Dynamics of HIV infection studied with Cellular Automata and conformon-P systems

David W. Corne Pierluigi Frisco
School of Mathematical and Computer Sciences,
Heriot-Watt University, Edinburgh, EH14 4AS
{dwcorne/pier}@macs.hw.ac.uk

Recently a cellular automaton (CA) has been used to model the dynamics of HIV infection, with interesting results. We replicate and further test this model, and we introduce an alternative model based on conformon-P (cP) systems.

We find (in common with other recently published comments) that the CA model is very sensitive to initial conditions and produces appropriate qualitative dynamics only for a narrow range of rule probabilities.

In contrast, the cP system model is robust to a wide range of conditions and parameters, with more reproducible qualitative agreement to the overall dynamics and to the densities of healthy and infected cells observed *in vivo*.

1 Introduction

The infection by the human immuno-deficiency virus (HIV), the cause of acquired immunodeficiency syndrome (AIDS), has been widely studied both in the laboratory and with computer models in order to understand the different aspects that regulate the virus-host interaction.

Several mathematical models have been proposed (for example [30,43,20]) but all of them fail to describe some aspects of the infection. Typically, models proposed so far have problems in maintaining biological plausibility at the same time as producing a qualitative match to the known dynamics of the concentrations of healthy, and dead cells, the source of the main difficulty being that characteristic dynamics occur at two quite distinct timescales (over the first few weeks, and over several years, respectively). A recent example is that of [20], who integrates a CA approach with a graph percolation model, focusing on HIV's long-term dynamics and the distribution of incubation periods, but this model does not account for the characteristic dynamics of the first few weeks of infection.

However, the recent model reported by Dos Santos & Coutinho in [38], based on cellular automata, clearly shows the different time scales of the infection and has a broad qualitative agreement to the density of healthy and infected

cells observed *in vivo*. The ability of this model to demonstrate the qualitative agreement over two time scales sets it apart as a promising candidate for further research. However, in [39] it is noted that this qualitative agreement is reached only if some parameters are chosen in a small interval. If some of the parameters are chosen outside this interval, then the cellular automata model of [38] does not follow the dynamics of what is observed *in vivo*. Dos Santos and Coutinho's approach is therefore clearly of interest (as is also pointed out in [39]), but it is unclear whether it fulfils the requirement for a model of HIV dynamics that is both *robust* and qualitatively accurate.

In this paper we set out to investigate a new approach – conformon-P systems – that has emerged from theoretical computer science (specifically from the area of *Membrane computing*) and that can be used for modelling natural processes. We therefore developed a cP system based model of HIV infection dynamics, and we also replicated the Dos Santos and Coutinho CA model [38] and performed many tests to determine their relative robustness to parameters and the degree to which their results qualitatively echo the dynamics seen in real data.

The paper proceeds as follows. In section 2 we further describe the two techniques, cellular automata and cP systems, focusing on a full explanation of the latter, which are less well known as a modeling paradigm. Section 3 then describes the particulars of implementing a Dos Santos and Coutinho style HIV infection model in each of these two systems. In section 4 we describe our experiments and show the results, and we discuss these results in section 5.

2 The modeling platforms

2.1 Cellular automata

Cellular Automata (CA) are a regularly used platform for modelling, and are increasingly explored as modelling tools in the context of natural phenomena that exhibit characteristic spatiotemporal dynamics [40,4]. They became popular in the 80s as tools for studying self-organisation in artificial systems [44] and were primary in the development of the field known as *Artificial Life* [24], while more recently there is a steady stream of research articles reporting the use of various types of CA to model a variety of dynamic processes. Of interest here, for example, are their use in modelling the spread of infection [1,27,38,20,41].

A CA consists of a finite number of cells (invariably arranged in a regular spatial grid), each of which can be in one of a finite (typically small) number

of specific states. For example, in a model of disease in which the CA's cells happen to correspond to biological cells, identified states might be *healthy*, *infected*, and *dead*. In the usual approach, at each time step t the status of the CA is characterised by its state vector; that is, the state of each of the cells. In the simplest type of CA, the state vector at time $t + 1$ is obtained from that at time t by the operation of a single rule applied in parallel (synchronously) to each cell. The rule specifies how the state of a cell will change as a function of its current state and the states of the cells in its neighbourhood (see figure 6).

In many applications, including that addressed here, it is appropriate for the rule to be probabilistic. That is, for each combination of a cell's current state and neighbourhood configuration, the rule provides a probability distribution over possible next states. Also, it is natural to think of and implement this rule as a number of individual nonconflicting rules, one for each possible 'current state'. Finally, CAs may also be 'asynchronous', in which individual cells are chosen randomly (usually) and updated one at a time.

The straightforward nature of the time evolution of a CA, combined with its emphasis on local interactions, has made it an accessible and attractive tool for modelling many biological processes.

2.2 Conformon-P systems

The subdivision of a cell into compartments delimited by membranes has been an inspiration to G. Păun for the definition of a new class of (distributed and parallel) models of computation called *membrane systems* [32]. The hierarchical structure, the locality of interactions, the inherent parallelism, and also the capacity (in the less basic models) for membrane division, represent the distinguishing hallmarks of membrane systems.

Research on membrane systems, also called *P systems* (where 'P' stays for 'Păun'), has really flourished [33,45].

In [11] a variant of membrane systems called *conformon-P (cP) systems* was introduced. This variant, later studied also in [12,7,9,10], is based on simple and basic concepts inspired by a theoretical model of the living cell centred around *conformon* [18,19].

The concept of conformon was introduced in molecular biology independently in [15] and [42]. The common part of the two definitions is the conformational deformation of (macro) molecules in a cell.

A cP system has conformons, a name-value pair, as objects. If V is an al-

phabet (a finite set of letters) and $\mathbb{N}_0$ is the set of natural numbers (with 0 included), then we can define a conformon as $[\alpha, a]$, where $\alpha \in V$ and $a \in \mathbb{N}_0$, we will say that α is the *name* and a is the *value* of the conformon $[\alpha, a]$. If, for instance, $V = \{A, B, C, \ldots\}$, then $[A, 5], [C, 0], [Z, 14]$ are conformons. To indicate several instances of the same conformon we will write (c, k), where c is a conformon and $k \in \mathbb{N} \cup \{+\infty\}$. So, for instance, we can write $([A, 1], 5), ([B, 0], +\infty)$, etc.

Two conformons can interact according to an *interaction rule*. An interaction rule is of the form $r : \alpha \xrightarrow{n} \beta$, where r is the label of the rule, $\alpha, \beta \in V$ and $n \in \mathbb{N}_0$, and it says that a conformon with name α can give n from its value to the value of a conformon having name β. A rule r can be applied only if the value of the conformon with name α is greater or equal to n. If, for instance, there are conformons $[G, 5]$ and $[R, 9]$ and the rule $r : G \xrightarrow{3} R$, the application of r leads to $[G, 2]$ and $[R, 12]$.

The compartments (membranes) present in a cP system have a label, every label being different. Compartments can be unidirectionally connected to each other and for each connection there is a *predicate*. A predicate is an element of the set $\{\geq n, \leq n \mid n \in \mathbb{N}_0\}$. Examples of predicates are: $\geq 5, \leq 2$, etc.. If, for instance, there are two compartments (with labels) m_1 and m_2 and there is a connection from m_1 to m_2 having predicate ≥ 4, then conformons having value greater or equal to 4 can pass from m_1 to m_2. In a time unit any number of conformons can move between two connected membranes as long as the predicate on the connection is satisfied. Notice that we have *unidirectional connections* that is: m_1 connected to m_2 does not imply that m_2 is connected to m_1. Moreover, each connection has its own predicate. If, for instance, m_1 is connected to m_2 and m_2 is connected to m_1, the two connections can have different predicates.

A simple cP system is illustrated in Figure 1.

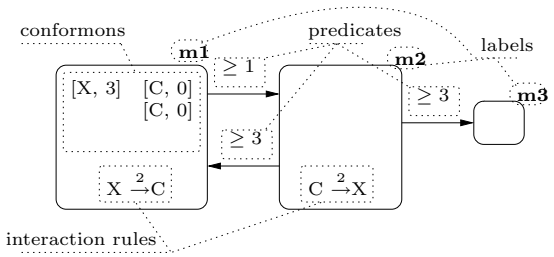

Fig. 1. A cP system.

Conformon-P systems do not work under the requirement of *maximal parallelism*, typical to the majority of the variants of P systems but, when their activity is simulated by a computer, then probabilities are associated with interaction rules and connections between membranes.

Here we do not introduce the details (output membrane, acknowledgement membrane, etc.) present in the definition of cP systems when used as a computing platform. The investigations on these systems led to the characterization of an initial computational hierarchy composed of: *basic cP system*, *proper cP system*, *cP system with priorities*, and *cP system with unbounded value*.

Variations on the structural and related constraints of cP systems have clear influences on their computational power. For example, it has been proved that the computational power of cP systems with priorities and cP systems with unbounded value are equivalent to that of so-called program machines [28], that the power of cP systems is equivalent to that of partially blind program machines, and that basic cP systems can only generate finite languages.

The use of (variants of) P systems to model various biological processes is not new [45], this line of research in P systems was initiated by D. Besozzi [3]. Conformon-P systems have been used to model simple biological processes [10]. Here we focus on modeling and we use a *grid of cP systems*, i.e. a complex cP system composed by *cells*, each cell being a simple cP system connected to some other cells.

In Figure 2 a grid of cP systems is depicted.

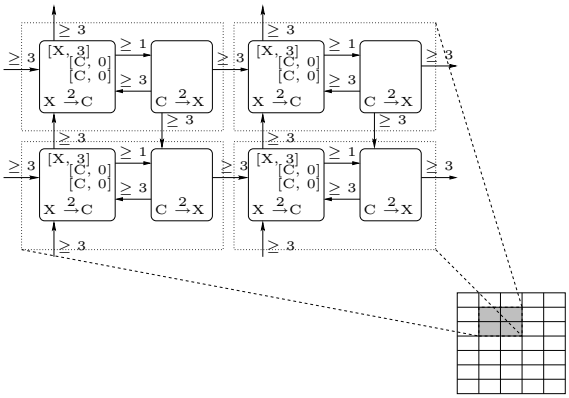

Fig. 2. A grid of cP systems.

This way to create P systems (to our knowledge new in membrane computing) resembles CA. In Section 5.1 we will highlight the differences between these two formal systems.

2.2.1 Some modules for conformon-P systems

In the following we will use the concept of *module*: a group of membranes with conformons and interaction rules in a cP system able to perform a specific

task.

An example of module is a *splitter* [7]: a module that, when a conformon $[X, x]$ with $x \in \{x_1, \ldots, x_h\}, x_i < x_{i+1},\ 1 \leq i \leq h-1$ is associated with a specific membrane of it, it can pass such a conformon to other specific membranes according to the value x. In Figure 3 a splitter is depicted.

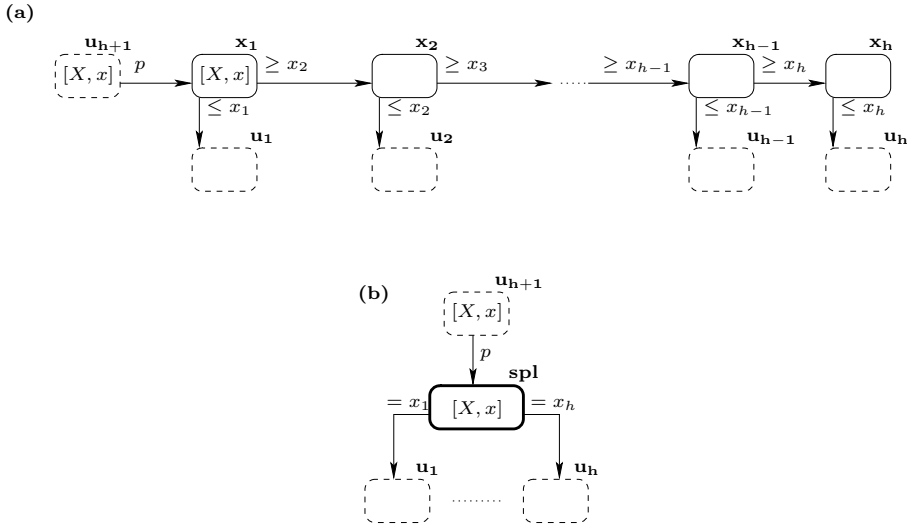

Fig. 3. A detailed Splitter **(a)** and its module representation **(b)**

The links between cells present in the cP systems described in Section 3.1 have predicates of the kind $[A, a]$ (a conformon). This is a shorthand for a *separator* module [7]: when conformons of type $[X_i, x],\ 1 \leq i \leq h,\ x \geq 1$ are associated with specific membranes of it, a separator can pass them to specific different membranes according to their name content. So if there is an edge between membrane 1 and membrane 2 having $[A, a]$ as predicate, it means that only the conformons $[A, a]$ can pass from membrane 1 to membrane 2. In Figure 4 a separator is depicted.

The combination of splitters and separators allows us to define the following kind of interaction rules:

- $A^{(\alpha)} \xrightarrow{\gamma} B_{(\beta)}$ where $\alpha, \beta, \gamma \in \mathbb{N}_0$, meaning that a conformon with name A can interact with B passing γ only if the value of A and B before the interaction is α and β respectively;
- $A \xrightarrow{\gamma} B_{(\beta)}$ where $\beta, \gamma \in \mathbb{N}_0$, meaning that a conformon with name A can interact with B passing γ only if the value of B before the interaction is β;
- $A^{(\alpha)} \xrightarrow{\gamma} B$ where $\alpha, \gamma \in \mathbb{N}_0$, meaning that a conformon with name A can interact with B passing γ only if the value of A before the interaction is α.

A module has to be regarded as a function in a computer program. Moreover, a module can be analysed according to its processing features and interconnected

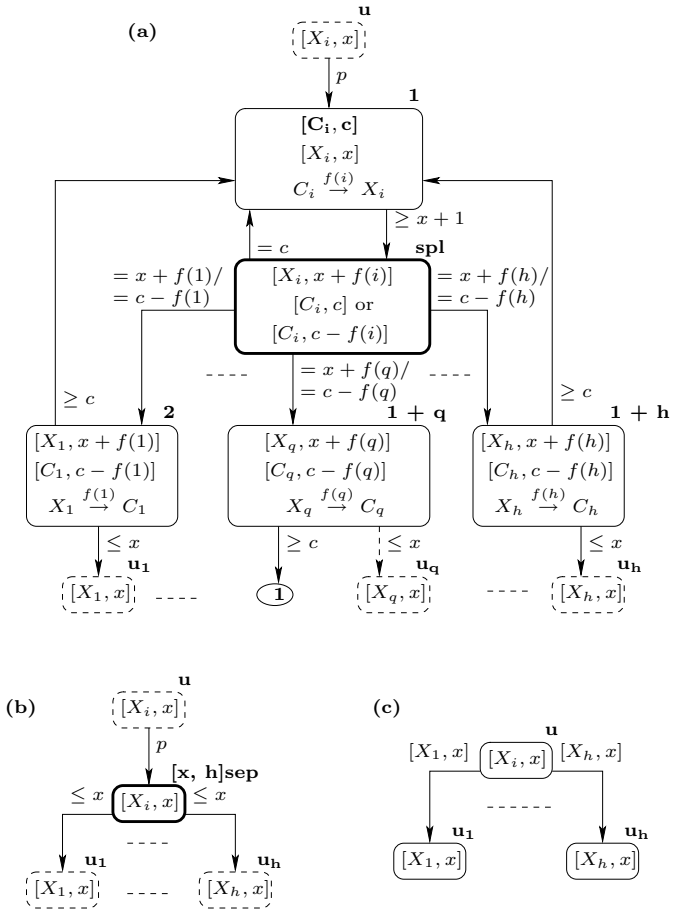

Fig. 4. A detailed Separator **(a)** and its module representations **(b)** and **(c)**

loops present in it; this allows us to classify processes modelled by cP systems as a function of the used modules. This last point will be discussed in Section 5.1

3 The process and the models

We modeled the dynamics of HIV infection, following the description of this process as indicated in [38,39].

The dynamics observed in HIV infections can be divided into three phases. Initially the amount of virus in the host grows in an exponential way, then the viral load drops to what is known as the "set point", finally the amount of virus in the host increases slowly, accelerating near the onset of AIDS. The first two phases occur in the first weeks following the infection, the third phase can last years. This is plotted in Figure 5 where each unit in the x axes represent a week in time.

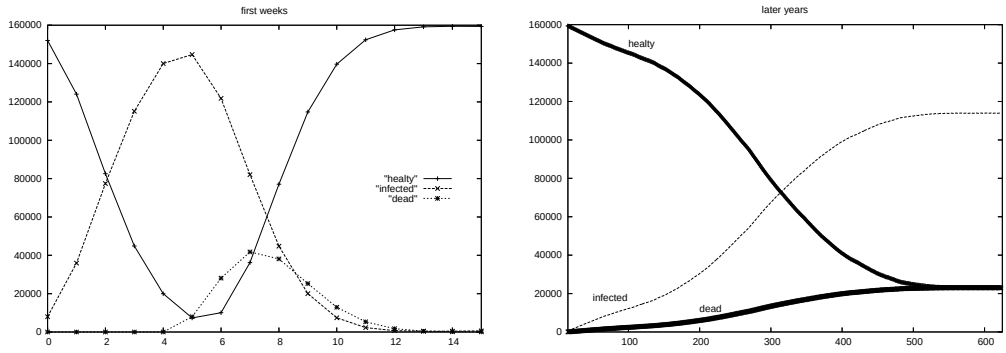

Fig. 5. Typical dynamics of HIV infection.

In [38] this process was modeled with a CA in which each cell could be in any of four possible states: *healthy*, *A-infected*, *AA-infected*, and *dead*. In the (random) initial configuration a cell had probability p_{HIV} to be *A-infected*, otherwise it is *healthy*.

Our CA model followed this implementation with minor differences, each of which lead to equivalent behaviour. The main difference was that the state *A-infected* was broken down into four separate states, *A-infected1*, *A-infected2*, *A-infected3*, and *A-infected4*. In this way the CA controlled the change from *A-infected* to *AA-infected* over $\tau = 4$ time steps without the need to introduce an additional counter-based mechanism in the CA's operation. This was not really *necessary*, but it enabled us to use a model that consisted of a simpler and more straightforward CA framework, and reflected our desire to compare relatively 'pure' versions of the CA and cP system approaches.

Where the term *A-infected* is used in the rules below, it is shorthand for the combined set of *A-infected1*, *A-infected2*, *A-infected3*, and *A-infected4* cells in the neighbourhood.

I if a *healthy* cell has at least one *A-infected* neighbour, then it becomes *A-infected1* at the next time step;
II if a *healthy* cell has no *A-infected* neighbours but at least R *AA-infected* neighbours, then it becomes *A-infected1* at the next time step;
III an *A-infected1* cell becomes *A-infected2* at the next time step;
IV an *A-infected2* cell becomes *A-infected3* at the next time step;
V an *A-infected3* cell becomes *A-infected4* at the next time step;
VI an *A-infected4* cell becomes *AA-infected* at the next time step;
VII an *AA-infected* cell becomes *dead* at the next time step;
VIII a *dead* cell becomes, at the next time step, either *healthy* (with probability $p_{repl} \times (1 - p_{infec})$), or *A-infected1* (with probability $p_{repl} \times p_{infec}$), or stays *dead* (with probability $1 - p_{repl}$).

The biological reasoning behind these rules is explained in [38]. Essentially, rules I and II model the basic spread of viral infection from cells to neigh-

bouring cells; rules III–VII model the short life of an infected cell, and the last rule models the body's continual replenishment of new healthy cells but maintaining a small probability of infection.

In [38] the following parameters were chosen: $p_{HIV} = 0.05, p_{repl} = 0.99, p_{infec} = 10^{-5}$, and $R = 5$; additionally, [38] had a parameter τ for indicating the number of timesteps after which an *A-infected* cell became *AA-infected*, and maintained $\tau = 4$ throughout their work. This parameter setting is reflected in our use of four separate *A-infected* states. They experimented with grids of size ranging from 300×300 to 1000×1000, and the averaged results of 500 simulations reported in [38] on toroidal grids ranging from 700×700 show a qualitative agreement to the density of healthy and infected cells observed *in vivo*.

In [39] it is shown that this qualitative agreement is reached only for values of the parameters close to the ones just indicated. If either $p_{HIV} < 10^{-2}$ or p_{infec} is chosen in the range 10^{-2} to 10^{-4}, then the CA model of [38] does not follow the dynamics of what is observed *in vivo*.

3.1 Grid of Conformon-P systems

The construction of the grid of cP systems used by us closely follows the CA defined in [38]. Each cell on the grid contains only one membrane (so in the following, cell and membrane will have the same meaning). The set of rules associate with each membrane can be divided in two parts: *part 1* and *part 2* (see Appendix A). The rules in each part are similar except in the probabilities associated with them. In the following we will consider and describe the rules in *part 1*, later on we will extend this description to the rules in *part 2*.

Each cell can be in one of five states: *1-healthy*, *A-infected*, *AA-infected*, *predead*, and *dead* (in respect to the rules in *part 1*) identified by the presence of the conformons: $[H, 1], [A, 1], [AA, 1], [PD, 1]$, and $[D, 1]$ respectively. If, for instance, a cell is in an *healthy* state, then it will contain $[H, 1], [A, 0], [AA, 0]$, $[PD, 0]$, and $[D, 0]$ (similarly for the other cases). In the initial configuration each cell contains the conformons $([R, 1], +\infty), ([V, 10], +\infty), ([E, 0], +\infty)$, and $([W, 0], +\infty)$.

If a cell is *A-infected*, then it can generate $[V, 11]$ (meaning: if a cell is *A-infected* it can generate a virus). This is performed by the rules:

$$\textbf{1: } R \xrightarrow{1} A_{(1)} \qquad\qquad \textbf{2: } A^{(2)} \xrightarrow{1} V_{(10)}$$

Notice that $[V, 10]$ does not represent a virus, $[V, 11]$ does.

$[V, 11]$ conformons can pass from a cell to any other in its neighbourhood (meaning: viruses can spread between cells).

An *1-healthy* cell can become *A-infected* if it contains a virus. This is performed by the rules:

$$\textbf{3: } V \xrightarrow{11} H_{(1)} \qquad \textbf{4: } H^{(12)} \xrightarrow{12} A_{(0)} \qquad \textbf{5: } A^{(12)} \xrightarrow{11} W_{(0)}$$

An *AA-infected* cell can generate $[E, 1]$ conformons. These conformons can pass to other cells and interact such that $[E, 4]$ conformons are created. When a $[E, 4]$ conformon is present in an *healthy* cell, then it can become *A-infected*.

This process mimics rule II in Section 3 and it is performed by:

$$\textbf{6: } R \xrightarrow{1} AA_{(1)} \qquad \textbf{7: } AA^{(2)} \xrightarrow{1} E_{(0)} \qquad \textbf{8: } E^{(1)} \xrightarrow{1} E_{(1)} \qquad \textbf{9: } E^{(2)} \xrightarrow{2} E_{(2)}$$

$$\textbf{10: } E \xrightarrow{4} H_{(1)} \qquad \textbf{11: } H^{(5)} \xrightarrow{5} A_{(0)} \qquad \textbf{12: } A^{(5)} \xrightarrow{4} W_{(0)}$$

and by the fact that $[E, 1]$ can pass from one cell to any other in its neighbourhood. From the rules 7, 8, and 9 it should be clear that only $[E, 1], [E, 2]$, and $[E, 4]$ can be present in the system.

An *A-infected* cell can become *AA-infected* by the application of the rule:

$$\textbf{13: } A^{(1)} \xrightarrow{1} AA_{(0)}$$

An *AA-infected* cell can become *dead*. Before doing so it goes into the *pre-dead* state in which the $[V, 11], [E, 1], [E, 2]$, and $[E, 4]$ conformons present in it are removed. This is performed by the rules:

$$\textbf{14: } AA^{(11)} \xrightarrow{1} PD_{(0)} \quad \textbf{15: } V^{(11)} \xrightarrow{1} PD_{(1)} \quad \textbf{16: } E \xrightarrow{1} PD_{(1)} \quad \textbf{17: } E \xrightarrow{2} PD_{(1)}$$

$$\textbf{18: } E \xrightarrow{4} PD_{(1)} \quad \textbf{19: } PD^{(1)} \xrightarrow{1} D_{(0)} \quad \textbf{20: } PD^{(2)} \xrightarrow{1} W_{(0)} \quad \textbf{21: } PD^{(3)} \xrightarrow{2} W_{(0)}$$

$$\textbf{22: } PD^{(5)} \xrightarrow{4} W_{(0)}$$

A *dead* cell can become *2-healthy* cell by the application of the rule

$$\textbf{23: } D^{(1)} \xrightarrow{1} H2_{(0)}$$

The R and W conformons do not have a direct relation with any aspect of HIV infection. In broad terms the R conformons can be regarded as 'food' molecules needed by a cell in a certain state to perform an action (for instance, if *A-infected* to generate a virus). The W conformons can be regarded as 'waste' molecules to which some conformons can pass part of their value. As W conformons only receive values from other conformons, their initial value

is not relevant for the simulation.

As the simulation of a cP system is probabilistic we cannot be sure that when a *dead* cell becomes *2-healthy* all the viruses and E conformons have been removed by the rules 15-18 and 20-22. In order to allow the removal of all viruses and $[E, 1]$ conformons from a *pre-dead* cell before it becomes *dead*, we associate with rules 15-18 and 20-22 an higher probability than the probability associated with rule 19 (see Appendix A). Anyhow, for lack of time, we did not study how effective this difference in probabilities is.

The state *2-healthy*, together with *A2-infected*, *AA2-infected*, *2-pre-dead*, and *2-dead* are managed by the rules in *part 2*. The rules in *part 2* are similar to the ones in *part 1* but they have $H2$ instead of H, $A2$ instead of A, $AA2$ instead of AA, $PD2$ instead of PD, and $D2$ instead of D.

Most importantly, the probabilities associated with the rules in *part 2* are lower than the probabilities associated with the rules in *part 1* (see Appendix A).

Considering what we said in Section 3, rules in *part 1* model the behaviour of the first two phases of the dynamics of HIV infection, while rules in *part 2* model the behaviour of the third phase.

It should be clear by the rules given above that a cell can only be in one of the 10 possible states: this depends on how the interaction rules have been devised.

4 Experiments and Results

The simulations performed with the CA were based on a toroidal 400 × 400 grid for every combination of the following parameters:

neighbourhood: von Neumann, Moore, diamond, as indicated in Figure 6, where the grey squares represent the domain and the black square the co-domain of the applied rules;

p_{HIV}: 0.05, 0.00005;

p_{infec}: 0.00001, 0.00005.

The simulations performed with the cP system were based on a toroidal 50 × 50 grid for every combination of the following parameters. As discussed in Section 5.1, the parameter p_{infec} is not directly equivalent to the similar

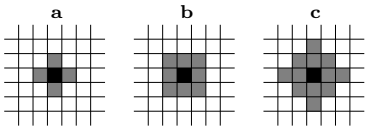

Fig. 6. The considered neighbourhoods: (a) von Neumann, (b) Moore, (c) diamond.

parameter as used in the CA, but the chosen values, in the context of how the cP system operates, lead to a like-for-like comparison:

neighbourhood: von Neumann, Moore, diamond, as indicated Figure 6 where a black cell can pass conformons to any of the grey cells;

p_{HIV}: 0.05, only 1 cell (i.e. 0.0004);

p_{infec}: 0.2, 1 (see Section 5.1 for a discussion on the choice of this parameter).

All simulations were run 10 times with different random number sequences.

We note that we also performed simulations of the CA model on a toroidal 50×50 grid, but these exhibited the three-phrase dynamics in only 3 of 120 runs. Finally, the choice of grid size for the CA was based on preliminary investigation using the [38] parameter set, which indicated that a toroidal 400×400 grid was the smallest at which the three-phase dynamics would be regularly observed (this dependence on grid size is material, and will be briefly discussed later). Meanwhile, the choice of grid size for the cP system was based on the pragmatics of the available software.

4.1 Cellular automata results

The outcome of the tests based on the CA can be divided into three categories: match with the expected results, lack of the first two phases, and lack of the third phase.

Only a few of the tests using the same parameters considered in [38] (Moore neighbourhood, $p_{HIV} = 0.05$, and $p_{infec} = 0.00001$) resulted in the expected curve. Similar results were present if the von Neumann neighbourhood was considered; in the diamond neighbourhood more of the tests showed the expected dynamics.

All the tests with $p_{HIV} = 0.00005$ rendered a result lacking the first two phases (this conforms with what was reported in [39]). The only choice of parameters that clearly qualitatively adhered to the expected results were when $p_{HIV} = 0.05$ and $p_{infec} = 0.00005$. When the qualitative agreement is present it strictly follows that indicated in [38] (see Figure 5), the only exception is when von Neumann neighbourhood is chosen together with $p_{HIV} = 0.05$ and $p_{infec} =$

0.00005.

4.2 Conformon-P systems results

In the vast majority of the tests we run the three phases of the dynamics of HIV are clearly visible (see Figure 7).

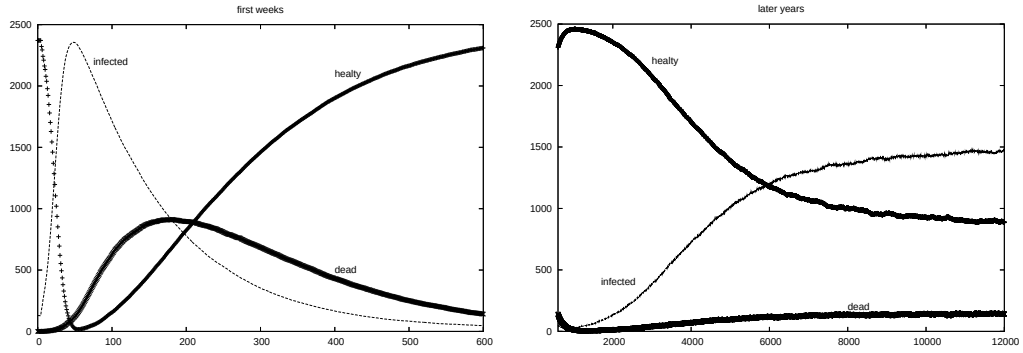

Fig. 7. Typical outcome for cP systems.

Only in two of the 120 tests the third phase is anomalous: all cells are *healthy*. This happened with the Moore neighbourhood, $p_{HIV} = 0.05, p_{infec} = 1$, and with diamond neighbourhood, $p_{HIV} = 1$ cell, $p_{infec} = 0.2$.

The outcome of the remaining tests does not have remarkable differences depending on the value of p_{infec}, it does have differences depending on the value of p_{HIV}. If only one *infected* cell is present in the initial configuration, then the curves of *infected* and *healthy* cells in the first two phases are smoother. The maximum value reached in these two phases by the number of *infected* cells is lower than when $p_{HIV} = 0.05$; in a similar way the minimum value reached in these two phases by the number of *healthy* cells is higher than when $p_{HIV} = 0.05$. These maximum and minimum depend on the considered neighbourhood: the smaller the number of neighbours, the smoother the function. This can be seen by the plot of the average results for the first two phases for each neighbourhood present in Figure 8.

When $p_{HIV} = 0.05$ the maximum reached by the number of *infected* cells in the first two phases is around 2300 (92% of the number of cells), while the minimum reached by the number of *healthy* cells is around 10 (0.4% of the number of cells). When only one *infected* cell is preset in the initial configuration in a von Neumann neighbourhood these values are around 1650 (66%) and 325 (13%) respectively; in a Moore neighbourhood these values are around 1750 (70%) and 230 (9.2%) respectively; in a diamond neighbourhood these values are around 1900 (76%) and 164 (6.5%).

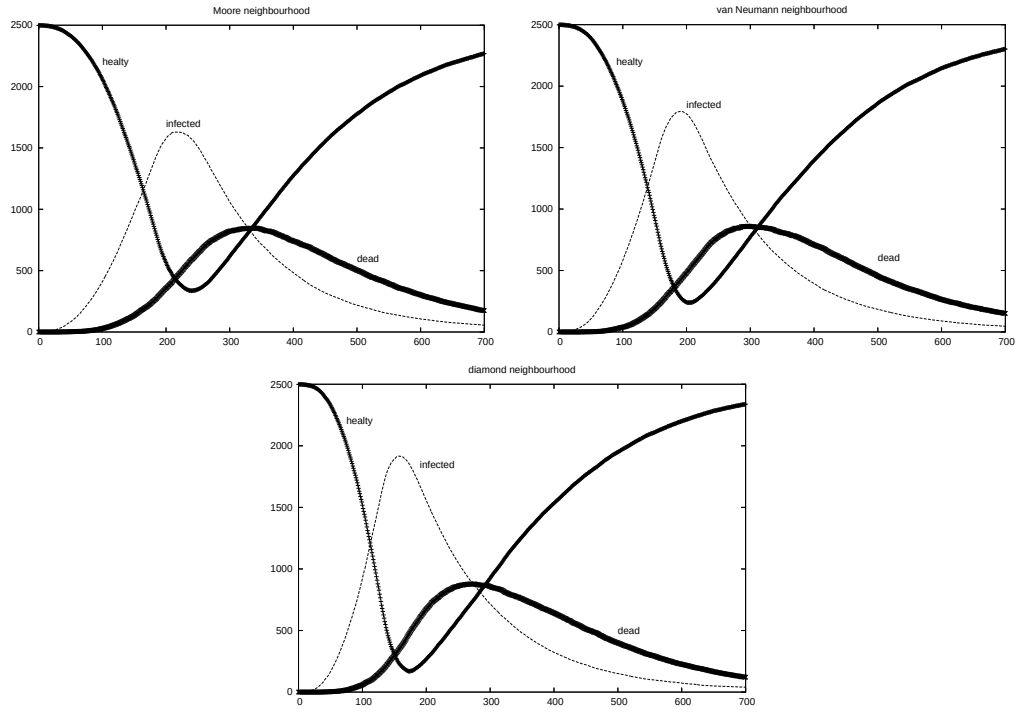

Fig. 8. Average dynamics of the three neighbourhoods of the first two phases of infection when the initial configuration has only one *infected* cell.

There is a qualitative agreement between our results for the density of *healthy*, *infected*, and *dead* cells and the *in vivo* observations. There are only two discrepancies:

(1) the time evolution does not follow what is observed *in vivo*: the first two phases (first 10 weeks of the dynamics) can be associated with the first 500 iterations, while in the third phase we can match 1000 iterations with one year;
(2) the number of *infected* and *healthy* cells at the end of the third phase is not in accordance with the observed values. The averaged values reported in [38] see the *infected* cells being at 70% density while the *healthy* and *dead* cells are both at 15%. Our average results see the *infected* cells density at 60%, and the *healthy* cells density at 34%, the *dead* cells being 6% of the total population.

In order to address these points we believe we can 'tune' the cP system by changing some of the probabilities associated to the interaction rules in *part 2*.

We will study how the number of events (interaction and passage rules) needed to the cP system to simulate a step in the CA model influences the outcome of the cP system. This is related to the fact that in the CA model almost all rules are performed in one time step, while the number of events in the cP

system simulating the CA rules is not always the same. For instance, rule I of the CA model is simulated by several events while rule VI in the CP model is simulated by one interaction rule in the cP system.

These studies will be done in the context of our ongoing research in understanding how to craft cP systems operations to model a given pattern.

5 Concluding remarks

5.1 CA and conformon-P systems

As indicated in Section 2.2 grids of cP systems resemble CAs. In particular, the style of cP system described bears a loose resemblance to a Lattice Gas Cellular Automaton (LGCA); these are CA models which involve 'particles' moving around on a lattice, and, following the earliest such model [16] which was set up to study ergodicity in fluids, they are now commonly used in fluid dynamics [37].

In computer science terms, there is much that could be said about the relative raw computational abilities of CAs, LGCAs and cP systems (especially cP systems, since these are primarily abstract computing devices and have been mostly studied in that light). The differences between their computational capabilities are a complex function of their relative structure and operations, and it can also arguably be said that many of the distinctions that can be made are of limited interest to those whose aim is to model biological systems. However, there are three aspects that we argue are of importance to bioscientists and computer scientists alike.

The first, which we will not cover further in this paper, concerns understanding the kinds of computation performed in nature; it is of interest, for example, to discover what is the 'simplest' (in some sense) computational model that can reliably replicate a certain observed natural process.

The second is the degree to which a modelling approach provides a framework that can be easily used and has robust, repeatable behaviour.

The third concerns the theoretical bases and the possibilities to analyse the approach; if the approach comes along with a strong theoretical background and associated analysis techniques, it becomes easier to reason about the space and time requirements of planned simulations, certain predictable aspects of the results, and so forth.

Cellular Automata are certainly highly accessible modelling tools and we be-

lieve this has much to do with their popularity. However they are not always the appropriate choice. The lack of robustness in the present case, suggests that the CA model does not adequately capture the dynamics of the process under study, although we have not examined robustness to parameters and reliability at considerably larger grid sizes.

In contrast, when modelling a natural process, the translation of plausible rules and related knowledge into the cP system framework is less direct and consequently less accessible, however we believe that the results indicated above show that it can be worth the effort.

Most importantly the cP systems can be analysed with several complexity measures. Dynamical properties of cP systems can be easily studied with the use of Petri nets. In contrast, the analyses of CAs to date ([4], chapter 4) have been done with a variety of approximations that yield statistical distributions for the state vector (see section 2.1) that are plausible only in simple cases.

Analyses of the dynamical properties of cP systems are not yet mature, however. There is only an initial study of the dynamical behaviour of membrane systems [46,34,17,22,9] and no study of the dynamics of a *grid* of P systems has been done to date. Meanwhile, in [9] the dynamics present in cP systems are analysed with Petri nets [36].

As cP systems allow a precise analysis of the modeled processes and as their dynamics can be easily studied with Petri nets, then we expect in future to have a deeper and more precise understanding of the processes modeled by (grids of) cP systems. We could, for instance, classify processes according to the modules present in the (grid of) cP systems used to simulated them.

Another measure of behaviour for cP system is *loop* number and topology. A loop is related to the path followed by a conformon (where only the name but not the value is considered) during a computation. In Figure 1, for instance, the X conformon can pass from membrane m1 to membrane m2 and then back to membrane m1. So, in that system, the X conformon can move in a loop. In [8] it is indicated that two loops are *connected* if one can be completed only if the other is taking place. We are going to study how the number of loops, their connections and topology are related to the behaviour of models of (grids of) cP systems.

One more measure of complexity that we will consider will be applied only to grids of cP systems and it concerns changing them (their interaction rules, passage rules, and conformons) according to some probabilistic variables. Starting from a ('original') grid of cP systems having a specific behaviour we will create other ('clones') similar grids of cP systems in which the presence of interaction rules, passage rules, and conformons is probabilistic. Then we will link the differences in behaviours of the 'clones' to the 'original' to the used probabil-

ities. This line of research is inspired by S. Kauffman's work on *autocatalytic set theory* (see, for instance [21]) described at the time as a novel approach concerning the origin of life.

If we consider the grid of cP systems described in this paper, then we could ask the following questions: How does the outcome of the model change in respect to probabilities associated to interaction rules/passage rules/conformons? Are there probabilities thresholds (as the ones found by Kauffman) below which a given grid does not exhibit the desired outcome? How are the probabilities of interaction rules, passage rules, and conformon related in systems exhibiting the desired outcome?

We used the simulator for cP systems described in [10]. This simulator was not created having in mind grids of these systems, so it is not optimized in this respect (for instance, the rules present in each membrane in the grid are present in as many copies as the number of cells in the grid while they could be present in just one copy). This 'bug' did not allow us to use grids of bigger size for the simulations. Anyhow, it is noteworthy that the size of the grid for cP systems, 50×50, was sufficient to yield robust behaviour in the present study. In contrast, tests of the CA on a 50×50 grid almost invariably led to all cells in the *healthy* state after the first few generations, which was maintained throughout the run.

Concerning grid size in particular, we make a note here about the point in phase three at which the *infected* cell density begins rapidly to rise and the *healthy* cell density begins rapidly to fall (corresponding to the onset of AIDS). In the Dos Santos and Coutinho model, it is noted [38,39] that this point is closely associated with the chance appearance of a particular small 'pattern' of cell states in the grid, which leads directly to expanding 'waves' of *infected* cells (see figure 7, left, which shows a corresponding pattern from our CA model). The chance of the 'seed' pattern occurring is clearly related to size of the grid and the probabilities in the rules, as well as dependent on the synchronous nature of the CA, and seemingly tied to the context of a regular lattice. In contrast, the cP system model does not seem to rely on such factors. Figure 7 (right) shows the grid of a cP system run, sampled at a point corresponding to when the CA run in the left figure was sampled, the point at which the density of *infected* cells became higher than the density of *healthy* cells.

What in the CA is a single event (i.e., a *healthy* cell becoming *A-infected1*), in the cP system is distributed over a sequence of events (interaction and passage rules) that can be argued to more finely mimic what happens in reality. We believe this hints at the reasons for the considerable difference in robustness between the two systems. In effect, the cP system model is more fine-grained in its treatment of time.

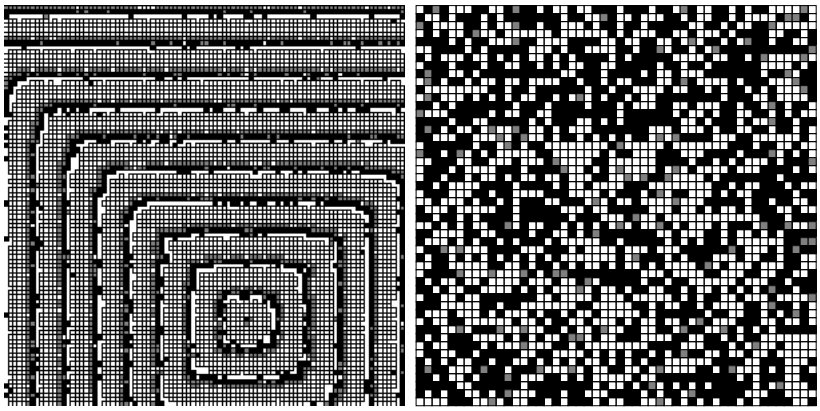

Fig. 9. *Healthy*=white, *infected*=grey, *dead*=black.

Anyhow, the high level of accuracy possible with cP systems does not have to be considered an absolute advantage. It is well known that a good model needs to have an appropriate balance between the level of details and abstractions [29]. We think that the use of modules in cP systems allows to easily tune this balance.

As indicated in Section 2.2.1 a module can be regarded as a function in a computer program. So it is possible to describe how (a module of) a cP system can perform a specific action. This action can then be regarded as an atomic event (eventually with a probability associated with it) in a more abstract cP system. Of course we do not think that this feature is a prerogative of cP systems. What we do think is that, in respect to other formal systems, cP systems are based on simpler operations and concepts (interaction between conformons, locality of interaction, and passage of conformons between local compartments) that anyhow do not limit their potential.

More then with modules the way to operate of cP systems can be complemented by algorithms approximating such processes as the ones described in [13,14,25]. A similar approach has been already investigated in P systems (for instance, [2,31]).

We did not pursue yet a precise study in trying to understand how the sequence of operations (interaction and passage rules) and the probabilities associated with them are related to the dynamics of HIV infection. These rules and component probabilities were designed, in the context of how the cP system operates, to match the corresponding parameter settings in the CA. However we need to more clearly set out and specify the relationship between the formulation of these operations and the process being modelled. This is our next first priority in this line of research. This understanding will allow us to model in a more accurate way processes and to understand better their dynamics. It is for this reason that in order to increase p_{infec} we only changed

the probabilities associated with rules 27 and 32.

5.2 Why conformon-P systems?

The tile of this section repeats a question that it is often asked to one of the authors. Here we give a comprehensive answer.

In Membrane Computing the main trend in analysing the computing power of a biological process has been to introduce a new variant of membrane systems, not to modify an already existing model. This has led to a proliferation of very specific variants at the expense of a global understanding of the features present in the processes having similar properties.

Conformon-P systems originate by the combination of a theoretical model of the living cell and membrane computing. Their definition is simple and very general and can easily fit the description of many biological process (they are not limited to cellular processes) and be used for the creation of abstract computing systems. One of their strengths relies in their abstract description: a conformons can be anything in a process. This is very different from other models of membrane systems defined and tailored for a specific process.

Because of their abstract description the laws deducted from a cP system with some feature remain valid for all the cP systems with the same features independently from the process they model. On the other hand the laws deducted from a variant of membrane systems defined for a specific process are only valid in the limits of that process.

In recent years a lot of effort has been devoted to looking at nature through 'glasses of computation': processes present in nature have been classified according to their computational potential or interpreted as computation [5,35,23]. This research has led to very promising results but limited to the investigated processes.

There is no 'global framework' able to unify in a structured way the various degrees of 'computation' present in nature. This global framework has to be powerful (able to simulate computation at any level), flexible (able to abstract a wide range of processes), and subject to the study of several complexity measures.

We believe that conformon-P systems could be such a 'global framework'.

Acknowledgements

Authors have been listed following alphabetical ordering on the family-name.

The work of P. Frisco has been partially supported by the research grant NAL/01143/G of The Nuffield Foundation.

References

[1] E. Ahmed, H. N. Agiza, and S. Z. Hassan. On modelling hepatitis B transmission using cellular automata. *J. Stat. Phys.*, 92(3/4), 1998.

[2] F. Bernardini, M. Gheorghe, N. Krasnogor, R. C. Muniyandi, M. J. Pérez-Jímenez, and F. J. Romero-Campero. On P systems as a modeling tool for biochemical systems. In Freund et al. [6], pages 114–133.

[3] P. Besozzi. *Computational and modelling power of P systems.* PhD thesis, Universitá degli Studi di Milano, Italy, 2004.

[4] A. Deutsch and S. Dormann. *Cellular Automaton Modeling of Biological Pattern Formation: Characterization, Applications, and Analysis.* Birkäusen, Boston, 2004.

[5] A. Ehrenfeucht, T. Harju, I. Petre, D. M. Prescott, and G. Rozenberg. *Computation in Living Cells. Gene Assembly in Ciliates.* Springer-Verlag, Berlin, Heidelberg, New York, 2003.

[6] R. Freund, G. Lojka, M. Oswald, and G. Păun, editors. *Proceedings of the 6^{th} International Workshop on Membrane Computing WMC6*, volume 3850 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, Heidelberg, New York, 2006.

[7] P. Frisco. The conformon-P system: A molecular and cell biology-inspired computability model. *Theoretical Computer Science*, 312(2-3):295–319, 2004.

[8] P. Frisco. Infinite hierarchies of conformon-P systems. In *Seventh Workshop on Membrane Computing (WMC7), Leiden, The Netherlands, July 17-21 2006*, 2006.

[9] P. Frisco. P systems, Petri nets, and Program machines. In Freund et al. [6], pages 209–223.

[10] P. Frisco and R. T. Gibson. A simulator and an evolution program for conformon-P systems. In *SYNASC 2005, 7^{th} International Symposium on Simbolic and Numeric Algorithms for Scientific Computing*, pages 427–430. IEEE Computer Society, 2005. Workshop on Theory and Applications of P Systems, TAPS, Timisoara (Romania), September 26-27, 2005.

[11] P. Frisco and S. Ji. Conformons-P systems. In M. Hagiya and A. Ohuchi, editors, *DNA8, 8th International Meeting on DNA Based Computers, Hokkaido University, Sapporo, Japan, June 10-13*, volume 2568 of *Lecture Notes in Computer Science*, pages 291–301. Springer-Verlag, Berlin, Heidelberg, New York, 2002.

[12] P. Frisco and S. Ji. Towards a hierarchy of info-energy P systems. volume 2597 of *Lecture Notes in Computer Science*, pages 302–318. Springer-Verlag, Berlin, Heidelberg, New York, 2002.

[13] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J.Comp. Phys*, 22:403–434, 1976.

[14] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem*, 81:2340–2361, 1977.

[15] D. E. Green and S. Ji. The electromechanical model of mitochondrial structure and function. *Molecular Basis of Electron Transport*, pages 1–44, 1972. J. Schultz, B. F. Cameron (eds).

[16] J. Hardy, Y. Pomeau, and O. de Pazzis. Time evolution of a two-dimensional model system, 1. Invariant states and time correlation functions. *J. Math Phys.*, 14:1746–1759, 1973.

[17] O. H. Ibarra, Z. Dang, and O. Egecioglu. Catalytic P systems, semilinear sets, and vector addition systems. *Theoretical Computer Science*, 312(1-2):379–399, 2004.

[18] S. Ji. The Bhopalator: a molecular model of the living cell based on the concepts of conformons and dissipative structures. *Journal of Theoretical Biology*, 116:395–426, 1985.

[19] S. Ji. The Bhopalator: an information/energy dual model of the living cell (II). *Fundamenta Informaticae*, 49(1-3):147–165, 2002.

[20] C. Kamp and S. Bornholdt. From HIV infection to AIDS: a dynamically induced percolation transition? *Proceedings of the Royal Society B: Biological Sciences*, 269(1504):2035–2040, 2002.

[21] S. Kauffman. *At Home in the Universe.* Oxford University Press, New York, 1996.

[22] J. Kleijn, M. Koutny, and G. Rozenberg. Towards a Petri net semantics for membrane systems. In Freund et al. [6], pages 292–309.

[23] L. F. Landweber and E. Winfree, editors. *Evolution as computation.* Natural computing series. Springer Verlag, Berlin, Heidelberg, New York, 1999.

[24] C. G. Langton. Studying artificial life with cellular automata. *Physica D*, 2(1-3):120–149, 1986.

[25] T. Lu, D. Volfson, L. Tsimring, and J. Hasty. Cellular growth and division in the Gillespie algorithm. *IEE Systems Biology*, 1:121–127, 2004.

[26] C. Martín-Vide, G. Mauri, G. Păun, G. Rozenberg, and A. Salomaa, editors. *Membrane Computing, International Workshop, WMC 2003, Tarragona, Spain, July 17-22, 2003, Revised Papers*, volume 2933 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, Heidelberg, New York, 2003.

[27] M. L. Martins, G. Ceotto, S. G. Alves, C. C. B. Bufon, J. M. Silva, and F. F. Laranjeira. Cellular automata model for citrus variegated chlorosis. *Phys. Rev. E*, 62(5):7024–7030, 2000.

[28] M. L. Minsky. *Computation: Finite and Infinite Machines*. Automatic computation. Prentice-Hall, 1967.

[29] F. Morrison. *The Art of Modeling Dynamic Systems*. Wiley-Interscience, 1991.

[30] A. S. Perelson and P. W. Nelson. Mathematical analysis of HIV-1 dynamics in vivo. *SIAM Review*, 41(1):3–44, 1999.

[31] M. J. Pérez-Jímenez and F. J. Romero-Campero. P systems, a new computationl modelling tool for systems biology. *Transactions on Computational Systems Biology VI LNBI*, 4220:176–197, 2005.

[32] G. Păun. Computing with membranes. *Journal of Computer and System Sciences*, 1(61):108–143, 2000.

[33] G. Păun. *Membrane Computing. An Introduction*. Springer-Verlag, Berlin, Heidelberg, New York, 2002.

[34] Z. Qi and J. You. P systems and Petri nets. In Martín-Vide et al. [26], pages 387–403.

[35] A. Regev and E. Shapiro. Cells as computation. *Nature*, 419:343, 2002.

[36] W. Reisig and G. Rozenberg, editors. *Lectures on Petri Nets I: Basic Models*, volume 1491 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, Heidelberg, New York, 1998.

[37] D. H. Rothman and S. Zaleski. *Lattice Gas Cellular Automata: Simple Models of Complex Hydrodynamics*. Cambridge University Press, 2004.

[38] R. M. Dos Santos and S. Coutinho. Dynamics of HIV infection: a cellular automata approach. *Physical review letters*, 87(16):168102, 2001.

[39] M. C. Strain and H. Levine. Comment on "Dynamics of HIV infection: a cellular automata approach". *Physical review letters*, 89(21):219805, 2002.

[40] T. Toffoli and N. Margolus. *Cellular Automata Machines: A New Environment for Modeling*. MIT press, 1987.

[41] S. Venkatachalam and A. Mikler. Towards computational epidemiology: Using stochastic cellular automata in modeling spread of diseases. In *Proceedings of the 4^{th} Annual International Conference on Statistics, Mathematics and Related Fields*, 2005.

[42] M. V. Volkenstein. The conformon. *Journal of Theoretical Biology*, 34:193–195, 1972.

[43] D. Wodarz and M. A. Nowak. Mathematical models of HIV pathogenesis and treatment. *Bioessays*, 24(12):1178–1187, 2002.

[44] S. Wolfram. Statistical mechanics of cellular automata. *Rev. Mod. Phys.*, 55:601–644, 1983.

[45] C. Zandron. P-systems web page: http://psystems.disco.unimib.it.

[46] S. Dal Zilio and E. Formenti. On the dynamics of PB systems. In Martín-Vide et al. [26], pages 197–208.

A Rules, links, and probabilities

part 1			*part 2*		
label	**rule**	**prob.**	**label**	**rule**	**prob.**
1	$R \xrightarrow{1} A_{(1)}$	1	**24**	$R \xrightarrow{1} A2_{(1)}$	0.2
2	$A^{(2)} \xrightarrow{1} V_{(10)}$	1	**25**	$A2^{(2)} \xrightarrow{1} V_{(10)}$	0.0075
3	$V \xrightarrow{11} H_{(1)}$	1	**26**	$V \xrightarrow{11} H2_{(1)}$	0.0075
4	$H^{(12)} \xrightarrow{12} A_{(0)}$	1	**27**	$H2^{(12)} \xrightarrow{12} A2_{(0)}$	0.2
5	$A^{(12)} \xrightarrow{11} W_{(0)}$	1	**28**	$A2^{(12)} \xrightarrow{11} W_{(0)}$	0.2
6	$R \xrightarrow{1} AA_{(1)}$	1	**29**	$R \xrightarrow{1} AA2_{(1)}$	0.2
7	$AA^{(2)} \xrightarrow{1} E_{(0)}$	1	**30**	$AA2^{(2)} \xrightarrow{1} E_{(0)}$	0.2
8	$E^{(1)} \xrightarrow{1} E_{(1)}$	1			
9	$E^{(2)} \xrightarrow{2} E_{(2)}$	1			
10	$E \xrightarrow{4} H_{(1)}$	1	**31**	$E \xrightarrow{4} H2_{(1)}$	0.0075
11	$H^{(5)} \xrightarrow{5} A_{(0)}$	1	**32**	$H2^{(5)} \xrightarrow{5} A2_{(0)}$	0.2
12	$A^{(5)} \xrightarrow{4} W_{(0)}$	1	**33**	$A2^{(5)} \xrightarrow{4} W_{(0)}$	0.2
13	$A^{(1)} \xrightarrow{1} AA_{(0)}$	0.25	**34**	$A2^{(1)} \xrightarrow{1} AA2_{(0)}$	0.0625
14	$AA^{(11)} \xrightarrow{1} PD_{(0)}$	1	**35**	$AA2^{(11)} \xrightarrow{1} PD2_{(0)}$	0.2
15	$V^{(11)} \xrightarrow{1} PD_{(1)}$	1	**36**	$V^{(11)} \xrightarrow{1} PD2_{(1)}$	0.2
16	$E \xrightarrow{1} PD_{(1)}$	1	**37**	$E \xrightarrow{1} PD2_{(1)}$	0.2
17	$E \xrightarrow{2} PD_{(1)}$	1	**38**	$E \xrightarrow{2} PD2_{(1)}$	0.2
18	$E \xrightarrow{4} PD_{(1)}$	1	**39**	$E \xrightarrow{4} PD2_{(1)}$	0.2
19	$PD^{(1)} \xrightarrow{1} D_{(0)}$	0.008	**40**	$PD2^{(1)} \xrightarrow{1} D2_{(0)}$	0.08
20	$PD^{(2)} \xrightarrow{1} W_{(0)}$	1	**41**	$PD2^{(2)} \xrightarrow{1} W_{(0)}$	0.2
21	$PD^{(3)} \xrightarrow{2} W_{(0)}$	1	**42**	$PD2^{(3)} \xrightarrow{2} W_{(0)}$	0.2
22	$PD^{(5)} \xrightarrow{4} W_{(0)}$	1	**43**	$PD2^{(5)} \xrightarrow{4} W_{(0)}$	0.2
23	$D^{(1)} \xrightarrow{1} H2_{(0)}$	0.008	**44**	$D2^{(1)} \xrightarrow{1} H2_{(0)}$	0.02

Links:
$[V, 11]$ can pass with probability 1 from any cell to any of its neighbours; $[E, 1]$ can pass with probability 0.01 from any cell to any of its neighbours.