replicated, but consumed in the application of the rule

$$\mathrm{Cell}_{\mathrm{dup,mark}}(\mathrm{Pro}(x), y) \xrightarrow{r_{\mathrm{replication}}} \mathrm{Cell}_{\mathrm{mark}}(\mathrm{Pro}(x), \mathrm{Pro}(\mathrm{2FreeBS}), y)$$

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

where, for example, we write $\mathrm{Cell}_{\mathrm{dup,mark}}$ instead of $\mathrm{Cell}_{\{\mathrm{dup,mark}\}}$ (and we employ similar notation below).

Next, we model cell division *per se*. Following the discussion in Section 3, we cannot use the analogous rule to (1), *viz.*

$$\mathrm{Cell}_{\mathrm{mark}}(\mathrm{Pro}(z), \mathrm{Pro}(z'), x, y) \xrightarrow{r_{\mathrm{division}}} \mathrm{Cell}_{\mathrm{mark,dup}}(\mathrm{Pro}(z), x), \mathrm{Cell}(\mathrm{Pro}(z'), y),$$

as it does not satisfy the uniqueness condition. We therefore use a set of rules with a similar effect. Much the same problem arose previously when applying bigraphs in a biological setting (Krivine *et al.* 2008), but a slightly different solution was adopted in that case.

We use the expressivity of multi-level multiset rules to partition the population of each species of a cell randomly between its two children. To that end, we introduce a new unary function symbol CellPrec to model cell precursors, and add the rule

$$\mathrm{Cell}_{\mathrm{mark}}(\mathrm{Pro}(z), \mathrm{Pro}(z'), x) \xrightarrow{r_{\mathrm{division}}} \mathrm{Cell}(\mathrm{CellPrec}(\mathrm{Pro}(z)), \mathrm{CellPrec}(\mathrm{Pro}(z')), x).$$

The mark is removed to prevent any of the above rules from being applied while the cell species are being partitioned, that is, during cell division.

The next three rules serve to partition the populations of the species in Cell (x in the preceding rule) between the two precursors CellPrec:

$$\begin{aligned}
\mathrm{Cell}(\mathrm{CellPrec}(y), \mathrm{mRNA}, x) &\xrightarrow{r_{\mathrm{partition}}} \mathrm{Cell}(\mathrm{CellPrec}(\mathrm{mRNA}, y), x)\\
\mathrm{Cell}(\mathrm{CellPrec}(y), \mathrm{yfp}, x) &\xrightarrow{r_{\mathrm{partition}}} \mathrm{Cell}(\mathrm{CellPrec}(\mathrm{yfp}, y), x)\\
\mathrm{Cell}(\mathrm{CellPrec}(y), \mathrm{cfp}, x) &\xrightarrow{r_{\mathrm{partition}}} \mathrm{Cell}(\mathrm{CellPrec}(\mathrm{cfp}, y), x).
\end{aligned}$$

We want cell division to happen instantaneously, so the rate $r_{\mathrm{partition}}$ must be very high with respect to all the other rates. It is important to understand that this high rate does not slow down the other parts of the simulation: these rules are only applicable in the presence of a CellPrec, which happens only when the other rules do not apply as the mark parameter is absent.

After the precursors are introduced, we need $|x|$ rule applications to partition all the species in x. After these simulation steps, we want the two precursors to become two new cells, so we might expect to use the rule

$$\mathrm{Cell}(\mathrm{CellPrec}(x), \mathrm{CellPrec}(y)) \xrightarrow{r_{\mathrm{separation}}} \mathrm{Cell}_{\mathrm{dup,mark}}(x), \mathrm{Cell}(y).$$

However, such a rule is forbidden by the generality condition of Section 3: there is no variable in the left-hand side Cell. Instead, we use the following separation rule:

$$\mathrm{Cell}(\mathrm{CellPrec}(x), \mathrm{CellPrec}(y), z) \xrightarrow{r_{\mathrm{separation}}} \mathrm{Cell}_{\mathrm{dup,mark}}(x, z), \mathrm{Cell}(y).$$

and make sure that z is very unlikely to contain any species by choosing $r_{\mathrm{separation}}$ to be very small compared with $r_{\mathrm{partition}}$. As the separation rule has constant stochastic activity, this ensures that it is very unlikely to be triggered if the partition rule can still be applied. The simulation is not slowed down, as there are at most $|x|$ steps of partition, after which separation is the only applicable rule. The cell division can therefore be simulated in $|x| + 1$ steps.

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

Table 2. *The stochastic rates used for the simulation*

$r_{\text{free-occ}}$	1.0	$r_{\text{m-degrad}}$	0.00556
$r_{\text{occ-free}}$	0.001	$r_{\text{replication}}$	$8.33\ 10^{-4}$
r_{transcr}	0.1	r_{division}	$4.17\ 10^{-4}$
r_{transl}	0.0167	$r_{\text{partition}}$	10^{12}
$r_{\text{c-degrad}}$	0	$r_{\text{separation}}$	10^{5}

In practice, we want this process to happen instantly in simulated time, so we must fix both $r_{\text{separation}}$ and $r_{\text{partition}}$ to be very large compared with the other rates of the system. In summary, we want

$$r_{\text{partition}} >> r_{\text{separation}} >> r^*$$

where r^* is the maximum all the other rates of the system.

We ran a simulation of the above set of rules with an initial population of one cell containing 2500 repressor proteins,

$$\text{Cell}_{\text{mark,dup}}(2500\text{yfp}),$$

and with the rates, in s^{-1}, given in Table 2. Figure 4 shows both the number of cells and the number of molecules of yfp and cfp in the run, plotted against time measured in seconds.

As expected, the evolution of the number of cells is linear. The run is similar to the one presented in the original article. With each division, the concentration of yfp decreases until the production of cfp becomes possible.

This second worked example gives some feeling for the strengths and weaknesses of multi-level multiset rewriting. It allows a very direct expression of nested dynamic compartments, for both cell division and promoter representation, and an intuitive rule-based system expression. On the other hand, the very strict conditions on the rules make modelling cell division awkward to express: we will return to this point in Section 6.

5. Other formalisms

We begin this section by looking at a term rewriting formulation. There is an obvious notion of stochastic term rewriting, which, surprisingly, does not seem to have appeared in the literature. Suppose we have a a rule R of the form

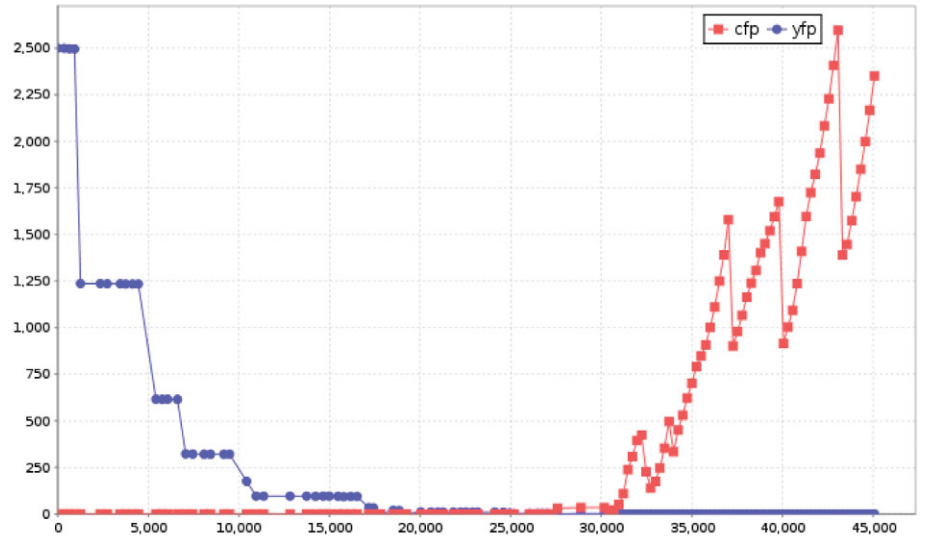
$$l \xrightarrow{k} r$$

where l and r are terms over a given signature, with $\text{Var}(r) \subseteq \text{Var}(l)$, and k is a non-negative real. We can then define a stochastic rate matrix on ground terms by setting

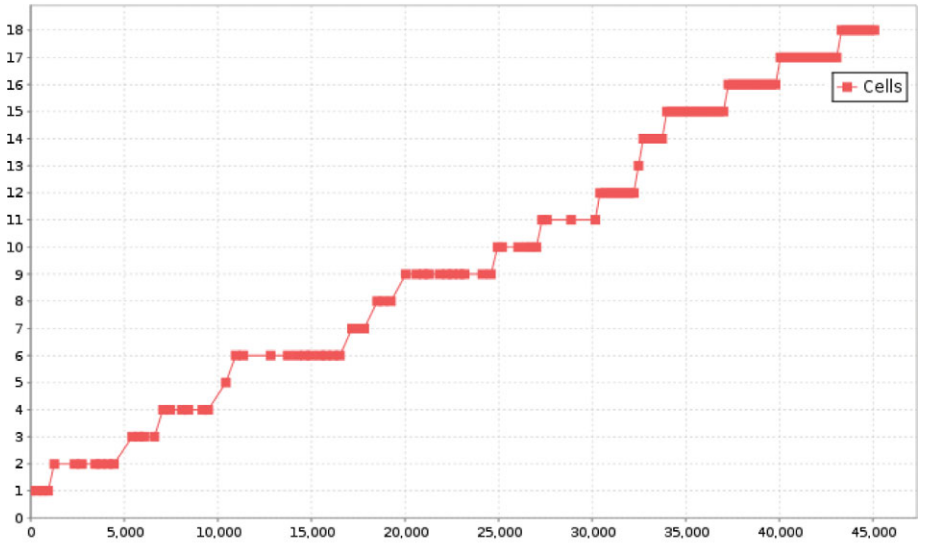
$$Q_R(t, t') = k \,|\{C[\,] \mid \exists \sigma.\, t = C[l\sigma] \wedge t' = C[r\sigma]\}|$$

for distinct t and t'.

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press



(a) yfp and cfp



(b) cells

Fig. 4. (Colour online) A run of the simulation

However, to account for our multi-level multiset terms, we need instead to work modulo a suitable equational theory. So, given the two disjoint sets, Spec of species and Agent of agents, we consider the equational signature with the species and 0 (not a species) as constants and the agents as unary operation symbols, and with a binary operation symbol $+$. We work modulo the equational theory T that says $+$ is associative and commutative,

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

and has a zero, 0. Our terms can then be seen as normal forms for the algebraic terms modulo T. More precisely, we define a map N to the former from the latter by

$$\begin{aligned} \mathrm{N}(x) &= x \\ \mathrm{N}(S) &= S \\ \mathrm{N}(A(t)) &= A(\mathrm{N}(t)) \\ \mathrm{N}(t+u) &= \mathrm{N}(t), \mathrm{N}(u). \end{aligned}$$

Then N is onto and we have

$$\vdash_{\mathrm{T}} t = u \quad \Longleftrightarrow \quad \mathrm{N}(t) = \mathrm{N}(u).$$

Clearly, a term and its normal form have the same variables. We say that an occurrence of a variable in a term is at *top level* if it is not within any unary operation symbol. Then $\mathrm{N}(t)$ obeys the above three conditions if and only if no variable occurs more than once in it, and there is exactly one top-level variable occurrence, with the same being true of every term u such that $A(u)$ is a subterm of t, for some unary operation symbol A.

Given a rule $R = l \xrightarrow{k} r$ as above in the present signature, we can define a rewriting relation, modulo T, between ground terms in a standard way, by putting

$$t \longrightarrow_R t' \iff \exists C[\,], \sigma.\ \vdash_{\mathrm{T}} t = C[l\sigma] \wedge \vdash_{\mathrm{T}} t' = C[r\sigma].$$

Then, setting $\mathrm{N}(l \xrightarrow{k} r) = \mathrm{N}(l) \xrightarrow{k} \mathrm{N}(r)$ and assuming this rule obeys the above conditions on multi-level multiset rules, we have, for any ground terms t and t',

$$t \longrightarrow_R t' \iff \mathrm{N}(t) \longrightarrow_{\mathrm{N}(R)} \mathrm{N}(t').$$

However, we do not know how to write the transition matrix between ground algebraic terms, modulo T, other than to use the normal form and the above definition of stochastic rates for multi-level multiset rules. We regard it as an interesting open problem to formulate a notion of stochastic rewriting for algebraic terms modulo equational theories. It may well be that one needs to restrict the class of equational theories considered; a possible such class is that of the *balanced theories*, which are those given by equations in which the same variables occur on each side – see Manes (1998).

The algebraic approach gives a possibility of generality for the development of useful term rewriting formalisms for computational systems biology. For example, one can argue that agents and compartments are different. Consider the term $A(B(x) + B(y) + z)$. Here we have two distinct B-agents with contents x and y within an agent A. However, if we instead take B to be a compartment name, then the agent A should have only one compartment with a given name; it is therefore natural to make the identification $B(x) + B(y) = B(x + y)$, that B commutes with the AC operation. So an extension to our formalism of potential interest would be to add unary compartment operations that commute with the AC operation.

We next give a reformulation in terms of forests. We follow very largely in Milner's footsteps, and in Krivine *et al.*'s for the stochastic aspects: we essentially specialise to place graphs, making slight adaptations to allow for the presence of species. This last

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

is a minor difference, as one can always simulate species by agents that never contain anything.

There are other small differences. For example, we may impose slightly different conditions, and we do not bring in any categorical ideas, though they are certainly there in the background. Where appropriate, we will add comments on particular relationships to the bigraphical approach.

We fix a countably infinite set $\mathscr{V}$. An *(n-ary) concrete*(Spec, Agent)-*forest*, which we will call an (n-ary) concrete forest for short, is a structure

$$(V, \mathrm{p}, \lambda)$$

where

— $V \subseteq_{\mathrm{fin}} \mathscr{V}$ is a finite set of *nodes*,
— $\mathrm{p} : [n] \cup V \to V \cup [1]$ is the *parent* map, and
— $\lambda : v \to \mathrm{Spec} \cup \mathrm{Agent}$ is the *labelling map*

where $[m] =_{\mathrm{def}} \{0, \ldots, m-1\}$, for $m \geqslant 0$. The parent map is required to be acyclic, meaning that if $\mathrm{p}^i(v) = v$, then $i = 0$, and species can only label roots, which are nodes v such that $\mathrm{p}^{-1}(v) = \varnothing$.

If we drop species, these are the special case of Milner's concrete place graphs from n to 1 in which the controls have arity 0. It will prove convenient to confuse the set of nodes V with the entire structure (V, p, λ). We write $V : n$ to indicate that V is n-ary; we say that it is *ground* if $n = 0$; and we say it is an *atom* if $\mathrm{p}^{-1}(0)$ is a singleton. A homomorphism

$$h : (V, \mathrm{p}, \lambda) \longrightarrow (V', \mathrm{p}', \lambda')$$

of n-ary concrete forests is a map from V to V' that respects structure in the evident sense that $h(\mathrm{p}(x)) = \mathrm{p}'(h(x))$ for any $x \in [n] \cup V$, and $\lambda(v) = \lambda'(h(v))$ for any $v \in V$. We work with isomorphism equivalence classes $[(V, \mathrm{p}, \lambda)]$ of forests; following Milner, we call them *abstract forests*. This is helpful for definitions as we can always pick disjoint representatives of different equivalence classes. We say an abstract forest is n-ary (or ground) if any of its members is, and write $[V] : n$.

We next define the *composition* of n-ary forests with ground forests, beginning with concrete ones. Given pairwise disjoint n-ary (V, p, λ) and ground $(V_i, \mathrm{p}_i, \lambda_i)$, for $i = 0, \ldots, n-1$, we define their composition

$$(V, \mathrm{p}, \lambda)((V_0, \mathrm{p}, \lambda), \ldots, (V_{n-1}, \mathrm{p}, \lambda))$$

to be the ground forest

$$(V \cup \bigcup_{i=0}^{n-1} V_i, \mathrm{p}', \lambda')$$

where

$$\mathrm{p}'(v) = \begin{cases} v' & (v \in V_i, \mathrm{p}_i(v) = v' \in V_i) \\ \mathrm{p}(i) & (v \in V_i, \mathrm{p}_i(v) = i) \\ \mathrm{p}(v) & (v \in V) \end{cases}$$

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

and

$$\lambda'(v) = \begin{cases} \lambda_i(v) & (v \in V_i) \\ \lambda(v) & (v \in V). \end{cases}$$

Note that if for a given concrete V', n-ary V and ground V_i $(i = 0, \ldots, n-1)$ we have $V' = V(V_0, \ldots, V_{n-1})$, then the V_i are uniquely determined.

It is not hard to see that if

$$[V] = [V']$$
$$[V_i] = [V'_i]$$

for $V, V' : 1$ and $V_i, V'_i : 0$ $(i = 0, n-1)$, then

$$V(V_0, \ldots, V_{n-1}) = V'(V'_0, \ldots, V'_{n-1}).$$

So we can define the composition

$$[[(V, \mathrm{p}, \lambda)]([(V_0, \mathrm{p}, \lambda)], \ldots, [(V_{n-1}, \mathrm{p}, \lambda)])$$

of an n-ary abstract forest $[(V, \mathrm{p}, \lambda)]$ with n ground ones $[(V_i, \mathrm{p}_i, \lambda_i)]$ $(i = 0, \ldots, n-1)$ to be

$$[(V, \mathrm{p}, \lambda)((V_0, \mathrm{p}, \lambda)], \ldots, [(V_{n-1}, \mathrm{p}, \lambda))],$$

with the understanding that disjoint members of the equivalence classes have been chosen.

A *stochastic rule* R is an n-ary equivalence class $[V_l]$ and an m-ary equivalence class $[V_r]$ with a map $\eta : [m] \to [n]$ and a *rate* $k \in \mathbb{R}_0$, written as

$$[V_l] \xrightarrow[\eta]{k} [V'_r]$$

where we impose the following two additional conditions:

— restricted to $[n]$, the parent map of V_l is a bijection; and
— η is 1-1.

Regarding the first condition, in Krivine *et al.* (2008), the left-hand sides of rules are required to be *solid*, which here amounts to the condition that, restricted to $[n]$, the parent map of V_l is 1–1 and does not have 0 in its range. However, we do have 0 in its range, and this causes an ambiguity in the application of rules of the kind discussed in Section 3 – the difficulty is again handled by the introduction of a suitable notion of a wide context.

The requirement that, restricted to $[n]$, the parent map is onto, is, as will be seen, the correlate of the generality condition. Presumably we could just as well have worked in the Krivine *et al.* style from the beginning, when an effectively equivalent generality condition would be imposed.

Regarding the second condition, we follow Milner (2009) rather than Krivine *et al.*, as the latter's rules are linear in a certain sense, which here amounts to the requirement that η is a bijection. In the context of multi-level multiset rewriting, this would correspond to adding the 'no vanishing' condition, which we do not wish to impose.

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

Qualitatively, we can assign to the rule a transition relation between ground $[V]$ defined by

$$[V] \longrightarrow_R [V'] \iff \exists [C] : 1, [V_0] : 0, \ldots, [V_{n-1}] : 0. \\ [V] = [C]([V_l]([V_0], \ldots, [V_{n-1}])) \wedge \\ [V'] = [C]([V_r]([V_{\eta(0)}], \ldots, [V_{\eta(m-1)}])).$$

Quantitatively, as already remarked, we first need to restrict the 'contexts' $[C] : 1$. We say that a concrete forest $C : 1$ is *wide* if 0 is the only child of its parent, and that an abstract forest $[C] : 1$ is wide if C is. Then we can assign to the rule R a stochastic transition matrix Q_R where, off the diagonal,

$$Q_R([V], [V']) = k \mid \{(C : 1, V'' : n, V_0 : 0, \ldots, V_{n-1} : 0) \mid C \text{ is wide}, \\ V = C(V''(V_0, \ldots, V_{n-1})) \wedge [V_l] = [V''], \\ [V'] = [C]([V_r]([V_{\eta(0)}], \ldots, [V_{\eta(m-1)}]))\} \mid .$$

Here we count the number of factorisations of an element of $[V]$. Note that the V_i are determined, so do not enter into the count: only their existence is required. Unlike Krivine *et al.*, we do not divide by the number of symmetries of the left-hand-side of the instance of the rule at hand: *cf.* the discussion of symmetry in Section 3.

We next give an equivalent definition that will help establish the relation with the multi-level multiset approach. For any concrete forest V, we define an equivalence relation on tuples $(C, V_0, \ldots, V_{n-1})$ such that $V = C(V_0, \ldots, V_{n-1})$, where $C : 1$ and $V_i : 0$ (for $i = 0, \ldots, n-1$), by

$$(C, V_0, \ldots, V_{n-1}) \sim_V (C', V'_0, \ldots, V'_{n-1}) \iff [C] = [C'] \wedge \bigwedge_{i=0}^{n-1} [V_i] = [V'_i],$$

and write $[C, V_0, \ldots, V_{n-1}]_V$ for the corresponding equivalence class.

Proposition 5.1.

$$Q_R([V], [V']) = k \sum_{\substack{[C, V_{\text{red}}]_V \\ C \text{ wide} \\ [V'] = [C]([V_r]([V_{\eta(0)}], \ldots, [V_{\eta(m-1)}])) \\ [V_l] = [\overline{V_l}]}} |[C, V_{\text{red}}]_V| \sum_{[\overline{V_l}, V_0, \ldots, V_{n-1}]_{V_{\text{red}}}} |[C, V_0, \ldots, V_{n-1}]_{V_{\text{red}}}| .$$

As before, given a finite set of rules $\mathscr{R}$, we can then define a transition relation and stochastic matrix by putting

$$[V] \longrightarrow_{\mathscr{R}} [V'] \iff \exists R \in \mathscr{R}. [V] \longrightarrow_R [V'] \qquad Q_{\mathscr{R}}([V], [V']) = \sum_{R \in \mathscr{R}} Q_R([V], [V']).$$

We next need some algebra on abstract forests. We begin with two constants: $\text{merge}_1 : 1$ is the equivalence class of the unary forest with empty node set; and, for every $S \in \text{Spec}$, $\underline{S}$ is the equivalence class of the nullary forest with a single node labelled by S. Next, for any agent A, we define a unary function $\underline{A}$ on abstract forests as follows. For any forest

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

$(V, \mathrm{p}, \lambda) : m$, we define $\underline{A}([V]) : m$ to be $[(V \cup \{*\}, \mathrm{p}', \lambda')]$ where

$$\mathrm{p}'(v) = \begin{cases} v' & (v \in V, \mathrm{p}(v) = v' \in V) \\ * & (v \in V, \mathrm{p}(v) = 0) \\ 0 & (v = *) \end{cases}$$

and

$$\lambda'(v) = \begin{cases} A & (v = *) \\ \lambda(v) & (v \in V). \end{cases}$$

Next, for any $n \geqslant 0$ and bijection $\theta : \sum_{i=0}^{n-1} [m_i] \cong [\sum_{i=0}^{n-1} m_i]$, we define an n-ary *summation* operation by putting, for abstract forests $[(V_i, \mathrm{p}_i, \lambda_i)] : m_i$ where $i = 0, \ldots, n-1$,

$$\sum_{i=0,n-1}^{(\theta)} [V_i] = \left[\left(\dot{\bigcup_{i=0}^{n-1}} V_i, \mathrm{p}, \lambda \right) \right]$$

where

$$\mathrm{p}(v) = \begin{cases} \mathrm{p}_i(j) & (v = \theta(i, j), j \in [m_i]) \\ \mathrm{p}_i(v) & (v \in V_i) \end{cases}$$

and

$$\lambda(v) = \lambda_i(v) \quad (v \in V_i).$$

This operation is essentially commutative, by which we mean that for any permutation $\pi : [i] \cong [i]$, we have

$$\sum_{i=0,n-1}^{(\theta)} [V_i] = \sum_{i=0,n-1}^{(\theta')} [V_{\pi i}]$$

where $\theta'(i, j) = \theta(\pi^{-1} i, j)$, for $0 \leqslant i \leqslant n-1$, $j \in m_{\pi i}$. We say that $[V'_{\pi(0)}], \ldots, [V'_{\pi(n-1)}], \theta'$ is a *reindexing* of $[V_0], \ldots, [V_{n-1}], \theta$.

Proposition 5.2. Let $[V] : n$ be an abstract forest. Then, it is either:

(a) An atom, in which case one of the following three mutually exclusive possibilities holds:

((1)) $[V] = \mathrm{merge}_1$

((2)) $[V] = \underline{S}$ for a unique $S \in \mathrm{Spec}$

((3)) $[V] = \underline{A}([V'])$ for a unique $A \in \mathrm{Agent}$, and abstract forest $[V']$.

Or:

(b) Not an atom, in which case

$$[V] = \sum_{i=0,n-1}^{(\theta)} [V_i]$$

for a unique $n \geqslant 0$ and unique, up to reindexing, atoms $[V_0], \ldots, [V_{n-1}], \theta$.

We now turn to linking the multi-level multiset and forest formalisms. First we need to translate terms to forests; we will only consider terms in which no variable occurs more than once. Variables in one will correspond to numbers in the other. We assume a fixed ordering $z_0, \ldots, z_n, \ldots$ of all variables, and say that i is the index of z_i. For any term t, we

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

write $\mathrm{Ind}(t)$ for the set of indices of its variables, and set $\mathrm{ar}(t) = |\mathrm{Ind}(t)|$. The translation assigns to every pair t, ρ, consisting of a term and a bijection $\rho : \mathrm{Ind}(t) \cong [\mathrm{ar}(t)]$, an abstract forest $\mathrm{F}_\rho(t) : \mathrm{ar}(t)$. For any $I \subseteq_{\mathrm{fin}} \mathbb{N}$, we fix a bijection $\rho_I : I \cong [|I|]$, and write ρ_0 for $\rho_{[1]}(= 0 \mapsto 0)$.

For atomic terms, we put

$$\begin{aligned} F_\rho(z_i) &= \mathrm{merge}_1 \\ F_\rho(S) &= \underline{S} \\ F_\rho(A(t)) &= \underline{A}(F_\rho(t)), \end{aligned}$$

and for non-atomic terms, we put

$$F_\rho(t_0, \ldots, t_{n-1}) = \sum_{i=0,n-1}^{(\theta)} F_{\rho_{\mathrm{Ind}(t_i)}}(t_i)$$

where $\theta(i, j) = \rho(\rho^{-1}_{\mathrm{Ind}(t_i)}(j))$, for $0 \leqslant i \leqslant n-1$, $j \in \mathrm{Ind}(t_i)$. We see from the above remarks that this is well defined. We will omit the (trivial) ρ when translating ground terms.

It follows from Proposition 5.2 that for any $\rho : I \cong [n]$, the function $F_\rho(-)$ is a bijection between terms t with $I = \mathrm{Ind}(t)$ satisfying the uniqueness condition and the $\mathrm{ar}(t)$-ary abstract forests. We note the following further correspondences, where $F_\rho(t) = [(V, \mathrm{p}, \lambda)]$:

— t satisfies the unicity condition if and only if $\mathrm{p}\restriction[\mathrm{ar}(t)]$ is 1-1.
— t satisfies the generality condition if and only if $\mathrm{p}\restriction[\mathrm{ar}(t)]$ is onto.

The translation maps substitution to composition in the following sense. Let t be a term with variables $z_{l_0}, \ldots, z_{l_{n-1}}$, where $n = \mathrm{ar}(t)$, and let $t_0, \ldots, t_{n-1}$ be ground. Then

$$F(t[t_0/z_{l_0}, \ldots, t_0/z_{l_{n-1}}]) = F_\rho(F(t_0), \ldots, F(t_{n-1})).$$

Having mapped terms to terms, we can now map mutilevel multiset rules to forest rules. To any rule

$$R = l \xrightarrow{k} r$$

and bijections $\rho_l : \mathrm{Ind}(l) \cong [\mathrm{ar}(l)]$ and $\rho_r : \mathrm{Ind}(r) \cong [\mathrm{ar}(r)]$, we assign the rule

$$F_{\rho_l, \rho_r}(R) = F_{\rho_l}(l) \xrightarrow[\eta]{k} F_{\rho_r}(r)$$

where $\eta = \rho_l \rho_r^{-1}$. This map is a surjection from mutilevel multiset rules and pairs of such bijections to forest rules. According to the next proposition, the two rules are qualitatively equivalent.

Proposition 5.3. For all ground terms t, t',

$$t \longrightarrow_R t' \iff F(t) \longrightarrow_{F_{\rho_l, \rho_r}(R)} F(t').$$

Proof. Let

$$\begin{aligned} \mathrm{Ind}(l) &= \{i_1, \ldots, i_n\} \\ \mathrm{Ind}(l) &= \{j_1, \ldots, j_m\}, \end{aligned}$$

where $n = \mathrm{ar}(l)$ and $m = \mathrm{ar}(r)$.

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

First we suppose $t \longrightarrow_R t'$. Then t has the form $C[u]$, and there is a substitution σ that is a match of l against u, and is such that $t' = C[r\sigma]$. Then,

$$t = C[l\sigma] = C[z_0][l[\sigma(i_0)/z_{i_0}, \ldots, \sigma(i_{n-1})/z_{i_{n-1}}]/z_0]$$

and

$$t' = C[r\sigma] = C[z_0][u[\sigma(j_0)/z_{j_0}, \ldots, \sigma(j_{m-1})/z_{j_{m-1}}]/z_0].$$

It follows that

$$F(t) = F_{\rho_0}(C[z_0])(F_{\rho_l}(l)(F(\sigma(\rho_l^{-1}(0))), \ldots, F(\sigma(\rho_l^{-1}(n-1))))))$$

and

$$F(t') = F_{\rho_0}(C[z_0])(F_{\rho_r}(r)(F(\sigma(\rho_r^{-1}(0))), \ldots, F(\sigma(\rho_r^{-1}(m-1))))))).$$

As we also have

$$F(\sigma(\rho_r^{-1}(j))) = F(\sigma(\rho_l^{-1}(\eta(j)))),$$

it follows that

$$F(t) \longrightarrow_{F_{\rho_l,\rho_r}(R)} F(t'),$$

as required.

Conversely, we now suppose

$$F(t) \longrightarrow_{F_{\rho_l,\rho_r}(R)} F(t').$$

Then

$$F(t) = [V]([F_{\rho_l}(l)]([V_0], \ldots, [V_{n-1}]))$$

and

$$F(t') = [V]([F_{\rho_r}(r)]([V_{\eta(0)}], \ldots, [V_{\eta(m-1)}]))$$

for some $[V] : 1$ and $[V_0] : 0, \ldots, [V_{n-1}] : 0$. As F_{ρ_0} is a bijection between terms t with $\mathrm{Ind}(t) = \{0\}$ satisfying the uniqueness condition and the unary abstract forests, there is a ground context $C[\,]$ such that $F_{\rho_0}(C[z_1]) = [V]$. Similarly, there are ground terms t_i such that $F(t_i) = [V_i]$, for $i = 0, \ldots, n-1$. We then have

$$[V]([F_{\rho_l}(l)]([V_0], \ldots, [V_{n-1}])) = F(C[l[t_0/z_{\rho_l^{-1}(0)}, \ldots, t_{n-1}/z_{\rho_l^{-1}(n-1)}]]).$$

So, as F is injective, setting

$$\sigma = [t_0/z_{\rho_l^{-1}(0)}, \ldots, t_{n-1}/z_{\rho_l^{-1}(n-1)}],$$

we see that $t = C[l\sigma]$. Similarly, we have

$$t' = C[r[t_{\eta(0)}/z_{\rho_r^{-1}(0)}, \ldots, t_{\eta(n-1)}/z_{\rho_r^{-1}(m-1)}]] = C[r\sigma],$$

and the result follows. □

We now turn to showing that the two rules are also quantitatively equivalent. First note that for any context $C[\,]$, if $F_{\rho_0}(C[z_0]) = [C']$, then $C[\,]$ is wide if and only if C' is. Next we need two lemmas.

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

Lemma 5.4. For any ground term t, wide context W[ ] and term u,

$$\mathrm{occ}_t(\mathrm{W}[\,], u) = |\{(V_1 : 1, V_0 : 0) \mid V = V_1(V_0), [V_1] = F_{\rho_0}(\mathrm{W}[z_0]), [V_0] = F(u)\}|$$

where $V : 0$ is any concrete forest such that $F(t) = [V]$.

Lemma 5.5. Let t be a ground term and l be a term with free variables $z_{i_0}, \ldots, z_{i_{n-1}}$ and satisfying the above three conditions. Then for any ground terms $t_0, \ldots, t_{n-1}$ and any $\rho : \mathrm{Ind}(l) \cong [n]$, we have

$$m(l;t)([t_0/z_{i_0}, \ldots, t_{n-1}/z_{i_{n-1}}]) = |\{(V_l : n, V_0 : 0, \ldots, V_{n-1} : 0) \mid V = V_l(V_0, \ldots, V_{n-1}), [V_l] = F_\rho(l), \text{ and } [V_i] = F(t_i)(i = 0, \ldots, n-1)\}|$$

where $V : 0$ is any concrete forest such that $F(t) = [V]$.

Using Proposition 5.1, and Lemmas 5.4 and 5.5, we then obtain the following proposition.

Proposition 5.6. For all ground terms t, t',

$$Q_R(t, t') = Q_{F_{\rho_l, \rho_r}(R)}(F(t), F(t')).$$

We conclude this section by briefly considering the relation between our system and the Stochastic Calculus of Wrapped Compartments of Coppo *et al.* (2010a). This formalism introduces *wrapped compartments*, which are denoted using an infix operation $\rfloor$. The compartment notation allows us to represent both the content of a membrane, on the left of $\rfloor$, and the content of a compartment, on the right of $\rfloor$. The content of the membrane is a multiset of *atoms*, which correspond to our species. The content of the compartment is a multiset of atoms and nested compartments.

This formalism was extended in Coppo *et al.* (2010b) by allowing a compartment to be labelled. For example, a cell containing an empty nucleus and having an atom a on its membrane would be denoted by

$$(a\rfloor(\rfloor)^{Nucleus})^{Cell}.$$

We refer to this extended formalism as *SCWC*.

Both SCWC and our formalism allow one to express nested compartments containing species, and each can be encoded by the other. Regarding species as atoms and agents as compartment labels, a translation function C from our formalism into SCWC can be defined by

$$C(M, A_0(t_0), \ldots, A_{n-1}(t_{n-1})) = (M, \rfloor(\rfloor \mathrm{C}(t_0))^{A_0}, \ldots, (\rfloor \mathrm{C}(t_{n-1}))^{A_{n-1}})^*$$

for any multiset of species $M = S_0, \ldots, S_{m-1}$, and where $* \notin$ Agent. It is notable that we have to make a choice for this translation: we have put the species on the membrane, but we could have put them inside the compartment, or we could even have chosen which to do according to the species. Because of this, the translation is not onto, though it is 1–1.

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

In the other direction, regarding atoms as species and compartment labels as agents, a translation function T from SCWC into our formalism can be defined by

$$\mathrm{T}(a) = a$$
$$\mathrm{T}((S_0, \ldots, S_{m-1} \rfloor t_0, \ldots, t_{n-1})^l) = l(M(\mathrm{T}(S_0, \ldots, S_{m-1})), \mathrm{T}(t_0), \ldots, \mathrm{T}(t_{n-1}))$$

where M is an agent that is not a compartment label; it is thought of as a *membrane agent*. The translation T from SCWC to our formalism is again 1–1, but again not onto as there is no restriction on terms saying that only species can appear in the membrane agent M.

By using the algebraic approach, we can get much closer to SCWC. Consider a two-sorted equational theory with: sorts m and c for membrane and compartment; an AC operation with a zero for each sort; a unary operation over c for each compartment label; a set of constants of sort m for the atoms that can be on a membrane; and a set of constants of sort c for the atoms that can be in a compartment. If we allow overloading, in particular, if we allow the two sets of constants to be the same, then SCWC terms can be seen as normal forms for the terms of type c.

The restrictions imposed on rules in SCWC are similar to the ones we impose: for example, forms of uniqueness and unicity are imposed on the left-hand sides of rules, but the generality condition is only partially imposed. Based on the above translations, we conjecture that, once the differences in the conditions imposed have been reconciled, mutual simulation results, including stochastic rates, can be established.

Despite their (presumed) equivalence, the two formalisms have somewhat different orientations. SCWC has an elegant representation of membranes, whereas our formalism assigns no particular role or structure to them. Our formalism may therefore permit a more natural modelling of those parts of biological structures that do not involve membranes. As the translation from SCWC shows, we further lose little, if any, naturality compared with the expression of systems with membranes.

6. Discussion and conclusions

There are several possibilities for future development. With respect to formalisms, it would be useful to have a notion of a multi-level Petri net to enable the graphical presentation of our multi-level multiset rule systems; indeed, even such a notion for rules with terms of height $\leqslant 1$ would be very helpful. In terms of generality, it would be interesting to develop the algebraic approach discussed above: one should investigate both general theory and particular systems; the addition of compartments would be of particular immediate interest.

In another direction, it would be useful to have a type system. So far, for example, there is nothing that prevents cells being inside cells inside cells, and so on, to arbitrary depths. One could imagine such a type system based on a forest, or dag, of types. Continuing the linguistic thought, very large lists of rules become difficult to understand and maintain, and often obscure the underlying structure. This might be alleviated by a suitable module system. It would be interesting to design a language for multi-level systems along the

https://doi.org/10.1017/S0960129512000199 Published online by Cambridge University Press

lines of LBS (Pedersen and Plotkin 2010), which is a modular language for the rule-based description of intracellular systems.

Facilities for parameterisation would also be useful. At the species level, one could follow LBS and use parametrised species $S_{x_1,\ldots,x_n}$, where the x_i run over suitable parameter spaces describing, for example, modification states (phosphorylation, ubiquitination, and so on). Similarly, one could make use of parametrised agents $A_{x_1,\ldots,x_n}$, where the parameters might now, for example, deal with cell fate, volume or location. Rules could make use of these parameters: for example, the stochastic rates could depend upon them, or boolean conditions on them could determine their applicability.

Finally, it is most important to have facilities for complexes. A simple possibility would be to add another multiset operation to deal with complexes. A more sophisticated, and more powerful, approach would be to describe complexes as connected graphs, following the lead of κ (Danos and Laneve 2003). Together with the multi-level multisets, we would then have something very similar indeed to Milner's bigraphs, whose potential for biological application has, as mentioned above, already been noted in Krivine *et al.* (2008).

Acknowledgments

We are very grateful to Vincent Danos and Jean Krivine for discussions on rule-based modelling, and to Peter Swain and Andrea Weisse for discussions on multi-level systems.

References

Agha, G. A., Meseguer, J. and Sen, K. (2006) PMaude: Rewrite-based specification language for probabilistic object systems. *Electronic Notes in Theoretical Computer Science* **153** (2) 213–239.

Amir-Kroll, H., Sadot, A., Cohen, I. R. and Harel, D. (2008) GemCell: A generic platform for modeling multi-cellular biological systems. In: Priami, C. (ed.) Converging Sciences: Informatics and Biology. *Theoretical Computer Science* **391** (3) 276–290.

Baader, F. and Nipkow, T. (1999) *Term Rewriting and All That*, Cambridge University Press.

Barbuti, R., Caravagna, G., Maggiolo-Schettini, A., Milazzo, P. and Pardini, G. (2008a) The calculus of looping sequences. In: Bernardo, M., Degano, P. and Zavattaro, G. (eds.) Formal Methods for Computational Systems Biology. *Springer-Verlag Lecture Notes in Computer Science* **5016** 387–423.

Barbuti, R., Maggiolo-Schettini, A., Milazzo, P., Tiberi, P. and Troina, A. (2008b) Stochastic CLS for the modeling and simulation of biological systems. *Transactions on Computational Systems Biology IX* **5121** 86–113.

Bauer, A. L., Beauchemin, C. A. A. and Perelson, A. S. (2009) Agent-based modeling of host pathogen systems: The successes and challenges. *Information Sciences* **179** (10) 1379–1389.

Bezem, M., Klop, J. W. and de Vrijer, R. (eds.)‡ (2003) Term Rewriting Systems. *Cambridge Tracts in Theoretical Computer Science* **55**, Cambridge University Press.

Bournez, O. and Hoyrup, M. (2003) Rewriting logic and probabilities. In: Nieuwenhuis, R. (ed.) Proceedings 14th International Conference on Rewriting Techniques and Applications. *Springer-Verlag Lecture Notes in Computer Science* **2706** 61–75.

‡ Also known as 'Terese'.

Bournez, O. and Kirchner, C. (2002) Probabilistic rewrite strategies. Applications to ELAN. In: Tison, S. (ed.) Proceedings 13^{th} International Conference on Rewriting Techniques and Applications. *Springer-Verlag Lecture Notes in Computer Science* **2378** 252–266.

Cardelli, L. (2005) Brane calculi. In: Danos, V. and Schächter, V. (eds.) International Conference on Computational Methods in Systems Biology. Revised Selected Papers. *Springer-Verlag Lecture Notes in Computer Science* **3082** 257–27.

Cardelli, L. (2008) Bitonal membrane systems: interactions of biological membranes. *Theoretical Computer Science* **404** (1-2) 5–18.

Chickarmane, V., Roeder, A. H. K., Tarr, P. T., Cunha, A., Tobin, C. and Meyerowitz, E. M. (2010) Computational morphodynamics: a modeling framework to understand plant growth. *Annual Review of Plant Biology* **6** 65–87.

Cickovski, T., Aras, K., Swat, M., Merks, R. M. H., Glimm, T., George, H., Hentschel, E., Alber, M. S., Glazier, J. A., Newman, S. A. and Izaguirre, J. A. (2007) From genes to organisms via the cell: a problem-solving environment for multicellular development. *Computing in Science and Engineering* **9** (4) 50–60.

Coppo, M., Damiani, F., Drocco, M., Grassi, E. and Troina, A. (2010a) Stochastic calculus of wrapped compartments. In: Di Pierro, A. and Norman, G. (eds.) Proceedings 8^{th} Workshop on Quantitative Aspects of Programming Languages. *Electronic Proceedings in Theoretical Computer Science* **28** 82–98.

Coppo, M., Damiani, F., Drocco, M., Grassi, E., Sciacca, E., Spinella, S. and Troina, A. (2010b) Hybrid calculus of wrapped compartments. In: Ciobanu, G. and Koutny, M. (eds.) Proceedings 4^{th} International Meeting on Membrane Computing and Biologically Inspired Process Calculi. *Electronic Proceedings in Theoretical Computer Science* **40** 102–120.

Danos, V. and Laneve, C. (2003) Core formal molecular biology. In: Degano, P. (ed.) Proceedings 12^{th} European Symposium on Programming. *Springer-Verlag Lecture Notes in Computer Science* **2618** 302–318.

Frisco, P. (2009) *Computing with Cells: Advances in Membrane Computing*, Oxford University Press.

Gillespie, D. (1977) Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* **81** 2340–2361.

Grieneisen, V. A. and Scheres, B. (2009) Back to the future: evolution of computational models in plant morphogenesis. In: Lohmann, J. and Nemhauser, J. (eds.) Cell signalling and gene regulation. *Current Opinion in Plant Biology* **12** (5) 606–614.

Harel, D. and Kugler, H. (2010) Some Thoughts on the Semantics of Biocharts. In: Manna, Z. and Peled, D. (eds.) Time for Verification: Essays in Memory of Amir Pnueli. *Springer-Verlag Lecture Notes in Computer Science* **6200** 185–194.

Haseltine, E. L., Rawlings, J. B. and Yin, J. (2005) Dynamics of viral infections: incorporating both the intracellular and extracellular levels. In: Maranas, C. and Hatzimanikatis, V. (eds.) Computational Challenges in Biology. *Computers and Chemical Engineering* **29** (3) 675–686.

Krivine, J., Milner, R. and Troina, A. (2008) Stochastic bigraphs. *Electronic Notes in Theoretical Computer Science* **218** 73–96.

Kugler, H., Larjo, A. and Harel, D. (2010) Biocharts: a visual formalism for complex biological systems. *Journal of the Royal Society Interface* **7** (48) 1015–1024.

Manes, E. G. (1998) Implementing collection classes with monads. *Mathematical Structures in Computer Science* **8** (3) 231–276.

Meier-Schellersheim, M., Fraser, I. D. and Klauschen, F. (2009) Multi-scale modeling in cell biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **1** (1) 4–14.

Meier-Schellersheim, M., Xu, X., Angermann, B., Kunkel, E. J., Jin, T. and Germain, R. N. (2006) Key role of local regulation in chemosensing revealed by a new molecular interaction-based modeling method. *PLOS Computational Biology* **2** (7) e82.

Milner, R. (2009) *The Space and Motion of Communicating Agents*, Cambridge University Press.

Mjolsness, E. and Yosiphon, G. (2006) Stochastic process semantics for dynamical grammars. *Annals of Mathematics and Artificial Intelligence* **47** 329–5.

Noble, D. (2002) Modeling the heart—from genes to cells to the whole heart. *Science* **295** 1678–1682.

Păun, G. (2001) P systems with active membranes: attacking NP-complete problems. *Journal of Automata, Languages and Combinatorics* **6** (1) 75–90.

Păun, G. (2008) Membrane computing and brane calculi. Old, new, and future bridges. *Theoretical Computer Science* **404** (1-2) 19–25.

Pedersen, M. and Plotkin, G. D. (2010) A language for biochemical systems: design and formal specification. In: Priami, C., Breitling, R., Gilbert, D., Heiner, M. and Uhrmacher, A. M. (eds.) Transactions on Computational Systems Biology XII. Special Issue on Modeling Methodologies. *Springer-Verlag Lecture Notes in Computer Science* **5945** 77–145.

Priami, C., Regev, A., Shapiro, E. Y. and Silverman, W. (2001) Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Information Processing Letters* **80** (1) 25–31.

Regev, A., Panina, E. M., Silverman, W., Cardelli, L. and Shapiro, E. Y. (2004) BioAmbients: an abstraction for biological compartments. *Theoretical Computer Science* **325** (1) 141–167.

Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. and Elowitz, M. B. (2005) Gene regulation at the single-cell level. *Science* **307** (5717) 1962–1965.

Spicher, A., Michel, O., Cieslak, M., Giavitto, J-L. and Prusinkiewicz, P. (2008) Stochastic P systems and the simulation of biochemical processes with dynamic compartments. *BioSystems* **91** 458–472.

Srivastava, R., You, L., Summers, J. and Yin, J. (2002) Stochastic vs. deterministic modeling of intracellular viral kinetics. *Journal of Theoretical Biology* **218** (3) 309–321.