Autocatalytic Closure and the Evolution of Cellular Information Processing Networks

by

James Decraene

B.Sc.

Supervisor: Prof. Barry McMullin

Internal examiner: Dr. Darragh O'Brien

External examiner: Prof. Hughes Bersini

A thesis submitted to the School of Electronic Engineering

Dublin City University for the degree of

Doctor of Philosophy

June 2009

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy (PhD.) is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ ID No.: 55156720

James Decraene

Date: _____

Abstract

Cellular Information Processing Networks (CIPNs) are chemical networks of interacting molecules occurring in living cells. Through complex molecular interactions, CIPNs are able to coordinate critical cellular activities in response to internal and external stimuli. We hypothesise that CIPNs may be abstractly regarded as subsets of collectively autocatalytic (i.e., organisationally closed) reaction networks. These closure properties would subsequently interact with the evolution and adaptation of CIPNs capable of distinct information processing abilities. This hypothesis is motivated by the fact that CIPNs may require a mechanism enabling the self-maintenance of core components of the network when subjected to internal and external perturbations and during cellular divisions. Indeed, partially replicated or defective CIPNs may lead to the malfunctioning and premature death of the cell.

In this thesis, we evaluate different existing computational approaches to model and evolve chemical reaction networks *in silico*. Following this literature review, we propose an evolutionary simulation platform capable of evolving artificial CIPNs from a bottom-up perspective. This system is a novel agent-based Artificial Chemistry (AC) which employs a term rewriting system called the Molecular Classifier System (MCS.bl). The latter is derived from the Holland broadcast language formalism.

Our first series of experiments focuses on the emergence and evolution of self-maintaining molecular organisations in the MCS.bl. Such experiments naturally relate to similar studies conducted in ACs such as Tierra, Alchemy and α-universes. Our results demonstrate some counter-intuitive outcomes, not indicated in previous literature. We examine each of these "unexpected" evolutionary dynamics (including an elongation catastrophe phenomenon) which presented various degenerate evolutionary trajectories. To address these robustness and evolvability issues, we evaluate

several model variants of the MCS.bl. This investigation illuminates the key properties required to allow the self-maintenance and stable evolution of closed reaction networks in ACs. We demonstrate how the elongation catastrophe phenomenon can be prevented using a multi-level selectional model of the MCS.bl (which acts both at the molecular and cellular level). Using this multi-level selectional MCS.bl which was implemented as a parallel system, we successfully evolve an artificial CIPN to perform a simple pre-specified information processing task. We also demonstrate how signalling crosstalk may enable the cooperation of distinct closed CIPNs when mixed together in the same reaction space. We finally present the evolution of closed crosstalking and multitasking CIPNs exhibiting a higher level of complexity.

Dedication

Pour Jean-Louis: mon père, ami et mentor

Acknowledgments

Je remercie: Magali qui fut toujours à mes côtés durant les moments les plus difficiles et sans qui cette thèse n'aurait vu le jour, Barry for his excellent and wise supervision, mes parents et ma soeur pour leur soutien constant et leur confiance en moi, George and Ciarán for all the good times during those past four years et finalement l'Union européenne[1] pour avoir financé mon doctorat.

[1]The work presented in this thesis was funded by the ESIGNET project (Evolving Cell Signalling Networks *in silico*, a European Integrated Project in the EU FP6 NEST initiative, contract no. 12789).

Contents

Chapter 1

Introduction

Complex Adaptive Systems (CAS) are dynamical networks of interacting agents which as a whole determine the behaviour, adaptivity and cognitive ability of the system (Holland, 1992b). CAS are ubiquitous and occur in a variety of natural and artificial systems (e.g., biological cells, stock markets, the biosphere). Realising and evolving CAS *in silico* may provide new critical tools for understanding, predicting and building CAS.

However modelling and evolving CAS remains problematic as the traditional analytical and statistical approaches (which may be coupled with Evolutionary Computation techniques) appear to limit the study of CAS (Holland, 2006). Indeed no computational techniques have to date successfully supported open-ended evolution as occurring in *natural* CAS. Achieving open-ended evolution is a critical problem which is related to many other grand challenges in the field of Artificial Life (Bedau et al., 2000; Gershenson and Lenaerts, 2008).

This thesis addresses one aspect of this issue on the evolutionary growth of complexity by examining the significance of autocatalytic closure for the evolution of CAS complexity. A specific subclass of CAS is addressed in this investigation: Cellular Information Processing Networks (CIPNs) which are chemical networks occurring in living cells capable of information processing. This thesis examines the evolution of organisationally closed CIPNs *in-silico*.

The remainder of this chapter is organised as follows. A presentation of autocatalytic closure (a concept originating from Kaufffman's autocatalytic set theory) is first provided. A brief description of CIPNs is then given. Finally the potential applications, aims, objectives and layout of this thesis are enumerated.

1.1 Autocatalytic set theory

The autocatalytic set theory was proposed by Kauffman (1993) to explain the emergence and early evolution of life. An autocatalytic set is a collection of molecular species where each is capable of supporting the catalysis of another species in the set. It is argued that given a critical mass of molecular species, the spontaneous emergence and self-organisation of an autocatalytic set may occur.

Although individual species are not capable of self-replication, *a contrario* to RNA world models (Gilbert, 1986; Gesteland et al., 2005) or hypercycles (Eigen, 1971; Eigen and Schuster, 1977), the set of species as a whole is able to catalyse its own production. Such a molecular set or reaction network is said to be *collectively* autocatalytic/self-replicating, see Fig. 1.1.

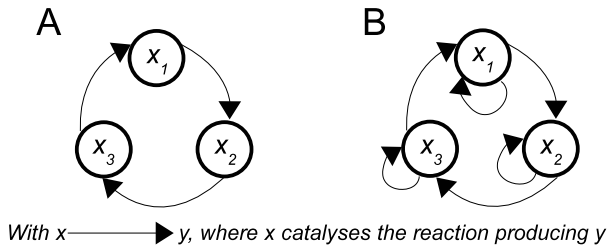

Figure 1.1: A: A simple collectively autocatalytic reaction network. The species x_1 catalyses the production of x_2. Similarly the production of x_3 and x_1 can be catalysed by x_2 and x_3 respectively, forming/closing the autocatalytic loop. B: A three element hypercycle, in contrast to collectively autocatalytic reaction networks, individual molecular species are capable of self-production in hypercycles.

An autocatalytic set is organisationally closed when the production of each species contained in the set may be catalysed by another member of the set. This *virtuous* catalytic cycle enables a "closed set" to self-maintain/repair when subjected

to internal and external perturbations and during cellular divisions (e.g., molecular mutations, removal/addition of molecular species through diffusion) given a continuous inflow of "food molecules".

Kauffman argued that self-organisation and autocatalytic closure are key principles allowing the spontaneous emergence and maintenance of order given an original chaotic system. When subjected to a Darwinian evolutionary regime where natural selection and variation/heredity of phenotypic traits may occur, closed reaction networks may evolve and *gradually* grow in structural and functional complexity leading to higher-order organisms.

Although the autocatalytic set theory was initially proposed to address biological organisations, it has also been metaphorically extended to the study of social and economic systems (Gabora, 2008; Kauffman, 1995).

1.2 Cellular Information Processing Networks

Cellular Information Processing Networks (CIPNs) are biochemical systems of interacting molecules occurring in living cells. CIPNs are responsible for coordinating the cellular activities in response to internal and external stimuli.

As signal processing systems, CIPNs can be regarded as special purpose computers (Bray, 1995). In contrast to conventional silicon-based computers, the information processing in CIPNs is not realised by electronic circuits, but by chemically reacting molecules in the cell. There is an almost infinite variety of potential molecular species, each of which would have distinct chemical functionality and could engage in interactions with other molecules with varying degrees of specificity.

An example of CIPN are Cell Signalling networks[1] (Helmreich, 2001; Krauss, 2003) such as the chemotaxis signalling pathway (Stock et al., 1992) which may

[1]The work presented in this thesis was funded by the ESIGNET project (Evolving Cell Signalling Networks *in silico*, a European Integrated Project in the EU FP6 NEST initiative, contract no. 12789). The ESIGNET project aimed at realising and evolving artificial cell signalling networks to perform computational functions.

occur in simple organisms (e.g., bacteria). This CIPN enables bacteria to move toward higher concentrations of specific chemicals or flee from toxic chemicals in their surroundings. this

In the simplest cases, CIPNs can be approximately modelled by systems of continuous differential equations, where the state variables are the concentrations of the distinct species of interacting molecules. As an "information processing" device, this is most naturally compared to a traditional *analog* computer. Analog computers are precisely designed to model the operation of a target dynamical system, by creating an "analogous" system which shares (approximately) the same dynamics. Electronic analog computers (based on the "operational amplifier" as the core computational device) have long been displaced by digital computers, programmed to numerically solve the relevant dynamical equations, due to their much greater ease of programming and stability.

While CIPNs are typically treated in this "aggregate" manner, where the information is carried by molecular concentration, one can also consider the finer grained behaviours of individual molecules which are computational in nature. Thus a single enzyme molecule can be regarded as carrying out pattern matching to identify and bind target substrates, and then executing a discrete computational operation in transforming these into the product molecule(s). This has clear parallels with a wide variety of so-called "rewriting systems" in computational theory. However, it also clearly differs in important ways, such as:

- Operation is stochastic rather than deterministic.
- Operation is intrinsically *reflexive* in that all molecules can, in principle, function as both "rules" (enzymes) and "strings" (substrates/products).

Dittrich (2004) provides a more extended discussion of the potential of such "chemical computing".

In this thesis, we hypothesise that CIPNs can be considered as subsets of organisationally closed reaction networks. A motivation for this hypothesis is that closure properties may be necessary to ensure the self-maintenance and robustness of CIPNs when subjected to internal and external perturbations and during cellular divisions.

Moreover by exploiting the principles of Darwinian evolution, it is intended to evolve closed chemical networks of higher complexity which are capable of performing information processing tasks distinctive of CIPNs.

1.3 Applications

The potential applications of realising and evolving CIPNs include:

- *Engineering crosstalk:* A natural phenomenon occurring in CIPNs called "crosstalk" and its potential contributions to engineering are discussed. Crosstalk occurs when signals from different pathways become mixed together. This arises very naturally in CIPNs due to the fact that the molecules from all pathways may share the same physical reaction space (the cell). Depending on the relative specificities of the reactions there is then an automatic potential for any given molecular species to contribute to signal levels in multiple pathways. In traditional communications and signal processing engineering, crosstalk is regarded as a defect: an *unintended* interaction between signals, that therefore has the potential to cause system malfunction. This can also clearly be the case with crosstalk in real biochemical networks, for example cells may become cancerous due to undesired crosstalk connections (Mukai et al., 2005; Yee and Lee, 2000). However, in the specific case of CIPN's, crosstalk also has additional potential functionalities, which may actually be constructive:

  - Even where an interfering signal is, in effect, adding uncorrelated "noise" to a functional signal, this may sometimes improve the overall system

behaviour. This is well known in conventional control systems engineering in the form of so-called "dither" (Korn and Korn, 1956). Molecular biologists indicated that noise is an inevitable by-product of inherent molecular interactions, and that in fact noise is essential for development (Volfson et al., 2005).

- The crosstalk mechanism may also provide a very generic way of creating a large space of possible modifications or interactions between chemical pathways. Thus, although many cases of crosstalk may be immediately negative in their impact, crosstalk may still be a key mechanism in enabling incremental evolutionary search for more elaborate or complex CIPNs. For example, Genoud and Metraux (1999) presented a number of crosstalk connections between real biochemical networks occurring in plants in which these "interferences" provided a relatively *rapid* and *efficient* mechanism for optimizing non-cognitive behaviour in response to various combinations of stimuli. Crosstalk may also provide the necessary signal that enables desired outcome to occur, an example of this could be coordinating the cell cycle (Goto et al., 2005).

Both above cases of crosstalk may give new insights on the use of crosstalk in control engineering.

- *CIPNs as information processing devices*: Nature is a source of inspiration for information processing techniques which have been successfully applied to a wide variety of complex application domains. In keeping with this we examine the possibility of utilising CIPNs for information processing purposes. Realising and evolving artificial CIPNs may provide new computational paradigms for a variety of application areas. Early work conducted by Bray (1995) showed that molecules could be regarded as information processing devices, these molecules would perform simple computational tasks. Examples of such

information processing functions are: signal acceleration (Mangan and Alon, 2003), signal amplification (Binder and Heinrich, 2004) or decision making (Xiong and Ferrell, 2003). A review on the computational abilities of signalling networks can be found in Sauro (2004). Identified information processing processes occurring in CIPNs indicate that complex operational features have been designed in CIPNs through natural evolution. Moreover, there may be applications where a molecular level analog computer, in the form of a CIPN, may have distinct advantages. Specifically, CIPNs may offer high speed and small size that cannot be realised with solid state electronic technology. More critically, where it is required to interface information processing with chemical interaction, a CIPN may bypass difficult stages of signal transduction that would otherwise be required. This could have direct application in so-called "smart drugs" and other bio-medical interventions.

- *Open-ended evolution:* Our project finally addresses the conditions allowing open-ended evolution to occur. Achieving an open-ended growth of complexity is a long-standing grand challenge related to the evolution of artificial systems (Bedau et al., 2000). Although many computational evolutionary techniques have been proposed to evolve artificial systems, no systems have to date managed to support the open-ended evolutionary growth of complexity as occurring in the real world. Evolutionary systems would typically plateau and avert the perpetual emergence of complex behaviours (Groß and McMullin, 2002).

  Understanding the key components[2] enabling open-ended evolution may en-

[2]Example key components are the fitness functions devised in evolutionary systems. Two principal approaches to fitness functions are distinguished: 1) Explicit fitness functions which are *explicitly* defined/engineered and govern the agents' genotype/phenotype mapping. These functions are usually fixed and do not evolve over time. Systems which rely on explicit fitness functions (and similar engineered elements) are declared as "top-down" evolutionary approaches. 2) Implicit fitness functions which, in contrast, are not engineered/explicitly devised in the system. Here, agents determine "by themselves" their fitness according to their intrinsic properties and interactions with the environment/other agents. Systems which rely on implicit fitness functions are declared as "bottom-up" evolutionary approaches.

able one to apply this knowledge to real-world domains such as solving optimisation problems. Traditional Evolutionary Computation techniques are ultimately limited by the capabilities of human engineers. A *genuine* open-ended and artificial evolutionary system (Pattee, 1973) would abolish this barrier and allow for the perpetual creation of novel solutions in an ever-changing environment.

1.4 Aims and objectives

We first enumerate the principal research questions of this thesis:

1. What are the minimal conditions necessary to obtain the open-ended evolutionary growth of complexity?
2. What is the significance of autocatalytic closure to the evolution of complexity?
3. Can CIPNs be considered as subsets of closed chemical reaction networks?
4. Can we evolve closed CIPNs, of higher complexity, to achieve pre-specified information processing tasks?

To address more specifically the above singular research questions, we investigate the following research workpackages and objectives which are identified as follows:

- *Modelling Chemical Networks:* To identify the state of the art in the multidisciplinary field of scientific modelling applied to the study of biochemical networks. A review of the main existing families of modelling techniques according to a range of selected and relevant criteria will be provided.
- *Evolving CIPNs:* To determine a satisfactory evolutionary framework to examine autocatalytic closure and the evolution of Cellular Information Processing Networks. A selection of techniques applied to evolving and examining closure dynamics of chemical networks will be evaluated. Two families of evolutionary

methods are distinguished: top-down/Evolutionary Computation techniques and bottom-up/Artificial Chemistries methods. A comparison between top-down and bottom-up evolutionary approaches will be conducted.

- *Evolutionary simulation platform:* To propose a novel evolutionary simulation platform capable of evolving closed reaction networks to carry-out pre-specified information processing tasks. This stochastic system will account for the reflexive nature of molecular species which are regarded as condition/action rules. A novel agent-based Artificial Chemistry will be constructed. This system is termed the MCS.bl and employs a term-rewriting formalism (the broadcast language which was devised by Holland, 1975, 1992a) to specify the molecular species and reactions.

- *Closure in reaction networks:* To provide complementary insights on the evolutionary dynamics (e.g., spontaneous emergence, self-maintenance) of closed reaction networks in Artificial Chemistries. A series of experiments focusing on the emergence, self-maintenance and evolution of closed reaction networks using the MCS.bl will be carried out.

- *Parallelism in Artificial Chemistries:* To address the concurrent nature of chemical processes in the MCS.bl. The effects of parallelism upon evolutionary dynamics in Artificial Chemistries will be explored. A parallel version of the MCS.bl using distributed computing facilities will be implemented.

- *Evolutionary capability:* To contribute to the understanding of evolutionary capability in Artificial Chemistries. The effects of compartmentalisation, molecular diffusion and cellular division over the system's evolutionary capability will be examined. A series of evolutionary experiments will be performed using the parallel version of the MCS.bl in which compartmentalisation, molecular diffusion and cellular division features are introduced.

- *Evolution of closed reaction networks:* To demonstrate the evolution of closed reaction networks capable of performing pre-specified information processing tasks. An additional series of evolutionary experiments will be conducted using the cellular model of the MCS.bl. A novel cellular division criterion is devised to drive the evolution of the closed reaction networks.

- *Crosstalk and the evolution of complexity:* To demonstrate the constructive role of crosstalk in enabling the evolutionary growth of complexity in closed reaction networks. A series of evolutionary experiments will be carried out in which crosstalking networks are evolved to carry-out pre-specified multitasking functions.

1.5 Structure of the thesis

The chapters of this thesis are summarised as follows:

Chapter 2 A review of the main computational techniques (i.e., deterministic, stochastic, probabilistic, algebraic and agent-based) to model chemical networks is given. Based on this evaluation, a suitable framework is identified to represent, simulate and analyse chemical networks in this project.

Chapter 3 An evaluation of several evolutionary techniques to evolve Cellular Information Processing Networks (CIPNs) is presented. Two families of evolutionary techniques are distinguished: top-down/Evolutionary Computation approaches and bottom-up/Artificial Chemistries. The outcome of this review is to identify an adequate framework to examine closure and the evolution of CIPNs.

Chapter 4 A novel Artificial Chemistry (AC) termed the Molecular Classifier System - broadcast language (MCS.bl) is described. The latter is an agent-based AC which employs the Holland broadcast language, a term-rewriting formalism, to specify the molecular species and reactions.

Chapter 5 A first series of experiments which focuses on the spontaneous emergence and self-maintenance of closed reaction networks in the MCS.bl is presented. Unexpected degenerative evolutionary dynamics caused by the emergence of parasitic and elongator molecular species are also examined.

Chapter 6 To address the evolutionary degeneration issues of the MCS.bl, two multi-level selectional model variants which introduce compartmentalisation are presented. These novel MCS.bl implementations exploit distributed computing facilities. A static reactor model with molecular diffusion and a cellular model are independently evaluated.

Chapter 7 The cellular model of the MCS.bl is employed to evolve closed reaction networks to carry-out a pre-specified information processing task. The potential role of crosstalk in enabling the evolutionary growth of complexity in chemical networks is investigated. This chapter demonstrates the evolution of crosstalking closed reaction networks of higher complexity.

Chapter 8 Finally, the contributions and the future work of this thesis are discussed.

Chapter 2

Modelling Chemical Reaction Networks

In Section 1.3 we outlined the potential applications of realising and evolving CIPNs *in silico* as information processing devices. In order to evolve such CIPNs it is necessary to specify, represent, simulate and analyse these chemical reaction networks through the use of modelling techniques. In this chapter a review of different existing techniques for modelling chemical reaction networks is conducted. The outcome of this review is to select a suitable approach to model CIPNs in our project. This multi-disciplinary review results from the collaborative effort that we initiated with the bio-analysis group at the Friedrich Schiller University of Jena (UJ) in Germany (an ESIGNET partner). The survey on Markov chains, chemical master equations and SBML/CellML was conducted by the UJ group (which includes Dr. Thomas Hinze, Thorsten Lenser and Dr. Peter Dittrich). Bayesian networks, term rewriting systems, Petri nets, cellulat and agent-based/learning classifier systems techniques were reviewed by myself. Differential equations and π-calculus were examined by both the UJ group and myself. The evaluation criteria and comparison table were equally devised and realised by the UJ and Dublin City University based group (including Dr. George G. Mitchell and myself). This chapter combines some of the materials published in several ESIGNET deliverables and at various international conferences (ESIGNET, 2006a,b; Decraene et al., 2006, 2007b, 2008a).

2.1 Introduction

A variety of modelling techniques for biological reaction networks have been established in recent years (Alon, 2007). We identify several main branches of modelling techniques:

- *Deterministic*: Chemical reactions are approximated as continuous deterministic processes at the macroscopic/system level. The system's variable states are uniquely determined by the pre-specified parameters describing the reactions (e.g., molecular concentration, reaction rates, etc.) and initial states of these variables. Given an initial set of pre-specified parameters, deterministic models enable one to monitor, predict and describe the dynamics of the system over time and/or space. Examples of deterministic modelling techniques include: ordinary/partial differential equations (Zwillinger, 1992; Polyanin and Zaitsev, 2002; Eungdamrong and Iyengar, 2004; Huang and Ferrell, 1996), Michaelis-Menten models (Heinrich and Schuster, 1996) and power-law models (Vera et al., 2007).

- *Stochastic*: In contrast with deterministic approaches, stochastic models explicitly account for the uncertainty that is involved in molecular processes. The system's variable states are determined by the pre-specified system's parameters and through the use of random variables. By addressing randomness or variability, stochastic models provide a more detailed representation of the system's potential dynamics (and not only the average behaviour as in deterministic approaches). Multiple executions of a stochastic model generate unique (from one another) dynamics/observations. The latter can be used to estimate probability distributions of the system's potential states (assisting in the construction of probabilistic models, see below). Examples of stochastic

modelling techniques include: Markov chains (Gomez et al., 2001) and chemical master equations (Gillespie, 2001).

- *Probabilistic*: Here, the description of stochastic processes/data is addressed in terms of probability. Probabilistic modelling techniques are deterministic approaches which may infer probabilistic relationships between molecular species/system's states from empirical observations. In contrast with stochastic approaches, a probabilistic model is a statistical inference and description technique which does not represent the underlying stochastic molecular mechanics. Given the initial states of the molecular species, these approaches provide a probability-based description of the system's states. The predictive power of these techniques relies on the probabilistic distributions inferred by the model upon a range of *in vivo/silico* experimental observations (i.e., the training set). An example of probability modelling technique include: Bayesian networks (Sachs et al., 2002) and hidden Markov models (Goutsias, 2006).

- *Algebraic*: Modelling discrete characteristics of chemical reaction networks is principally achieved with algebraic approaches. A common basic assumption for these approaches is a finite or recursive enumerable number of elementary objects. Each object is considered as the smallest unit that can be processed by the system model. In particular, a definition of objects determines the granularity and abstraction level of corresponding models (hierarchically composed of objects, classes of objects, and temporal interaction rules). Both biomolecules and processes can form these objects. Interaction between these objects is usually specified by a relationship between system configurations. The whole system description is based on discrete transitions. This allows structural and comparative analysis of both system composition and behaviour, independent of numerical simulation results. Examples of algebraic modelling techniques include: P-systems (Paun and Rozenberg,

2002; Paun et al., 2006), broadcast language (Holland, 1992a), Alchemy (Fontana and Buss, 1994a), Boolean networks (Genoud and Metraux, 1999), π-calculus (Regev et al., 2001) and Petri nets (Reddy et al., 1993).

- *Agent-based*: Agent-based models (ABMs) extend the algebraic framework by introducing richer features in the computational units (i.e., agents). ABMs are commonly implemented with Object-Oriented programming environments in which agents are instantiations of object classes. The latter is a collection of properties (e.g., size, location, concentration, etc.) and methods (e.g., move, die, react, etc.). Agent-based simulations typically involve a large number of molecular and/or cellular agents which are executed in a concurrent or pseudo-concurrent manner. Each agent possesses its own distinct state variables, can be dynamically created/deleted and is capable of interacting with the other agents. The agents' computational methods may include stochastic processes resulting in a stochastic behaviour at the system level. Examples of agent-based modelling techniques include: Stochsim (Le Novère and Shimizu, 2001), Cellulat (Gonzalez et al., 2003) and AgentCell (Emonet et al., 2005). A review of agent-based techniques is given by Chavali et al. (2008).

Deterministic and stochastic approaches are the most frequently employed and studied approaches in the field, whereas the attention given to the use of probabilistic, algebraic and agent-based approaches is more recent but rapidly growing.

Our review of modelling techniques is not exhaustive. We select and review a limited number of techniques which exemplify the above families of modelling approaches. Following on from this we propose a model comparison table. We finally relate this evaluation with the requirements of our project to select a suitable technique with regard to the modelling, simulating and analysing of organisationally closed CIPNs.

2.2 Survey of modelling approaches

We review a selection of modelling techniques used in the study of CIPNs: differential equations, Markov chains, chemical master equations, Bayesian networks, Term Rewriting Systems, Petri nets, π-calculus, Cellulat and Agent-based Learning Classifier Systems. We then present the Systems Biology Markup Language (SBML) and CellML which allow one to specify and disseminate biochemical network models using a standardised language. These markup languages also permit the migration of reaction network models between differing modelling approaches.

2.2.1 Differential equations

Chemical reactions are approximated as continuous deterministic processes at the macroscopic level. Differential equations provide a global understanding of a system and are commonly employed to model chemical reaction networks (Zwillinger, 1992; Polyanin and Zaitsev, 2002; Eungdamrong and Iyengar, 2004; Huang and Ferrell, 1996). Given an initial set of pre-specified properties describing the reactions (e.g., molecular concentration, reaction rates, etc.), this modelling approach enables one to monitor, predict and describe the dynamics of the system over time and/or space.

Here, state variables represent the concentrations of molecular species occurring in a well-stirred reactor with no in/out-flows. The following equation governs the dynamics of each species S whose rate of change in concentration $[S]$ depends on the production and consumption rates v_p and v_c:

$$\frac{d[S](t)}{dt} = v_p([S](t)) - v_c([S](t)). \tag{2.1}$$

In mass-action kinetics, these rates result from the reactant concentrations, their stoichiometric factors $a_{\{i,j\}} \in \mathbb{N}$ (reactants), $b_{i,j} \in \mathbb{N}$ (products) and kinetic constants $k_j \in \mathbb{R}_+$ assigned to each reaction quantifying its speed. For a reaction system with a total number of n species and r reactions

$$
\begin{aligned}
a_{11}S_1 + a_{12}S_2 + \ldots + a_{1n}S_n &\xrightarrow{k_1} b_{11}S_1 + b_{12}S_2 + \ldots + b_{1n}S_n \\
a_{21}S_1 + a_{22}S_2 + \ldots + a_{2n}S_n &\xrightarrow{k_2} b_{21}S_1 + b_{22}S_2 + \ldots + b_{2n}S_n \\
&\vdots \\
a_{r1}S_1 + a_{r2}S_2 + \ldots + a_{rn}S_n &\xrightarrow{k_r} b_{r1}S_1 + b_{r2}S_2 + \ldots + b_{rn}S_n
\end{aligned}
$$

the corresponding ordinary differential equations (ODEs) read:

$$
\frac{\mathrm{d}\,[S_i]}{\mathrm{d}\,t} = \sum_{j=1}^{r} \left(k_j \cdot (b_{ji} - a_{ji}) \prod_{h=1}^{n} [S_h]^{a_{jr}} \right)
$$

In order to obtain a concrete trajectory, all initial concentrations $[S_i](0) \in \mathbb{R}_+$, $i = 1, \ldots, n$ have to be specified. Solving this ODE system together with given initial values allows us to describe the temporal behaviour of the reaction system (Dittrich et al., 2001).

Reaction-diffusion models take into account the spatial location of molecules and allow species concentrations in different spatial locations to vary continuously. These models are specified with sets of Partial Differential Equations (PDEs) (Fritz, 1982). Solutions to PDEs derived from reaction-diffusion models provide an approximation of the species concentrations as a function $[S](t, x)$ of both time t and space x:

$$
\frac{\partial [S](t,x)}{\partial t} = D \frac{\partial^2 [S](t,x)}{\partial x^2} - v([S](t,x)) \frac{\partial [S](t,x)}{\partial x} + v_p([S](t,x)) - v_c([S](t,x)) \quad (2.2)
$$

Equation 2.2 is an example PDE where the variables and functions represent: $[S]$ concentration of species S, $D \in \mathbb{R}_+$ diffusion coefficient, $v([S](t,x))$ convective velocity, and $v_p([S](t,x)), v_c([S](t,x))$ production and consumption rates.

Differential equations (especially ODEs) are the most commonly employed tech-

niques to model biochemical systems due to their strong establishment in the sciences. Nevertheless using these methods (particularly PDEs) may also represent a significant mathematical challenge when attempting to solve large systems of nonlinear differential equations. Moreover, it has been argued that the main challenge of this approach is the limited ability to describe biochemical systems with low species concentrations (Fontana and Buss, 1996). Chemical kinetic models specify the cell with limited structural descriptions. Biological systems are made of collections of objects whose identities are maintained and continuously evolve. These evolving properties may include the activation state, concentration, or the location.

2.2.2 Markov chains

Another method to examine biochemical systems is to express them as Markov chains (Gomez et al., 2001), in which the state of the chain represents either approximations or exact number of the molecules present. Reactions are modelled as transitions between these states. The system is memoryless ("Markovian") since the future development only depends on the present, not on the past. Therefore, the term Markov chain denotes time-discrete systems which are defined as a sequence of random variables $X_1, X_2, X_3, ...$ with the Markov property, i.e., $P(X_{t+1} = x|X_t = x_t, X_{t-1} = x_{t-1}, ..., X_1 = x_1) = P(X_{t+1} = x|X_t = x_t)$.

Provided there is no feedback in the system, the analysis of Markov chains is well developed, and the steady-state probability distribution of the process can be derived. Feedback, which is an inherent feature of many reaction networks, poses problems for analysis since a steady-state distribution of the system does not have to exist in this case.

Many straightforward, yet interesting simulation techniques which utilise the Markov property are based on explicit collisions between randomly selected molecules. This technique has the advantage of being easy to implement in a non-spatial case, and yet simple to extend to spatial simulations. A representative exam-

ple of this type of algorithm is given by StochSim (Le Novère and Shimizu, 2001).

2.2.3 Chemical master equation

Where the model's time is continuous rather than discrete, the Markov chain is replaced by a "continuous-time Markov process". Here, the system again has a finite, discrete set of states, but now a continuous time index t exists. For simplicity, we focus on the case in which each state is given by the number of molecules per molecular species (i.e., a vector $x \in \mathbb{N}^k$). At any given point in time, the system occupies each state with a certain probability, yielding a probability distribution over all the states. The Chemical Master Equation (CME) provides a means to describe the temporal change of this distribution exactly for the case of a well-stirred and homogeneous reactor space (Van Kampen, 2007). Since chemical systems can be considered as Markovian, the CME approach is a special case of the continuous-time Markov chains.

Gillespie (1976); Gillespie et al. (1977) proposed two precise "Stochastic Simulation Algorithms" (SSA) to simulate instances of the random process defined by the CME. These algorithms are widely used in the stochastic simulation of biochemical reactions (Meng, 2004) due to their significant efficiency in terms of computational cost. The principal factors in SSAs are reaction propensities f_μ, i.e., the likelihood of a reaction μ to occur in the next (small) time step dt. These are computed from the mesoscopic rate constants and the number of molecules available as substrates to the reaction. From these, the next reaction and the time for that reaction have to be decided. This is done by using two random numbers. From the CME, it can be shown that the probability density function for reaction μ to occur as the next reaction after time τ is $P(\mu, \tau) = f_\mu exp(-\tau \sum_j f_j)$, which is the basic equation SSAs are built on.

Gillespie's original work has been extended several times, most notably by the "Next Reaction Method" (Gibson and Bruck, 2000). This reduces the complexity

from linear to logarithmic time in the number of reactions. Another technique is given by the "tau-leap methods" (Gillespie, 2001; Chatterjee et al., 2005), which approximates the exact solutions obtained from SSAs. For larger numbers of molecules and reactions, however, these algorithms still suffer from high computational requirements. Bernstein (2005) extended the Gillespie algorithm to reaction-diffusion equations by dividing the reaction volume into several compartments and modelling diffusion between them.

2.2.4 Bayesian networks

A Bayesian network (BN) is a directed acyclic graph commonly used as a probabilistic modelling tool (Pearl, 1988). Modelling chemical networks with BNs was introduced by Sachs et al. (2002). In a BN, variables (a molecular property) are represented as nodes in the graph. Directed edges express the dependence relation between nodes. A variable can be either discrete or continuous and may form a hypothesis, a known value (e.g., a concentration) obtained by experimental measurement or a latent variable. Variables which are not connected by edges are "conditionally independent".

If the state of a variable is known then the state of other variables can be predicted. This is accomplished through the use of:

$$p(x) = \sum_{y} p(x, y) \tag{2.3}$$

This formula sums the probabilities of all routes through the graph, thus allowing one to predict, with some probability distributions, the state of an unknown variable x. Continuous values for probabilities could be specified with a probability density function (e.g., Needham et al., 2006 employs Gaussian distributions).

BNs have been used to reverse-engineer and infer the structure of biochemical networks (Sachs et al., 2002; Kim et al., 2003; Needham et al., 2006). However, the setting of probabilities (learning) of BNs requires static experimental data, oth-

erwise this may result in increasing the complexity of the task (Li and Lu, 2005; Chickering, 1996). The solid foundation of BNs in statistics enables the handling of the stochastic behaviour of real chemical networks and noisy experimental measurements (de Jong, 2002). Another attribute of using BNs is that they can be employed when incomplete or only steady-state data on the reaction network are available. In this common case, kinetic models have been found to be less useful (Woolf et al., 2005). Pe'er (2005) discussed the various techniques to infer BN models from experimental data.

2.2.5 Term rewriting systems

Regulated term rewriting is a basic principle of information processing. Biomolecules, their polymeric subunits or groups of similar biomolecules are interpreted as objects encoded by character strings (terms). Sets of term rewriting rules describe possible interactions among objects and system components (e.g., pathways or membrane structures). Each application of a rule performs a discrete step of a process. The terms as a whole contain all information about the system status. Term rewriting systems can run in a massively parallel manner considering nondeterministic recombinations. Classes of grammar systems, P-systems (Paun and Rozenberg, 2002), broadcast language (Holland, 1975, 1992a) and Alchemy based on the lambda calculus fall into this category (Fontana and Buss, 1994a). We demonstrate this modelling approach with the broadcast language (BL).

Holland originally proposed the BL formalism to assist his research on the "adaptive plan". Holland argued that the BL provides a straightforward representation for a variety of natural models such as biochemical networks.

The BL basic components are called *broadcast units* which are strings formed from the set of "monomers" $\Lambda = \{0,\ 1,\ *,\ :,\ \Diamond,\ \triangledown,\ \blacktriangledown,\ \triangle,\ p,\ '\}$. Molecular species are broadcast units which can be viewed as *condition/action* rules. Whenever a broadcast unit conditional statement (pattern matching expression) is satisfied,

the computational action statement is executed, i.e., when an enzyme broadcast unit detects, in the environment, the presence of one or more specific substrate signal(s) then the broadcast unit broadcasts an output product signal. General signal processing can also be performed with broadcast units: e.g., a broadcast unit may detect a signal I and broadcast a signal I', so that I' is some modification of the signal I. The broadcast monomers/symbols encode for the pattern matching and computational/enzymatic functions of molecular species. In addition, broadcast symbols may act as both operators and operands addressing the reflexive nature of molecular species (i.e., a molecule may act as both an enzyme and/or substrate).

Limited stochastic elements are involved in the computational functions of broadcast units which result in a semi-stochastic behaviour at the system level. The modelling of a genetic regulatory networks (which addressed only the regulatory/qualitative aspects of CIPNs) using the BL was proposed by Decraene et al. (2007b). Although possible, no quantitative studies have been previously reported to have been conducted with the BL prior to the work described in this thesis. Finally the BL formalism does not account for spatial information.

2.2.6 Petri nets

Petri nets (PNs) are a graph-oriented formalism originally from formal software engineering. Developed in the early 1960s (Petri, 1962; Peterson, 1981), Petri nets provide a means to model and analyse systems, which comprise of properties such as concurrency and synchronisation. Petri nets consist of "places", "transitions", and "arcs". "Arcs" are used to connect the "transitions" and "places", "input arcs" connect "places" with "transitions", while "output arcs" start at a "transition" and end at a "place".

The modelling of biochemical networks with Petri nets was introduced by Reddy et al. (1993). Here, place nodes are used to represent molecular species (enzymes, compounds, ions etc.) and transition nodes to denote chemical reactions.

Other elements can be defined to specify in detail the chemical reactions to occur (Pinney et al., 2003).

Ordinary Petri nets provide an accessible modelling tool with well-established analysis techniques. For this reason, the use of Petri nets for qualitative analysis of biochemical network is growing. However, due to their timeless nature, Petri nets are limited regarding dynamic network analysis.

2.2.7 π-calculus

The π-Calculus is a process calculus, which is a formal method for modelling concurrent communicating processes (Hoare, 1983; Milner, 1999). The π-Calculus provides a framework for the representation, simulation, analysis and verification of such systems. The π-calculus allows the application of algebraic reasoning in order to determine the equivalence between processes.

When modelling biochemical networks using π-Calculus, molecules and their individual domains are treated as computational concurrent processes (Regev et al., 2001). Complementary structural and chemical determinants correspond to communication channels. Chemical interactions and subsequent modifications coincide with communication and channel transmission.

The π-Calculus provides a highly detailed description of network nodes. However, the basic π-Calculus gives only a semi-quantitative view. A significant factor to be considered is the lack of an associated temporal dimension and as a result all interactions can occur with the same probability/rate. Extensions of the basic π-calculus address this limitation (Regev and Shapiro, 2004; Blossey et al., 2008).

2.2.8 Agent-based models

In an agent-based model (ABM), several computational objects called *agents* are simulated to reproduce real phenomena within an artificial environment. ABMs originate from the late forties with the development of Cellular Automata

(von Neumann, 1949) and have been extensively used in the following fields: complex systems, multi-agent systems, and evolutionary programming (Luck et al., 2004; Winikoff and Padgham, 2004). An ABM is typically implemented with an object-oriented framework (Rumbaugh et al., 1991; Bersini, 2008). Each agent or class is defined with particular properties and methods. Agents are situated in space and time, interactions between with each other may occur following rules. Global and complex behaviour may emerge from these local agent-agent interactions and properties.

ABMs provide a flexible framework to: specify and refine with ease rules governing agent behaviours and interactions (e.g., using production rules or Boolean logic), secondly, to model emergent system or global behaviours (Ausk et al., 2006). Preliminary works to model bio-chemical networks using ABMs appeared in the late nineties (Schwab and Pienta, 1997; Fisher et al., 1999). ABMs consider the cell and its components as agents with cognitive capabilities. Two distinct ABM approaches are presented:

1. In Cellulat, which was developed by Pérez et al. (2002); Gonzalez et al. (2003), a cell is seen as a collection of adaptive autonomous agents. Communication between agents is performed via propagating signals on a shared data structure, named "blackboard" referring to the blackboard architecture (Nii, 1986a,b). An agent receives a signal or a combination of signals from a designated blackboard level and transduces these into another signal (or set of signals) on the same or different blackboard level. Transduction mechanisms of the signal depend of the cognitive capabilities of the agent. A blackboard level could represent extracellular, membrane, cytosol or nucleus region, this enables the modelling of spatial organisation.

2. A second ABM is described where Learning Classifier Systems (LCS) are used to specify the agents' behaviour and interactions. LCS are systems constructed

from condition-action rules called *classifiers*. LCS can be seen as a simplification of the broadcast language where classifiers are binary strings that can be viewed as IF/THEN statements. Holland's initial work was modified a number of times and at present many different varieties of learning classifier systems are available (Lanzi et al., 2002; Bull and Kovacs, 2005).

LCS are commonly used as a machine learning technique. However Holland (2001) proposed an agent-based model where the agents' behaviour and adaptation are determined by the use of LCS. This work argued that LCS could be used to evolve a simple repertoire of condition-action rules to a more complex goal directed set of rules.

In typical biochemical networks, interactions between molecules follow the same condition-action mechanisms. Thus Holland suggested that this approach could be used to model and simulate CIPNs. His proposition to design chemical networks was to start with a LCS-based "over-general" model of a biological phenomenon (e.g., transformation of a healthy cell to a cancer cell). Then this general phenomenon could be refined through several iterations. At each iteration, the details (e.g., compartment level) of the occurring interactions can be specified. These iterations were continued until the desired network level/granularity was reached, where the submolecular objects are specified (e.g., protein ligand, receptor, ions etc.). This refining process highlights the top-down/hierarchical approach and descriptive power of LCS to model and simulate complex CIPNs. Moreover this approach can be naturally coupled with Genetic Algorithms. This evolutionary feature may allow one to examine phylogenetic relationships between different reaction networks (where the signalling differences may be due to random molecular mutations). However no actual implementation and experimental examination of this system have ever been reported, therefore this proposal and associated potential

benefits remain conjectural.

2.2.9 SBML & CellML

Modelling techniques may be employed in conjunction with a markup language to store generated models. The use of a standard format facilitates the analysis, visualisation, simulation and exchange of biochemical network models within the modelling community, providing opportunities for refinement and incorporation of new knowledge. So far, two approaches have emerged, resulting in the model-description languages SBML (Systems Biology Markup Language) (Hucka et al., 2004) and CellML (Lloyd et al., 2004), both based on the XML markup language (Bray et al., 2000).

- In SBML, a biochemical network is described in terms of the molecules taking part in it - termed species - and the reactions taking place between them. The present amount of each species can be expressed either in terms of its concentration or of the number of molecules present. Each reaction has an associated kinetic law, which defines the rate of the reaction depending on the present amount of its substrates. Additionally, the model can be subdivided into a fixed set of well-stirred compartments to include a non-hierarchical spatial component. Nevertheless SMBL models cannot specify fluxes between compartments at present (i.e., in SBML level 2 version 4 release 1).

- In CellML, a more general approach is taken, in which a model consists of components and connections between components. Each component can contain variables and a reaction between them, and connections are used to transfer the value of variables from one component to another.

Although CellML is following a slightly more general approach, it is not as widely used as SBML, for which a large collection of software tools is available (see www.sbml.org for a list of these tools). Additionally, the first model repositories have started to use SBML as a representation language, e.g., see the BIOMODELS

database at `www.ebi.ac.uk/biomodels`. Therefore, SBML can be seen as the first emerging specification standard for biological models at the cellular level. Finally the use of such a common language provides the ability to analyse and complement intersecting information on differing compatible modelling techniques.

2.3 Comparison of approaches

In this section, we compare the previously introduced methods to model CIPNs by using a set of defined criteria. Following this, a comparison table is presented to summarise this review. The intention is to determine a suitable modelling technique to be employed in our project. This selection will be discussed after the presentation of this comparison table. We identify evaluation criteria with regards to stochasticity, time, granularity, space, topology and modularity.

2.3.1 Evaluation criteria

Relevant criteria are outlined here in order to compare the modelling techniques presented in Section 2.2:

- **Stochasticity**: This property reflects the range of possible processing scenarios that may be identified by the model.
  - *Deterministic*: The system behaviour purely depends on inherent data. No external or statistical fluctuation may occur and influence the system's dynamics. The system may only operate along one known path.
  - *Nondeterministic*: A number of alternative paths for system processing may exist which can be completely explored by the model. All possible scenarios are taken into account by the model in which no unanticipated events may affect the system's dynamics.
  - *Stochastic*: In contrast, stochastic models select one possible path in a random manner that can be based on a given probability distribution.

This implies uncertainty (external and statistical fluctuation may be accounted for) and inhibits repeatability of systems runs.

- **Time**: This property describes how time is represented within the model.
  - *Atemporal*: When executed, the model remains static and introduces no temporal consideration.
  - *Events*: A sequence of pre-identified events defines the granularity of time. An event is an action within the system which characterises the progress of the system processing. Events are not necessarily equidistant in time. Dependencies between processes, their synchronization and concurrency may also be based on the interplay of events.
  - *Discrete*: Temporal changes are characterized by fixed periodic intervals. A discrete time interval defines the smallest unit measuring the system's dynamic behaviour. Discrete time points allow one to express recursive formulation of the system processing. Discrete time may be referred as a global clock for the system.
  - *Continuous*: Infinitesimal time intervals allow the finest granularity for measuring time represented by real numbers. Computer-based simulation techniques, by their nature, require an approximate discretisation of points in time.
- **Granularity**: This property designates how the molecules or particles are represented in the model. It refers to the abstraction level of their specification. The finer the granularity the more detailed the system that can be described. Granularity also constrains the level of monitoring capabilities.
  - *Submolecular*: This level allows one to compose molecules by atomic specifiers or functional units (e.g., protein domains).

  - *Molecular*: Molecules are considered as the smallest expressible unit. A mapping between the chemical substance and the assigned identifier (e.g., symbol) is either assumed or abstracted.
  - *Species*: An enumerable amount of molecules having the same chemical substance is regarded as a species. This level of granularity enables one to quantify a molecular species as a whole within the system, however one cannot isolate an individual molecule of a given species.
  - *Concentration*: Allows one to quantify the relative amount of a particular molecular species existing in a system. As represented by real numbers, transforming absolute molecular amounts into concentrations can require an approximation. Concentrations can be viewed as an approximation of the molecular species quantities.

- **Space**: When handling molecules of given granularity within a model, a system component which is analogous to a reactor is assumed. This component can provide space if the positioning of the molecules (within the reaction system) is taken into consideration.
  - *Implicit*: Particle or molecule identifiers include spatial information, e.g., using an index. System components that control the evolution can be equipped with regulation schemes for updating this information. Here, a homogeneous distribution of the molecules within the reactor is assumed. In this "well-stirred" reactor, no boundaries are specified, and there is no explicit definition of space in the model.
  - *Compartmental*: A hierarchically nested or graph-based number of explicit compartments is distinguished. Each molecule is assigned to one of the specified compartments and can move from one compartment to another. Within each compartment, no further specification of molecular positioning is defined.

  - *Grid*: Apart from the compartmental structure, a spatial geometry is used to locate molecules more precisely. This way, discrete spatial distributions of molecules can be mapped using the model.
  - *Continuous*: The finest granularity of defining space is given by continuous values. Here, each molecule can be positioned arbitrarily within the reactor. Analogous to continuous time, computer-based simulations may require discretisation which would imply approximation.

- **Topology**: This designates the ability of the model to dynamically modify its structural components (e.g., pathway structure, dependencies between compartments, active membranes, receptor dynamics).
  - *Fixed*: A static system structure is assumed.
  - *Dynamic*: Principles or rules are defined that allow the system structure to change over time and space. These rules are a part of the model description.

- **Modularity**: This refers to the ability of the model to subdivide a given biological reaction system into functional sub-units (i.e., modules). The subdivision process is carried out through algorithmic strategies applied on the model. Modules are determined/classified according to specific properties (e.g., network topology/clusters, functions) across these sub-units. Modularity may facilitate the study of a system by examining sub-units independently instead of the system as a whole.
  - *No*: The whole reaction system is regarded as a monolithic entity which currently prevents the identification of sub-units.
  - *Hierarchical structure*: The sub-units are represented as nodes forming a tree-based structure. Modules communicate with each others (e.g.,

transmission of molecules from one sub-unit to another) via specified interfaces, typically through diffusion over transduction/communication channels.

- *Graph-based structure*: These structures are a generalisation of tree-based structures which does not necessarily account for a hierarchical organisation.

2.3.2 Comparison table and discussion

As a summary of previous sections, a comparison table is presented (Table 2.1) which uses the criteria that were discussed above. The table provides an immediate comparison of differing modelling techniques and allows one to identify desirable attributes which may be necessary for modelling a specific biochemical system.

	Differential equations (ODE)	Differential equations (PDE)	Markov chains	Master equation (CME)	Bayesian networks	Term Rewriting Systems (BL)	Petri nets	π-calculus	Cellulat	Agent-based LCS
Model type	deterministic	deterministic	stochastic	stochastic	probabilistic	algebraic	algebraic	algebraic	agent-based	agent-based
Time	cont.	cont.	discrete	cont.	atemporal	discrete	events	events	discrete	discrete
Granularity	conc.	conc.	species.	species	species	(sub)molecular	molecular	molecular	(sub)molecular	(sub)molecular
Space	implicit	cont.	implicit	implicit	implicit	implicit	implicit	implicit	compartmental	implicit
Topology	fixed	fixed	fixed	fixed	fixed	dynamic	fixed	fixed	dynamic	dynamic
Modularity	no	no	graph-b.	graph-b.	graph-b.	graph-b.	graph-b.	graph-b.	hierarch. or graph-based	hierarch.

Table 2.1: Comparison of CIPN modelling approaches with respect to previously defined classification scheme

Based on our review of modelling techniques which is summarised in Table 2.1, we discuss and relate this evaluation with the requirements of this project identified as follows:

- The specification and development of a software platform capable of modelling and simulating artificial CIPNs.
- The examination/traceability of the individual behaviour of molecular species/instances.
- The possibility to conduct both qualitative and quantitative analysis of the CIPN system's dynamics.
- The inclusion of stochastic processes reflecting the random nature of molecular collisions/reactions.
- The specification of submolecular properties (e.g., monomers).
- The ability to dynamically change the CIPN's topology over time (e.g., creation/modification or removal of molecular reactions due to molecular mutations).

Probabilistic models are employed to infer the model from experimental observations which conflicts with our attempt to construct a simulation platform. Examining organisational closure in CIPNs implies that the monitoring/tracing of the individual molecules is required. However deterministic and stochastic techniques treat the molecular species in an aggregate manner. As a result algebraic and agent-based approaches are more appropriate for the current endeavour.

On the other hand, the main drawback of algebraic and agent-based approaches is the fact that they may not always permit a detailed (i.e., both qualitatively and quantitatively) analysis of biochemical processes' dynamics. The latter is a key feature of deterministic and stochastic approaches. Moreover, these mathematically

grounded approaches have a strong theoretical scientific foundation associated with a plethora of analytical and simulation tools available.

Nevertheless we may also consider that we intend to realise and evolve *artificial* CIPNs. In other words, we do not intend to provide a simulation platform of *real* biochemical networks, in which case a precise account for the temporal dimension would have been necessary. Thus the semi-quantitative and discrete approach of algebraic and agent-based techniques is sufficient. Moreover, exporting algebraic/agent-based models into the SBML format allows us to conduct further analyses of the systems' dynamics using complementary SBML tools. In Chapter 7, we employ the SBML to generate and analyse the deterministic dynamics of reaction networks employed and generated in our evolutionary experiments.

Finally agent-based approaches offer most flexibility: ABMs provide a detailed and adaptable description of molecular species (including submolecular components) and of the system as a whole. Secondly it is possible to utilise existing algebraic methods (and exploit associated features) to specify molecular species and reactions within an ABM. Most ABMs account for the dynamic nature of biochemical systems in which the relationships between molecular species/modules may change over time due to internal or external perturbations (e.g., molecular mutations). Stochastic elements may also be involved in the specification of the agents' interactions with other species/systems' elements.

According to the requirements of this project, ABMs appear to represent a suitable and flexible technique to model, simulate and analyse CIPNs and their evolution.

2.4 Conclusion

We introduced the overall concepts associated with the main CIPN modelling branches: deterministic, stochastic, probabilistic, algebraic and agent-based approaches. To best illustrate each of these classes of modelling techniques, we pre-

sented and reviewed a selection of techniques.

We also identified two markup languages: SBML and CellML which are widely used to specify, disseminate and exchange CIPN models within the scientific community. These languages also permit the migration of specifications between differing modelling techniques. This migration capability allows one to widen the range of analytical studies of a given modelling technique. We noted that the SBML possesses a longer history than CellML and has subsequently become the standard language for storing CIPN models. We will therefore employ the SBML as a means to migrate and disseminate our generated CIPN models.

Following this, a model comparison table was presented and highlighted the modelling capabilities of the different techniques. We outlined the requirements of this project with regards to the modelling of CIPNs. We related and discussed these requirements with the review conducted in this chapter. We finally distinguished ABM approaches as a suitable and flexible modelling technique. We consequently select the ABM framework to model, simulate and analyse CIPNs in this thesis.

Chapter 3

Evolving Cellular Information Processing Networks

In the previous chapter we reviewed several computational techniques for modelling chemical reaction networks. We identified agent-based approaches as suitable methods to model, analyse and simulate Cellular Information Processing Networks (CIPNs) according to the requirements of our project. Following on from this, we now present a literature review which focuses on the evolution of CIPNs *in silico*. This review is composed of two main parts. We first examine Evolutionary Computation (EC) techniques applied to the artificial evolution of reaction networks. Secondly we present Artificial Chemistry based approaches which address the emergence and evolution of collectively autocatalytic (organisationally closed) reaction networks.

This chapter is tightly coupled with Chapter 2 as the evolutionary methods presented here rely on modelling frameworks presented earlier. Based on both examinations conducted in the previous and current chapter, we intend to identify an adequate evolutionary simulation framework to examine closure and the evolution of artificial CIPNs.

3.1 Introduction

At present no previous work *directly* related to closure and the evolution of CIPNs has been reported in the literature. Nevertheless we distinguish two related fields

of study: Evolutionary Computation (EC) (Fogel, 2007) and Artificial Chemistry (AC) based studies (Dittrich et al., 2001). Both EC and AC models address complementary aspects of our research:

- *The EC approach*: Several investigations have been conducted on the evolution of artificial biochemical networks capable of performing computational tasks. This area of research focuses on the development of novel computational paradigms. EC techniques (e.g., Genetic Algorithms, Genetic Programming) are commonly utilised to evolve models of simplified reaction networks. The objective fitness function of these approaches is typically to mirror a computational function. Within a given reaction network, distinct arbitrary molecular species are designated as input and output signals. Properties of the reaction network (e.g., topology, kinetic rates, etc.) are subjected to mutations which ultimately affect the signal-processing and response level of output signal(s). Here the reaction network's computational function is modelled as a traditional information processing system involving an input-process-output relationship. The *explicit* definition of the fitness function directs the evolutionary process of the system, which we regard as a top-down evolutionary approach. No EC-based approaches have addressed closure properties in reaction networks nor the "cellular" nature of CIPNs: These investigations aim at evolving reaction networks without any consideration about the container/reactor space.

- *The AC approach*: Another distinctive approach to the evolution of biochemical networks is addressed through the use of Artificial Chemistries (ACs). In contrast to the computation-oriented EC approaches, AC-based research focuses on the self-organisation, emergence and evolution of molecular organisations. ACs are an abstraction of chemical organisations which attempt to understand and engineer the conditions for life (as it could be) and open-ended evolutionary growth of complexity. No explicit fitness function is devised in

ACs, the system's dynamics are driven by the intrinsic properties of the molecular species (e.g., enzymatic function, binding specificity). These properties may provide the system with the ability to self-sustain/maintain. These abilities are the *implicit* fitness function driving an AC's (evolutionary) dynamics. Moreover, typical ACs aim at evolving minimalist molecular organisations towards networks of higher complexity where novel phenomena may occur at the system level. We thus regard ACs as a bottom-up evolutionary approach. Limited AC-based researches (Tominaga et al., 2007) have been conducted towards implementing molecular computing devices compared to the EC counterpart. The computational function of ACs is modelled as a perturbation of the closed system's dynamics, without any designated input/output signals. Nevertheless, ACs have intensively focused on organisational closure, which according to our hypothesis, may be a property of CIPNs.

We present a selection of evolutionary systems which illustrate both top-down and bottom-up approaches.

3.2 Evolutionary Computation

Evolutionary Computation techniques are non-deterministic search algorithms inspired by neo-Darwinian principles. Within the EC field, three major families of computational techniques may be identified: Genetic Algorithms (GA, Holland 1992a), Genetic Programming (GP, Cramer 1985) and Evolution Strategy (ES, Rechenberg 1973). These techniques differ on the specification and implementation of common system properties: representation, recombination, mutation and selection. These methods have been successfully applied to a wide range of optimisation problems. Such techniques have been combined and employed in evolutionary systems applied to the evolution of biochemical networks *in silico*. We review a selection of three distinct EC-based systems.

3.2.1 Bray and Lay

Bray and Lay (1994) proposed the first significant investigation on evolving reaction networks through EC. A simple computational and idealised signalling pathway model (Fig. 3.1) was proposed to investigate the evolution of biomolecular receptors found in real cells. This model includes two cellular receptors R_1 and R_2, an extracellular ligand L (the input signal), an intracellular target molecular species T and a phosphorylated target species Tp (the output signal).

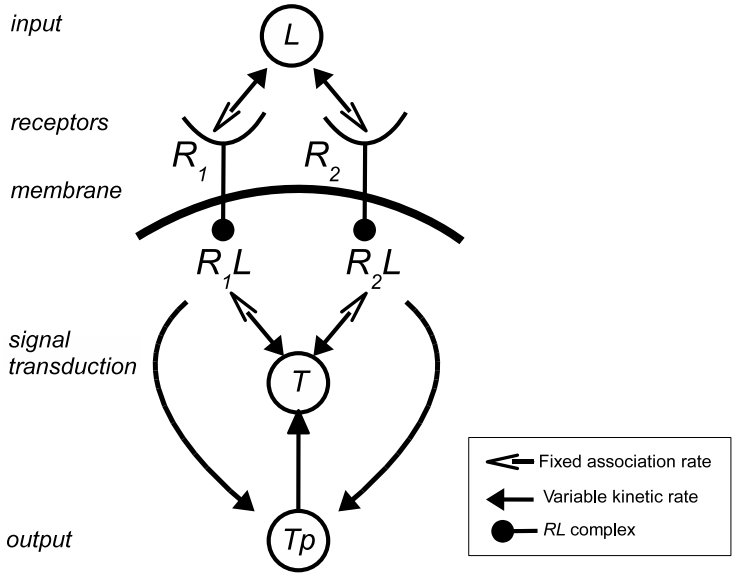

Figure 3.1: Artificial signalling pathway model adapted from Bray and Lay (1994).

Seven chemical reactions could occur between these molecular species and are

governed by mass-action kinetics:

$$R_1 + L \Leftrightarrow R_1L \tag{3.1}$$

$$R_1L + T \Leftrightarrow R_1LT \tag{3.2}$$

$$R_1LT \Rightarrow R_1L + Tp \tag{3.3}$$

$$R_2 + L \Leftrightarrow R_2L \tag{3.4}$$

$$R_2L + T \Leftrightarrow R_2LT \tag{3.5}$$

$$R_2LT \Rightarrow R_2L + Tp \tag{3.6}$$

$$Tp \Rightarrow T \tag{3.7}$$

The model was trained to exhibit a specified response to pulses of stimulus species L. The topology of the reaction network was fixed, i.e., no creation or removal of reactions could occur. An ES approach was employed in which mutations were applied upon the seven variable reaction rates. This mutation operator was utilised to generate offspring of the unique, initial seed reaction network (where initial reaction rates are equal and set arbitrarily). Selection was conducted as follows:

- The numerical integration of the differential equations (i.e., the network's genotype) generated from the above reactions was carried-out.
- The results of this integration (i.e., the network's phenotype) provided the time course and concentration level of output signal. This information was used to determine the network's fitness value.
- The fittest offspring network (i.e., which best exhibited the specified and desired behaviour) was selected and used for the next iterative generation.

Using the above ES algorithm, Bray and Lay successfully trained this model to exhibit a specified response, see Fig. 3.2 .

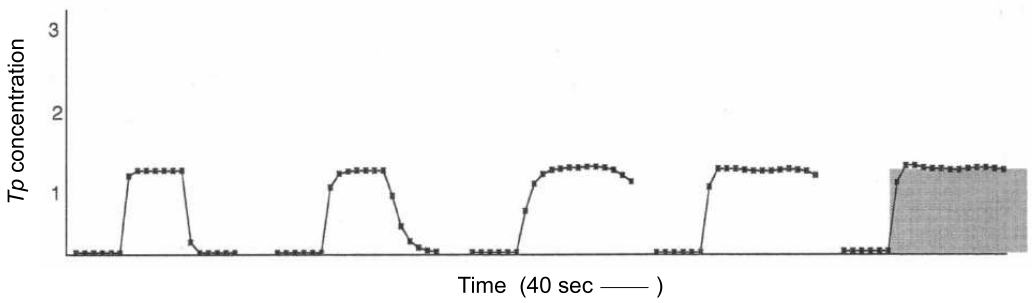

Figure 3.2: Results of an example simulation, taken from Bray and Lay (1994), in which the network was trained to exhibit a specified time course and concentration level of Tp (this target response is here represented by the shaded gray area). Each individual curve depicts the concentration of Tp produced in response to a squarewave pulse of stimulus L, applied for 40 seconds. The response of the network evolved from a square wave pulse that matches the stimulus level, to one that remains elevated for a longer period after the stimulus was applied, thus matching the pre-specified target response. This evolved reaction network behaves as an "ON switch".

In this approach, mutation and structural effects were limited, only the kinetic parameters were subjected to variations. The evolutionary process led to the adaptation of the given model to mirror a specified behaviour. Bray and Lay demonstrated that under an evolutionary regime, it was possible to evolve an artificial reaction network to perform a simple computational task (i.e., a switch). However it was also reported that the current model could not be evolved to perform other computational functions such as a sigmoidal or first derivative function. Following this work, Bray (1995) later discussed the analogy of molecular species as computational units occurring in living cells. Such *in-silico* evolutionary experiments, aiming at realising computational functions in chemical networks, have to date not been reported to be conducted/validated in wet lab conditions.

3.2.2 Lakhesis

The computational capabilities of Bray and Lay's approach were limited with regards to performing mathematical operations. However this work was extended and involved more elaborated EC techniques. Deckard and Sauro (2004) proposed the

Lakhesis system as a modelling and evolutionary simulation platform. The motivation to this work was to investigate computational properties and capabilities of biochemical reaction networks.

Similarly to Bray and Lay's system, chemical reactions are governed by mass-action kinetics and are modelled as differential equations. Input and output molecular species were identified and fixed within reaction networks modelled with Lakhesis.

However, a number of EC related specifications differ between Lakhesis and Bray and Lay's system:

- The initial population contains a multitude of randomly generated reaction networks (where the initial number of molecular species and associated reactions are set at random). In each of these networks, input and output signal(s) are designated and fixed.

- The number of molecular species n may vary throughout the initial reaction networks. However within a given reaction network, this number is fixed and may not vary through mutational changes.

- The number of chemical reactions r within a reaction network is subjected to variation. However a reaction network may not contain more than r_{max} reactions.

- The networks' fitness values are obtained by computing the steady state solution of the systems' set of ODEs. Networks that exhibit steady states and the least deviant behaviour (according to the objective function) are the fittest.

- During each iteration, the subset containing the fittest network candidates is selected and remains within the population for the next iteration. The unselected reaction networks are removed from the population.

- Mutation operators are devised and introduce variations upon the network's topology *and* reaction rates, see Fig. 3.3.

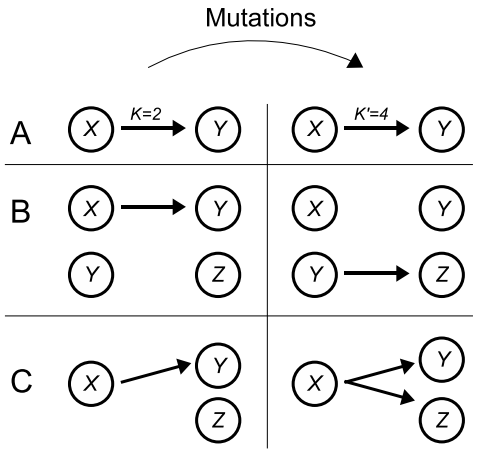

Figure 3.3: Mutation operators in Lakhesis adapted from Deckard and Sauro (2004). A: Change of reaction rate, B: addition/removal of reaction, C: new reaction type. These mutations occur with different probabilities, A mutations occurring the most often as their effects are less disruptive than B and C mutations.

The above EC approach extends Bray and Lay's work by introducing richer mutation operators. As a result, a greater variety of reaction networks could be evaluated during the evolutionary process.

Using this system, Deckard and Sauro successfully evolved reaction networks to perform a range of mathematical functions such as: multiplication by constants, square roots, cube roots and natural logarithms. Such computations could not be evolved using Bray and Lay's approach which appeared to be limited due to the restricted number of signalling molecular species and lack of feedback loops (Bray and Lay, 1994).

Moreover, in particular reaction networks capable of performing the square root function, a subpart or *module* of the network was identified as being able to solve quadratic equations. Such a modular structure is of interest as it is a common feature of CIPNs. This work may thus provide further insights on the evolution of modularity in CIPNs.

Deckard and Sauro's approach was recently adopted and developed by Lenser et al. (2007). A multi-level Evolutionary Algorithm (EA) was proposed to

conduct similar studies on the evolution of reaction networks capable of signal-processing functions.

3.2.3 Learning Classifier Systems

In previous top-down evolutionary approaches, molecular species were treated in an aggregate manner at the system level. The internal structure and intrinsic properties of molecular species were not specified and did not affect the reaction network's topology nor dynamics.

We present an alternative top-down approach in which the individual behaviour of molecules is considered. As previously presented in Section 2.2.8, Holland (2001) proposed an agent-based model where the agents' behaviour and adaptation are determined through the use of Learning Classifier Systems. Holland suggested this ABM as a suitable modelling framework for the study of CIPNs. The modelling of CIPNs was conducted in a top-down fashion, where an iterative refinement process was employed to specify the different levels of CIPN interactions. No implementation of this system was performed, nevertheless Holland proposed a toy model to illustrate this evolutionary approach. This toy model provided an existence proof that his system could be used to evolve a simple repertoire of condition-action rules to a more complex goal directed set of rules. We describe the application of this ABM approach to evolve CIPNs.

Cells are autonomous agents which possess five principle components:

1. *A set of detectors*: Cellular agents are situated in space and may probe their surrounding environment to detect input messages (e.g., nutrients, hormone molecules, toxic molecules). This sensor apparatus (i.e., cellular receptor) constrains the agent to only interact with its local environment. Detected messages are stored in the list of messages.

2. *A list of messages*: Messages (i.e., input signals or stimuli molecules) are received from the environment as binary encoded data. Input signals are then

stored in an internal data structure termed the *message list*.

3. *A list of classifiers*: Enzymatic operations are addressed with classifiers which are strings formed over the ternary alphabet {0,1,#}. Classifiers are condition/action rules. The condition statement refers to the "binding" conditions where input messages are evaluated. This binding condition (a pattern matching expression) determines the network's topology, i.e., the reactions/-connections between the molecular species. The action statement describes the enzymatic/computational function which results in the production of an output signal (i.e., intra-cellular signalling molecule). The latter may interact with the agent's effectors (see below). When the character # occurs in the condition part of a classifier, # acts as a single character wildcard which allows for the potential matching of a greater number of input strings, e.g., the string 10# can match both inputs 100 or 101. When occurring in the action part of a classifier, # may also copy the matched character into the output signal, e.g., let us consider the rule IF 1#0 THEN 00# and the input 110, the condition 1#0 is satisfied by 110, the # occurring in the action part 00# is here equal to the third character of 110, as a result the product molecule 000 is generated. This character may thus provide for string/signal processing capabilities.

4. *A set of effectors*: Similarly to classifiers, effectors are conditional rules. When effectors (e.g., flagella) are satisfied by output messages generated by classifiers, an action causing some changes to the environment is performed (e.g., move, produce inter-cellular signalling molecules).

5. *A set of reservoirs*: When a classifier evaluates a binding input signal, an appropriate response is determined. This response is indicated by the classifier's action expression. If this action exhibits a specified desired behaviour then a reward mechanism is implemented through the use of a scalar reinforcement algorithm. Each classifier has an associated fitness measure, quantifying the

usefulness of the rule in attracting external reward. The rewards are resources which fill the reservoir. These resources, which deplete over time, are needed by the cell to persist in its environment.

Fig. 3.4 provides a schematic view of Holland's system where a cellular agent is depicted.

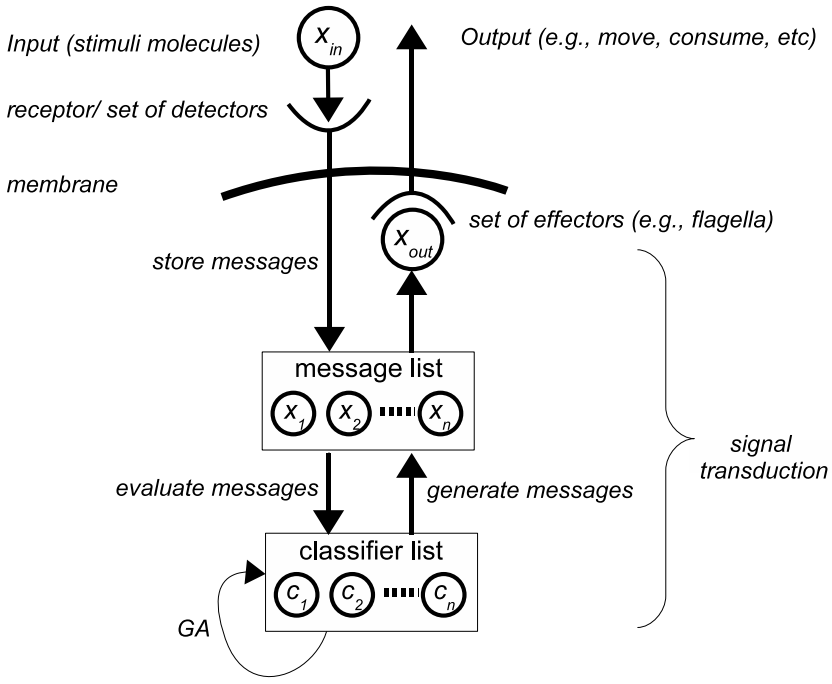

Figure 3.4: Schematic of Holland's agent-based Learning Classifier System

A credit assignment algorithm (such as the bucket brigade algorithm) is employed to reward and strengthen efficient classifiers/rules. Moreover a rule discovery mechanism is specified which is responsible for generating *potentially* more efficient rules. GAs are employed to generate new classifiers where selection is carried out upon the classifiers' fitness measure. Both algorithms are pre-specified and do not evolve. Similarly to Bray and Deckard's approach, these explicit fitness functions are used to drive the evolution of the agent. This ABM is therefore a top-down evolutionary approach.

Although Holland addressed the inner structure and behaviour of individual

molecules through the use of messages and classifiers, Holland distinguished a demarcation between substrate/signalling molecules (messages) and enzymatic molecules (classifiers). However, molecular species are *reflexive* by nature and may act as both substrates and enzymes.

Nevertheless, in contrast to Bray and Deckard, Holland addressed certain distinctive features of CIPNs:

- CIPNs are reaction networks capable of signal-processing which are contained within a cellular structure. The cellular membrane marks the distinction between the environment and the intra-cellular milieu in which CIPNs occur.

- The network's topology is determined by the intrinsic properties of molecular species, here modelled as pattern matching expressions.

- The hierarchical organisation of the cell/CIPN couple is considered. CIPNs do not directly interact with the environment but only operate within the cell. Whereas the cell may interact with the environment through the use of detectors (e.g., membrane receptors) and effectors (e.g., flagella). Two levels of interaction, i.e., molecular and cellular are identified which characterise CIPNs/cells.

Holland's proposal complements Bray and Deckard's approaches by introducing characteristic features of CIPNs. However some aspects of CIPNs are still not considered in this approach, e.g., reflexive nature of molecular species, cellular division and a potential mechanism enabling the (self)-maintenance of CIPNs.

3.3 Artificial Chemistries

The above EC-based research did not address closure dynamics in reaction networks. We present Artificial Chemistry based investigations which consider both closure dynamics and individual behaviour of molecular species. Artificial Chemistries (AC)

are an abstraction of real chemical processes which aim at understanding the dynamics of complex molecular organisations (Dittrich et al., 2001). An AC typically consists of a set of computational "molecules" and of a set of rules. The rules describe the reactions that may occur between the simulated molecules. The rules are applied according to an AC-specific algorithm which also characterises the reaction space. Artificial reactions lead to the production, modification or destruction of molecules.

During an AC simulation, several phenomena of interest may arise such as the emergence and evolution of closed biochemical organisations. Although most ACs have been employed to investigate artificial/simplified models of chemical reaction networks, some AC systems have been specifically devised to study chemical reaction networks from a more biologically realistic perspective (Lenaerts and Bersini, 2009; Tominaga et al., 2009).

Nevertheless there is, to our knowledge, no ACs which were specifically developed to examine closure and the evolution of CIPNs. We review a number of selected and related ACs in which the spontaneous emergence and evolution of organisationally closed reaction networks were examined.

3.3.1 Alchemy

Fontana and Buss (1994a,b) developed Alchemy to study the emergence of self-maintaining organisations in biochemical systems. Alchemy employs the λ-calculus (Church, 1932) formalism to specify molecular species and reactions.

Molecules are specified as λ-expressions which upon reacting with each other may generate product molecules. Reactions occur in a well-stirred flow reactor where molecules may collide with each other at random. A collision between the molecules A and B generates a function $A(B)$ which is then subjected to a reduction process. The latter produces the normal form C of $A(B)$, C is the product molecule. The reduction process involves a number of necessary reduction steps. If the normal

form is not obtained after k reduction steps then the reaction is elastic, i.e, no product molecules are generated. This reduction to normal form determines the possible reactions between the molecular species. Therefore, as in LCS, the intrinsic properties of molecular species identify the reaction network's topology.

Elastic reactions may also occur when the following condition is satisfied: Let us consider the molecular species X and B, if the normal form $X(B)$ is directly obtained without any reduction steps, then the reaction $X(B)$ is elastic. In other words, molecular species which are not capable of enzymatic/computational transformations are not allowed in Alchemy. This filter applies when molecules are randomly generated in Alchemy simulations.

As no mutations may occur in Alchemy, the molecular diversity is determined by the initial randomly generated molecular population and subsequent catalytic reactions that may occur between the molecules.

Moreover, in contrast to EC-based systems, no explicit fitness function is defined. The dynamics of the system are only driven by the individual behaviour of the molecular species.

A series of experiments was conducted in which different forms and levels of organisation could be distinguished:

1. *Level 0 organisations*: At the simplest level, the system is initialised with random and unique molecular species. When executed, this system quasi-deterministically converges to a state where a single autocatalytic molecular species dominates the whole population. These molecules can self-replicate when colliding with a copy of themselves, i.e., if we consider two instances (molecules) A_1 and A_2 of species A, we have $A_1(A_2) = A_3$. This phenomenology where molecular species can replicate themselves is referred to as *level 0* ($L0$). The self-replicase species are called $L0$-objects. An example $L0$-object is $\lambda x.x$. $L0$-organisations have the form of hypercycles which were introduced by Eigen (1971); Eigen and Schuster (1977). A three-element hypercycle is

presented in Fig. 3.5. Under perturbation/mutation effects (e.g., addition/removal of molecules), such organisations are known to be *fragile* and collapse to a single replicase molecular species (i.e., a single-element hypercycle).

2. *Level 1 organisations*: At the next level, an additional filter is defined to prevent self-replication reactions (where the product molecule is syntactically identical to the substrate and/or enzyme molecules) from occurring. When this filter is applied in Alchemy simulations, the emergence of a novel form of organisation can be observed: a collectively autocatalyctic set of molecular species. This form of organisation is referred to as an $L1$-organisation. In this type of reaction network, a distinct molecular species is not capable of self-replicating, but is capable of catalysing the production of another species. A closed cycle of complementary productions enables the maintenance of each molecular species present in the reaction network. Each molecular species is thus necessary for the production of another species and ultimately responsible for the maintenance of this *virtuous* self-maintaining cycle. An example $L1$-organisation is depicted in Fig. 3.5. In contrast to $L0$-organisations, it has been reported that $L1$ sets are relatively more robust to perturbations, having the ability to self-repair. $L1$-organisations are examples of autocatalytic sets (Kauffman, 1993), i.e., $L1$-sets are no longer hypercycles but collectively autocatalytic (organisationally closed) reaction networks.

3. *Level 2 organisations*: In the last series of experiments, the system is seeded with two distinct $L1$ organisations (obtained from previous independent $L1$ experiments) in which two phenomena may be observed:

    - If no molecular species may interact between both $L1$ sets then one of the two $L1$ organisations displaces the other one.
    - If some molecular interactions occur between both $L1$ sets then a *level 2* organisation emerges. An $L2$ set is a metaorganisation which contains

both $L1$ organisations as subnetworks (i.e., modules). In addition a set of molecular species is generated as a result of the molecular interactions occurring between the $L1$ sets. These molecular species do not belong to the closed cycles of both $L1$ sets. Moreover, these species do not form a closed cycle and thus cannot self-maintain. However the metaorganisation containing both $L1$ sets and these extra molecular species is closed. The $L2$ set is therefore able to maintain *all* molecular species present in the network. These molecular species, occurring outside the $L1$ sets, enabled the stabilisation and integration of both $L1$ sets into a higher order $L2$ organisation.

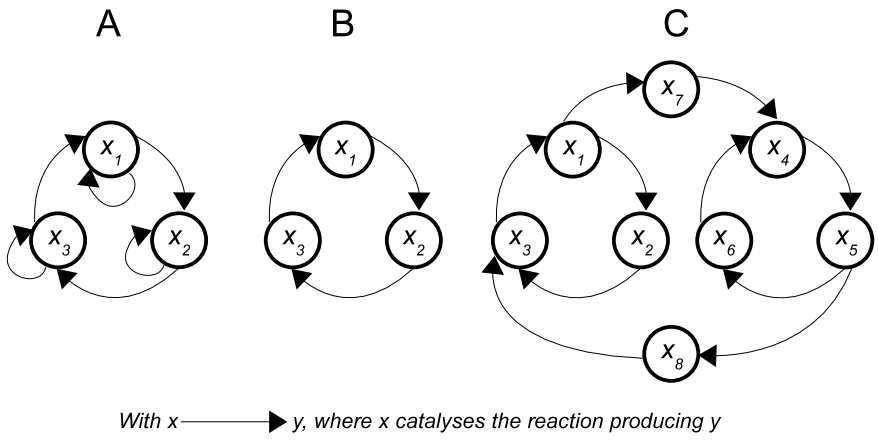

Figure 3.5: A: Level 0 organisation (a three-element hypercycle). B: Level 1 organisation. C: Level 2 organisation containing two Level 1 organisations

The above phenomena resulted only from self-organisation dynamics. Contrary to EC-based research, this work did not attempt to perform or investigate computation in reaction networks. Nevertheless some CIPN aspects can be identified in this investigation:

- The $L2$ experiments exhibited reaction networks where a modular structure can be observed. Modules (i.e., $L1$ sets) can be identified and are contained within a higher structure or metamodule. This hierarchical organisation is

characteristic of CIPNs where each module may possess a distinctive function.

- The signalling pathway formed by the"glue" molecular species (which resulted from the molecular interactions occurring between the $L1$ sets) may be acting as a *control* or *regulation* function. These molecular species prevent the $L1$ sets from displacing each other by regulating the reactions occurring in each $L1$ set. For example let us consider the $L2$ set C, A and B are $L1$ sets, and G the set of glue molecules so that $A \cup B \cup G = C$. If A grow faster than B then this has an affect on G which would subsequently grows faster. However if G grows faster then B would also benefit from this growth. As a result, A and B are stabilised.

In this approach, only catalytic molecular species are allowed. Fontana and Buss reported that if molecular species acting only as substrate molecules are not filtered out then no spontaneous emergence of closed organisations could be obtained. When allowed, an accumulation of such inert molecular species is observed. These molecular species being inert, cannot contribute to the production of other molecular species. Fontana and Buss argued that this accumulation of "waste" molecular species disrupts the metabolic processes and prevents self-maintaining metabolic cycles from emerging.

Although the Alchemy system presented significant results related to closure in reaction networks. The introduction of non-catalytic molecular species prevents autocatalytic closure from emerging. However CIPNs typically involve such molecular species during signal transduction. Finally no random variations at the molecular level were devised in Alchemy. Fontana and Buss's work did not address the evolution but the emergence of self-maintaining organisations in initially unorganised reaction networks. The Alchemy approach may thus not be directly applicable to the study of closure and the evolution of CIPNs.

3.3.2 α-universes

α-universes are a class of AC which were proposed by Holland (1976) to investigate the spontaneous emergence of life. Holland aimed at demonstrating that emergence of self-replicating systems was possible given an initial unorganised system state. This investigation was addressed through a quantitative analysis of the emergence time of autocatalytic systems/reaction networks. In this investigation, a specific α-universe was described by Holland and is presented in the remainder of this section.

The model is specified as a one dimensional and circular linear string of "nodes". The latter may be empty or contain an "atom" from the set $\{0, 1, :, N_0, N_1\}$. This instance of α-universe is a mass conservation model, i.e., the number of existing atoms is fixed, atoms cannot be created or removed, they are simply rearranged in space.

Adjacent atoms may be bonded to constitute structures (i.e., molecules) which are separated from each other by empty nodes. Each atom is associated with a level of bonding (i.e., weak or strong) to the adjacent atom, if any, at its right. A bond between an atom and an empty node is weak by convention.

Two classes of operator, respectively called "primitive" and "emergent", are distinguished and operate stochastically:

1. Two primitive operators are devised allowing the spontaneous stochastic recombination of the molecules' structures: the `EXCHANGE` operator produces diffusion-like movements of atoms whereas the `BOND-MODIFICATION` operator causes random changes in the strength of bonding between atoms. These operators are context-insensitive and allow for random variations to occur at the molecular level (i.e., acting like mutation operators). A feature which was not devised in Alchemy.

2. Two emergent operators are defined and are sequences over $\{0, 1, :\}$. The latter are regarded as functional elements whereas $\{N_0, N_1\}$ are viewed as

"nucleotides". The colon symbol is employed to separate the sequences into the operator's arguments. The latter designate the type of the operator (i.e., `COPY` or `DECODE`) and the prefix of the "substrate molecule". These operators are context-sensitive, i.e., these operators/functional molecules "emerge" according to the specific arrangement of atoms in the sequences. The `DECODE` operator translates sequences over $\{N_0, N_1\}$ into sequences in the alphabet $\{0, 1, :\}$ through the use of a mapping table. The `COPY` produces copies of sequences over $\{N_0, N_1\}$.

A distinction is thus made between molecules that can only act as substrates and those which possess functional capabilities. Non-functional molecules were not present in Alchemy as it would have prevented the emergence of self-maintaining organisations (Section 3.3.1).

The molecular computational operations can be viewed as a simplification of Learning Classifier Systems (i.e., production rules). As in the agent-based LCS approach, the network's topology is determined by the binding/pattern matching conditions of the functional molecules. However the collision rule differs and accounts for space (favouring collisions between molecules that are nearest to each other). A reaction occurs if the functional molecule's binding condition (as devised in the molecule/operator's arguments) is satisfied by the structure of the substrate molecule.

As proposed, this system does not allow for the implementation of individually autocatalytic/self-replicating molecular species. However Holland argued that an infinite class of collectively autocatalytic reaction networks could be identified in α-universes. Although no experiments were conducted, Holland presented a theoretical analysis of the expected emergence time of such closed reaction networks.

Actual experiments (McMullin, 1992) demonstrated that closed reaction networks could not self-maintain due to side reactions not predicted by Holland. It was

also observed that necessary "food molecules" (i.e., nutrients) would rapidly deplete and prevent any further reactions from occurring. It has however been suggested that under flow conditions (providing an inflow of nutrients), it could be envisaged to obtain such closed reaction networks to self-maintain (McMullin, 2000).

3.3.3 Tierra

Alchemy and α-universes focused on the spontaneous emergence of organisations given an initial unorganised system. In contrast to these approaches, Ray (1991) did not focus on prebiotic conditions but on the requirements allowing for the spontaneous evolutionary process of increasing diversity and complexity of organisms. To assist this research, Ray proposed Tierra and intended to explore open-ended evolution using this system.

In the Tierra metaphor, a virtual computer is employed to represent the universe. The computer memory designates the universe's one-dimensional space. Within the virtual computer memory, we may distinguish patterns of computer codes which identify the digital organisms/agents. The latter are represented as concurrent computer programs which compete with each other for CPU time (energy) and access to memory (nutrients). Interactions between these computer processes may lead to the creation, modification or destruction of processes. These evolvable digital organisms/processes can be regarded as molecules which may interact/react with each other leading to the production/destruction of molecules/computer processes. These reactions occur under flow conditions. The molecular computational processes also follow the virtual computer metaphor and consist of computer instruction operators (32 distinct operators/atomic elements are devised). The agents are evolvable in the sense that they can be modified through the use of stochastic operators (e.g. error-prone replications, spontaneous mutations).

Parallelism is simulated in the Tierra virtual computer where the number of CPUs is equal to the number of existing agents. Each of these CPUs execute a

small time slice in turn as devised by the time-slicing algorithm. CPUs perpetually carry out a "fetch, decode, execute, increment instruction pointer" cycle where the instruction pointer indicates a memory address in the virtual computer memory.

When an agent is interpreted for computation, specific operators may be employed to point at specific regions of the computer memory. In particular cases these pointers may jump to memory areas where another agent (or raw material) is already present. In such cases, interactions (i.e., chemical reactions) between the enzyme and substrate species may occur. These instruction pointer mechanisms identify the reactions that may occur between species, and therefore determine the reaction network's topology.

In the Tierra system, it was necessary to construct and feed the system with at least a single autocatalytic/self-replicator molecule called the "ancestor". This introduction of an ancestor species was necessary as no spontaneous emergence of such species could be observed (a variant of the Tierra system called Amoeba was specifically developed by Pargellis, 2001 to address this issue).

The reaction space's size is fixed and may contain up to 60000 instructions. When 80% of the space if filled up, then a reaper is activated and kills molecules (by deallocating their memory) in a particular order specified in the reaper's queue. This mechanism introduces mortality and prevents the saturation of the reaction space, in which case no further reactions could occur.

Example evolutionary dynamics observed when a single ancestor molecule (which contains 80 instructions) is inserted in this reaction space are as follows:

- The self-replicator species would rapidly fill up the reaction space. Then through mutations, mutants of the self-replicator species would emerge and were potentially more efficient (having a faster reaction speed) than the ancestor molecules. In such cases, the mutants possessed a selective advantage and displaced the original ancestor molecular species. Series of such selective

displacements involving mutant species could be observed.

- A second more complex phenomenon includes the emergence of pseudo-collectively autocatalytic sets of molecules. This phenomenon was due to the emergence of mutant species which were acting as "parasites". These parasites were not able to self-replicate, but relied on "host" species to be replicated. Thus the replication of each molecular species, including both the parasitic and the self-replicator species, was possible in such reaction networks. The latter can therefore be considered as organisationally closed. However as autocatalytic species are still present, the term *collectively* autocatalytic cannot be rigorously applied here.

- Another dynamic of interest is distinguished with the emergence of "social hyper-parasites". In reaction networks composed of such entities, no autocatalytic molecular species were present. An aggregation of molecules was observed in which each of them supported the production of another molecule in this set. Therefore these social hyper-parasite species constituted a collectively autocatalytic reaction network. It has been reported that such organisations possessed a selective advantage which would drive the extinctions of self-replicators, parasistes and hyper-parasites. The latter are not described here, Ray (1992) provides further details about these species.

Other emerging phenomena were observed such as the emergence of immunity to parasitism, circumvention of immunity to parasitism, hyper-parasitism, Lotka-Volterra cycles and cheaters with hyper-hyper-parasites. This set of behaviour demonstrates the capability of Tierra to exhibit intricate evolutionary dynamics and phenomena. An increase in complexity and diversity was observed, which was potentially facilitated by the mutation operators and the rich set of enzymatic/computational operators.

On top of exhibiting the emergence of closed reaction networks, this work is of interest to the evolution of CIPNs as it demonstrates that under an evolutionary regime, complex behaviours at the macroscopic level, e.g., emergence of ecologies, exhibition of punctuated equilibrium dynamics (Gould, 2002), may emerge. Similarly to the Tierran societies, cells may exhibit complex behaviours that are not predictable.

Nevertheless we may also argue that the set (and nature) of the instruction operators is already *quite* complex. Indeed machine instructions that are equivalent to the Tierra operators are capable of universal computation. However the set of enzymatic operators occurring in CIPNs *may not* share such a level of computational capability.

3.4 Other systems

Our review of evolutionary approaches applied to biochemical networks is far from exhaustive. Many other systems have been developed and address various aspects of the emergence, self-organisation, self-maintenance and evolution of biochemical organisations. Examples of related studies include:

- Kauffman (1986); Farmer et al. (1986a) conducted an early study on the spontaneous emergence of autocatalytic sets. A simplified protein-network model was proposed in which only cleavage and condensation reactions could occur. This work focused on the reaction graph's connectivity as the key feature enabling the emergence of autocatalytic sets. Given a well-stirred prebiotic soup containing such simplified proteins generated from the random assembly of monomers, Kauffman et al. demonstrated that as the diversity of molecular species increases (which indirectly affects the reaction network's level of connectivity), the probability of an autocatalytic set to spontaneously emerge increases accordingly. Kauffman's work thus suggests that the emergence of such autocatalytic sets is feasible under relatively reasonable conditions.

- Typogenetics is a simplified model of DNA replication presented as a typographical formal system (Hofstadter, 1979). The novelty of this system relied on the introduction of a primary/informational and secondary/functional structures. Codons of "nucleotides" would code for specific functional operators though the use of a mapping table. It was proposed that Typogenetics was a suitable approach to investigate Artificial Life (Morris, 1987). Studies on self-replicators and hypercycles were also conducted using this system (Kvasnicka and Pospichal, 2001; Wee and Lee, 2005; Gwak and Wee, 2007)

- Avida (Adami and Brown, 1994) is a popular variant of the Tierra system. In contrast to Ray's system, the update rules (as devised by the time-slicing algorithm) are not fixed and may vary according to the nature of the organisms' genomes. Avida agents are executed on individual virtual CPUs whose "speed" may vary from one another. As Tierra, Avida was also employed to investigate open-ended evolution (Lenski et al., 1999, 2003).

- Echo (Holland, 1990, 1994, 1996) is an agent-based system proposed to investigate the abstract class of Complex Adaptive Agents (which may be applied to a wide variety of artificial and natural systems). Echo employs simplified LCS-like rules to determine the agents' behaviour. This model was inspired by ecological research in which typical agent interactions include combat, trade and mating.

- Farmer et al. (1986b) proposed a dynamic network model of Artificial Immune Systems (AIS) based on the network theory of Jerne (1974). Farmer et al. demonstrated a strong analogy between their proposed model and Holland's LCS in which molecular (antibody) species are classifiers, the latter's conditions and actions are respectively epitopes and paratopes, finally the classifiers' fitness/strength designate the molecular species' concentration. This work on adaptable/evolvable biochemical networks differs from other LCS-based inves-

tigations by explicitly addressing the molecular species' concentration with the classifiers' fitness as devised by the credit assignment algorithm (Section 3.2.3). This algorithm (e.g., the bucket brigade algorithm) governs the changes in concentration of the different molecular species and can be thus regarded as the reactor model. However, in contrast with above presented systems, AIS-based research has focused on problem-solving applications using AIS as adaptive machine learning techniques (Bersini and Varela, 1990; Bersini and Carneiro, 2006).

Although the specification and implementation of the above models may vary significantly compared to selected reviewed EC and AC models, the observed evolutionary dynamics are essentially equivalent (e.g., emergence of closed networks, increase in diversity and complexity of species, emergence of complex behaviours at the molecular population level). However the evolution of reaction networks using these systems would always plateau (where no further increase in complexity nor emerging behaviour would be observed) during long term evolution. To date, no evolutionary system has managed to demonstrate an open-ended evolutionary growth of complexity.

In the following section we evaluate the strengths and weaknesses of EC/top-down and AC/bottom up evolutionary approaches.

3.5 Top-down versus bottom-up approaches

We discuss the pros and cons of the above top-down and bottom-up evolutionary approaches. We distinguish several evaluation criteria that are related to closure and the evolution of CIPNs:

- *Granularity*: EC models (excluding Holland's agent-based LCS) consider molecular species in an aggregate manner at the system level. Bray and Deckard's approaches employed differential equations to model the artificial

reaction networks. As reviewed in Chapter 2, these modelling techniques do not allow for the examination of the individual behaviour of molecular species. However examining closure and the evolution of CIPNs requires the ability to trace the behaviour (subjected to an evolutionary process) of distinct species. Presented AC models rely on modelling techniques which address the individual and inner structure (including submolecular elements and/or reflexive nature) of molecular species.

- *Cell-level interactions*: Bray and Deckard's approaches evolved reaction networks in which the nature of the container (e.g., a cell) was not considered. By contrast, Holland's approaches considered cell-level properties leading to systems possessing two distinct levels of interaction. However further key properties of cells were still not addressed. For example no systems implemented cellular division where molecular species are randomly selected and distributed to offspring cells. This stochastic process may therefore dramatically affect the behaviour/performance of offspring cells/agents. EC models did not attempt to examine self-maintenance properties of biochemical networks. On the other hand, ACs addressed the self-organisation and self-maintenance properties of reaction networks. However the potential functions of the system itself (i.e., the cell) and cell-level interactions were not investigated.

- *Computation*: In terms of traditional computational/signal processing functions, EC models were successfully employed to evolve a range of mathematical functions (e.g., square/cubic root, normal logarithm, etc.). Artificial reaction networks were also evolved with success to solve quadratic equations. Such results have not been obtained using ACs. Evolving mathematical functions in EC models is facilitated by the explicit definition of an objective function. The latter allows one to specify precise computational functions to be mirrored.

  No ACs have to date been devised to evolve reaction networks capable of

distinct signal-processing functions. ACs are mainly employed to investigate the emergence and evolution of behaviours exhibited by living systems. However these targeted natural behaviours exclude signal-processing functions that some natural systems are capable of (e.g., CIPNs, neural networks).

Nevertheless some AC models are computational universal, therefore such ACs should be able to generate computational functions of any complexity. Although the evolution of computational functions using ACs has not been investigated, the modelling of molecular computing devices using ACs has been addressed (Tominaga et al., 2007). This work supports the suggested ability of AC systems to perform computational functions.

- *Closure*: According to Kauffman (1993), the network's topology is a critical property which may allow or not the spontaneous emergence of collectively autocatalytic reaction networks. Given a randomly generated reaction network, a level of connectivity/reactions between the molecular species is necessary to obtain spontaneously a collectively autocatalytic reaction network.

  In ACs, the inner structure (binding rules) of molecular species determines the reaction network's topology. As the structure of species is dynamic, the network's topology may also change over time. Varying these properties dynamically allows ACs to exhibit such spontaneous phenomena involving closure in reaction networks.

  Although in some EC models, the network's topology may be dynamic, no consideration for organisational closure was given. However this could have been addressed in these models where the objective fitness function could be modified to account for closure properties. This engineered top-down trick remains hypothetical as further examinations would be necessary.

- *Evolution*: In EC models, the evolutionary process is driven by explicitly devised fitness functions. These fitness functions are pre-specified and do

not evolve. As discussed by Groß and McMullin (2002), these attributes may stifle the occurrence of "perpetual novelty" during evolution. Therefore the performance of the system (as defined by the fitness function) may be limited during long term evolution.

An alternative would be to define a meta Evolutionary Algorithm (EA) that would evolve these mechanisms, however this would present one with yet another problem: how to specify the EA fitness function? The latter is fixed and is potentially another point where novelty may be stifled, recreating this credit assignment problem.

To avoid this infinitely recursive problem, ACs rely on implicit fitness functions that the agents devise themselves. Nevertheless, in any given ACs, the evolution of complexity appears to plateau during long-term evolution. Although the use of implicit fitness functions seems more appropriate, further conditions for open-ended evolution exist. The identification and understanding of these conditions remain, to date, critical and challenging problems in both artificial and natural systems (Gershenson and Lenaerts, 2008).

Based on the above evaluation, we do not identify a clear suitable evolutionary framework to investigate closure and the evolution of CIPNs. However complementary desired features are present in these systems. AC-based/bottom-up evolutionary frameworks appeared to present essential features with regards to granularity, evolution and closure.

Moreover ACs that are modelled as agent-based systems, and implemented using an object-oriented programming environment (Bersini, 1999, 2000), offer flexibility. The latter allows us to propose a novel AC which would also account for CIPN-specific properties and information processing capabilities. This novel AC is described in the next concluding section.

3.6 Conclusion

We reviewed a selection of top-down and bottom-up evolutionary systems which addressed complementary aspects with regards to computational abilities, emergence, self-organisations, self-maintenance and evolution of CIPNs. However, this review revealed that no existing evolutionary frameworks can be directly applied to investigate closure and the evolution of CIPNs.

Nevertheless we identified agent-based ACs as a relatively more promising and flexible evolutionary approach. We thus propose the development of a novel agent-based AC which is inspired by, and combines features of, the ACs previously evaluated. This AC will incorporate the following features:

- This AC will be modelled through the use of an algebraic/agent-based approach. The inner structure of molecular species will be thus specified. This modelling approach will allow for the examination of individual behaviour of molecular species.

- A production rule (i.e., condition/action rules) formalism will be adopted to implement the molecular computational processes (binding conditions and enzymatic operations). However no demarcation between operands and operators will be distinguished addressing the reflexive nature of molecular species.

- The set of primitive computational operators will be constructed in a minimalist fashion (*a contrario* to Tierra which involved computer-like instructions). This will allow us to minimise the initial complexity of the model, and potentially not bias emerging phenomena that may occur.

- Reactions will occur in a well stirred reactor.

- Molecular collisions and enzymatic/computational operations will involve stochastic elements resulting in a stochastic behaviour at the system level.

- In accordance with traditional ACs, we will examine CIPNs from a bottom-up approach in which no explicit fitness function will be devised.

- Molecular species will be subjected to random changes (mutations) introducing greater genotypic/phenotypic diversity during evolution.

- Mutations will affect the binding conditions/structures of molecular species. As a result the reaction network's topology will be dynamic which may facilitate the emergence of closed reaction networks.

- Molecular species will be contained within a single reactor which will ultimately be evolved to carry-out signal-processing functions. This single reactor model is presented and evaluated in Chapter 4 and 5 respectively.

- Compartmentalised containers will be introduced. Diffusion/exchange of molecules between these multiple compartments will be devised. This static reactors model with molecular diffusion is investigated in Chapter 6.

- Similarly to biological cells, compartmentalised containers (i.e., cells) will be able to "grow and divide". This cellular model is also examined in Chapter 6.

In the next chapter, we present in detail our novel proposed AC, called the Molecular Classifier Systems (MCS.bl), specifically devised to address closure and the evolution of CIPNs.

Chapter 4

The Artificial Chemistry

We present our Artificial Chemistry (AC) called the Molecular Classifier System (MCS.bl) which is derived from the Holland broadcast language (BL). This evolutionary simulation platform is implemented as an agent-based system in which the agents are artificial molecules. Chemical reactions between the artificial molecular species refer to the interactions between agents. In this chapter, we introduce our motivations for utilising the MCS.bl. We then present the class of Molecular Classifier Systems and our implementation of the Holland broadcast language. We finally describe the system's algorithm and summarise this chapter.

4.1 Motivations

In chapter 2, we concluded that algebraic and agent-based frameworks provide most flexibility. Because of their discrete composition of structural entities, they can act at different levels of abstraction ranging from sub-molecular interactions up to summarised system global function. Moreover, introducing analytical or stochastic information is enabled through the use of transformation techniques. Based on this review we propose to use an agent-based approach in which agents (molecular species) and interactions (chemical reactions) are modelled as algebraic expressions.

In chapter 3, we argued that bottom-up evolutionary approaches (i.e., Artificial Chemistries) offer the most adequate framework to study the evolution of molecular

organisations. These systems, *a contrario* to top-down approaches, do not rely on explicitly defined fitness functions which may stifle the occurrence of "perpetual novelty" during evolution (Groß and McMullin, 2002). Furthermore ACs are commonly implemented as agent-based systems, these approaches may thus benefit from the advantages, as highlighted earlier, of agent-based modelling techniques. Therefore we propose to employ an agent-based AC to model *and* evolve artificial molecular organisations.

To specify the molecular species and reactions, we employ the broadcast language. The latter is a term rewriting system which was proposed by Holland but has never been implemented nor evaluated (Section 2.2.5). The benefits of using the broadcast language here are twofold:

1. The broadcast language, being a term-rewriting approach, provides a flexible modelling tool addressing the discrete and reflexive nature of molecular species. The development of complementary tools enables the translation of BL models to the SBML format. Through the use of this standard format, it is possible to transform the BL models into ODE systems, which allows us to conduct further analytical studies and make the generated models widely accessible.

2. We provide the first implementation and evaluation of the broadcast language. This study complements the proposal originally made by Holland in the mid-seventies (Holland, 1992a). Our implementation may also provide a starting point for conducting further studies in allied areas such as in Evolutionary Computation and Genetic Programming.

The above points suggest that the BL is a suitable modelling technique to be utilised in our investigation. However the original Holland broadcast language contained a number of features which posed some semantic ambiguities (Decraene, 2006). To facilitate our investigation, we propose a simplification of the broadcast language in which we remove the problematic features (Decraene et al., 2007b).

These system modifications are summarised in Appendix B.

Finally to address the stochastic nature of molecular reactions, we integrate the BL within the Molecular Classifier System (MCS) approach, a class of string-rewriting based ACs. In keeping with the conclusions of our literature review, we propose an agent-based AC where the agents and interactions are specified in the broadcast language. This novel AC is derived from the MCS and (simplified) BL formalisms which are both now presented.

4.2 The Molecular Classifier Systems

We define the Molecular Classifier System (MCS) as a class of string-rewriting based Artificial Chemistries. This approach is inspired by Holland's Learning Classifier Systems (LCS). Both the MCS and LCS formalisms rely on the `IF THEN` metaphor: `IF` a condition is satisfied (e.g., some molecules collide and bind with each other) `THEN` an action is executed (e.g., a product molecule is generated). In LCS, a demarcation is distinguished between *rules* and *messages*, however operations in biochemical networks are intrinsically *reflexive* in the sense that all molecules can function as both rules (enzymes) and messages (substrates). The MCS addresses these issues by removing this rule/message demarcation found in the LCS.

The behaviour of the condition/binding properties and action/enzymatic functions is specified by a "chemical" language defined in the MCS. The chemical language defines and constrains the complexity of the chemical reactions that may be represented and simulated with the MCS. For example, a MCS model using a limited number of computational functions may only faithfully represent very simplistic chemical reactions.

Before describing the nature of the enzymatic functions (action part of a molecule), the binding properties of the molecules must be identified. In the MCS approach, a reaction between molecules may only occur if the informational string of a substrate molecule satisfies/binds with the conditional part ("binding site") of an

enzyme molecule. The condition part refers to the binding properties of a molecule whereas action refers to the computational ("enzymatic") function. This pattern matching implies a notion of *binding specificity*. A molecule's binding site having a high specificity would significantly limit the range of molecular species that may bind to it. Whereas a greater range of species would bind to binding sites exhibiting low specificity.

Reactions initially occur within a single reactor whose carrying capacity is limited (i.e., a reactor may contain a fixed maximum number of n_{max} molecules). When two molecules can bind and consequently react with each other, the action part of one of the molecules is used to carry out the enzymatic operations upon the binding molecule (substrate). The symbols contained in the MCS action part are processed in a sequential order (parsed from left to right). This operation results in producing an offspring molecule whose nature depends on the symbols' functionality. This is analogous to the action part of a LCS rule used by Holland (Holland, 2001).

When a successful catalytic reaction occurs, a product species is inserted into the reactor. If the latter is full (i.e., if the reactor contains n_{max} molecules) then a molecule is selected at random (other than the reactants) and removed from the reactor. Moreover all reactants are catalytic in the sense that they are not consumed during reactions.

A differing implementation of the MCS was proposed to investigate protocell computation (McMullin et al., 2007). In that study, a protocell is modelled as a container for artificial molecules. The latter may interact with each other to generate new molecular offspring. The chemical language used in that instance of the MCS employs a minimal set of computational components which only allows the modelling of replicase molecules. To represent, simulate and evolve CIPNs, more computational functions are necessary. To allow a richer repertoire of chemical reactions, we employ here a simplification of the Holland BL to specify and model the artificial molecular species and reactions. In the remainder of this chapter, we

first present some system features which are generic to the MCS. Following this, our version of the BL and the MCS.bl system are described.

4.2.1 The single reactor model

In the MCS class, the alphabet of the employed chemical language is denoted by Λ. $S = \{s_1, \ldots, s_i, \ldots, s_\alpha\}$ is the set of strings (with string length L and a maximal fixed length $Lmax$) over Λ, $\alpha = \sum_{L=1}^{Lmax} |\Lambda|^L$. S also constitutes the set of all possible molecular species that may appear in the MCS. $R = \{r_1, \ldots, r_j, \ldots, r_\beta\}$ is the set of all reactions that may occur between molecular species $s_i \in S$, $\beta \leq \alpha^2$ with α^2 being the total number of species-pair combinations. Chemical reactions are bimolecular (i.e., involve the interaction of two distinct molecules) and are noted as follows:

$$r_j = \; x \; + \; y \rightarrow z \tag{4.1}$$

Eq. 4.1 depicts an example reaction r_j where x, y and z are molecular species in S. In this notation, the *order* of the reactants is considered, i.e., the first term x always designates an active broadcast device species (enzyme), the second term y an input broadcast device species (substrate molecule) and z an output broadcast device species (product molecule).

Chemical reactions are asymmetric in the sense that commuting x and y designates a different reaction, for example:

$$r_1 \;\; = \;\; s_1 \; + \; s_2 \rightarrow s_3 \tag{4.2}$$

$$r_2 \;\; = \;\; s_2 \; + \; s_1 \rightarrow s_4 \tag{4.3}$$

r_1 and r_2 are two distinct chemical reactions, where s_1 is employed as an enzyme and s_2 as a substrate in Eq. 4.2. Whereas in Eq. 4.3, s_2 is utilised as an enzyme and s_1 is the substrate. s_3 and s_4 are two different product species.

$$r'_j = \ x \ + \ y \rightarrow \emptyset \tag{4.4}$$

Eq. 4.4 describes an example elastic reaction (i.e., a reaction which does not lead to the production of an output broadcast device).

The system's state U can be described by its finite collection of molecular instances denoted by m_k at time t. For example $U(t) = (m_1, \ldots, m_k, \ldots, m_{n_{max}})$ describes the list of molecules occurring in U at time step t, n_{max} is finite. A distinct/molecule of the species class s_i is denoted by m_k^i. Multiple molecules which are syntactically identical belong to the same molecular species class.

All molecules are contained in a single reactor in which they are "competing" with each other. Reactions result from successful molecular interactions which occur at random. During these random molecular collisions, two molecules m_e and m_s are identified where m_e is treated as the enzyme molecule and m_s as the substrate molecule.

If m_e can bind/react with m_s then a reaction successfully occurs: A product molecule m_p is inserted in the reactor whereas another molecule m_x (where $x \neq e \wedge x \neq s$) selected at random is removed from the reactor space (designating the system outflow). In a MCS simulation, reactions may thus be described as follows:

$$m_e \ + \ m_s \ + \ m_x \ \rightarrow m_e \ + \ m_s + \ m_p$$

Figure 4.1 depicts the flow of a MCS simulation.

This single reactor model was inspired by the Alchemy system in which a similar approach was employed.

4.2.2 Mutation

We define the different operators which allow molecular variations to occur in the MCS. Two types of "mutation" are identified:

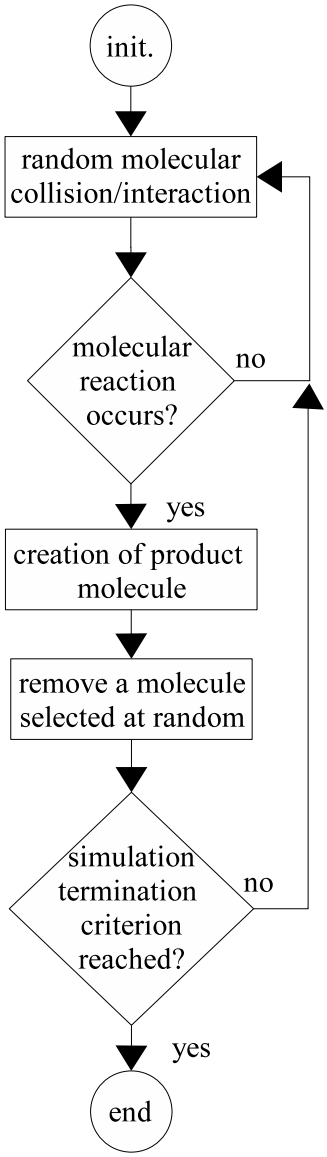

Figure 4.1: Program flow of a MCS simulation

1. *Molecular mutations*:

   - When a new molecule is produced, a mutation is applied with probability p_{sym} to each of its symbols. Therefore, the longer the molecule, the higher the probability of one or more mutations occurring.
   - Three subtypes of molecular mutation are distinguished and are applied with equal probabilities at each symbol position:
     - *Symbol flipping:* The current symbol is replaced with a symbol picked

uniformly at random from Λ in which the current symbol is excluded.

- *Symbol insertion:* A symbol is picked uniformly at random from Λ and inserted after the current symbol.
- *Symbol deletion:* The current symbol is removed.

2. *Spontaneous mutations*: To maintain diversity in the event of low ongoing reaction activity, a spontaneous mutation mechanism occurring every x timesteps is also available. A subset r_{mut} of the population is selected at random and one of the three types of mutation mutation (chosen as above) is then applied to a single symbol picked uniformly at random in each molecule of this subset.

4.3 The broadcast language

Prior to the development of the MCS.bl, we investigated the original broadcast language proposed by Holland[1]. As no implementation of the broadcast language was publicly available, we proposed the first complete specification and implementation of this formalism (Decraene, 2006). Using this system, we successfully constructed a NAND gate (Decraene et al., 2007a) and a static Genetic Regulatory Network model (Decraene et al., 2007b). In the remainder of this section we present our simplified version of the broadcast language which is utilised to specify the molecular species and reactions in our agent-based Artificial Chemistry.

4.3.1 Introduction

We present our simplification of the Holland broadcast language. Artificial molecules (broadcast language strings) are referred to as broadcast devices, see Figure 4.2. A

[1]The broadcast language is a programming formalism devised by Holland in 1975, which aimed at allowing Genetic Algorithms (GAs) to use an *adaptable representation*. A GA may provide an efficient method for adaptation but still depends on the efficiency of the fitness function used. During long-term evolution, this efficiency could be limited by the fixed representation used by the GA to encode the problem. When a fitness function is very complex, it may be desirable to adapt the problem representation employed by the fitness function. By adapting the representation, the broadcast language intended to overcome the deficiencies caused by fixed problem representation in GAs.

broadcast device is parsed into zero, one or more broadcast units, where each unit represents a single condition/action rule. When an action statement is executed, a new molecule is generated or "broadcast" in the reaction space. A broadcast device containing no broadcast units is called a null broadcast device and can function only as a substrate molecule (i.e., possessing no enzymatic/computational functions).

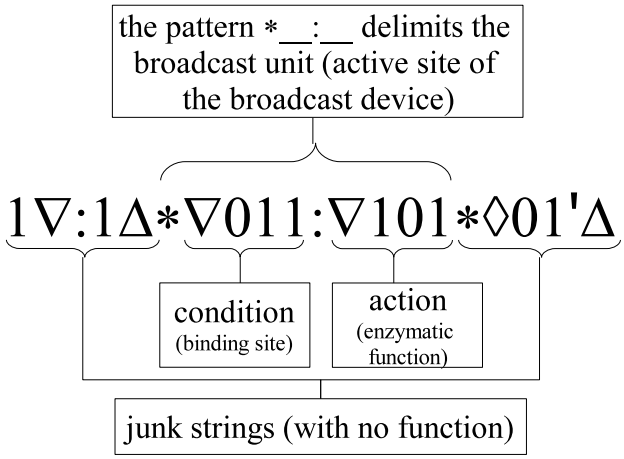

Figure 4.2: An example broadcast device

Whenever a broadcast unit conditional statement is satisfied, the action statement is executed. This is analogous to an enzyme which would form a product molecule upon the binding of a specific substrate molecule to its binding region. In this metaphor the active site (where catalysis occurs) of the enzyme can be thought of as a broadcast unit, a substrate molecule would be a binding/input broadcast device, the active site's binding region would refer to the broadcast unit conditional statement, the product molecule is the output broadcast device and finally the environment would be the reaction space (e.g., the cell). Figure 4.3 depicts an example chemical reaction in the BL.

Some broadcast units may generate an output broadcast device that may itself contain zero, one or more broadcast units. Similarly, a broadcast device can be interpreted as a substrate molecule that can be catalysed by another broadcast device. As a result, a broadcast unit may produce an output broadcast device

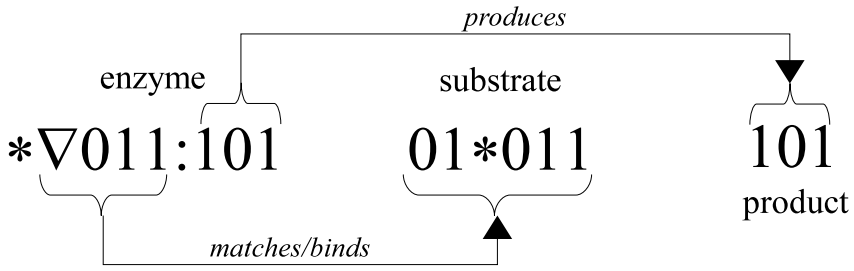

Figure 4.3: Example reaction

which results from some modifications of the input broadcast device (i.e., signal processing).

Biology	Broadcast language
sequence of amino acids from $\{A, R, N, D, C, E, \ldots\}$	string of symbols from $\Lambda = \{0,\ 1,\ *,\ :,\ \Diamond, \nabla,\ ,\ \triangle,\ '\}$
substrate molecule	input broadcast device
product molecule	output broadcast device
molecule with no enzymatic function	null broadcast device
enzyme's active site	broadcast unit
enzyme molecule	broadcast device
cellular milieu	list of strings from Λ

Table 4.1: Comparison of biological and broadcast language terminology

As a summary, Table 4.1 presents a comparison between the biological and the broadcast system terminology. A detailed description of the broadcast language's syntax and semantics follows.

4.3.2 The syntax

Our specification of the broadcast language partially adheres to the original proposal presented by Holland. As mentioned earlier, a number of features have been removed to facilitate the evaluation of this system. Moreover additional details have been introduced to complement Holland's proposal and resolve some identified ambiguous issues. We now describe the different structures constituting the broadcast language: the symbols, broadcast units and broadcast devices. The interpretation

of the symbols, broadcast units and broadcast devices will follow.

The broadcast language alphabet Λ is finite and contains eight *symbols*. The symbols constitute the atomic elements of the language.

$$\Lambda = \{0,\ 1,\ *,\ :,\ \Diamond,\ \triangledown,\ \triangle,\ '\}$$

Let s_i be an arbitrary string (i.e., a molecular species) from S. A symbol occurring in s_i is said to be quoted if it is preceded by the symbol $'$. $A \subseteq S$ is the set of strings over Λ which do not contain unquoted occurrences of symbol : and $*$. The set A does not contain null strings. $S^* \subseteq S$ is the set of strings over Λ which are of the form $*a_1 : a_2$, where $a_1 \in A \wedge a_2 \in A$. The broadcast device's active sites are called *broadcast units* which are arbitrary strings from $S^* = \{u_1, \ldots, u_l, \ldots, u_o\}$, o is finite. The minimal length to realise a broadcast unit is $BD_{Lmin} = 4$ where $length(*a_1 : a_2) = BD_{Lmin} \wedge length(a_1) = length(a_2) = 1$. Several broadcast units may be concatenated within a single broadcast device. BD_{Lmax} is the fixed maximum string length of broadcast devices. A broadcast device m_k may contain $0 \leq n_u \leq \frac{BD_{Lmax}}{BD_{Lmin}}$ broadcast units.

If $n_u = 0$ then m_k does not contain any broadcast units and m_k is then called a *null* broadcast device. A null broadcast device may *only* be interpreted as an input broadcast device and is not capable of any enzymatic/computational functions. A broadcast device which is not null is said to be *active* and may generate an output broadcast device (resulting from the computational function specified in the broadcast unit's action statement) upon the binding of an appropriate input broadcast device.

Some example broadcast devices are shown in Figure 4.4. We may note that m_3 is a null broadcast device.

A broadcast device m_k is parsed into broadcast units as follows:

- Any prefix symbols occurring to the left of the leftmost unquoted $*$ are ignored

$$
\begin{aligned}
m_1 &= 10 * 11' * \triangle 0 : 1\triangle * : 11\triangledown : 11\triangledown \\
m_2 &= 011' * *\triangledown : \Diamond 1011\triangle \\
m_3 &= 11' * \triangle 0 : 1\triangle' * : 11\triangle : 0001\Diamond
\end{aligned}
$$

Figure 4.4: Example broadcast devices

(junk string).

- The first broadcast unit is designated from the leftmost unquoted $*$ to (not including) the next unquoted $*$ on the right *if any*.
- Following broadcast units are obtained by repeating the above procedure for each successive unquoted $*$ from the left.

For example the broadcast device m_1:

$$m_1 = 10\underline{*11' * \triangle 0 : 1\triangle} * \triangle 00\underline{*11\triangledown : 11\triangle} : 1$$

designates two distinct broadcast units u_1 and u_2:

$$
\begin{aligned}
u_1 &= *11' * \triangle 0 : 1\triangle \\
u_2 &= *11\triangledown : 11\triangle
\end{aligned}
$$

4.3.3 The semantics

We describe the interpretation of the broadcast units and symbols.

Broadcast Units

Let us consider the broadcast unit $u_l = *11' * \triangle 0 : 1\triangle$. The string $u_{l_{IN}} = *11' * \triangle 0$ stands for the broadcast's unit conditional statement (binding region) and may be translated into a pattern matching expression. Whereas the string $u_{l_{OUT}} = 1\triangle$ refers to the broadcast unit's action statement and encodes for the computational function of u_l.

When a broadcast unit's action expression is interpreted for producing an output broadcast device, a quote is removed from each quoted symbol. This quote mechanism allows one to prevent symbols from interpretation and to be passed into the output broadcast device (see next section for an example).

In some cases, despite a broadcast unit u_l being *syntactically correct*, its action statement may not be executed. This inability to generate a product broadcast device (leading to an elastic reaction) may result from the nature of symbols present in the action expression. These cases are described in detail in the following section.

The symbols

The interpretation of each symbol in $\Lambda = \{0,\ 1,\ *,\ :,\ \Diamond,\ \nabla,\ \triangle,\ '\}$ is now presented. Within active broadcast devices, we may identify *ignored* symbols. These symbols do not hold any functions in the binding and enzymatic operations of a given broadcast device. These substrings are analogous to non-coding DNA strings (junk strings). We illustrate the usage of each symbol through the depiction of example reactions:

- The quote symbol $'$ is used to "quote" a symbol in the arguments of a broadcast unit. The specific function of a quoted symbol is ignored when interpreted, regardless of the exact position of the quoted symbol in $u_{l_{IN}}$ or $u_{l_{OUT}}$. Nevertheless quoting 1s or 0s does not affect the function of these particular symbols.

  For example:

$$r_1 = *11'\triangle 0 : 1'1 + 11\triangle 0 \rightarrow 11$$

$$r_2 = *11'\triangle 0 : 11 + 1100 \rightarrow \emptyset$$

  whereas

$$r_3 = *11\triangle 0 : 1'\triangle + 1100 \rightarrow 1\triangle$$

In the above examples, quoting the symbol $\triangle$ prevents the interpretation of its specific function (described below) when occurring in either $u_{l_{IN}}$ or $u_{l_{OUT}}$.

- The star symbol * is the broadcast unit separator. This symbol, when unquoted, indicates that the following symbols until the next unquoted $*$ (if any) *may* be interpreted as a broadcast unit. If a broadcast device species s_i does not contain any unquoted $*$ then s_i is a null broadcast device.

- The colon symbol : is used as a punctuation mark to separate the parts (condition and action expressions) of a broadcast unit. Similarly to *, the colon is a structural symbol which is necessary to constitute a broadcast unit. If a string of symbols following an unquoted * does not contain any unquoted occurrences of :, then this string cannot be interpreted as a broadcast unit. If more than one unquoted : is found in a broadcast unit then the second : and anything to the right of it are ignored.

  A broadcast unit is identified by the pattern $*u_{l_{IN}} : u_{l_{OUT}}$ where $u_{l_{IN}}$ and $u_{l_{OUT}}$ are arbitrary strings from A. If $u_{l_{IN}} \notin A \vee u_{l_{OUT}} \notin A$ then the string $*u_{l_{IN}} : u_{l_{OUT}}$ is not a broadcast unit. $u_{l_{IN}}$ refers to the conditional statement of u_l or pattern matching expression whereas $u_{l_{OUT}}$ is the computational or enzymatic function of u_l. $u_{l_{IN}}$ and $u_{l_{OUT}}$ are also called the "arguments" of u_l.

  For example:

$$
\begin{aligned}
r_4 &= *10111 : 00 \ + \ 10111 \rightarrow 00 \\
r_5 &= 10111 : 00 \ + \ 10110 \rightarrow \emptyset \\
r_6 &= *1011100 \ + \ 10110 \rightarrow \emptyset \\
r_7 &= * : 1011100 \ + \ 10110 \rightarrow \emptyset
\end{aligned}
$$

  In r_4, the species $*10111 : 00$ conforms to the pattern $*u_{l_{IN}} : u_{l_{OUT}}$ and is therefore an active broadcast device which contains a single broadcast unit.

The condition statement $u_{l_{IN}} = 10111$ matches the substrate species 10111. As a result, the action statement $u_{l_{OUT}} = 00$ generates the product species 00. In r_5, r_6 and r_7, the enzyme species do not conform to the pattern $*u_{l_{IN}} : u_{l_{OUT}}$. These species are thus null broadcast devices resulting in r_5, r_6 and r_7 being elastic reactions.

- The symbols 0 and 1 possess different functions when located in either the condition or action statements of a broadcast unit. When 1 (or 0) occurs in $u_{l_{IN}}$ at the position *pos*, a substrate species s_k would bind to $u_{l_{IN}}$ if s_k presents a 1 (or 0) at the relative position *pos*. If 1 or 0 occurs in the action statement, a 1 or 0 is generated in the product species string accordingly.

  A string such as 010110 can be regarded as the *tag* of a particular broadcast device. This tag can be employed by a broadcast unit to react with specific broadcast devices.

  For example:

$$\begin{aligned} r_8 &= *10111 : 100 \ + \ 10111 \rightarrow 100 \\ r_9 &= *10111 : 100 \ + \ 10110 \rightarrow \emptyset \end{aligned}$$

- Ordinarily the lozenge symbol $\Diamond$ acts as a single character wildcard. When this symbol is met in the conditional statement of a broadcast unit, it indicates that an input broadcast device "colliding" with the broadcast unit may present any single symbol at this position. This symbol occurring in the input broadcast device does not affect its acceptance or rejection by the broadcast unit. In no circumstances, $\Diamond$ may match a *null* substring.

For example:

$$\begin{aligned} r_{10} &= *10\Diamond 11 : 00 \ + \ 10011 \rightarrow 00 \\ r_{11} &= *10\Diamond 11 : 00 \ + \ 10 : 11 \rightarrow 00 \\ r_{12} &= *10\Diamond 11 : 00 \ + \ 11 : 11 \rightarrow \emptyset \\ r_{13} &= *10\Diamond : 00 \ + \ 10 \rightarrow \emptyset \end{aligned}$$

In r_{10}, the enzyme species $s_1 = *10\Diamond 11 : 00$ would successfully react with any input broadcast devices of the form $10\Diamond 11$ where $\Diamond$ indicates a single arbitrary symbol from Λ.

However, $\Diamond$ may in some cases act as a multiple character wildcard. If $\Diamond$ occurs at the rightmost position of $u_{l_{IN}}$, then it indicates that an input broadcast device reacting with u_l may present any suffix (i.e., any string of symbols from Λ) without affecting its acceptance or rejection by u_l.

For example:

$$\begin{aligned} r_{14} &= *1011\Diamond : 100 \ + \ 1011\underline{000} \rightarrow 100 \\ r_{15} &= *1011\Diamond : 100 \ + \ 1011\underline{010101010} \rightarrow 100 \end{aligned}$$

Underlined strings designate the input broadcast device's substring that is matched by the multiple/single character wildcard. If an unquoted occurrence of $\Diamond$ occurs in $u_{l_{OUT}}$ then this symbol is ignored when $u_{l_{OUT}}$ is executed.

- The reversed triangle symbol $\triangledown$ is a multiple character wildcard which may also act as a *variable holder*. If this symbol occurs at the leftmost or rightmost position of $u_{l_{IN}}$, then an input broadcast device may present any arbitrary initial (prefix) or terminal (suffix) string of symbols and will successfully bind to $u_{l_{IN}}$.

If ∇ occurs in both $u_{l_{IN}}$ and $u_{l_{OUT}}$ then ∇ holds for *value* the string of matched prefix or suffix substring. When the computational function $u_{l_{OUT}}$ is interpreted for execution, any occurrence of ∇ is replaced with its value. This allows one to pass a string of symbols from the input broadcast device to the output broadcast device (i.e., signal processing).

For example:

$$r_{16} = \ *10\nabla : \nabla \ + \ 10\underline{011} \ \rightarrow 011$$

with $\nabla = 011$ whereas

$$r_{17} = \ *10\nabla : \nabla\nabla \ + \ 10\underline{0100101} \ \rightarrow 01001010100101$$

with $\nabla = 0100101$.

$$r_{18} = \ *10\nabla : \nabla\nabla \ + \ 110100101 \ \rightarrow \emptyset$$

If ∇ does not occur at the first or last position of $u_{l_{IN}}$ then ∇ is ignored. If two ∇ symbols simultaneously occur at the first *and* last position of $u_{l_{IN}}$ then the rightmost occurrence is ignored.

- The triangle symbol $\triangle$ is employed in the same manner as ∇ but designates a *single* arbitrary symbol whose position can be anywhere in both arguments of a given broadcast unit. For example:

$$r_{19} = \ *11\triangle 0 : 1\triangle \ + \ 11\underline{0}0 \rightarrow 10$$

with $\triangle = 0$ whereas

$$r_{20} = \ *11\triangle 0 : 1\triangle \ + \ 11\underline{1}0 \rightarrow 11$$

with $\triangle = 1$.

Moreover, if a broadcast unit contains more than one (unquoted) $\triangle$ symbol then only the leftmost occurrence of $\triangle$ is operative and is to be interpreted by the broadcast unit. The other occurrences of $\triangle$ found in u_{lIN} are ignored. A $\triangle$ symbol may occur anywhere within u_{lIN}.

If u_{lOUT} contains an unquoted occurrence of $\triangle$ or ∇ which is not present in u_{lIN}, then this symbol cannot be interpreted and is ignored.

Table 4.2 presents a number of example reactions that can be realised with the MCS.bl.

Enzyme	Substrate	Product	Reaction
$*\nabla 1: \ \nabla 0$	$1:0$	$\emptyset$	elastic reaction
$*\nabla 1: \ '*\nabla$	$0:1$	$*0:1$	activation
$*'*0\nabla:0\nabla$	$*0:1$	$0:1$	inhibition
$*\nabla:\nabla$	$*00:11$	$*00:11$	replication
$*\nabla 0:\nabla 0$	$*\nabla 0:\nabla 0$	$*\nabla 0:\nabla 0$	self-replication
$*\nabla 1:\nabla 10$	$*0:1$	$*0:10$	concatenation
$*\nabla 1:\nabla$	$*0:1$	$*0:$	cleavage

Table 4.2: Example reactions realised with the MCS.bl

Finally, to clarify the relationship between the MCS.bl, MCS, LCS, broadcast language and Alchemy, an overview is provided and depicted in Fig. 4.5.

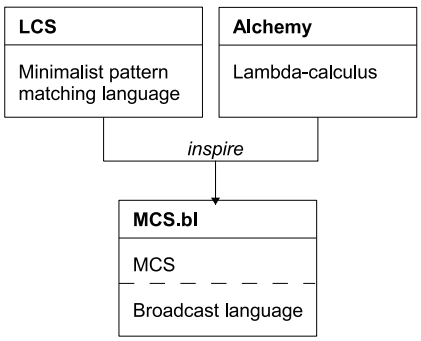

Figure 4.5: Overview of the MCS.bl and related systems. Alchemy and the Learning Classifier System (LCS) are inspirational methods to the MCS.bl, whereas the latter is based on the Molecular Classifier System (MCS) and Holland's broadcast language .

4.4 Summary

We enumerated our motivations to devise and employ the MCS.bl. The latter is an agent-based Artificial Chemistry which is derived from both the class of Molecular Classifier Systems and the Holland broadcast language. The MCS class was introduced and its generic system features (i.e., the MCS reactor model and mutational operators) were described. Our simplification of the Holland broadcast language was finally presented and illustrated with a range of example reactions. In the following chapter we explore the emergence and self-maintenance of closed-reaction networks in the MCS.bl formalism.

Chapter 5

Emergence and Self-Maintenance of Closed Reaction Networks

In the previous chapter we presented an agent-based Artificial Chemistry: the MCS.bl. We employ this system to investigate organisational closure and the evolution of Cellular Information Processing Networks. Using this formalism we now present our first series of experiments. These experiments examine both the spontaneous emergence and the self-maintenance of closed reaction networks in the MCS.bl. The content presented in this chapter was published (Decraene et al., 2008b) at the Eleventh International Conference on the Simulation and Synthesis of Living Systems (Alife'08) .

5.1 Introduction

In keeping with our bottom-up approach to investigating closure in artificial CIPNs, we first examine the conditions required for the spontaneous emergence and self-maintenance of minimal closed reaction networks in the MCS.bl[1]. This key initial step is necessary to examine and understand the evolutionary dynamics of simple closed reaction networks in our AC. This fundamental knowledge will allow us to

[1]The MCS.bl implementation is object-oriented and was conducted using the C++ language. The MCS.bl software package and sources can be downloaded at `http://esignet.net/dokumente/upload/WP13`

study later the emergence and evolution of artificial CIPN functions in these closed reaction networks.

According to our definition of closure, the *simplest* form of closed reaction network that can be identified in the MCS.bl are autocatalytic organisations of the following form:

$$r_j = \ 2\ s_i\ \rightarrow 3\ s_i$$

s_1 is an autocatalytic molecular species also called a *self-replicase*. If we consider a reaction network C containing only the molecular species s_1 then C is organisationally closed. As reactions are bimolecular in the MCS.bl, this reaction involves a trans-acting replicase (requiring the interaction of two distinct molecules to achieve replication). This self-replication scheme differs from the more traditional $s_i \rightarrow 2\ s_i$ with s_i growing exponentially. Here, a "survival of the common" dynamics applies in which the growth of s_i is hyperbolic (Szathmáry and Maynard Smith, 1997). The domination of a given species is dependent on both its intrinsic fitness *and* its relative concentration in the molecular population. For another species to displace the dominant one, a significant difference in intrinsic fitness and/or a high/higher initial concentration is necessary.

Both the spontaneous emergence and self-maintenance of such individually autocatalytic molecular species were reported as easily obtained in Alchemy. In the original Tierra system, where autocatalytic reactions are first order (exponential) and not hyperbolic, the spontaneous emergence of autocatalytic molecular species was not expected or reported; however it did arise in the related Amoeba system, specifically devised for this purpose (Pargellis, 2001).

In this chapter, we first present a series of evolutionary experiments focusing on the spontaneous emergence of autocatalytic molecular species. Following this, we identify the minimal self-replicase s_{R0} that can be specified in the MCS.bl. We examine the system's dynamics when s_{R0} is manually introduced in a population of

randomly generated molecular species. We then investigate the effects of the self-replicases' binding specificity over the system's dynamics. Finally, we describe the system's evolutionary dynamics when a hand-designed self-replicase having a high binding specificity is employed.

5.2 Spontaneous emergence of self-replicases

To examine the spontaneous emergence of autocatalytic species, we perform a first experiment in which the molecular species are generated from the random assembly of monomers. An artefact of the BL's syntax is that it is moderately difficult to observe the spontaneous emergence of an individually autocatalytic molecule. Specifically, there are 4^8 $(65,536)$ distinct molecules of length 4 symbols (the minimal length to construct a functional/enzymatic molecule), of which only a single one ($s_{R_0} = *\nabla : \nabla$) is autocatalytic.

Although the probability of spontaneously obtaining such autocatalytic molecules from random assembly of monomers is therefore quite low in MCS.bl, the intuition is that, *once* such a molecular species does appear, it should be able to rapidly fill the reaction space. This phenomenon was indeed observed in Alchemy and is expected to occur in MCS.bl.

An evolutionary experiment is conducted and uses the following parameters:

- Each simulation run is initialised with 100 randomly generated, 10-symbol long, molecules.
- $n_{max} = 1000$ (i.e., the population initially grows without any displacement; but once the total number of molecules reached 1000 it is limited to this value, by displacing one random molecule for each new molecule generated, as previously described).
- 30 simulation runs are performed, each for 5000000 molecular interactions (i.e., collisions).

- Molecular mutations, as presented in Section 4.2.2, occur with $p_{sym} = 0.001$ and $r_{mut} = 0.001$.

To identify spontaneously emerging self-replicases, every molecule was tested at each timestep for self-replication functionality. The spontaneously emerging self-replicases identified in these 30 simulation runs are listed in Table 5.1.

Self − replicases		
$00'\triangle$	$*$	$\nabla : \triangle\nabla\nabla * 0$
$1\nabla 0$	$*$	$\nabla : \nabla$
$00'\triangle$	$*$	$\nabla : \triangle\nabla\Diamond\nabla * 0$
$1\triangle\nabla 0$	$*$	$\nabla : \nabla$
$: 1$	$*$	$\nabla : \nabla\Diamond : 1 * \nabla : \nabla\Diamond$
$: 0\nabla\nabla$	$*$	$\nabla : \triangle\triangle\nabla$
$:$	$*$	$\nabla : \nabla\nabla * 01$
$*\nabla$	$*$	$\nabla : \nabla\triangle\nabla\triangle\triangle$
$1\Diamond\nabla :$	$*$	$\nabla\nabla : \nabla :$
$*\nabla :$	$*$	$\nabla : \nabla\triangle\nabla\triangle\triangle$
	$*$	$\nabla : \nabla$
	$*$	$\nabla\nabla : \nabla$
	$*$	$\nabla 0\nabla\nabla : \nabla 0$
$\Diamond\nabla * \nabla$	$*$	$\nabla : \Diamond\nabla\nabla$
$\triangle 1$	$*$	$\nabla : \nabla\Diamond$

Table 5.1: Spontaneously emergent self-replicases in MCS.bl

Table 5.1 shows that 15 syntactically distinct self-replicases appeared. In the 30 experimental runs, the highest molecular count achieved by any of these spontaneously occurring self-replicases was just a *single* isolated molecule.

Although these self-replicases are syntactically different, note that it is a property of the BL syntax that some symbols are ignored when functionally interpreted (they are, in a certain sense, "junk" symbols). Thus, although 15 distinct self-replicases were identified, it turns out that the core broadcast units (the "active sites", after discarding "junk" symbols) are, in fact, identical for 14 of these; and are all equivalent to the self-replicase, $s_{R_0} = *\nabla : \nabla$. Only the broadcast device $*\nabla 0\nabla\nabla : \nabla 0$ possesses a core broadcast unit of a different form, namely $*\nabla 0 : \nabla 0$. This is an

alternate form of s_{R_1}, having just the minimal binding specificity of one symbol, i.e., s_{R_1} may only replicate molecules whose sequence finishes with the symbol 0.

The spontaneous appearance of self-replicases was expected. Results indicated that (self-)replicases do emerge, however they never manage to grow in concentration and would quasi-deterministically be displaced by other molecules. As mentioned earlier, the highest molecular count achieved by any of these spontaneously emerging self-replicases was just a single instance. Nevertheless for a self-replication reaction to occur in the MCS.bl, the collision of two distinct instances of a self-replicase species is required. Therefore self-replication dynamics could not be observed.

We also propose the following potential underlying phenomena which may have discouraged the emergence and self-maintenance of autocatalytic molecular species in the MCS.bl:

- As already noted, the BL syntax does not strongly facilitate the spontaneous emergence of self-replicases. This syntactical constraint may discourage the spontaneous emergence of such species. The BL syntax may also have an impact on the robustness of these self-replicases against mutation effects. For example, a mutation may lead to the removal or replacement of a structural symbol such as * or : in a given active broadcast device. As a result, this molecule would lose its enzymatic function and become a null broadcast device.

- Secondly, if we consider that multiple concurrent instances of a self-replicase species successfully emerge, such molecular species are likely to possess a low molecular concentration when occurring. This low concentration diminishes the capacity of these molecular species to persist against side reactions and mutation events.

- Finally although molecular species with the ability to self-replicate do emerge, these species may also function as replicases being able to catalyse the replication of other species. The latter may be viewed as parasites if these species do

not contribute, in return, to the replication of the replicases. In this case the replicases' binding specificity (i.e., the range of species that can be replicated by the replicases) may affect the ability of the replicases to self-maintain in the population.

These three factors, when combined, may significantly have lowered the probability of having a self-replicase spontaneously emerge and self-sustain in the MCS.bl.

To investigate the above propositions, we conduct a second series of experiments in which we manually design and introduce a minimalist self-replicase. Following this, we will explore the role of the replicases' binding specificity upon the system's dynamics.

5.3 No selective advantages for universal replicases

We present a second experiment in which a minimal self-replicase is devised and manually introduced in a reactor in addition to randomly generated molecules.

The behaviour of the minimal self-replicase s_{R_0}, which recurrently emerged in the previous experiment, is as follows. The matching condition is defined by a single symbol, ∇, which designates a multiple character wildcard. This indicates that s_{R_0} may bind to any molecular species. In addition when reactions occur between s_{R_0} and substrate species s_i, ∇ is assigned a value, being the matched substring of s_i. In this case, this will be the complete string s_i. A unique symbol ∇ also constitutes the action part of s_{R_0}. This specifies that the output string of s_{R_0} is exactly the string bound by the ∇ in the condition part, i.e., a copy of s_i's string. Therefore the broadcast device s_{R_0} is actually a "universal" replicase; which, by definition, means that it is also a *self*-replicase (in the special case that it binds to another instance of itself, i.e., $s_i = s_{R_0}$). The "specificity" of s_{R_0} is said to be *null*.

Figure 5.1 presents a first experiment examining the behaviour of s_{R_0} averaged over 30 simulation runs. In this experiment, the following parameters are employed:

- The reactor is seeded with 900 randomly generated molecules, each of length 10 symbols. These initial molecules are independently and randomly generated for each simulation run.

- In addition, 100 instances of s_{R_0} are inserted.

- n_{max} designates the fixed maximum number of molecules that may be contained in the universe, $n_{max} = 1000$.

- Molecular interactions occur as follows: two molecules m_e and m_s are picked at random. m_e is considered as an enzyme and m_s as a substrate. If m_e can bind and react with m_s then a molecule m_p is produced and added to the population. A molecule m_x (other than the m_e, m_s and m_p) is picked at random and removed from the population.

- No mutation may occur in these experiments in order to facilitate our investigation on replicases.

A high initial molecular amount (100 instances) of s_{R_0} was chosen to satisfy the trans-replication constraint (i.e., at least two distinct molecules are required to achieve self-replication) and minimise early extinction due simply to stochastic fluctuation.

From Figure 5.1 it is clear that the species s_{R_0} never grows to take over the population; rather, it consistently diminishes, contrary to the original, informal, prediction. A formal explanation of this outcome is given by modelling the system with the (approximate, continuous) catalytic network equation (Stadler et al., 1993). The state of the system is described by the concentration vector $x = (x_1, \ldots, x_n)$ with $x_1 + \ldots + x_n = 1$ and $x_i > 0$, where x_i refers to the concentration of a molecular species (or collection of "chemically equivalent" species) s_i. The general dynamic behaviour is then given by:

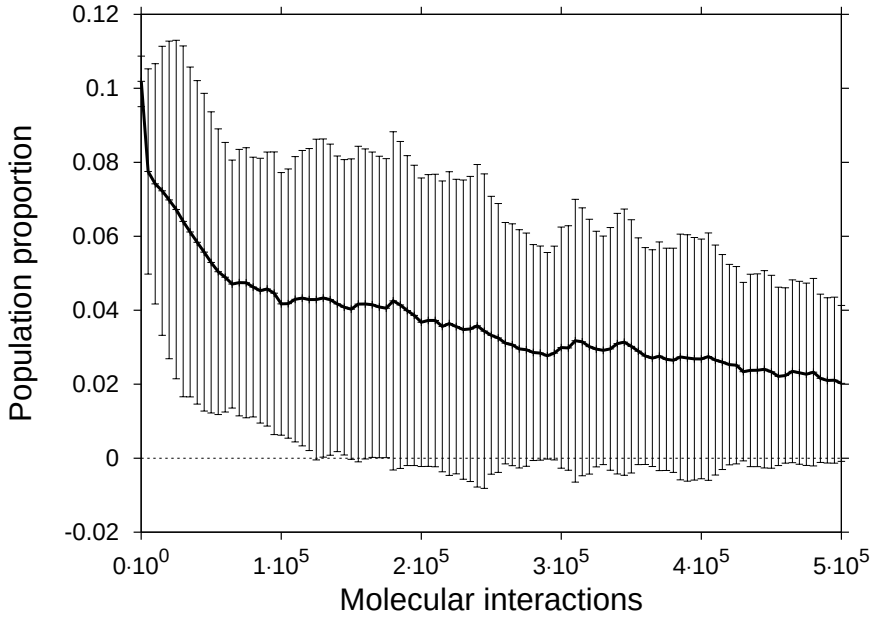

Figure 5.1: Relative population growth of replicators s_{R_0} averaged over 30 simulation runs. Solid line is average concentration; error bars denote standard deviation.

$$\dot{x_k} = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_{ij}^{k} x_i x_j - x_k \sum_{i,j,l=1}^{n} \alpha_{ij}^{l} x_i x_j \qquad (5.1)$$
$$\text{with } k = 1, \ldots, n$$

In Eq. 5.1, the second term represents the dilution flow: A molecule may be removed (at random) when a successful reaction occurs in the single reactor model (Section 4.2.1). α_{ij}^{k} are the rate constants for each reaction $s_i + s_j \rightarrow s_i + s_j + s_k$.

In this experiment, these simplify to:

$$\alpha_{ij}^{k} = \begin{cases} 1 \text{ if } s_i + s_j \rightarrow s_i + s_j + s_k \\ 0 \text{ otherwise} \end{cases} \qquad (5.2)$$

To explain the results presented in Fig.5.1, we propose a *simplified* analysis which focuses on the dynamics of universal replicases. Consider the case where only universal replicases (s_{R_0}) and the set of all non-enzymatic molecular species (S_{NE}) (that may only act as substrates) are present. This is clearly the *most* favourable

case for the growth of s_{R_0}. Denote the molecular concentrations of s_{R_0} and S_{NE} by x_1 and x_2 respectively. Then $\alpha^1_{ij} = 1$ if $i = 1, j = 1$; otherwise $\alpha^1_{ij} = 0$. Similarly, $\alpha^2_{ij} = 1$ if $i = 1, j = 2$; otherwise $\alpha^2_{ij} = 0$. Inserting into Eq. 5.1, we obtain:

$$\dot{x_1} = x_1^2 - x_1(x_1^2 + x_1 x_2) \tag{5.3}$$

But given that $x_2 = 1 - x_1$:

$$\begin{aligned} \dot{x_1} &= x_1^2 - x_1^3 - x_1^2 + x_1^3 \\ \dot{x_1} &= 0 \end{aligned} \tag{5.4}$$

whereas the growth rate of molecules S_{NE} is:

$$\dot{x_2} = x_1(1 - x_1) - (1 - x_1)[x_1^2 + x_1(1 - x_1)] \tag{5.5}$$

$$\begin{aligned} \dot{x_2} &= x_1 - x_1^2 - (1 - x_1)(x_1^2 + x_1 - x_1^2) \\ \dot{x_2} &= x_1 - x_1^2 - x_1 + x_1^2 \\ \dot{x_2} &= 0 \end{aligned} \tag{5.6}$$

Thus, both molecular species s_{R_0} and S_{NE} share a common zero "expected" growth. Under the stochastic conditions of the reactor this would yield a random drift in relative concentrations—as opposed to a quasi-deterministic growth of the s_{R_0} species.

In such systems driven by random drift dynamics, only two possible outcomes may be observed where the system reaches steady state. If both species s_{R_0} and S_{NE} are initialised with a common concentration, then both outcomes would deterministically occur with equal chances:

1. The universal replicase species eventually displaces all S_{NE} molecules in the reactor.

2. All replicases s_{R_0} *but one* are displaced by S_{NE} molecules. This single replicase molecule remains and cannot be displaced since: 1) Reactants are not consumed during reactions in the MCS.bl. 2) No side reactions, involving S_{NE} molecules only, may occur.

Qualitatively the above phenomenon is due to the fact that any (self-)replicase having low or zero specificity, such as s_{R_0}, will not only replicate itself but also replicate any other molecules; and therefore cannot selectively displace these molecules. But recall that this was the *best case* situation for growth of s_{R_0}, where none of the other molecules had any enzymatic activity. In the practical case of Figure 5.1 the collection of such additional side reactions will give a nett negative growth rate for s_{R_0}, which therefore, quasi-deterministically, decays until $x_1 = 0$.

With regards to the spontaneous emergence and domination of self-replicases given a set of randomly generated molecular species, this analysis is consistent with the results described in Section 5.2. Self-replicases of very low specificity (which do spontaneously occur) cannot grow to significant concentrations and would therefore be commonly displaced by other molecules.

In Section 5.2, we also mentioned that the replicases' binding specificity may potentially affect the system's dynamics. We examine the role of binding specificity in the following section.

5.4 Specificity and domination of the replicases

To investigate the role of binding specificity, we proceed to a series of experiments in which we incrementally increase the specificity of the (self-)replicases. Table 5.2 shows the different replicases employed in these experiments. s_{R_1} designates a molecular species that would only react with molecules whose strings end with the symbol "1". As the latter occurs at the rightmost position of s_{R_1}, it may react with itself, producing another instance of s_{R_1}. Similarly, s_{R_2} only binds to molecular strings containing the suffix 01. This tag forms a constraint on the replicases, allowing them

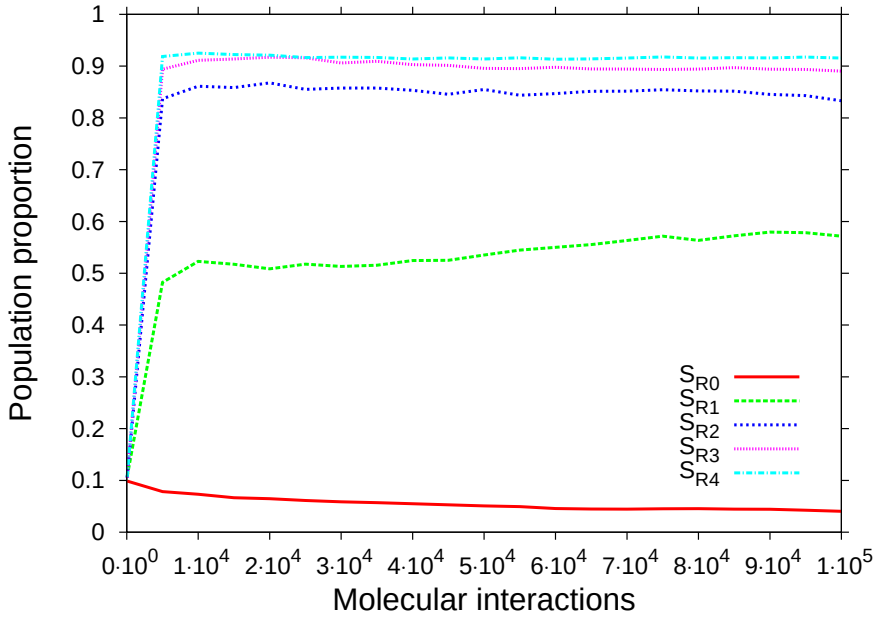

Figure 5.2: Population growth of replicators $s_{R_0}, s_{R_1}, s_{R_2}, s_{R_3}$ and s_{R_4}. Each line represents the average concentration of corresponding replicase over 30 simulation runs.

to react only with a progressively more restricted set of substrate molecular species. This impacts directly on these molecules' binding specificity.

Replicase	Broadcast device
s_{R_0}	$*\nabla : \nabla$
s_{R_1}	$*\nabla 1 : \nabla 1$
s_{R_2}	$*\nabla 01 : \nabla 01$
s_{R_3}	$*\nabla 101 : \nabla 101$
s_{R_4}	$*\nabla 0101 : \nabla 0101$

Table 5.2: (self-)replicases with increasing binding specificity

The results depicted in Figure 5.2 suggest the potential role of binding specificity in encouraging the domination of replicase species. The ability of a (self-)replicase to dominate the reaction space, in which a random initial population of molecules is generated, increases progressively with its binding specificity. Fig.5.3 depicts the growth of the replicases in each of the 30 simulation runs conducted independently in the 5 experimental series.

Fig.5.3 shows that as the replicases' binding specificity increases, the number of simulation runs in which x_1 reaches 1 increases accordingly. When non-universal

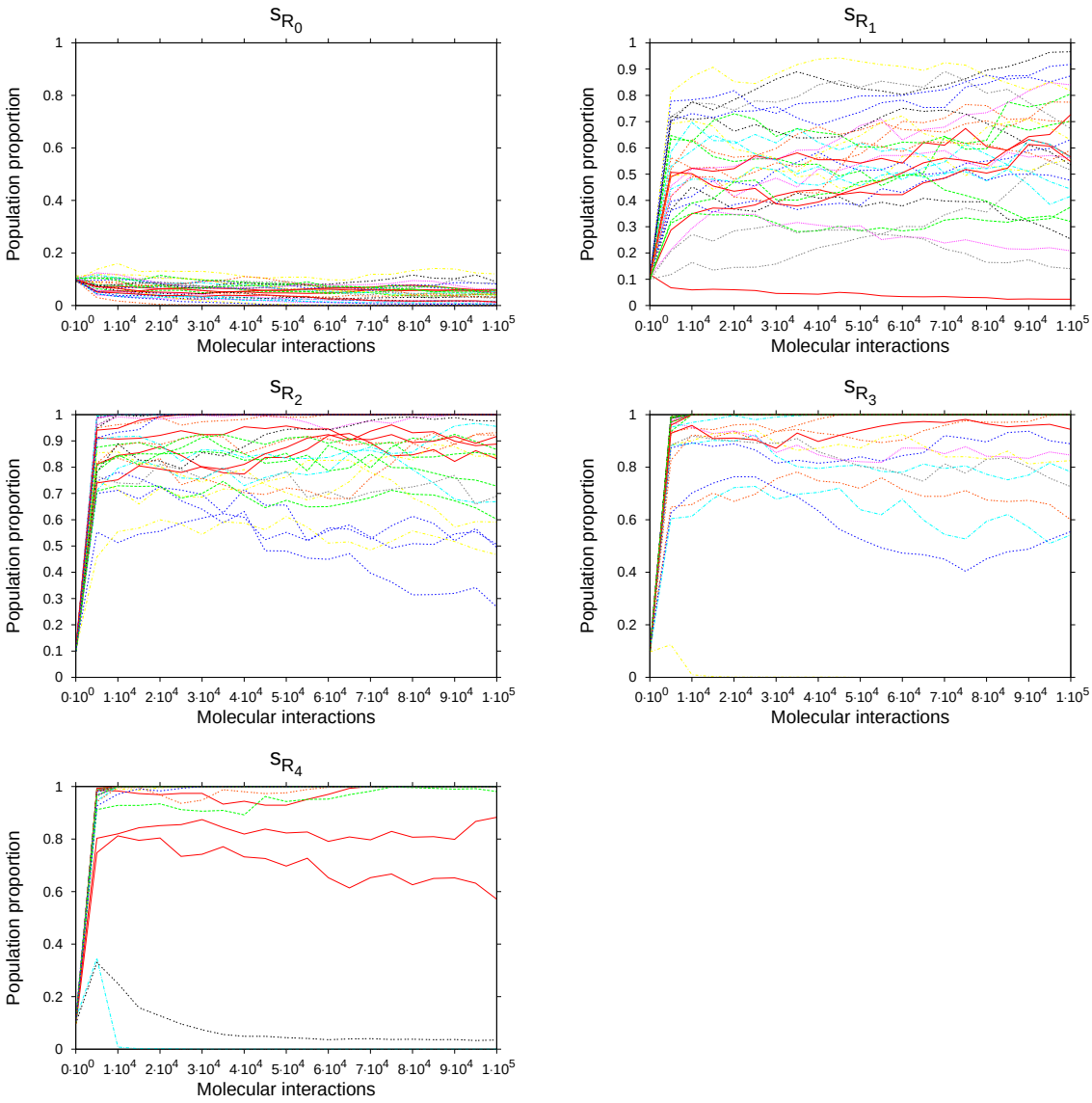

Figure 5.3: 5 series of experiment in which the binding specificity of the seed replicase molecules was incrementally increased using s_{R_0}, s_{R_1}, s_{R_2}, s_{R_3} and s_{R_4}. For each experiment, 30 simulation runs were conducted.

replicases are employed, an initial increase of x_1 is commonly observed. However this increase in x_1 appears, in particular simulation runs, to be followed by a random drift dynamics (observed at different x_1 level), preventing the replicases to fully dominate the reaction space, echoing the results presented in Section 5.3.

As in the previous section, we examine and explain this phenomenon through the use of an ODE model. To facilitate our investigation, this analysis employs a simplified model illustrating a *best case* scenario. Although being less intricate,

this model contains the core elements of the system employed in one of the above experiments (where s_{R_1} species are employed). In this model, we consider a reactor containing only the following molecular species:

- Replicases s_{R_1} which only replicate molecules terminating with the symbol "1" (which includes s_{R_1} molecules themselves).
- A variety of non-enzymatic molecules S_{NE} which are randomly generated. $S_{NE_1} \subseteq S_{NE}$ is the subset of molecules whose strings terminate with the designated symbol. These molecules contained in S_{NE_1} can be replicated by molecules s_{R_1}.

The concentration vector is given by $x = (x_1, x_2, \ldots, x_n)$ with $x_1+x_2+\ldots+x_n = 1$ where x_1 is the concentration of s_{R_1} and x_2 is the sum of concentrations of molecules in S_{NE_1}. The growth rate of the different molecular species in this reactor are as follows:

$$\dot{x_1} = x_1^2 - x_1(x_1^2 + x_1x_2) \tag{5.7}$$

$$\dot{x_1} = x_1^2 - x_1^3 - x_1^2x_2$$

$$\dot{x_1} = x_1^2(1 - x_1 - x_2) \tag{5.8}$$

The growth rate of molecules S_{NE_1} is:

$$\dot{x_2} = x_1x_2 - x_2(x_1^2 + x_1x_2) \tag{5.9}$$

$$\dot{x_2} = x_1x_2 - x_1^2x_2 - x_1x_2^2$$

$$\dot{x_2} = x_1x_2(1 - x_1 - x_2) \tag{5.10}$$

Since $x_1 + x_2 + \ldots + x_n = 1$, we have $x_1 + x_2 < 1$ and therefore $\dot{x_1} > 0$ and $\dot{x_2} > 0$.

$S_{NE_2} = S_{NE} - S_{NE_1}$ is the set of non-enzymatic molecules species that cannot be replicated by s_{R_1}. Let us set $x_3 = \sum_{i=3}^{n} x_i$, $\dot{x_3}$ is the growth rate of species S_{NE_2}

and is given by:

$$\dot{x_3} \;=\; 0 - x_3(x_1^2 + x_1 x_2) \tag{5.11}$$

In Eq. 5.11, we note that any given molecules $s_i \in S_{NE_2}$ possess a negative growth rate which indicate that these molecules would deterministically be displaced by molecules s_{R_1} and S_{NE_1}.

The displacement of S_{NE_2} molecules allows both s_{R_1} and S_{NE_1} to increase in concentration. However when S_{NE_2} molecules are fully displaced, we then obtain $x_1 + x_2 = 1$ (in contrast to the initial $x_1 + x_2 < 1$ condition). The system's dynamics are now equivalent to those described in Section 5.3 where both species have a common null growth rate. In this case where a random drift dynamics applies, the species having the highest relative concentration is more likely to displace the other one.

The replicases' binding specificity limited the initial concentration of S_{NE_1} species (resulting from the random initialisation of the molecular population). As the replicases's binding specificity increased, the initial concentration of S_{NE_1} decreased. Consequently the replicases having a higher *relative* specificity possessed a higher chance to displace the parasitic species (once S_{NE_2} species were first fully displaced).

The replicases' binding specificity conditioned the initial concentration of parasitic species which explains the behaviour observed in Figure 5.2, in which replicases with higher specificity are more likely to take over the reactor space. Therefore in this system, for replicase molecules to successfully dominate a randomly generated molecular population, a significant binding specificity is required.

Regarding the related spontaneous emergence and domination of autocatalytic species in the MCS.bl, given a set of randomly generated molecular species, it is progressively more difficult for self-replicases of higher specificity to *spontaneously*

arise by chance (due to their greater length, and relatively rare frequency as defined by the BL syntax). The spontaneous emergence of a "sustainable" self-replicase (i.e., of sufficient specificity to establish itself) remains theoretically possible in MCS.bl. However, both the experimental results and the analysis presented here suggest that the expected emergence time would be extremely (perhaps infeasibly) long. While we have not formally quantified this, it appears that MCS.bl therefore shares this property with the Tierra system.

It is conjectured that this binding specificity property may have been implicated in the dynamics of a variety of previously reported artificial chemistries. For example, Fontana and Buss reported (in level 0 Alchemy experiments) the recurrent emergence and domination of the universal replicase $\lambda x.x$ (Fontana and Buss, 1994a). Nevertheless Fontana and Buss also mentioned that if non-enzymatic species are not filtered out then an accumulation of such inert species would occur. This observation suggests that the lack of binding specificity may have, as well, affected the system's dynamics where non-enzymatic species prevented the domination of emerging universal replicases. As suggested, this underlying phenomenon has potentially been involved in previous AC-based research. Nevertheless it has never been examined and explicitly isolated in the manner presented here.

Finally, we may also consider the potential effects of molecular mutations. Even though replicase species having a significantly high binding specificity are employed, mutations may lead to the emergence of non-enzymatic mutant species possessing the replicases' tag (enabling these mutants to be replicated by the replicases). The replicases' binding specificity, which was originally high, would thus become again *relatively* low or null. As a result, the replicases' binding specificity may not prevent *potentially disruptive* effects from occurring. To test this hypothesis, we carry out a final evolutionary experiment "a la Tierra" in which a hand-designed self-replicase (i.e., an ancestor molecule) having a "high" binding specificity capable of self-sustaining is employed. This experiment is presented in the next section.

5.5 Rise and fall of the fittest

In the Tierra system, a hand-designed molecule called the "ancestor" is manually introduced into the space. This initially grows to saturate the available core memory. The population subsequently evolves into a variety of collectively autocatalytic reaction networks (where Tierra "creatures" or programs are here considered analogous to "molecules"). Accordingly, our next step is to mirror this methodology, and introduce a hand-designed self-replicase of relatively high specificity into the MCS.bl system.

Figure 5.4 presents an example of such an experiment in which ancestor molecules are manually introduced. The results indicate that MCS.bl does *not* exhibit an evolutionary dynamic at all comparable to Tierra in this case. This evolutionary dynamic was moreover systematically observed in repeated simulation runs. The ancestor self-replicators do, at first, quickly dominate the reaction space, just as expected. However, this population immediately collapses again. The average molecular length then increases dramatically, while the overall reaction rate (indicating the average rate of binding between random molecules in the population) also collapses. In this particular run, molecules are arbitrarily limited to a maximum length of $BD_{Lmax} = 1.0 \times 10^6$.

As with the experiments discussed earlier, these results were not expected. In fact, certain mutants of the original autocatalytic molecule developed a distinct advantage over the ancestor. That is, these mutants could be replicated by the ancestor molecules but only to the cost of these ancestors, i.e., an asymmetric relationship. Moreover, some of these mutants also lose their ability to self-replicate, explaining the rapid decrease in the global number of self-replicases. By exploiting their molecular tag and the ancestors, these non-autocatalytic molecules succeed in displacing the dominant ancestors.

To illustrate this phenomenon, we present a simple example of such a case in

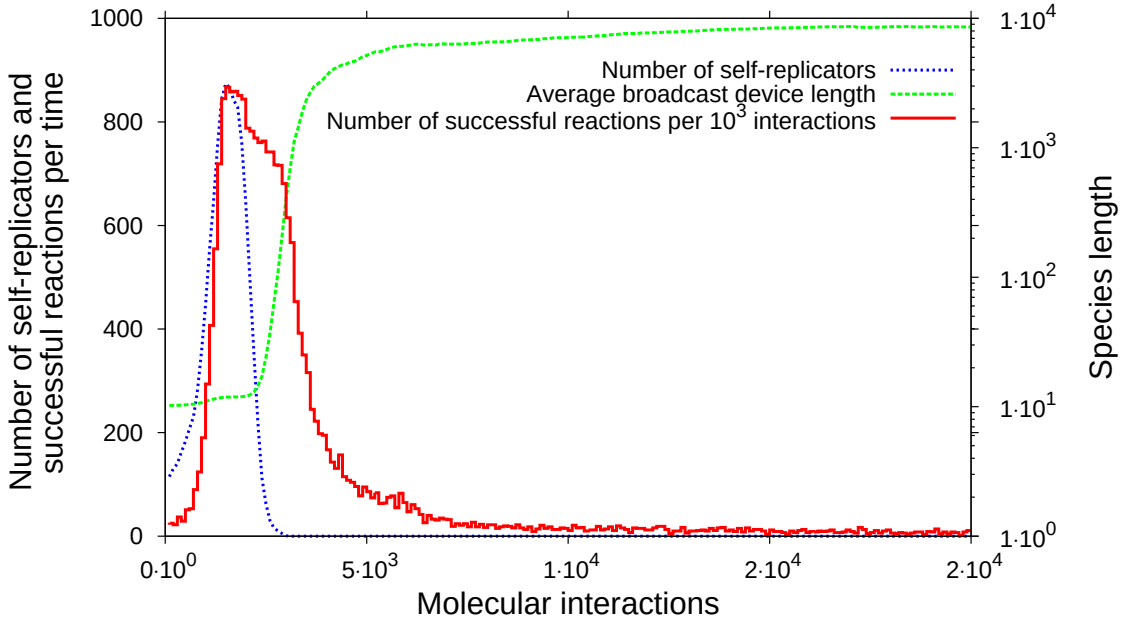

Figure 5.4: Effects of molecules length growth upon overall system reactions rates. In this experiment, an ancestor ($s_{R_4} = \nabla 0101 : \nabla 0101$) is inserted (with initial concentration $[s_{R_4}] = 0.1$) in addition to randomly generated molecules. Moreover mutation per molecule and per symbol is turned on with $p_{sym} = 0.001$ and $r_{mut} = 0.001$.

which we define two molecules: $s_{R_4} = *\nabla 0101 : *\nabla 0101$ and $s'_{R_4} = *\nabla 0101 : *\nabla 00101$. The latter is a readily accessible mutant of s_{R_4}. Once it appears, the mutant s'_{R_4} allows for a runaway degenerative scenario to occur. The possible reactions between species s_{R_4} and s'_{R_4} are as follows:

$$
\begin{aligned}
s_{R_4} + s_{R_4} &\longrightarrow 3s_{R_4} \\
s_{R_4} + s'_{R_4} &\longrightarrow s_{R_4} + 2s'_{R_4} \\
s'_{R_4} + s_{R_4} &\longrightarrow s_{R_4} + 2s'_{R_4} \\
s'_{R_4} + s'_{R_4} &\longrightarrow 2s'_{R_4} + s''_{R_4}
\end{aligned}
$$

The product s''_{R_4} is of the form $*\nabla 0101 : *\nabla 000101$ and similarly has a selective advantage over both s_{R_4} *and* s'_{R_4}. The reaction $s''_{R_4} + s''_{R_4}$ would result in the production of a molecule s'''_{R_4} of the form $*\nabla 0101 : *\nabla 00000101$ and clearly shows the

potential for unlimited elongation in molecule length. Of course, as molecule length increases, the per-molecule mutation rate also increases, leading to progressively more frequent disruptive changes to molecular structure. The observed consequences of these mutation effects are twofold:

- Molecules may become inactive (i.e., lose all enzymatic activity). This is a direct consequence of the BL syntax. A mutation leading to the removal or insertion of structural symbols such as $*$ or : would "break" the active site. This degenerative effect may be regarded as a consequence of syntactic "brittleness" of BL.

- The binding specificity may be increased. This arises when mutations lead to the insertion of informational symbols such as $0s$ and $1s$. As a result, although some molecules may still possess an active site capable of some enzymatic function, their high specificity decreases the variety of target molecules that it can bind to; ultimately meaning there may be few, if any, functional targets for it left in the population.

Both of these phenomena result in a continual decrease in the overall reaction rate until reactions effectively cease completely (i.e., system death). Figure 5.5 summarises this cascade of events. Note that this system level degeneration (the "elongation catastrophe") occurs precisely because of the stepwise emergence of molecules which are progressively "fitter" at the molecular level. Nevertheless this notion of fitness differs from its meaning usually implied in Artificial Chemistries. In typical ACs, molecular species or more specifically "digital organisms" improve in fitness by evolving their intrinsic properties. Such agents become fitter by incrementally ameliorating their ability to perform a target task, e.g., competing for CPU resources in Tierra. Here the molecules are fitter, not due to their intrinsic functionality, but due to their ability to exploit their interactions with the other molecules. This interpretation of fitness applies specifically in the case of catalytic reaction networks

in contrast with the more general class of self-replicating multi-agent systems.

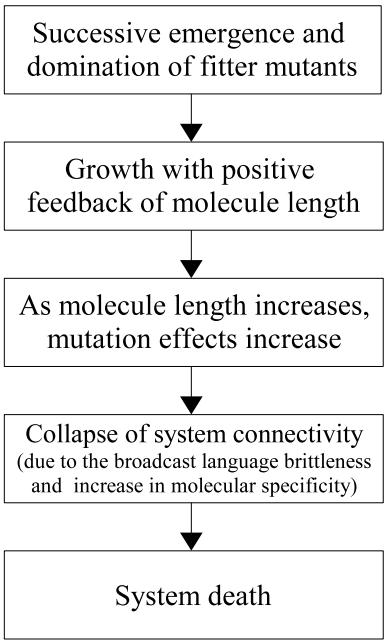

Figure 5.5: Elongation catastrophe in MCS.bl

5.6 Solving the elongation catastrophe problem

In this section, we describe different qualitative modifications conducted on the MCS.bl, which are aimed at preventing the elongation catastrophe from occurring. These various technical modifications are directed at limiting the string length of product molecules. Following this, the different outcomes are briefly presented.

1. The multiple symbol wildcard $\triangledown$ was altered so that it would not be able to pass an unlimited number of symbols from the input molecule (substrate) to the output molecule (product). An integer parameter $1 \leq c \leq c_{max}$ represents the number of symbols that can be matched and passed by $\triangledown$, i.e., the capacity of the wildcard. This capacity may be subjected to a form of "parametric" mutation, where its value would change randomly in $[1, c_{max}]$ over time.

2. Similarly to (1), a finite total number of "free" symbol objects (monomers) available in the reactor was defined. This reservoir of (untyped) monomers is

reduced when new molecules are produced, and increased when molecules are destroyed. If insufficient monomers are available to complete a reaction, the reaction fails. This should favour smaller molecules over longer ones suffering from elongation catastrophe.

3. Proposal (2) was extended, further constraints were defined to limit the number of particular symbols available in the universe. Different arbitrary symbol distributions were employed (e.g. structural and informational symbols such as $*$,:,1,0 could be made more common than multiple symbol wildcards such as ∇).

4. Another extension to (3) was to introduce a probability of successful reaction which would depend on the product's length. Smaller molecules could then be given a selective advantage over the longer ones.

In summary, the above system changes generally produced one of the following outcomes:

- Did not prevent the elongation catastrophe.

- The system evolved towards a population of inactive and relatively small ([1−4] symbols long) molecules. The system's global success reaction rate was also close to zero. Such an accumulation of inactive or inert molecules was also reported to occur in Alchemy when inert molecules were not filtered out.

- The system converged towards a population where enzymatic molecules were still present but could not react with any other molecules present in the reaction space. The specificity continuously increased until no further reactions occurred.

Thus, although a range of modifications were implemented, the different outcomes do not differ substantially from the degenerative cases presented above (Section 5.5).

5.7 Conclusion

We conducted a series of experiments using the MCS.bl system. These experiments focused on the emergence, self-maintenance and evolution of closed molecular organisations.

We first reported an experiment focusing on the spontaneous emergence of autocatalytic species in the MCS.bl. The results showed that, autocatalytic species do emerge but cannot significantly grow in concentration and eventually get displaced by other molecules. To explain this phenomenon, a number of potential explanations were proposed and developed in following experiments.

The second experiment introduced minimalist self-replicases that were manually inserted in the reaction space in addition to randomly generated molecular species. Using a simplified model, we showed that these species being universal replicases do not possess any selective advantages over non-enzymatic species. Under the stochastic conditions of the reactor this would yield a random drift in relative concentrations of universal replicases and non-enzymatics species.

We then exposed the effects of the replicases' binding-specificity over the system's dynamics. We showed that this binding specificity conditioned the capability of the replicases to displace the other molecular species present in a randomly generated molecular population. As the replicases' binding specificity increased, the capability of the replicases to displace the other species increased accordingly.

Similarly to the Tierra system, we hand-designed an ancestor molecular species, which possessed a high binding specificity ensuring its ability to self-maintain. We described the results obtained from an evolutionary experiment in which these ancestor molecules were employed. Our results presented unexpected degenerate evolutionary dynamics in which the closed reaction networks were not able to self-maintain. This degenerate scenario was due to an elongation catastrophe phenomenon. To address this robustness issue, we explored several model variants

which all failed to prevent the degenerate outcomes to occur. These results indicated counter-intuitive outcomes when compared with a variety of other AC systems in the literature.

In the next chapter, we explore additional models of the MCS.bl which are implemented on a parallel architecture. These models aim at enhancing the robustness and subsequently the evolvability of the MCS.bl system. On top of providing computational benefits, a distributed implementation allows us to explore novel system models. These modifications introduce a second level of selection at the reactor level. These extended MCS.bl systems are thus multi-level selectional models where selection occurs at both the molecular and reactor level.

Chapter 6

Evolutionary Capability in Multi-Level Selectional Models

The preliminary experiments presented in Chapter 5 suggested that the MCS.bl cannot support the self-maintenance of closed reaction networks when subjected to perturbations (i.e., mutations). Consequently, this lack of robustness prevented us from evolving closed reaction networks using the MCS.bl. Although several model variants were proposed to address these robustness and evolvability issues, none of these attempts inhibited the degenerative evolutionary dynamics.

In keeping with this effort to improve the MCS.bl framework, we propose further system modifications. However, contrary to previous proposed variants which focused on modifications at the molecular level, we suggest a novel MCS.bl implementation which introduces new features at the container level. This MCS.bl variant is implemented as a parallel system using distributed computing facilities. The latter were provided by both the ESIGNET project and the Irish Center for High-End Computing (ICHEC). The material presented in this chapter was partially published in several conference articles (Decraene et al., 2008b,a; Decraene, 2009).

6.1 Introduction

We present a novel implementation of the MCS.bl which exploits distributed computing facilities. On top of providing benefits in terms of experimental scalability (i.e., we can simulate more molecular species), employing high-computing resources enables us to explore a wider repertoire of AC models. Particularly, we identify two approaches involving compartmentalisation. In these models, molecular species are contained in compartments which result in *multi-level selectional* dynamics (Hogeweg and Takeuchi, 2003). Selection applies at both the molecular and compartment levels. These models presented significant results with regards to the self-maintenance of closed reaction networks when subjected to perturbations:

1. *Static reactors with molecular diffusion.* McCaskill et al. (2001) addressed evolutionary degeneration issues in a spatially resolved stochastic system in which molecular species are contained in distinct compartments. The migration or diffusion of molecules between compartments could occur given a specified diffusion coefficient.

   This model considered the effects of mutant species which relied on the host (non-mutant) autocatalytic species to be replicated. This form of parasitism destabilises the system's dynamics and prevents the self-maintenance of the closed reaction network. This degenerative phenomenon is similar to the elongation catastrophe described in Section 5.5, which was also caused by the emergence of such parasitic species. McCaskill et al. demonstrated analytically that for such a compartmentalised system, according to the level of mutation rate, a range of diffusion coefficients exists which enables stable cooperation to occur between the molecular species.

2. *Cellular model.* Similarly to McCaskill et al.'s approach, molecular species are contained in distinct compartments. However in cellular models, com-

partments have the ability to "grow and divide". A compartment grows (i.e., molecules are produced within the compartment) until a condition (specific to the cellular model) is satisfied. This condition triggers the division of the compartment. Half of the molecules are selected at random and transferred into another newly created compartment.

This division process is analogous to cellular division, i.e., we regard compartments as biological cells. It has been demonstrated that such cellular models enable the stabilisation and self-maintenance of closed reaction networks when subjected to parasitic phenomena (Szathmary and Demeter, 1987; Cronhjort and Blomberg, 1997; Hogeweg and Takeuchi, 2003). Here, parasitised compartments ultimately decay and may be replaced by non-parasitised ones which would result from the division of neighbouring compartments. Cellular models allow the isolation of parasited cells and prevents the invasion of parasite species over the whole cellular population.

We propose to investigate compartmentalisation and its potential benefits upon evolutionary capability in the MCS.bl system. Our distributed implementation of the MCS.bl will consequently incorporate compartmentalisation properties. Nevertheless we introduce a supplementary feature which addresses the concurrent nature of biochemical processes. Operations at the compartment level will be executed in a *genuine* parallel manner in contrast to traditional ABMs which rely on a synchronisation mechanism. Implementing the MCS.bl as a distributed/parallel system enables us to explore further selectional models which may resolve the lack of robustness presented in previous chapter. This work will also provide complementary insights on the effects of compartmentalisation and parallelism upon evolutionary capability in agent-based Artificial Chemistries.

In the remainder of this chapter, we present the details of our parallel architecture and its immediate effects on the MCS.bl. We then evaluate our compartmentalised

MCS.bl system in which molecular diffusion and cellular division are independently applied.

6.2 A parallel architecture

Agent-based models typically implement the computational agents on serial computers. Although being conceptually concurrent computational units, agents are not generally executed simultaneously in ABMs. A discrete clock mechanism is usually employed to synchronise the agents' interactions and state updates. The MCS.bl, presented in Chapter 4 and utilised in Chapter 5, employs a similar clock technique.

Ray (1992), Adami and Brown (1994) attempted to address parallelism by devising models in which agents are executed by multiple virtual CPUs. In Avida, each agent possessed its own memory space and was executed by its own virtual CPU (which speed may vary from other agents' CPUs). Whereas in Tierra, the multiple virtual CPUs executed, in turn, the code (the agents' genotypes) present in the virtual computer memory shared by the different agents. Nevertheless, as in most ABMs, a time-slicing algorithm was employed to simulate the parallel computational processes in both systems. Lenski et al. (2003) also conducted a series of experiments using Avida on a grid computer; but Lenski et al. still relied on a pseudo-parallel system where parallelism was only simulated.

We propose a parallel approach to ABMs, applied here to the MCS.bl model. A major difference between this extended MCS.bl and Tierra/Avida is that multiple compartments (each of which contains a population of molecular species) are introduced. This compartmentalisation is the key feature by which parallelism is addressed in the extended and distributed version of the MCS.bl:

- Molecular species are contained in distinct compartments each of which possesses a fixed maximum carrying capacity of n_{max} molecules. Within the compartments, molecular interactions are processed in a sequential manner (as in traditional ABMs) through the use of a time-slicing algorithm. Each com-

partment (conceptually regarded as a meta-agent) is executed on individual distinct CPUs in a parallel fashion.

- At the compartment level, we do not introduce any synchronisation mechanisms. Meta-agents are executed in parallel on separate CPUs. These meta-agents may interact with each other by communicating signals (the effects and nature of these signals are dependent of the specific compartment model utilised and will be presented later).

As introduced earlier, devising these compartment properties enables us to investigate further models which may resolve the MCS.bl's evolutionary degeneration issues.

6.2.1 Implementation

Simulations using the distributed/compartmentalised MCS.bl are run for a pre-specified length of time (defined in seconds). In a given simulation run, the number of molecular collisions/interactions occurring in each compartment may vary significantly. This variance depends on the level of computations and communications occurring in the compartments during the run.

The integer number N of compartments (i.e., CPUs) used in a simulation is fixed. Compartments are identified by a unique ID number ($1 \leq ID \leq N$). $C = c_1, \ldots, c_N$ is the finite set of all compartments occurring in the universe. Each compartment initialises simultaneously their respective random number generator[1]. These random number generators rely on random seeds which are determined by using the server local time value multiplied by the unique compartment ID.

The Message Passing Interface (MPI) is employed to handle the communications between the different CPU nodes/compartments. A simplex topology is utilised to

[1]The MCS.bl implementation employs the GNU Scientific Library random number generator. The latter is by default based on the Mersenne twister (MT19937) generator of Matsumoto and Nishimura (1998) and has a cycle length of $2^{19937} - 1$.

condition the interactions between compartments, i.e., the distance between any two-compartments is equal.

Compartments may communicate signals to other compartments. Signals are composed of a vector of molecules. Given a compartment c_1, if a condition (specific to the compartment model utilised) is satisfied then a communication occurs. c_1 selects the target compartment c_2 at random. Following this, c_1 sends a *non-blocking* signal to c_2 (i.e., c_1 continues to operate without having to wait for c_2 to receive the signal; this prevents communication deadlocks from happening). The exact nature and effects of the signals will be identified in the models' descriptions below.

When a compartment checks its "mailbox" for incoming signals, several signals may have been received. MPI buffers the incoming signals in an orderly manner (first in first out queue). At each timestep, a single incoming signal is processed by the compartment. Fig. 6.1 depicts this algorithm which is run simultaneously on each compartment processes.

6.2.2 Introducing chemical kinetics

In the non-parallel version of the MCS.bl, all reactions occurred in a sequential manner. A clock mechanism was utilised to sequence the processing of molecular interactions. At each discrete time step a single molecular collision/interaction occurred. Elastic and catalytic reactions would be equivalent from a temporal point of view. The computational time necessary to process a reaction was not accounted for and did not affect the system's dynamics. Although occurring in a simplified and constrained form, chemical "kinetics" were present in the original MCS.bl and dictated all reaction rates to be equal. This constraint is now being relaxed as follows.

In the distributed implementation of the MCS.bl, several reactions may be occurring *simultaneously*. Although molecular interactions are processed sequentially within a compartment, N reactions may potentially be processed at the same time

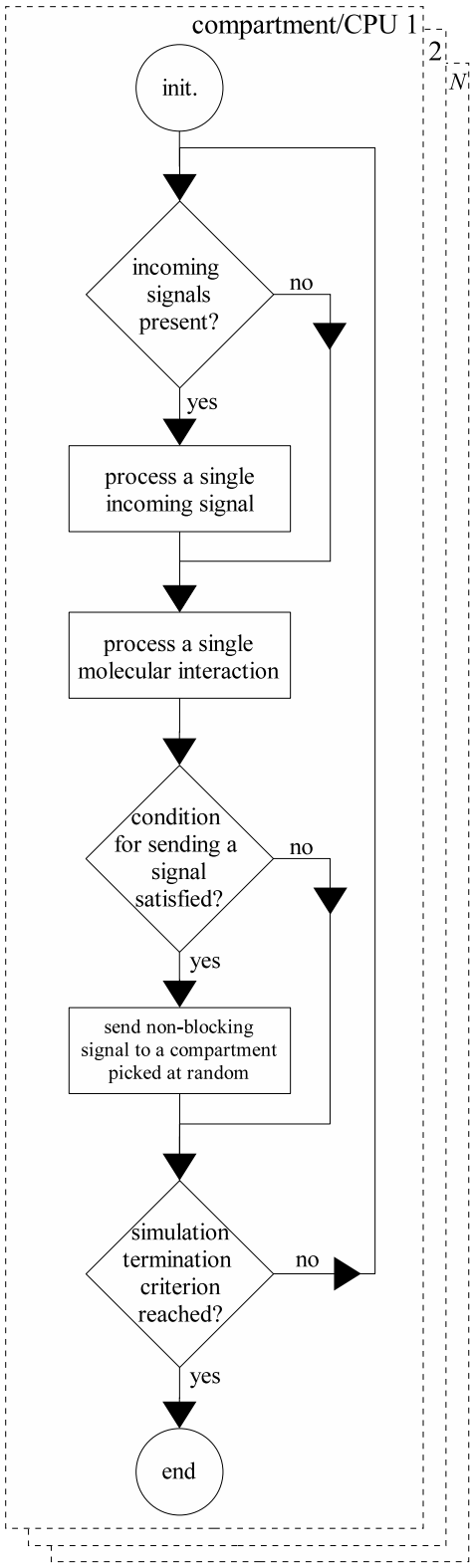

Figure 6.1: N compartments/processes running in parallel

throughout the N compartments. Catalytic reactions may result in the modification of the compartments' states. According to the state of a given compartment, a signal can be emitted and addressed to another compartment. These signals may further change the states of both the emitter and receiver compartments. Consequently the speed of chemical reactions occurring in the compartments may indirectly alter the system's dynamics at the compartment level.

Parallelism in this distributed implementation of the MCS.bl altered the system's chemical kinetics and introduced variations in the reaction rates.

6.3 Static reactors with molecular diffusion

We evaluate a first extended MCS.bl model in which both compartmentalisation and molecular diffusion are addressed.

6.3.1 Introduction

An analytically tractable model was proposed to address the inhibition of degenerative effects due to parasitism in closed reaction networks (McCaskill et al., 2001). McCaskill et al. demonstrated that given a compartmentalised model where migrations/diffusions of molecular species occur between the compartments, there exists, according to some parameters (i.e., the replication error rate and decay rate), an optimal diffusion coefficient which allows for the cooperation of molecular species and subsequently stabilises the self-maintaining cycle of the closed reaction network.

The following minimalist model (in which such disruptive parasitic effects are occurring) was employed by McCaskill et al. to examine the effects of compartmen-

talisation:

$$
\begin{aligned}
2X &\xrightarrow{Q} 3X \\
2X &\xrightarrow{R} 2X + Y \\
X + Y &\xrightarrow{2r} X + 2Y \\
X, Y &\xrightarrow{d} \emptyset
\end{aligned}
$$

X is an autocatalytic species which is being parasitised by Y. The latter is a mutant species of X whose replication can only be catalysed by X. Q is the fidelity rate of replication of X. $R = 1-Q$ is the replication error (mutation) rate leading to the production of mutant species Y. r is the recognition coefficient which determines the rate of successful bindings between X and Y. The spontaneous decay rate of X and Y is denoted by d. This decay rate is fixed and independent of the production rate Q. The compartments possess a maximal carrying capacity of n molecules. The diffusion coefficient between the compartments is denoted by m. The latter is the key parameter enabling the inhibition of parasitic effects. However m depends on the above various parameters for which minimal and maximal threshold values exist (e.g., the mutation rate has to be *tolerable*).

Through the use of a stochastic modelling technique (Master equation) which accounted for space, McCaskill et al. showed that a range of diffusion coefficients existed and permitted the survival of the catalyst X, despite being parasitised by Y. This range was dependent on the mutation and decay rates. As the mutation or decay rate increases, the range of diffusion coefficient decreases until it disappears.

A number of significant differences exist between McCaskill et al.'s stochastic model and the MCS.bl:

- The MCS.bl is a constructive AC where a rich variety of molecular species exists. If we consider a maximal species string length of molecular species of

$BD_{Lmax} = 500$, then there are 8^{500} possible combinations of molecular species, the nature of whose interactions may vary greatly. In McCaskill et al.'s model, only two species and five reactions are considered.

- The parasites involved in the degenerative evolutionary dynamics (Section 5.5) are capable of enzymatic functions. When these particular parasites react with copies of themselves, a further new enzymatic species is generated. We showed that this newly created species is also capable of parasitising both the parent parasite species and ancestor/seed catalysts. Whereas in McCaskill et al.'s model, the replication of mutant species Y can only be catalysed by X.

- Mutations are implemented differently in the MCS.bl. A per-symbol mutation probability is employed. As the length of a species increases, the probability of a mutation to occur increases accordingly. McCaskill et al. employed a fixed mutation rate which is not affected by the nature of the molecular species.

- Further differing details exist between the MCS.bl and McCaskill et al.'s model. These points will be addressed in the description of the extended MCS.bl model (Section 6.3.2).

Although significant differences exist between both models, the parasitic effects and degenerative phenomena studied in the original non-compartmentalised models are essentially similar: The exploitation of catalytic species by mutant parasitic species disrupts the self-maintaining cycle of the closed reaction network.

Using McCaskill et al.'s work on compartmentalisation and evolutionary capability as an inspiration, we propose an extended version of the MCS.bl which accounts for these compartmentalisation and diffusion properties. This extended MCS.bl will then be employed to conduct a set of evolutionary experimentations which will illuminate the effects of compartmentalisation and molecular diffusion upon the system's evolutionary capability.

6.3.2 The model

We present an extended MCS.bl model which accounts for compartmentalisation and molecular diffusion:

- In contrast to the model employed in Chapter 5, successful molecular reactions (generating new molecules) do not lead to the removal of another molecule at random when the reactor is full. Here, a molecular reaction creating a new molecule occurs only if the carrying capacity n_{max} of the reactor has not been reached. Reactors may frequently be filled when the molecular production rate is higher than the decay rate.

- Moreover a decay probability d is defined, this parameter addresses the continuous decay rate defined in McCaskill et al.'s approach. At each discrete timestep, a molecule is selected at random and may be removed, with the probability d, from the compartment. This spontaneous decay of molecules enables further reactions to occur in filled (i.e., saturated) reactors.

- In McCaskill et al.'s model, molecular diffusions occurred continuously in time according to the diffusion coefficient m. However time is discretised in the MCS.bl. Consequently exchanges of molecules occur in a sequential fashion according to the time-slicing algorithm. A probability of molecular diffusion p_m is introduced. At each timestep, an exchange of molecules between two compartments may occur with the probability p_m. Devising a probability instead of a fixed time interval was decided to avoid all compartments diffusing at the same time (in which cases the traceability of the system would not have been facilitated). Although the compartments are executed in parallel, all compartments could simultaneously diffuse in some cases, e.g., during the early phase of a simulation run where seed molecules and the nature of computations are equivalent in all compartments.

- When a compartment $c_i \in C$ diffuses, some molecules contained in c_i are selected at random. These selected molecules are removed from c_i and constitute the *diffusion initiation signal*. The size of this signal/vector of molecules is determined by the integer division of the container's current size c_{i_n} by the diffusion coefficient m, denoted by $\lfloor \frac{c_{i_n}}{m} \rfloor$. This signal is then sent from c_i to another compartment $c_j \in C$ which is selected at random. c_i continues to operate following this signal emission. When c_j receives the diffusion initiation signal from c_i, c_j similarly selects and removes a number of molecules according to $\lfloor \frac{c_{i_n}}{m} \rfloor$ (with c_{i_n} and c_{j_n} being potentially different). The molecules received from c_i are then inserted in the c_j compartment. The molecules removed from c_j constitutes the *response signal* which is sent back to c_i. Upon receiving this response signal from c_j, the molecules contained in this signal are similarly inserted into c_i. This concludes the exchange of molecules between c_i and c_j.

- The number of molecules exchanged between compartments is not fixed. Moreover this exchange of molecules may be asymmetric in some cases. Molecular exchanges are symmetric only when the two "communicating" compartments contain the same number of molecules. If we consider a molecular exchange between c_i and c_j with $c_{i_n} > c_{j_n}$ then c_i would emit a higher number of molecules than c_j with $\lfloor \frac{c_{i_n}}{m} \rfloor > \lfloor \frac{c_{j_n}}{m} \rfloor$. If $c_{i_n} > 0$ and $c_{j_n} = 0$ then molecules would be diffused only from c_i to c_j. Diffusion equilibrium may be reached during subsequent molecular exchanges when $c_{i_n} = c_{j_n}$.

In the following section, we conduct a series of evolutionary experiments using the above extended MCS.bl model.

6.3.3 The experiments

We present a series of experiments using the extended parallel version of the MCS.bl. These experiments aim at demonstrating the effects of compartmentalisation and molecular diffusion upon the system's evolutionary capability.

The differences outlined earlier prevent us from mapping directly the various parameters into the MCS.bl. As a result we cannot determine analytically the range of suitable parameters, if any, which may allow for the stabilisation of evolution in the MCS.bl. Nevertheless we attempted to select parameter values that would not *clearly* facilitate the disruptive parasitic effects to occur, e.g., p_{sym} was deliberately set to a relatively low value to diminish the frequency of emergence of potentially parasitic mutant species. The following set of fixed parameters is proposed and utilised in all experiments:

- 30 compartments are utilised and executed in parallel using 30 AMD Opteron 270 (2.0 GHZ) CPUs.
- Experiments are run for 3600 seconds (1 hour of "wall clock" time).
- The maximal compartment carrying capacity is $n_{max} = 1000$.
- The diffusion probability is set to $p_m = 0.05$.
- The spontaneous decay probability is set to $d = 0.1$.
- Similarly to experiences conducted in Chapter 5, the maximal species string length is set to $BD_{Lmax} = 500$.
- The per-symbol mutation is set to $p_{sym} = 1.0 \times 10^{-5}$.
- The global spontaneous mutation rate is set to $r_{mut} = 0$.
- As in the evolutionary experiments presented in Section 5.5, each compartment is seeded/initialised with the $s_{R_4} = \nabla 0101 : \nabla 0101$ molecular species. However we completely fill each compartment with 1000 instances of s_{R_4} species.

Considering the MCS.bl experiments conducted in Chapter 5, McCaskill et al.'s investigation and the specification of the extended MCS.bl, we propose the following ideal scenario/prediction where degenerative evolutionary outcomes may be inhibited due to compartmentalisation and molecular diffusion in the extended MCS.bl:

1. In this model, compartments which are infected by parasites would present a lower rate of successful catalytic reactions (as shown in Section 5.5).

2. As a result the molecular production rate would decrease accordingly.

3. If the production rate becomes lower than the decay rate then the compartment would start depleting.

4. Molecular diffusions occurring between non parasitised compartments (in which no depletion occurs) and infected/depleting ones would lead to asymmetric molecular exchanges, i.e., infected compartments would import more non-parasitic species and export less or no parasitic species (limiting the spread of parasites throughout the compartments).

5. Consecutive molecular exchanges would allow for the persistence and spreading of the non-parasitic catalysts throughout the compartment population.

In this idealised scenario, compartmentalisation would allow for the isolation of infected compartments. Consequently the closed reaction networks would be able to self-maintain when subjected to disruptive parasitic effects.

To test the above prediction, we conduct three series of experiment (where 5 simulations were run in each series) in which the following diffusion coefficients are employed: $m_1 = 0.01$, $m_2 = 0.05$ and $m_3 = 0.1$.

Fig. 6.2 present an overview of the dynamics of an example simulation run where the diffusion coefficient is set to $m_1 = 0.01$. We identify the following chain of events:

1. We first note an initial phase where the system is stable with an average species length of 12 symbols (i.e., the length of s_{R_4} species) and the average population size stagnating at nearly 1000 molecules. Most compartments are thus full during this phase (i.e., the molecular production rate is higher than the decay rate).

2. However from $t \approx 950$ we note that the average length of species starts to increase rapidly. This behaviour suggests that the elongation phenomenon is occurring.

3. Then 150 seconds later, at $t \approx 1100$, we observe a rapid decrease in the average population size throughout the 30 compartments. This indicates that the production rate has now become smaller than the decay rate, i.e., the compartments are depleting.

4. Nevertheless this decrease does not apply to the species length which continues to increase until $t \approx 1270$, where the species length reaches its peak with an average length of 278 symbols. During the phase $950 < t < 1270$, successive species having an increasing length emerged and displaced each other. As a

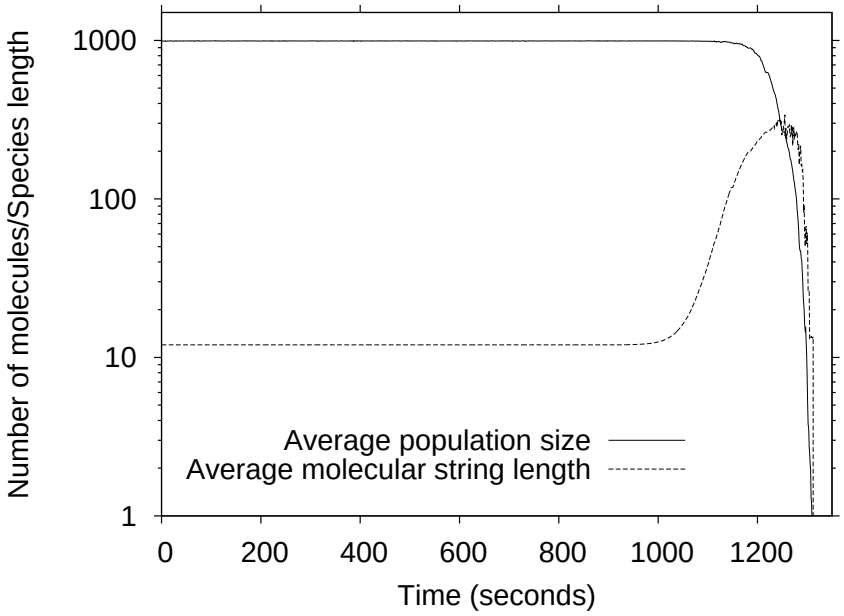

Figure 6.2: Dynamics of the compartments' average population size and molecular species string length (averaged over the species length averages of each compartment) with diffusion coefficient $m_1 = 0.01$. We depict the dynamics of the molecular population size as it provides an approximate indication about the molecular production rate against the decay rate. For example if the production rate is higher than the decay rate then the occupancy compartments would be maximised and would only fluctuate due to molecular diffusion. On the contrary if the production rate is lower than the decay rate then we would perceive a decrease in the average population size.

result we observe a linear increase in the average species length throughout all compartments.

5. Following the peak of the species length observed at $t \approx 1270$, the average species length is then rapidly decreasing (similarly to the average population size) until $t \approx 1310$ where the system becomes extinct, i.e., all molecules have decayed.

In Fig. 6.3, we focus on the extinction phase and present the standard deviations of the compartments' population size and species string length.

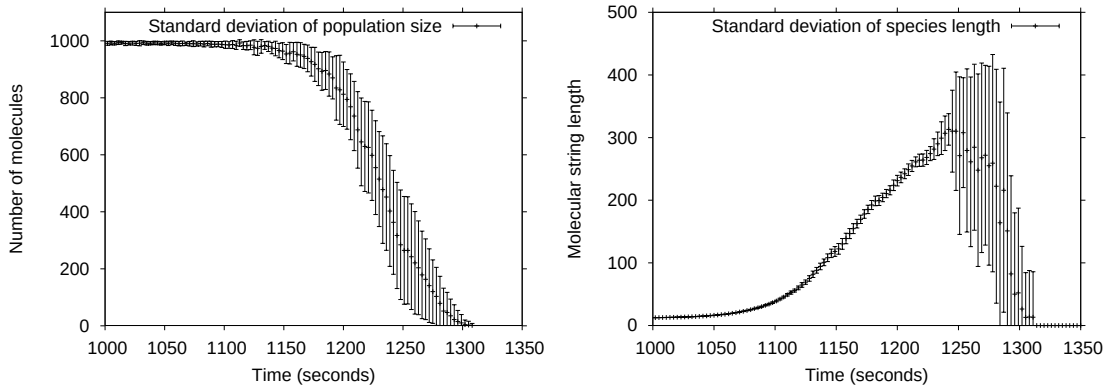

Figure 6.3: Extinction phase of example simulation 1 with diffusion coefficient $m_1 = 0.01$. For clarity purposes, both curves were plotted using a point interval of 3.

Fig. 6.3 complements the current analysis by providing more detailed information about the dynamics of each compartment:

1. During the early stable phase, we note that the molecular string length is relatively homogeneous throughout the 30 compartments with little variance occurring. This assertion applies for both the average population size and species string length.

2. However when $t \approx 1100$, we observe a divergence in the composition of the compartments (i.e., compartments with different population sizes exist). This variance is maintained until close to the end of the extinction phase.

3. Although the composition of compartments starts diverging at $t \approx 1100$, the average species length (which is globally increasing) is more or less homogeneous throughout the compartments until $t \approx 1230$. Thus during the phase $1000 < t < 1230$, the mutant species (here classified by their string length) are well diffused throughout the compartments. However as the average molecular population size is already decreasing (since $t > 1100$), it indicates that these mutant species have a production rate lower than the decay rate.

4. When $t > 1230$, a significant range of variances is observed in the species string length. This phenomenon suggests that only a few reactions leading to the creation of much longer species are succeeding in some compartments.

The above degenerative scenario is characteristic of the elongation catastrophe phenomenon. We first observed the successive emergence of mutant species having an increasing length. We distinguished successive displacements which led the system to a state where successful catalytic reactions occurred less often. However a decay rate is applied in the current system, and consequently the compartments started to deplete as the production rate became lower than the decay rate. As a result, we ultimately obtained the extinction of the system where all species have decayed.

The behaviour reported above was exhibited in all 5 simulation runs in which m_1 was employed. Moreover this degenerative dynamic was also observed in the supplementary experiments in which we set the diffusion coefficient to m_2 and m_3. The distinctive extinction phases of example simulation runs are shown in Fig. 6.4. All remaining simulation runs are presented in Appendix C.

In Fig. 6.4, we distinguish that the range of variances in the average molecular population size actually decreases in contrast to the increase trend reported earlier. Little variance of average molecular string length is observed over the whole simulation runs excepting during a few seconds where the systems collapse. These

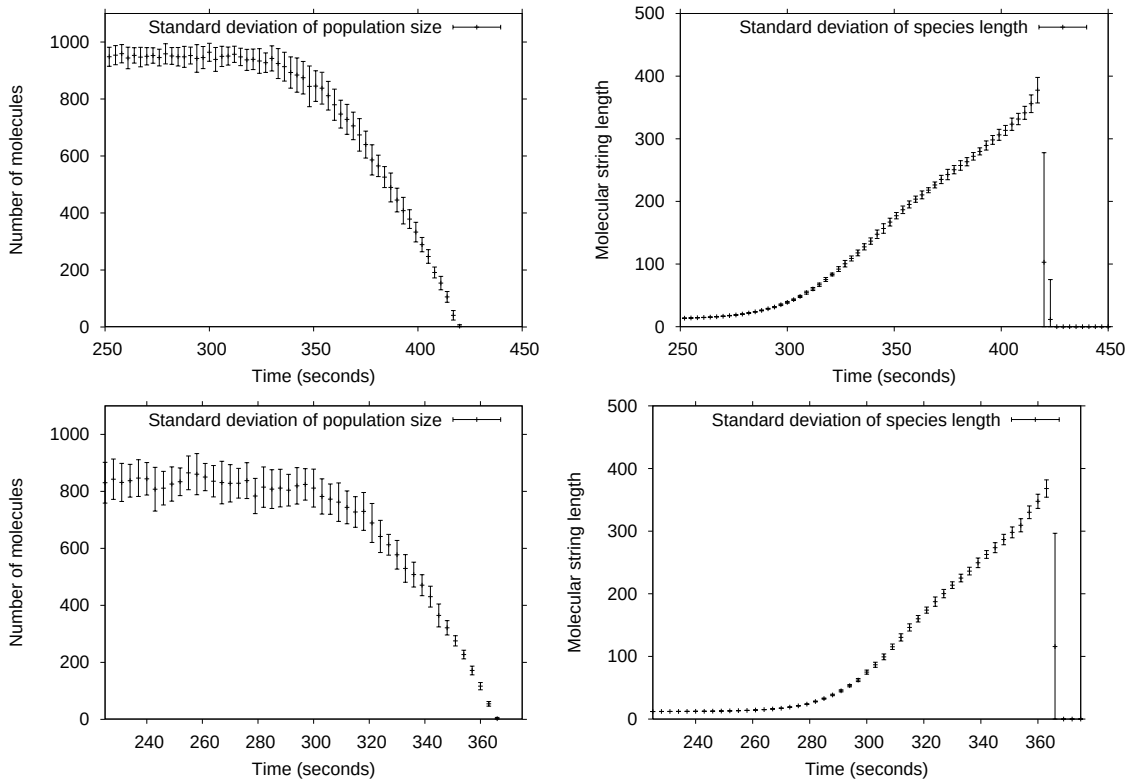

Figure 6.4: Example simulation runs 2 and 3 using the diffusion coefficients $m_2 = 0.05$ (top graphics) and $m_3 = 0.1$ (bottom graphics) respectively.

results suggest that higher diffusion coefficients allowed for the diffusion equilibrium to be achieved more rapidly. Consequently we identify a homogeneous molecular composition throughout all compartments over time.

Moreover we note a second trend related to the speed of the extinction phases which seems to be correlated with the coefficient diffusions. Table 6.1 presents the duration of the extinction phase for all experiments.

Series	1	2	3	4	5	Avg.
m_1	279	243	223	266	202	242.6
m_2	145	189	149	176	129	122.4
m_3	95	120	103	112	78	101.6

Table 6.1: Duration in seconds of extinction phases for the 3 series of experiments conducted with parameters m_1, m_2 and m_3. We devise an arbitrary threshold (i.e., when the average molecular string length is higher than 20) for determining the start of the extinction phases. This string length criterion is employed as it is a clear symptomatic evidence of the elongation catastrophe phenomenon.

In Table 6.1, we observe that as the diffusion coefficient increases, the duration of the extinction phase decreases. A higher diffusion coefficient accelerates the necessary duration for reaching diffusion equilibrium, which in this case has facilitated the spreading of mutant parasites throughout the compartments. Consequently the elongation catastrophe phenomena occurred at a faster pace. These results also suggest that if we were to decrease the key parameter (i.e., diffusion coefficient), this may only extend the duration of the extinction phase which would still occur quasi-deterministically upon the emergence of mutant-parasitic species.

The above results obtained using the static reactors MCS.bl model with molecular diffusion contradict the prediction that we proposed earlier. The selected range of parameters did not allow for the isolation of infected compartments.

Nevertheless according to our prediction, for an effective isolation and inhibition of the degenerative phenomenon to occur, we require the depletion of the infected compartments. This depletion is here happening *globally* at a late stage where the parasites have already invaded all compartments (through the exploitation of molecular diffusion). At that stage no healthy compartments remain which prevents any attempts to regulate and stabilise the system via diffusion.

For a rapid depletion to occur *locally* in infected compartments, we suggest that additional experiments should focus on modifying the diffusion probability which affects indirectly the diffusion coefficient and speed of diffusion equilibrium. A significant decrease of this diffusion probability may result in the partial depletion of infected compartments. This depletion would occur before the mutant species have the opportunity, through molecular diffusion, to spread over all compartments.

Nevertheless if this probability of diffusion is set too low, we would obtain a model where infected compartments would first decay entirely before being subjected to molecular diffusion. In such cases, there would be *no exchange* of molecules, the migration would only be one way. Such a model would then less account for molecular diffusion and the nature of the reactor model would not be as *continuous* (i.e., we

would observe the successive depletions and growths of molecular species in the compartments). This resulting model would become more similar to the *cellular* model approach (which will be described later) than McCaskill et al.'s original proposed model.

Further experiments are necessary to identify a *potentially* suitable range of parameters where the degenerative effects of parasitism can be controlled, whilst *exchanges* of molecules still occur between compartments.

6.3.4 Conclusion

We introduced the static reactors with molecular diffusion model which was examined by McCaskill et al. The latter demonstrated that for such models, according to the mutation and decay rates, there exists a range of diffusion coefficients which allow for the stabilisation of the self-maintaining cycle of closed reaction networks in which parasitic effects are occurring.

We presented the extended model of the MCS.bl which accounted for compartmentalisation and molecular diffusion. Using this system we conducted a series of evolutionary experiments to examine evolutionary capability.

Our experiments suggested that according to the parameters used, compartmentalisation and molecular diffusion do not allow for the inhibition of degenerative evolutionary phenomena.

However as we did not conduct a systematic evaluation of parameters due to time constraints and the multi-dimensional nature of the search space (example dimensions are the mutation and decay rate, diffusion probability and coefficient). We only explored one area of this vast search space where it may still be possible to find a suitable set of parameters which would stabilise evolution in the MCS.bl.

More specifically we discussed the potential role of the diffusion probability, a feature introduced in the extended MCS.bl but not present in McCaskill et al.'s continuous model. We argued that decreasing this parameter may allow for the local de-

pletion of infected compartments to occur. This isolation of infected compartments would prevent the parasitic species from invading the remaining non-parasitised compartments through molecular diffusion. Therefore this alternate solution may lead to the inhibition of degenerative dynamics due to parasitism.

Nevertheless, although this model with a low diffusion probability may control the elongation catastrophe phenomenon, this resistance against parasistism would not be directly due to molecular diffusions. In McCaskill et al.'s model, molecular diffusion was the key regulator where the number of molecules, contained in the reactors, remained more or less static. In this proposed model, degenerative outcomes would be prevented by allowing the parasitised compartments to deplete/decay before integrating species incoming from neighbouring and non-parasitised compartments. Parasitised compartments would first decay and molecular diffusion would occur in one direction, from non-decayed compartments to decayed ones. Following this unidirectional molecular exchange, the received species would increase in number in the formerly parasitised compartment. In such cases, the resulting model would thus be more adequately captured as a dynamic reactors model with unidirectional molecular diffusions. Such a model is, to some extent, quite analogous to the cellular model presented in the next section.

6.4 A cellular model

We present a second multi-level selectional model applied to the MCS.bl in which, similarly to biological cells, compartments may grow and divide. This version of the MCS.bl is independent from the previously extended model and does not include the properties (complementary to compartmentalisation) introduced in Section 6.3.

6.4.1 Introduction

Cronhjort and Blomberg (1997) proposed a deterministic model (using PDEs) where *dynamic* clusters (regarded here as compartments) of molecules would spontaneously

arise, grow, divide and die in a two-dimensional space.

In Cronhjort and Blomberg's model, molecules would aggregate spontaneously and form compartments without possessing a complex membrane structure. As mentioned earlier, compartments have also the ability to divide spontaneously. It was shown that, when parasites are introduced in a compartment, they rapidly destabilise the self-maintaining cycle of closed reaction networks contained in the compartment. This destabilisation ultimately leads to the decay of all molecules. Moreover when a compartment is infected with parasites, the latter spread to other compartments through parasitising the molecular species which are present *between* the compartments. By exploiting this propagation/diffusion technique, the parasites can invade and "kill" all remaining compartments.

However Cronhjort and Blomberg introduced a cut-off rule which restricted the spread of parasites throughout the compartments. This rule sets to zero (with a defined cut-off probability) the concentration of the molecular species which occur between the compartments (without affecting the molecular concentration in the compartments). Using a suitable cut-off value, infected compartments may become isolated in space. As a result, parasitic species decay locally without invading the rest of the compartments. Following this, an empty space emerges which may be occupied by a new compartment resulting from the division of a neighbouring "healthy" non-parasitised compartment.

In this model, compartmentalisation isolated the infected compartments which successfully prevented the invasion of parasitic species over the rest of the compartments. Cronhjort and Blomberg demonstrated that this cellular model may provide resistance against disruptive parasitic effects. We utilise this work as an inspiration and implement compartmentalisation and cellular division features in the MCS.bl. The goal is to realise a sufficiently robust MCS.bl system which would prevent the degenerative evolutionary dynamics from occurring.

Finally McCaskill et al. and Cronhjort and Blomberg's models share common

properties (e.g., definition of continuous production and decay rates, nature and interactions of catalytic/parasitic, etc) which differ from the MCS.bl. These differences are presented and addressed in the following section.

6.4.2 The model

In the remainder of this section, compartments which can grow and divide are referred to as "cells". We present an extended MCS.bl model which accounts for compartmentalisation and cellular division:

- In contrast to Cronhjort and Blomberg's model, we define clear compartmental boundaries. Molecular species are contained in distinct compartments/cells which are simulated on individual CPUs. The cells possess a maximal carrying capacity of n_{max} molecules.

- We employ a simplex topology to condition the interactions between cells as opposed to the two-dimensional space utilised by Cronhjort and Blomberg.

- Similarly to McCaskill et al.'s approach, catalytic reactions creating new molecules do not lead to the removal of another one in the cell. However, these reactions may occur only if the cell is not full (i.e., iff $c_i \in C, c_{i_n} < n_{max}$).

- When a cell c_i becomes full (i.e., $c_{i_n} = n_{max}$), a cellular division occurs as follows. c_i selects $\frac{n_{max}}{2}$ molecules at random. These molecules are removed from the cell and constitute the signal to be sent to a target cell $c_j \in C$. The latter is selected at random. c_i continues to operate with no further interactions (directly associated with the current division event) with c_j. When c_j receives the signal sent by c_i, all molecular species contained in c_j are removed. Following this, c_j inserts the molecules included in the signal into its own reaction space. This step concludes the cellular division process of c_i which offspring effectively displaced c_j.

- In contrast to both McCaskill et al. and Cronhjort and Blomberg's models, we do not define a decay rate or decay probability as we did in the previous extended MCS.bl model. In the static reactors model with molecular diffusion, reactions could occur in a given reactor only if the latter was not saturated. In such a saturated compartment, the spontaneous decay of molecules enabled further reactions to occur. In Cronhjort and Blomberg's model, the decay of a parasitised compartment led to the creation of a vacant space which then allowed for the formation of a new compartment (resulting from the division of a neighbouring non-parasitised compartment). Both the saturation of compartments and the displacement of decayed ones are dealt with by our specification of our cellular division mechanism. The latter prevents compartments from becoming saturated (i.e., half all of the molecules are "diffused" when n_{max} is reached) and also triggers the displacement of potentially decayed compartments. Consequently the specification of a spontaneous decay rate is not required in the current extended MCS.bl model.

- Finally, in contrast to Cronhjort and Blomberg's model, *any* cells (including parasitised ones) may still divide upon producing n_{max} molecules. Therefore even though parasites have invaded a given cell, the latter may still spread the parasitic species through cellular division.

In the following section we present and examine a series of experiments using the above extended MCS.bl model. Following this, we analyse the effects of chemical kinetics in this particular extension of the MCS.bl.

6.4.3 Cell-level mutations

During cellular divisions, molecules are randomly selected and transferred into the offspring cells. Due to the stochastic nature of these processes, some molecular species may not be transmitted to the offspring cells. Moreover, the concentration of the transmitted molecular species may also significantly vary.

Such "mutant" cells would contain novel reaction networks exhibiting differing dynamics. We refer to these error-prone transmissions as *cell-level mutations*.

Note that the term "mutation" is here relaxed and does not specifically refer to the commonly known phenomenon where variance may be introduced over *strings of monomers*. We nevertheless use this notion to capture the stochastic recombination involved in this process which may affect the dynamics/phenotype of the cellular species.

6.4.4 The experiments

We present a series of experiments using the extended parallel version of the MCS.bl. These experiments aim at investigating the effects of compartmentalisation and cellular division upon the system's evolutionary capability.

We propose the following set of fixed parameters which is utilised in all experiments:

- 30 compartments are utilised and executed in parallel using 30 AMD Opteron 270 (2.0 GHZ) CPUs.
- The maximal compartment carrying capacity is $n_{max} = 1000$.
- The maximal species string length is set to $BD_{Lmax} = 500$.
- The global spontaneous mutation rate is set to $r_{mut} = 0$.
- Each compartment is seeded/initialised with 250 instances of the $s_{R_4} = \triangledown 0101 : \triangledown 0101$ molecular species.

As mentioned in the conclusion of Section 6.3, we predict that in such compartmentalised models, infected compartments should ultimately decay. However in the current model no spontaneous decay is implemented. Infected cells would here present a low or null growth rate which would prevent the cell from dividing, as opposed to non-parasitised cells which may still grow, divide and displace any cells.

As shown by Cronhjort and Blomberg, this isolation of parasitised cells prevents parasitic species from spreading over the rest of the cells. Compartmentalisation and cellular division may then inhibit the degenerative phenomena from occurring in the MCS.bl. However a mutation error rate threshold may still exist. Indeed, if the mutation rate is too high, then parasitic species may emerge too rapidly in the cells and potentially kill all cells.

We conduct three series of 5 simulation runs each. In each of these series, we decrease the per-symbol mutation probability using the following parameters: $p_{sym1} = 1.0 \times 10^{-4}$, $p_{sym2} = 5.0 \times 10^{-5}$ and $p_{sym3} = 1.0 \times 10^{-5}$. As the mutation rate is decreased in the successive experiments, the execution time of the simulation runs is increased as follows: 3600, 7200 and 36000 seconds in series 1, 2 and 3 respectively.

In the first series of experiment (where p_{sym1} is employed, see Fig. 6.5) we note that the degenerative dynamics occur in the simulation runs 1, 3 and 4. The "extinction phases" observed in these simulation runs are all associated with a sudden increase in the molecular species' length, suggesting the elongation catastrophe phenomenon. However in runs 2 and 5, no extinction phenomena occur. In both simulation runs, we observe little variance in the reaction networks' level of catalytic activity which is described here by the average reaction success rate.

Nevertheless, in all simulation runs, we identify a comparable level of variance in the average species string length. This indicates that, as expected, parasitic species emerged and invaded a number of cells, as shown in Fig. 6.6 with the sporadic presence of cells containing species with a higher average string length (which causes the high variances/peaks in the average species length). Moreover we distinguish a clear initial lineage of cells where the average species length vary around 12 (the length of the seed species s_{R_4}).

In the second series of experiments, the mutation rate is decreased to $p_{sym2} = 5.0 \times 10^{-5}$, see Fig. 6.7. Here, we distinguish the survival of the cells in all simulation runs. Moreover, we observe a phenomenon which was not present in the first series:

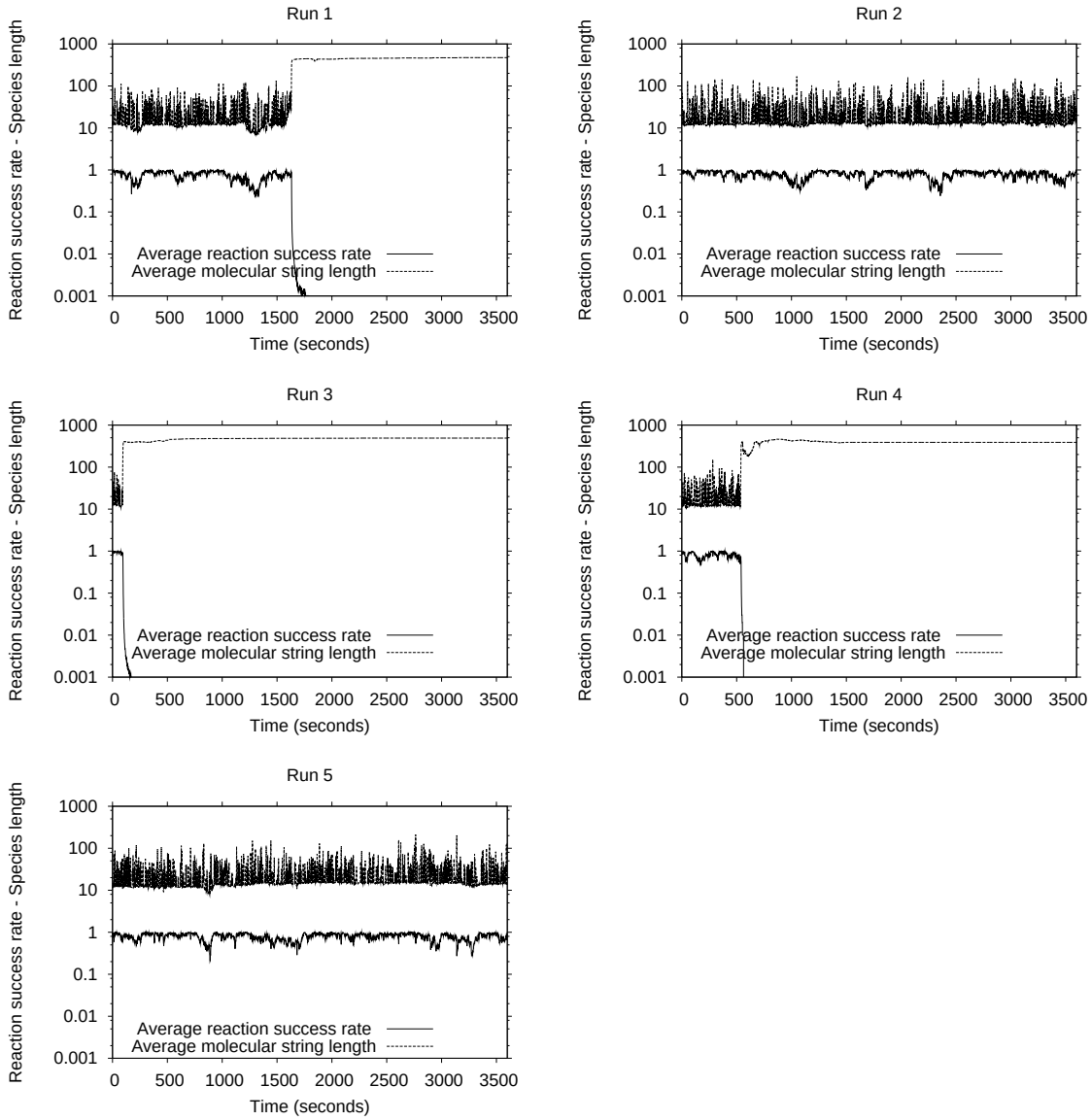

Figure 6.5: 5 example simulation runs with $p_{sym1} = 1.0 \times 10^{-4}$. The average reaction success rate represents the average number of catalytic reactions divided by the number of molecular interaction per second, e.g, if the reaction network is complete then the ratio is 1. In the contrary if only elastic reactions occur then the ratio is 0. As the cells are seeded with species s_{R_4}, the initial reaction networks are complete.

- In all runs we identify the emergence of cellular lineages in which the molecular species have a longer average string length (two example lineages are depicted in Fig. 6.8). However we do not observe an exponential growth of the species' string length as commonly observed.

- These species remain stable and self-maintain for a period of time after which

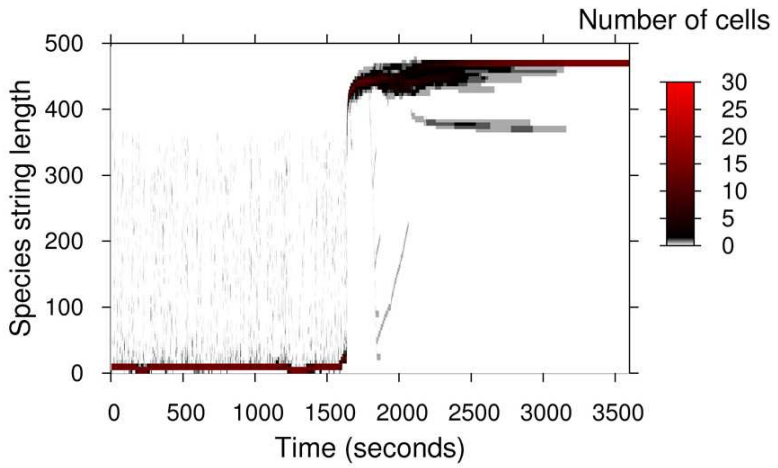

Figure 6.6: Classification and evolution of cellular lineages from simulation run 1. The cells are classified by the average length of molecular species at a given timestep. We employ a crude classification criterion to discriminate the parasitised cells from the non-parasitised ones.

they may eventually be displaced by other species having, similarly, a longer average string length.

- These cellular lineages are due to mutant species which were able to gain catalytic support from the ancestor species. However in contrast to the parasites involved in the elongation catastrophe phenomena, those species do not catalyse the production of further parasite species. Those species can in fact catalyse the production of species which are equivalent to the ancestors species from a phenotypic point of view.

- However an additional function of those species is also to elongate the substrate species. This elongation occurs in a linear fashion as opposed to the exponential string growth observed in the elongation catastrophe phenomenon.

- This slow/linear growth of species length had no immediate *negative* effects for the cells (in contrast to an exponential string growth which would rapidly cause the cell to degenerate) and therefore permitted the diffusion of these slightly longer species throughout the cellular population.

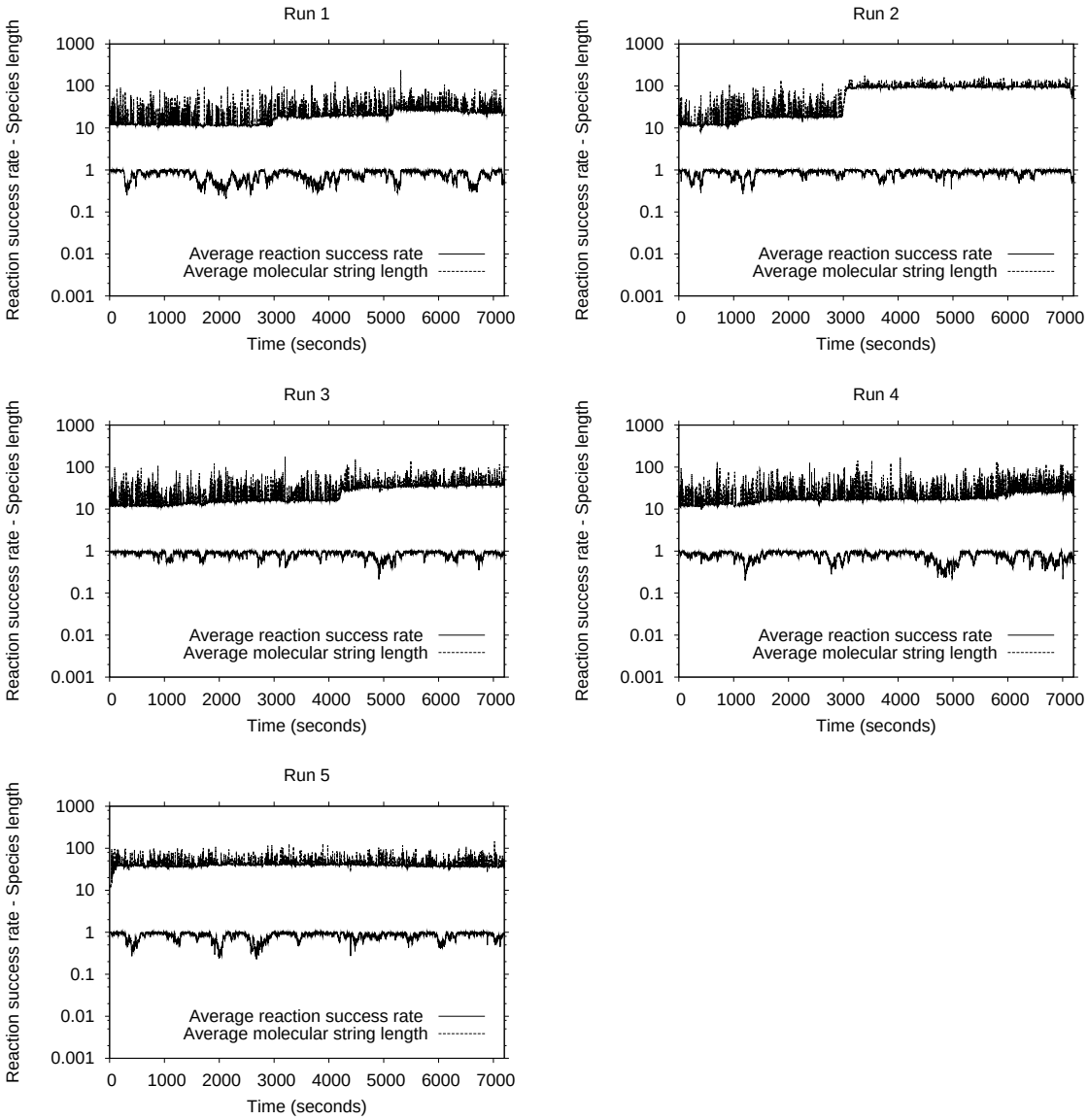

Figure 6.7: 5 example simulation runs with $p_{sym2} = 5.0 \times 10^{-5}$.

- Nevertheless as the species length increased, it soon became a selective disadvantage for cells to contain such "elongator" species. As a result, these cells were selected against and the increase in the average string length ceased. This fitness penalty did not apply to the cells from the start, when these elongators emerged (and shortly after), as the difference in the average molecular species length was then not significant from a computational point of view.

The difference in average string length in both lineages is due to the nature of

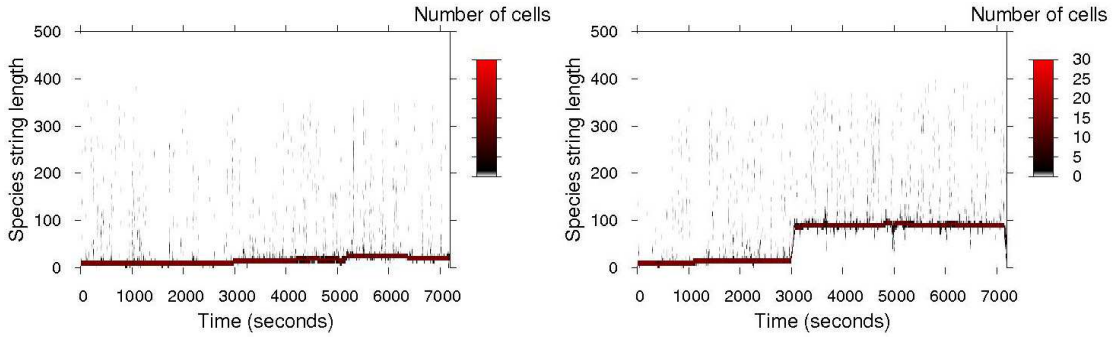

Figure 6.8: Evolution of cellular lineages in simulation run 1 (left) and 2 (right). A level of homogeneity in the molecular composition of most cells is observed at any given timesteps.

the patterns that were concatenated to substrate molecules. This pattern is longer in simulation runs 2 and 5 and explains the higher average molecular string length.

Note that the reproduction rate of the cells containing molecular species with a significantly longer string length should be lower (due to a higher computational demand) than the reproduction rate of cells containing shorter molecules. We may thus hypothesise that over time, these cell lineages containing longer molecular species might be displaced if a mutant molecular species emerges and possesses the following intrinsic properties:

1. The molecular string length is short enough to exhibit a significant difference in fitness as autocatalysis is here hyperbolic (Section 5.1).

2. The ability to counter-act the exploitation of the elongators by developing an immunity to these pseudo-parasites or being similarly able to exploit the elongators to be replicated/or produce smaller molecules.

The emergence of such molecular species has not been observed in any of the simulation runs.

Table 6.2 compares example elongators species triggering the elongation catastrophe (with exponential increase of string length) with those present in the above experiment (with linear increase of string length). Both species are mutants (result-

ing from a single mutation) of the self-replicase $s_{R4} = *\triangledown 0101 : \triangledown 0101$.

Elongator species s_1 with exponential length growth
$s_1^0 = *\triangledown 0101 : \triangledown 00101$
$2\ s_1^0 \rightarrow\ 2\ s_1^0\ +\ s_1^1$ with $s_1^1 = *\triangledown 0101 : \triangledown 000101$
$2\ s_1^1 \rightarrow\ 2\ s_1^1\ +\ s_1^2$ with $s_1^2 = *\triangledown 0101 : \triangledown 00000101$
$2\ s_1^2 \rightarrow\ 2\ s_1^2\ +\ s_1^3$ with $s_1^3 = *\triangledown 0101 : \triangledown 000000000101$
...
$length(s_1^i) = length(s_1^0) + i^2$
Moreover if we consider the reaction between two differing s_1 species generations with $i \neq j$ then: $s_1^i\ +\ s_1^j \rightarrow\ s_1^i\ +\ s_1^j +\ s_1^{i+j}$
Elongator species s_2 with linear length growth
$s_2^0 = *\triangledown 0101 : 0\triangledown 0101$
$2\ s_2^0 \rightarrow\ 2\ s_2^0\ +\ s_2^1$ with $s_2^1 = 0 * \triangledown 0101 : 0\triangledown 0101$
$2\ s_2^1 \rightarrow\ 2\ s_2^1\ +\ s_2^2$ with $s_2^2 = 00 * \triangledown 0101 : 0\triangledown 0101$
$2\ s_2^2 \rightarrow\ 2\ s_2^2\ +\ s_2^3$ with $s_2^3 = 000 * \triangledown 0101 : 0\triangledown 0101$
...
$length(s_2^i) = length(s_2^0) + i$
$s_2^i\ +\ s_2^j \rightarrow\ s_2^i\ +\ s_2^j +\ s_2^{j+1}$

Table 6.2: Example elongator species s_1 and s_2 with exponential and linear string length growth respectively. Both s_1^0 and s_2^0 represent the first generation of s_1 and s_2 species, whereas s_1^i and s_2^i are the $(i+1)^{th}$ generation.

Fig. 6.9 presents the final series of experiments where the mutation probability p_{sym3} is employed. We note that all cells successfully resisted against the disruptive parasitic effects. We also distinguish a net decrease in the density of parasitised cells (see Table 6.3) which consequently affected the average reaction success rate. The latter remained stable with little variance occurring (when compared with previous experiments) throughout the simulation runs.

Series	1	2	3	4	5	Avg.
p_{sym1}	0.141	0.145	0.204	0.165	0.145	0.160
p_{sym2}	0.060	0.029	0.056	0.071	0.020	0.047
p_{sym3}	0.021	0.016	0.012	0.021	0.013	0.017

Table 6.3: Density of high variances in the average molecular species length, i.e., number of peaks where the species length increased significantly (where the average species length at time t is 1.5 times higher than at $t-1$) per second.

The probability of mutation p_{sym3} employed in this final series of experiment also

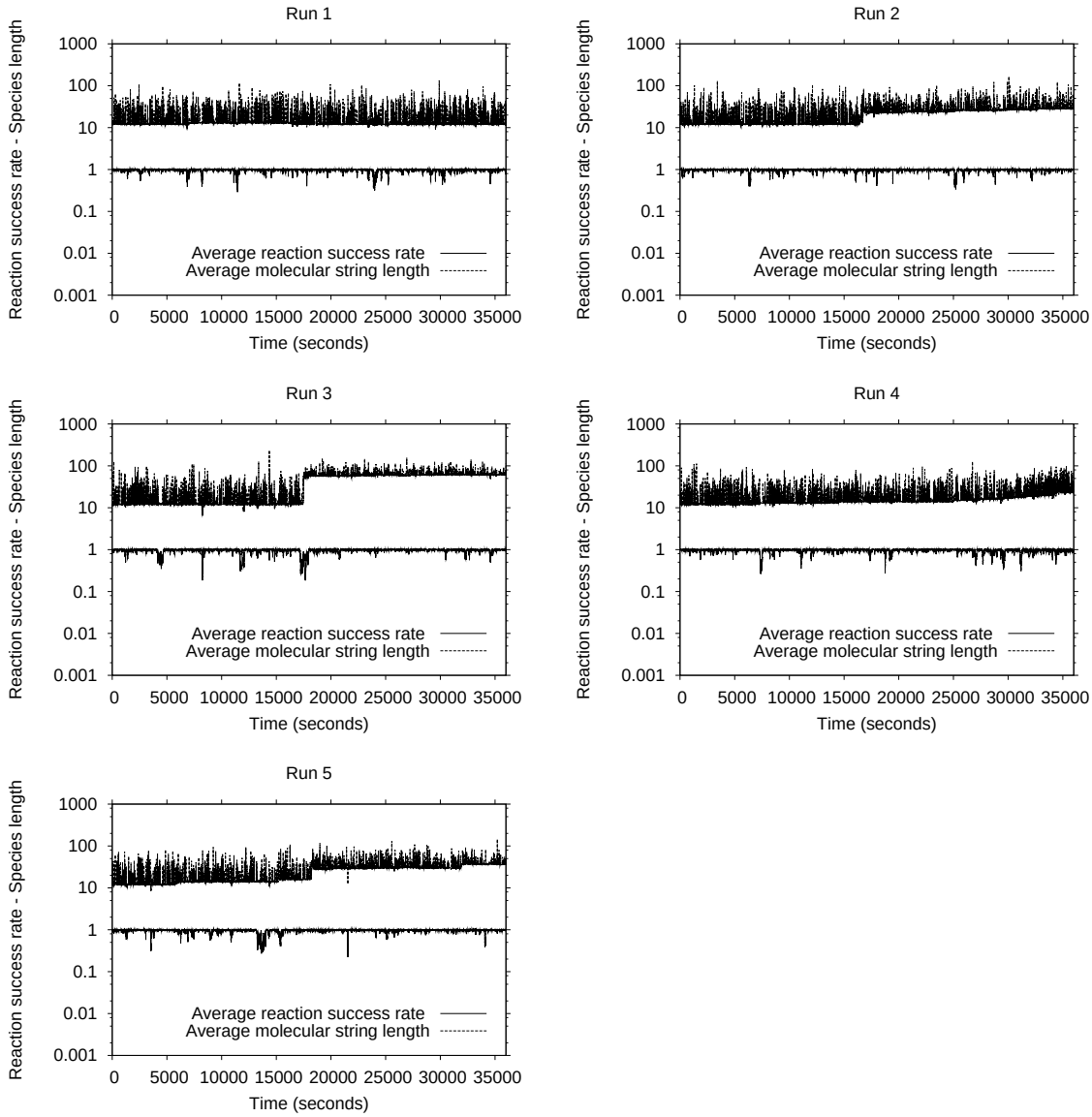

Figure 6.9: 5 example simulation runs with $p_{sym3} = 1.0 \times 10^{-5}$.

allowed for the self-maintenance of closed reaction networks in all cells throughout the simulation runs. Similarly to the second series of experiment, the emergence of non-lethal "elongator" species increasing the average string length is observed but no degenerative outcomes occur.

As the mutation rate decreased, the probability of emergence of mutant parasitic molecules decreased. As the latter diminished, the number of cells where the elongation phenomenon may occur *simultaneously* reduced, as shown in Table 6.3

with the number of peaks (high variance in average species length). These results suggest that a minimal number of infected cells is necessary to spread (rapidly) enough parasites throughout the cellular population, and consequently to provoke extinction at the system level. Moreover these results indicate that a mutation error threshold may exist in the range $p_{sym} < 1.0 \times 10^{-4}$ in the MCS.bl/cellular model. This final series of experiment suggests that the cellular model of the MCS.bl is able to provide the closed reaction networks with resistance against parasites and associated degenerative effects. However this control in evolutionary degeneration is possible given a tolerable rate of mutation.

In the next section, we analyse how chemical kinetics may also have contributed to this control of parasitic effects by providing a selective advantage to non-parasitised cells over the parasitised ones in specific cases.

6.4.5 Effects of chemical kinetics

We examine the effects of chemical kinetics upon the system's evolutionary dynamics. A side effect of the parallel implementation of the MCS.bl is the alteration of chemical kinetics, as presented in Section 6.2.2. We discuss here the potential role of chemical kinetics in improving resistance against parasites in particular cases.

As mentioned earlier, parasitised cells may still divide as they may *even* contain reaction networks that are still complete (i.e., all molecular collisions lead to successful creations of new parasitic molecules). Let us consider the cells $c_i \in C$ and $c_j \in C$ which both contain complete reaction networks. The cell c_i contains non-parasitic species only, as opposed to c_j in which we insert parasites only.

If all reaction rates are equal, then both c_i and c_j would exhibit an equal growth rate. As any reactions occurring in both cells would occur at the same pace, both cells would thus possess an equivalent growth rate and may ultimately divide simultaneously. Both cellular species would have equal chances to displace each other, this scenario applies *until* the parasitised cell has started to decay (with the reaction

success rates decreasing).

However, if we consider variations in the reaction rates, which is addressed here by the necessary computational time to process a reaction in the parallel version of the MCS.bl, then we argue that non-parasitised cells possess a selective advantage. In Section 5.5, we showed that the elongation catastrophe was due to mutant parasitic species which had the effect of increasing their string length repeatedly. However in the parallel system, as the length of species increases, the computation necessary to process reactions between these species increases accordingly. Reactions involving parasite species as the reactants would therefore be computationally more demanding. As time matters in this parallel system, the growth/production rate of parasitised cells would be lower than the rate of non-parasitised cells.

Therefore we hypothesise that, in the current extended version of the MCS.bl, parallelism improves resistance against parasitism where the elongation catastrophe phenomena occur.

To test the above hypothesis, we present a simple experiment in which we employ the following set of parameters:

- 30 compartments are utilised and executed in parallel using 30 AMD Opteron 270 (2.0 GHZ) CPUs.
- Experiments are run for 60 seconds.
- The maximal compartment carrying capacity is $n_{max} = 1000$.
- The maximal string length is set to $BD_{Lmax} = 500$.
- A single cell c_1 is seeded with 2 instances of $s_{R_0} = \nabla 0 : \nabla 0$.
- The remaining 29 cells $c_2, \ldots, c_{30}$ are seeded with 2 instances of $s'_{R_0} = \nabla 0 : \nabla 00$. s'_{R_0} triggers the elongation catastrophe.

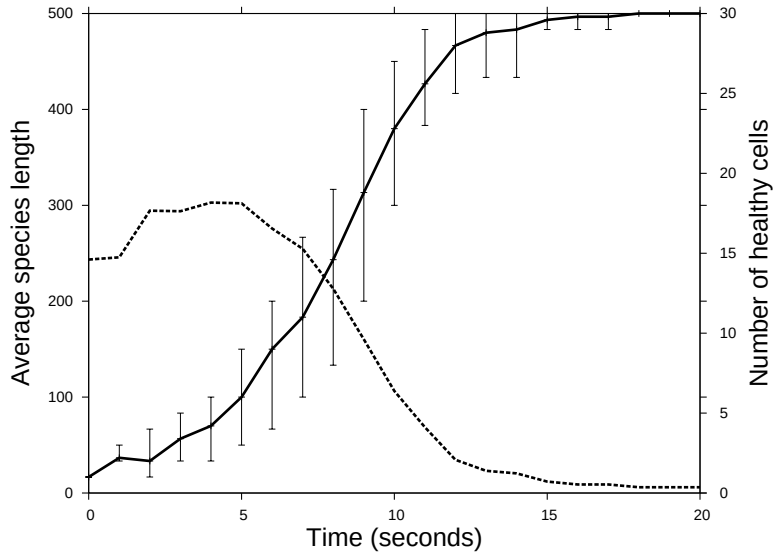

Figure 6.10: Dynamics of average number of non-parasitised cells (averaged over 5 independent experiments. The error-bars denote the minimal and maximal number of non-parasitised cells throughout all experiments at a given timestep. The reaction success rate, not described here, was maintained to one during all simulation runs (due to the completeness of the reaction networks.

- Both the global spontaneous mutation rate r_{mut} and mutation per symbol probability p_{sym} are set to 0.

Mutation is turned off in order to restrict the diversity of species to only s_{R_0} and the lineage of parasites species s'_{R_0}.

Only the single healthy cell c_1 was specified against 29 parasitised cells, which is clearly the most disadvantageous situation for the survival of c_1. In a non-parallel system, c_i would have only a $\frac{1}{30}$ chance not to be displaced during the first "round" of cellular divisions.

5 distinct simulations are run. The results are shown in Fig. 6.10.

In Fig. 6.10, an early phase is observed where the average species length is increasing, suggesting that the elongation catastrophe is occurring in the cells. However this trend diminishes (down to 6, i.e., the length of s_{R_0}) in contrast to the average number of non-parasitised cells which increases until all parasitised cells have been displaced. This displacement occurred due to the fast reproduction rate of non-parasitised cells which were provided here with a selective advantage over the

parasitised cells. Fast reactions permitted the non-parasitised cells to divide more rapidly and were subsequently given the opportunity to displace the parasitised cells.

This experiment demonstrates the potential improvement in evolutionary capability due to the new chemical kinetics properties introduced in the parallel MCS.bl. Nevertheless, this experiment was an ideal scenario where no mutation could occur and all cells where initiated at the same time. Consequently, the non-parasitised cell was always the first to trigger its cellular division and *successive ones* over the rest of the parasitised cells.

6.4.6 Conclusion

We introduced Cronhjort and Blomberg's cellular model in which compartmentalisation and cellular division prevented degenerative effects due to parasistism from occurring. Using this work as an inspiration, we proposed an extended parallel implementation of the MCS.bl which accounts for these cellular features.

Using this extended version of the MCS.bl , we conducted and examined a series of evolutionary experiments in which we varied the mutation rates. Our results showed that as the mutation rate decreased, the emergence and density of parasitised cells decreased. We showed that a level of parasitised cells is necessary for the parasites to spread rapidly enough throughout the cellular population and to eventually cause extinction at the system level. If no such rapid diffusions occur, then the parasitised cell(s) would decay locally and will be replaced by non-parasitised cells resulting from the division of other cells.

We also showed how chemical kinetics may improve, in particular cases, resistance against parasites. This feature resulted from the parallel implementation of the MCS.bl.

This cellular model successfully improved the MCS.bl's evolutionary capability and, according to the level of mutation rates, prevented the degenerative elongation catastrophe phenomena from occurring.

6.5 Summary

We presented two distinct multi-level selectional models applied to the MCS.bl. This work aimed at resolving the evolutionary degeneration issues of the MCS.bl.

We first examined a parallel implementation of the MCS.bl which was inspired by the static reactors model with molecular diffusion investigated by McCaskill et al. Our result suggested that the range of parameters selected in our experiments could not allow for the self-maintenance of closed reaction networks when subjected to perturbations. Nevertheless we mentioned that future work could illuminate further on the application of this model upon the MCS.bl to improve evolutionary capability.

Our second attempt addressed compartmentalisation and cellular division. These features were present in a model previously examined by Cronhjort and Blomberg. We conducted a series of evolutionary experiments using a novel parallel implementation of the MCS.bl where we introduced compartmentalisation and cellular division properties. Our results indicated that this cellular model was able to improve the resistance of closed reaction networks against the disruptive effects due to parasitism, including the elongation catastrophe phenomena.

The above investigations provide complementary insights on the potential effects of compartmentalisation over evolutionary capability in agent-based Artificial Chemistries (ACs). As no analytically tractable methods are currently available for the study of such complex ACs, empirical investigations are necessary and may provide guidance on the construction and analysis of future evolutionary systems.

The conclusive outcomes obtained with the cellular model provided us with a novel and robust MCS.bl system which we may now employ to conduct further evolutionary experiments. The latter will focus on the evolution of closed reaction networks to carry-out pre-specified information processing tasks. This follow-up investigation is presented in the next chapter.

Chapter 7

Evolving Closed Reaction Networks

In Chapter 6, we evaluated a cellular model applied to the parallel implementation of the MCS.bl. Our results indicated that this multi-level selectional model provided the MCS.bl system with the necessary robustness to control the degenerative effects due to parasitism. This control of evolutionary degenerations permitted the closed reaction networks to self-maintain when subjected to perturbations. This novel MCS.bl system enables us to conduct further experiments in which we aim at evolving closed reaction networks to carry out pre-specified information processing tasks. These evolutionary experiments are presented in this chapter which combines some of the material published at several international conferences (Decraene, 2009; Decraene et al., 2009).

7.1 Introduction

We examine the evolution of closed reaction networks to carry out pre-specified information processing tasks. This chapter is composed of two main sections:

1. We first describe an experiment in which we evolve closed reaction networks to perform a signal-amplification function. To drive the evolution of the closed reaction networks to achieve a specified task, we introduce and discuss a novel cellular division criterion.

2. We then examine the evolution of crosstalking closed reaction networks capable of multitasking. We aim at demonstrating the potential constructive role of crosstalk in enabling the evolution and realisation of closed reaction networks of higher complexity.

Both investigations are assisted with the parallel and cellular model of the MCS.bl which was presented and evaluated in Section 6.4.

7.2 Evolving a signal amplification closed reaction network

We introduce a novel cellular division criterion which aims at driving the evolution of closed reaction networks towards the achievement of a pre-specified task. We then summarise preliminary experiments in which self-replication reactions occur. Following this, we describe an experiment in which collectively autocatalytic reaction networks (where no self-replication reactions may occur) are evolved to carry out a simple target information processing task.

7.2.1 Introduction

In the experiments conducted with the cellular model of the MCS.bl in Section 6.4, a cell would divide only when n_{max} molecules have been produced (i.e., when the cell is full). These cellular divisions occurred regardless of the nature of the molecular species.

In order to evolve closed reaction networks to carry out pre-specified information processing tasks, we propose to modify the conditions triggering the cellular divisions. The latter determine the implicit fitness of the cellular species. In contrast to fitness functions *explicitly* devised in top-down evolutionary approaches, implicit fitness functions do not directly specify the genotype/phenotype mapping of the candidate species. Here, the cellular agents determine, by themselves, their own actions and ultimately their fitness with regards to the realisation of the target task. Defining new cellular division criteria allows one to indicate the desired target tasks,

but not the *actual computations*, that the reaction networks have to perform.

In the remainder of this section, we present a series of experiments in which we utilise a new cellular division criterion devised as follows. A cell divides when the amount of a specific target molecular species, denoted by s_T, reaches n_{target}. The cellular reproduction rate now depends on the molecular growth rate of s_T. The ability of the closed reaction network to both promote the growth of s_T and maintain closure defines the cell's fitness. The pre-specified task assigned to these closed reaction networks is to amplify the "signal" s_T. Similar *in-vivo* experiments were conducted to maximise the production of target molecular species which had for effect to increase the growth rate of the *E.coli* bacterium (Palsson, 2006). We may also interpret s_T as a necessary molecular species (e.g., a membrane species) to allow the cellular division to occur.

Using the above cellular division criterion, we conducted a series of preliminary experiments where self-replication reactions (i.e., reactions in which the product is syntactically identical to the enzyme and substrate molecules) occur. These self-replication reactions were allowed in the experiments conducted in Chapters 5 and 6. In the current evolutionary simulations (described in Appendix D.1), the cells were seeded with the self-replicase s_{R_4} species, the latter was also designated as the target species s_T. Our results indicated that the networks would converge toward molecular organisations which are n-element hypercycles where n would typically be lower than 5. As these networks of limited complexity were hypercycles, they were fragile and subsequently could not self-maintain under perturbations. These closed reaction networks would typically collapse to the single autocatalytic and target species s_{R_4}.

The above phenomenon was also encouraged by the objective task which was devised to optimise the production of s_{R_4}, thus reaction networks containing only s_{R_4} were already the optimal catalytic networks to produce the target species. As these experiments presented a limited interest with regards to the evolutionary growth

of complexity, we propose to conduct a similar experiment where self-replication reactions may not occur.

This second series of experiment involving *collectively autocatalytic* reaction networks employs a hand-designed seed reaction network which is presented in the next section.

7.2.2 The seed reaction network

As self-replication reactions are now disabled, we cannot seed the cells with an ancestor autocatalytic molecular species as in previous experiments. We also demonstrated that the spontaneous emergence of closed reaction networks was unlikely to occur in the MCS.bl given a randomly generated population of molecular species. We thus propose to hand-design a minimalist *collectively* autocatalytic reaction network which will be used as the seed network in the experiment.

The construction and exploitation of this seed reaction network are clearly characteristic of "top-down" approaches (Section 3.5). Although we advocate for minimising the use of top-down/engineered elements, this remains necessary as no alternative is currently available to explore the evolution of closed reaction networks using the MCS.bl.

We attempted to construct this seed collectively autocatalytic reaction network in a *minimalist* manner with regards to the complexity at both the molecular (i.e., using simplest/shortest molecular species) and network level (i.e., involving the least number of both molecular species and reactions). Moreover, we include additional constraints to the realisation of this reaction network:

- The reaction network is not, by design, the optimal catalytic network to realise the target task.
- Molecular species which can perpetually generate new species (such as elongator species) are not allowed. This filter was necessary as such molecular

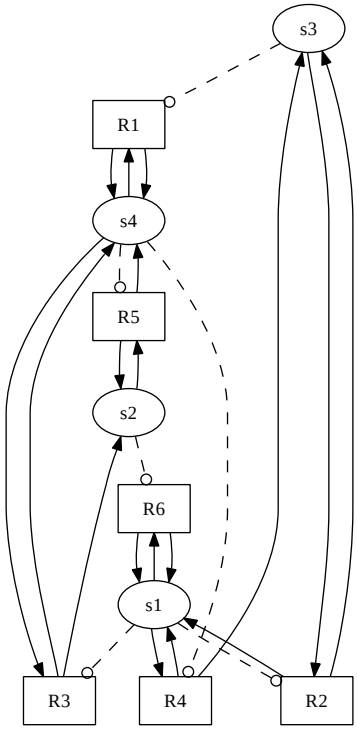

Figure 7.1: Bipartite reaction network of the seed closed network.

species trigger the spontaneous growth of the reaction network's topology/-complexity by default. Although a natural limitation in the growth of the molecular length would be observed (Section 6.4.4), these molecular species are filtered out as they do not facilitate and bias the current investigation on the evolutionary growth of complexity.

The construction of this reaction network is documented in Appendix D.2. This informal investigation suggested that the reaction network (depicted in Fig.7.1) is potentially the minimalist collectively autocatalytic reaction network, satisfying the above requirements, that can be realised in the MCS.bl. The molecular species composing this reaction network are listed in Table 7.1.

Figure 7.2 presents the deterministic dynamics (neglecting mutation) of the seed closed reaction network when the initial amount of each molecular species is set to 10. This graph was obtained by solving the ODE system generated from the SBML specification of the seed reaction network using the SBML ODE solver (Machné et al.,

Molecular species
$s_1 = *\nabla 0 : \nabla 1$
$s_2 = *\nabla 0 : \nabla 0$
$s_3 = *\nabla 1 : \nabla 0$
$s_4 = *\nabla 1 : \nabla 1$

Table 7.1: Molecular species present in the seed reaction networks.

2006). We also conducted a series of 100 simulation runs to measure in real time (i.e., in seconds) the growth of the different molecular species. In each of these independent simulation runs, a single reactor containing the seed reaction network was executed, with no mutations occurring, until 200 s_1 molecules were produced. Moreover as the nature (i.e., genotype and phenotype) of the molecular species s_1, s_2, s_3 and s_4 is quite similar (i.e., only the symbols 1 and 0 are permuted with each other in the different molecular species) the variations in reaction rates are negligible in these simulation runs. In the latter, the growth dynamics of the molecular species approximately match the deterministic dynamics depicted in Figure 7.2. The end simulation time was averaged over the 100 runs. This averaged measurement was then employed to rescale the deterministic time course of the different molecular species' growth. This scaling operation is also conducted in the next related Figures 7.6 and 7.7.

In the remainder of this chapter, cellular species are classified by the specific reaction network contained in the cells. The above collectively autocatalytic reaction network or *cellular species* is denoted by c_0 and employed as the seed cellular species in the evolutionary experiment presented in the next section.

7.2.3 Experiment

An evolutionary experiment is presented in which c_0 is employed as the seed reaction network and evolved to promote the growth of $s_T = s_1$. A prediction regarding this experiment is proposed as follows. We may first consider the existence of non-closed reaction networks that are more efficient than c_0 at producing s_T. Nevertheless,

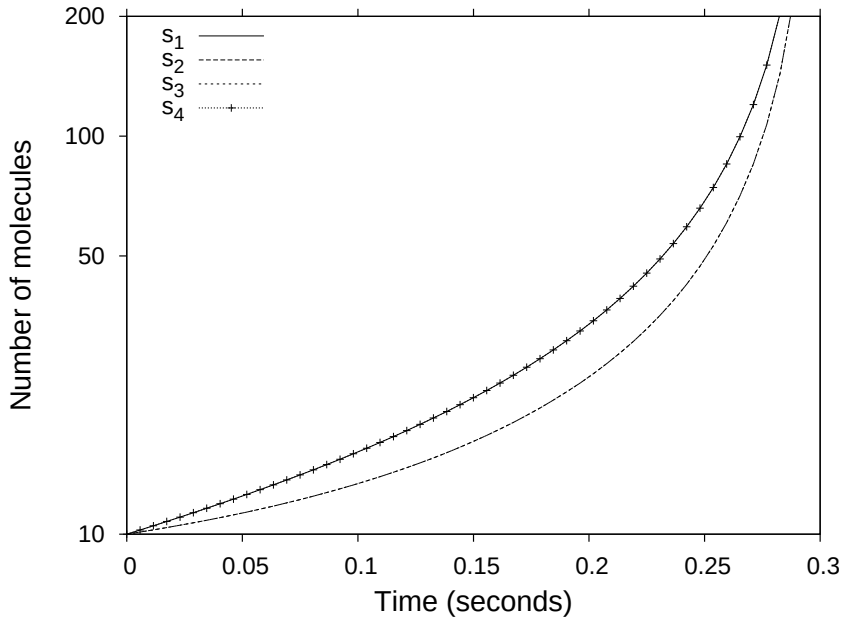

Figure 7.2: Deterministic dynamics of seed closed reaction network, the species $S1$ and $S4$ are overlapping (top line). $S2$ and $S3$ are similarly overlapping. Time was rescaled using experimental data.

in such networks, as the production of the molecules species required to generate s_T molecules cannot be maintained through closure, these molecular species would thus deplete over time. These networks, being not organisationally closed, would therefore not be able to self-maintain over successive cellular reproductions.

As a result, such non-closed reaction networks would possess a clear selective disadvantage against c_0. We may then predict that c_0 cannot be displaced by cellular species which contain non-closed reaction networks. In other words, if selective displacements occur, they would necessarily involve novel and organisationally closed reaction networks.

To test this prediction, an evolutionary experiment is conducted using c_0 and the following set of parameters:

- 31 cells are utilised and executed in parallel using 31 AMD Opteron 270 (2.0 GHZ) CPUs.
- Simulations are run for 3600 seconds.

- The maximal compartment carrying capacity is $n_{max} = 1.0 \times 10^6$.
- The target molecular species division threshold is set to $n_{target} = 200$.
- The maximal species string length is set to $BD_{Lmax} = 500$.
- The global spontaneous mutation rate is set to $r_{mut} = 0$.
- The per-symbol mutation probability is set to $p_{sym} = 1.0 \times 10^{-5}$.
- Each compartment is seeded/initialised with 10 instances of each species s_1, s_2, s_3 and s_4.
- The target molecular species is $s_T = s_1$.

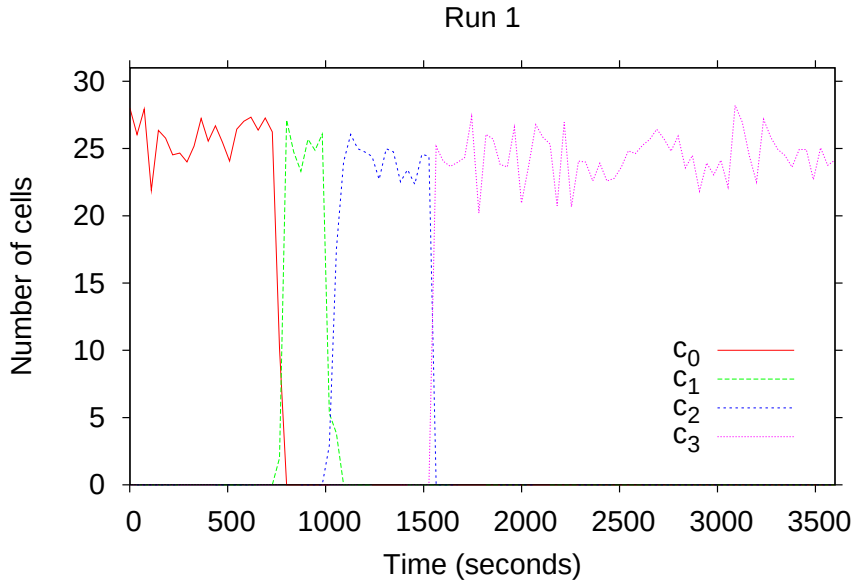

Figure 7.3: Dynamics of dominant cellular species. A spline function was employed to approximate the different curves.

During this experiment the average number of interactions per cell per hour was over 4.0×10^7. 1235 different and unique reaction networks were generated due to molecular and cellular mutations. The latter refer to mutations occurring at the cell-level which may result from the stochastic nature of cellular divisions (i.e., some molecular species may not be transferred into the offspring cells, resulting in mutant

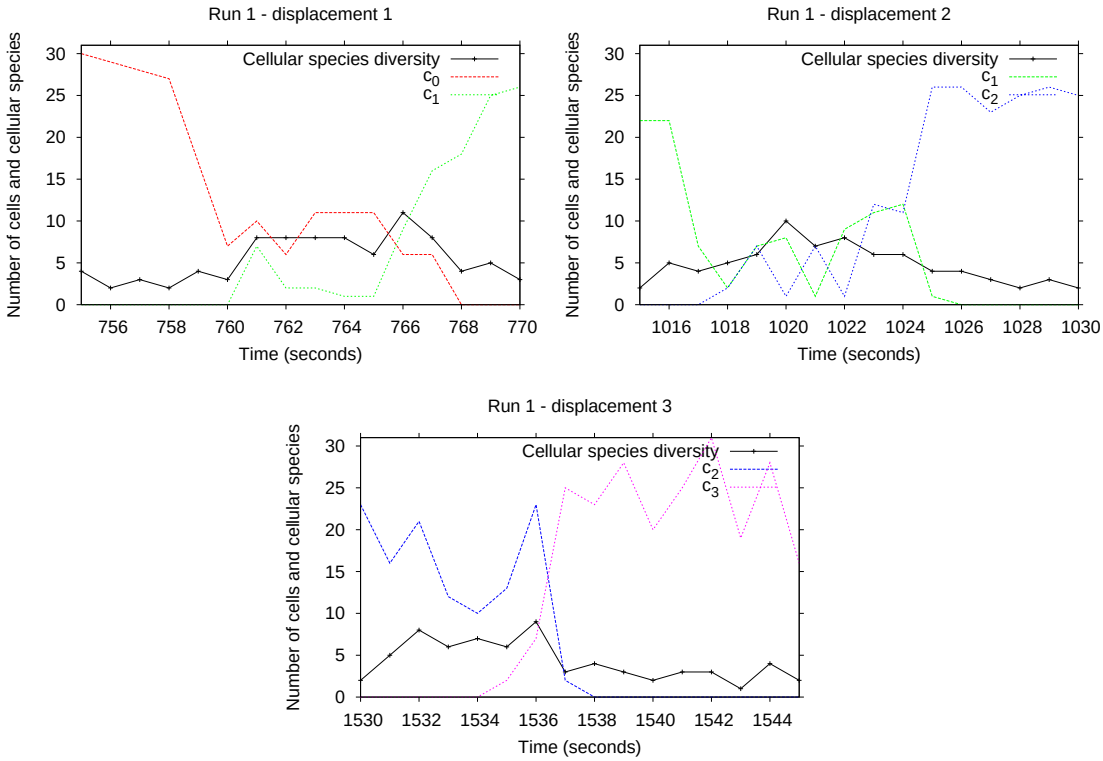

Figure 7.4: Successive displacements that occurred in simulation run 1.

cellular species). In Fig.7.3, we can distinguish that three successive displacements occurred (Figure7.4). The reaction networks contained in these three consecutive cellular species are listed in Table 7.2.

c_0	c_1	c_2	c_3
$s_1 = *\nabla 0 : \nabla 1$	s_1	s_1	s_1
$s_2 = *\nabla 0 : \nabla 0$	s_2	s_2	s_2
$s_3 = *\nabla 1 : \nabla 0$	$s_5 = *\nabla\diamond : \nabla 0$	$s_7 = *\nabla\diamond : \nabla\triangle 0$	$s_9 = *\nabla\triangle : \nabla 0$
$s_4 = *\nabla 1 : \nabla 1$	$s_6 = *\nabla\diamond : \nabla 1$	$s_8 = *\nabla\diamond : \nabla\triangle 1$	$s_{10} = *\nabla\triangle : \nabla 1$

Table 7.2: Molecular species contained in successive dominant closed reaction networks denoted by c_1, c_2 and c_3.

We first observe that, as predicted earlier, the successive dominant closed reaction networks successfully maintained closure. We first investigate the displacement that occurred between c_0 and c_1. We examine the reaction network contained in c_1 which is depicted in Figure 7.5. The dynamics of the different molecular species' growth are shown in Fig.7.6.

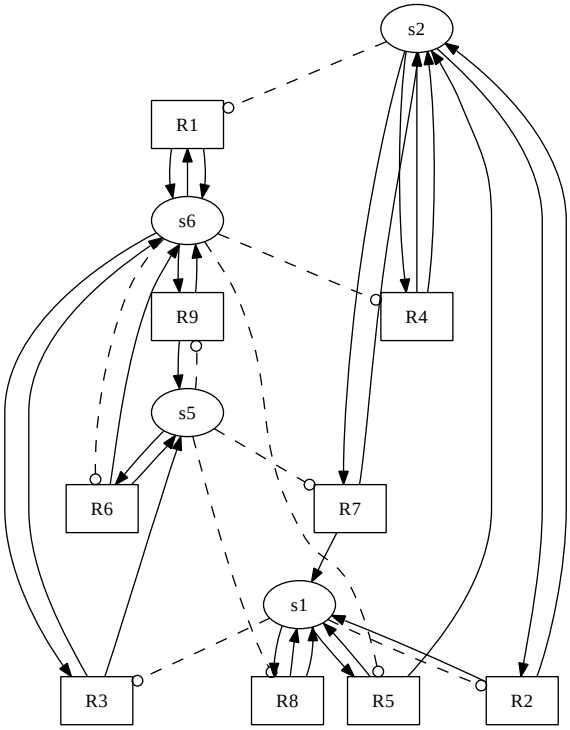

Figure 7.5: Evolved closed reaction network c_1 promoting growth of molecular species s_1.

In c_1, we note that both species s_3 and s_4 have been replaced by the new molecular species s_5 and s_6. These new molecular species increased the total number of possible reactions from 6 to 9 (see Fig.7.5), suggesting a higher overall catalytic activity. We compare the dynamics of this evolved closed reaction network against the seed closed reaction network, see Figure 7.6.

In Figure 7.6, it can be seen that s_1 reaches the division threshold at $t \approx 0.155$ whereas in the seed closed reaction network s_1 attains this threshold at $t \approx 0.280$. By producing the target species s_1 at a faster rate, the evolved network gained a selective advantage over the seed network. The emergence of the molecular species s_5 and s_6 had the effect of promoting the growth of s_1 whilst maintaining closure. The network closure properties evolved and permitted the network to promote the growth of species s_1.

Finally Figure 7.7 compares the different growth dynamics of s_1 using the differ-

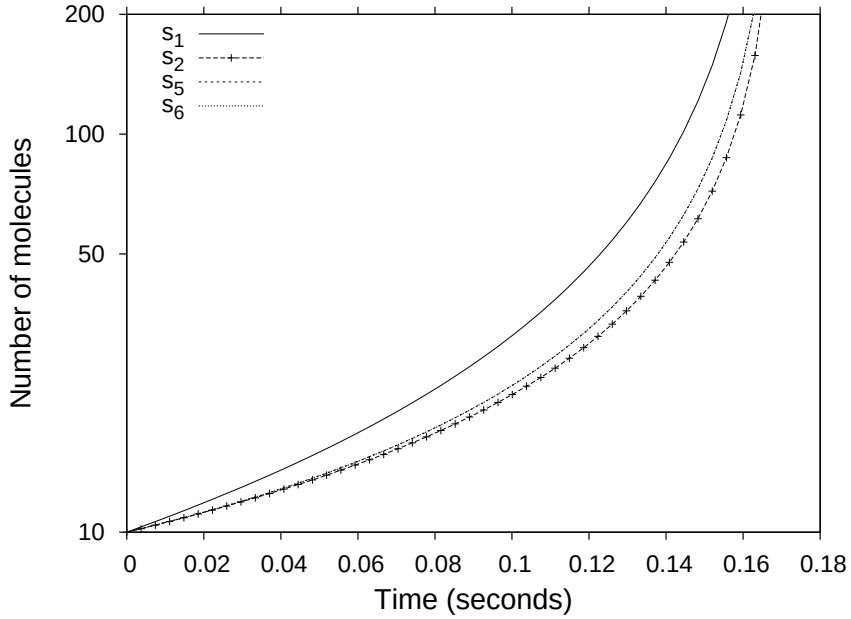

Figure 7.6: Dynamics of the evolved closed reaction network, the species s_5 and s_6 are overlapping (middle line), s_1 is the top line and s_2 is the bottom line. Each molecular species amount is initialized to 10.

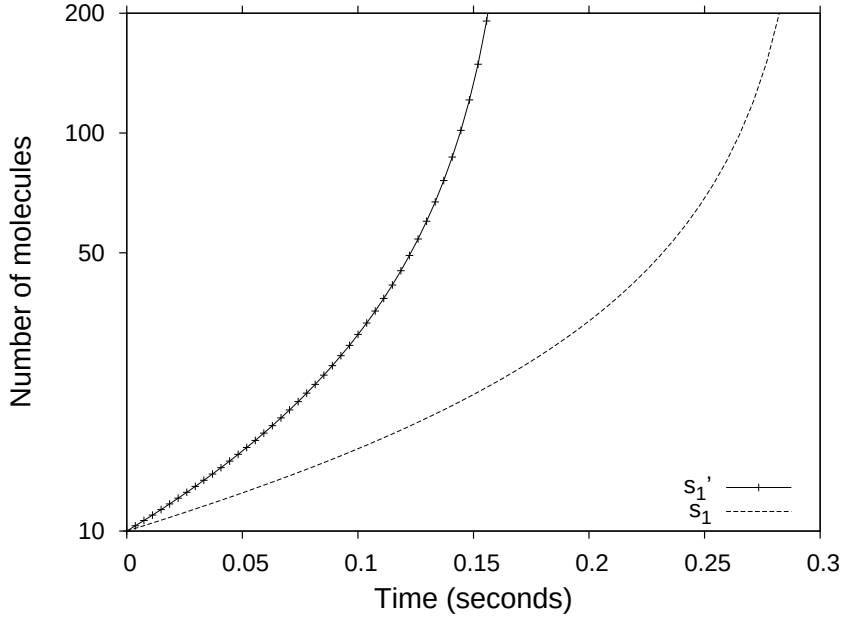

Figure 7.7: Comparison of s_1 species' growth in seed and evolved reaction networks. s_1 and s_1' depicts the dynamics of s_1 with initial molecular amount 10, using the seed and evolved reaction networks respectively.

ent networks and a common initial number of molecules. This comparison highlights the improvement in the networks' fitness (i.e., hastened the production of molecular species s_T).

We now examine the displacements that occurred between c_1, c_2 and c_3. Although the molecular species contained in c_1, c_2 and c_3 are different from a genotypic point of view, their phenotypes are similar. The symbols $\diamond$ and $\triangle$ act in the same manner when occurring in the condition statement. The symbols $\triangle$ occurring in the action statement of c_2 are ignored. Moreover, the genotypic differences yield negligible effects upon the reaction rates ($\triangle$ and $\diamond$ are computationally equivalent and the *ignored* symbol $\triangle$ occurring in c_2 has little computational impact). As a result the species contained in c_1, c_2 and c_3 are phenotypically equivalent. These cellular species are thus likely to possess an equivalent fitness.

This common level of fitness suggests that the displacements that occurred between the cellular species c_1, c_2 and c_3 are due to drifts. However, the *abrupt* transitions between these cellular lineages are symptomatic of selective displacements. Whereas displacements due to drift dynamics would take place at a slower speed. All displacements in this simulation run took between 5 and 10 seconds to occur. According to Figure 7.6, a cell may reproduce at least 6 times per second. At the population level ($\times$ 30 cells), this would increase to 180 cellular reproductions per second[1]. If we consider the approximate number of cellular reproductions that may occur in 5 to 10 seconds (between 800 and 1800) and the relatively small cellular population size employed (30 cells), then it may be envisaged that drift dynamics, as proposed by the neutral theory of molecular evolution formulated by Kimura (1983), only could result in those displacements (which would then appear to be abruptly occurring due to the employed timescale) between c_1, c_2 and c_3.

[1]In real experimental conditions, the cellular reproduction rate would usually be higher than 180 reproductions per second as the initial molecular amounts would differ from those employed in Fig.7.6. Through successive cellular reproductions, the molecular amount distribution may vary mechanically and lead to a faster reproduction speed, being typically twice faster with an initial molecular amount of 100 s_1 and fewer molecules s_2, s_5 and s_6.

Also note in Figure 7.4 that these displacement are associated with a relative increase in the cellular species diversity. This suggests that other cellular species, which are potentially mutants of the pair of dominant cellular species, may have contributed to these displacements.

Another possible explanation of those displacements, is that c_2 and c_3 may have incrementally increased their capacity to resist against potentially disruptive mutation effects (mutations occurring at both the cellular and molecular level). This hypothesis, where the networks have improved their robustness (Wagner, 2005), could be tested in future work where each cellular lineage would be isolated and examined in details when subjected to mutational perturbations.

Finally with regards to the evolutionary of growth complexity, we note that in this evolutionary experiment, the number of molecular reactions was increased from 6 to 9 when comparing the seed and evolved reaction networks. Nevertheless the complexity of the molecular species remains equivalent with an average string length of 6. Moreover both the seed and evolved reaction networks contain the same number of molecular species. Although the seed reaction network was successfully evolved and optimised to achieve the pre-specified task, we did not observe a significant growth of complexity in this experiment.

10 additional repetitions of the above experiment were conducted and are described in Appendix D.3. In four of these runs, we observed the emergence and domination of either c_1 or c_3. In four other runs, the emergence of reaction networks containing the molecular species of either c_1 or c_3 in addition to some other molecular species were noted. However these additional molecular species did not improve the fitness of the cellular species. It is thus conjectured that, given enough time, these reaction networks would displace these extra molecular species and collapse to c_1 or c_3. In the remaining two runs, the emergence of c_0 mutants with no phenotypic differences was observed.

7.2.4 Conclusion

We introduced a novel cellular division criterion as a means to drive the evolution of closed reaction networks to accomplish a pre-specified task. The latter was devised to promote the growth of a target molecular species (i.e., signal amplification). Preliminary experiments, where self-replication reactions were enabled, showed evolutionary dynamics of limited interest with regards to the evolutionary growth of complexity. We then proposed to conduct further experiments in which we disabled self-replication reactions. A hand-designed minimalist collectively autocatalytic reaction network was presented and used as the seed/ancestor network. A series of evolutionary experiments were conducted in which we identified the common emergence of a "fitter" reaction network. We examined this evolved reaction network and demonstrated its ability to produce the target species at a faster rate whilst maintaining closure. Although the seed closed reaction reaction was successfully evolved and optimised to achieve the pre-specified task, a significant evolutionary growth of complexity was not observed in this experiment. In the next section, we extend this work and explore further avenues for evolving closed reaction network of higher complexity.

7.3 Crosstalk and the evolution of complexity

In Section 7.2 , we successfully evolved a simple signal processing ability in a closed reaction network using the cellular and parallel implementation of the MCS.bl. This resulted in the optimisation of a *minimalist* closed Cellular Information Processing Network capable of a distinct signal-processing function.

Nevertheless, the previous experiment failed at exhibiting a clear evolutionary growth of complexity. In this section, we extend this preliminary work on the evolution of closed reaction networks and intend to evolve networks of higher complexity.

To assist this research on the evolutionary growth of complexity, we examine

a phenomenon occurring in real biochemical networks: crosstalk. Crosstalk phenomena arise very naturally in such networks due to the fact that molecules from different signalling pathways may share the same physical reaction space (the cell). Depending on the relative specificities of the reactions there is then an automatic potential for any given molecular species to contribute to signal levels in multiple pathways.

In Section 1.3, we proposed a potential benefit of crosstalk in chemical networks: Crosstalk is a key mechanism in enabling incremental evolutionary search for more complex closed reaction networks.

We argue that the above benefit may be achieved through crosstalk by allowing distinct closed reaction networks to cooperate with each other when occurring in the *same* reaction space. We thus propose that crosstalk may enable the merging of distinct crosstalking closed reaction networks to form a new closed reaction network of higher complexity.

This work is thus naturally related to the symbiogenesis theory which was originally postulated by Mereschkowsky (1910). According to this theory, separate organisms may merge with each other to form new organisms of higher complexity (Margulis, 1981; Margulis and Sagan, 2002). Barricelli was the first to conduct computer-based experiments in which symbiogenetic organisms were artificially evolved (Barricelli, 1957, 1963).

More specifically, the work presented in this chapter is inspired by seminal related experiments conducted in Alchemy (Fontana and Buss, 1994a). Based on this preliminary work, we hypothesise that crosstalk enables the cooperation and subsequently, the evolutionary growth of complexity of biochemical networks. We develop further this hypothesis using the MCS.bl.

7.3.1 Introduction

This investigation on crosstalk and complexity was inspired by specific experiments carried out by (Fontana and Buss, 1994a) with the Alchemy system. When mixing two collectively autocatalytic reaction networks (which were obtained from previous independent experiments) in the same reaction space, two outcomes could be observed according to the level of interaction (i.e., crosstalk) between the two reaction networks:

1. If no molecular interaction (i.e., no crosstalk) exist between the two networks then one would displace the other network.

2. If, on the contrary, some molecular interactions occur between the two crosstalking networks then a "meta" hierarchical closed reaction network emerges which contains and maintains both seed closed reaction networks.

These observations suggest that crosstalk may be responsible for the emergence of molecular organisations of higher complexity. To develop further this hypothesis we extend this seminal investigation using the MCS.bl. However a number of key differences exist between Alchemy and the MCS.bl:

- Alchemy is based on the $\lambda-$calculus formalism, whereas the MCS.bl employs the broadcast language. Although both agent-based ACs employ term-rewriting systems, the specification of molecular species and reactions (binding rules and enzymatic capabilities) varies greatly. For example, there is only a single level of enzymatic/computational transformation that can be defined in molecular reactions within the MCS.bl. Whereas in Alchemy, the analogous maximum number of reduction steps was set to 10000.

- We define mutation operators at both the molecular and cellular level. No evolutionary operators were specified in Alchemy. In Alchemy the molecular

diversity resulted from the initial randomly generated molecular population and subsequent catalytic molecular reactions. The MCS.bl introduces a greater space exploration of molecular species by implementing molecular mutations. This molecular and cellular variance allows for *evolutionary* dynamics to occur in the MCS.bl.

- Similarly to Alchemy, molecular species may interact and compete with each others. In addition to this first level of selection we introduced a higher level of selection: Molecules are contained in multiple reactors (i.e., cells) which are capable of competing with each others through cellular divisions. As demonstrated in Chapter 6 this second level of selection was necessary to allow closed reaction networks to self-maintain in the MCS.bl. As no mutations could occur in Alchemy, no evolutionary degenerations were observed in this system where only a single level selection was implemented.

- In our approach we evolve closed reaction networks to carry out pre-specified tasks. In Alchemy, the reaction networks were not driven to perform any target functions. Introducing a target task in the MCS.bl affects the implicit fitness function devised in this system. In addition to performing self-maintenance (as in Alchemy), closed reaction networks in the MCS.bl have to carry out a pre-specified task. Both the ability to perform self-maintenance and the pre-specified task affect the fitness of a given closed reaction network in the MCS.bl. The fitness landscape in the MCS.bl is thus multi-dimensional where the dimensions are:

  1. The cellular reproduction rate which depends on the ability of the cell to achieve the target task.
  2. The ability to maintain closure which is required to control potentially disruptive mutational effects. Note that closure properties may themselves evolve to better control evolutionary degenerations.

We now report three series of experiments addressing crosstalk and the evolutionary growth of complexity in closed reaction networks. Although these experiments may be directly related to level 2 organisation experiments conducted in Alchemy, they partially diverge from the Alchemy experiments as molecular and cellular mutations occur and introduce the potential for evolutionary dynamics. The first and second series of experiments involve non-crosstalking and crosstalking closed reaction networks respectively. In the first two experiments, only cell-level mutation applies[2]. Therefore the diversity of molecular species in these two experiments is limited. The only molecular species that may appear in the simulation runs, involving non-crosstalking networks, are the initial molecular species contained in the seed networks. Molecular-level mutations are disabled to facilitate the comparison with the Alchemy system in which no mutations were implemented. α being the total number of molecular species that may appear and ν the number of initial molecular species, we have $\alpha = \nu$. The crosstalking networks based experiment includes further molecular species which may result from novel reactions occurring between the crosstalking molecular species, therefore $\alpha \leq \frac{\nu^2}{2}$. Cell-level mutations may produce mutant cellular species in which only the specific assortment of molecular species (given this limited set of molecular species) may vary. The third experiment examines systems of crosstalking closed reaction networks where both cellular *and* molecular mutations occur. The potential diversity of molecular species in this final experimental series is thus significantly increased to $\alpha = \sum_{L=1}^{BD_{Lmax}} |\lambda|^L$.

7.3.2 The seed reaction networks

In the following experiments, no self-replication reactions (i.e., reactions in which the product is syntactically identical to the enzyme and substrate molecules) are allowed (as was the case in analogous Alchemy experiences). As briefly discussed in Sec-

[2]Cell-level mutations occur during cellular reproductions where the stochastic distribution of molecules may result in mutant cells. Cell-level mutations are an inherent feature of the compartmentalised MCS.bl and cannot be disabled.

tion 7.2.1, enabling self-replication would prevent any relatively complex molecular organisations from emerging. However we showed that the spontaneous emergence of closed reaction networks was unlikely to occur given a randomly generated population of molecular species. Thus, similarly to Section 7.2, we employ seed closed reaction networks to initialise the molecular populations in the experiments.

We define the different reaction networks X, Y and Z which are utilised throughout these series of experiments, see Table 7.3.

X	Y	Z
$s_1 = *\nabla 00 : \nabla 01$	$s_5 = *\nabla 10 : \nabla 11$	$s_9 = *\nabla 10 : \nabla 00$
$s_2 = *\nabla 00 : \nabla 00$	$s_6 = *\nabla 10 : \nabla 10$	$s_{10} = \nabla 1 * \nabla 00 : \nabla 10$
$s_3 = *\nabla 0\diamond : \nabla 00$	$s_7 = *\nabla 1\diamond : \nabla 10$	$s_{11} = *\nabla 10 : \nabla 10$
$s_4 = *\nabla 0\diamond : \nabla 01$	$s_8 = *\nabla 1\diamond : \nabla 11$	$s_{12} = \nabla 1 * \nabla 00 : \nabla 00$

Table 7.3: Molecular species contained in seed closed reaction networks X, Y and Z

No molecular species from X interact with any molecular species from Y and *vice versa*. X and Y are declared as *non-crosstalking* reaction networks. The species s_1, s_2, s_3 and s_4 from X may interact with species s_9 and s_{12} from Z, whereas species s_{10} and s_{12} may interact with s_2 and s_3 from X. X and Z are declared as *crosstalking* reaction networks.

X, Y and Z were obtained from previous experiments in which they were evolved to maximise the production of molecular species s_1, s_5 and s_9 respectively. Fig. 7.8 depicts the bipartite reaction network graphs of networks X, Y and Z. Note that X and Y possess the same network topology. The number of instances of a given molecular species $s_j \in S$ contained in a cell $c_i \in C$ is denoted as n_j^i. All simulations are run for a pre-defined amount of time $t_{max} = 3600$ (seconds in real time) using thirty AMD Opteron 270 (2.0 GHZ) CPUs.

7.3.3 Non-crosstalking networks

In this first series of experiment, we investigate the dynamics of a system in which the non-crosstalking closed reaction networks X and Y are used. 30 concurrent cells

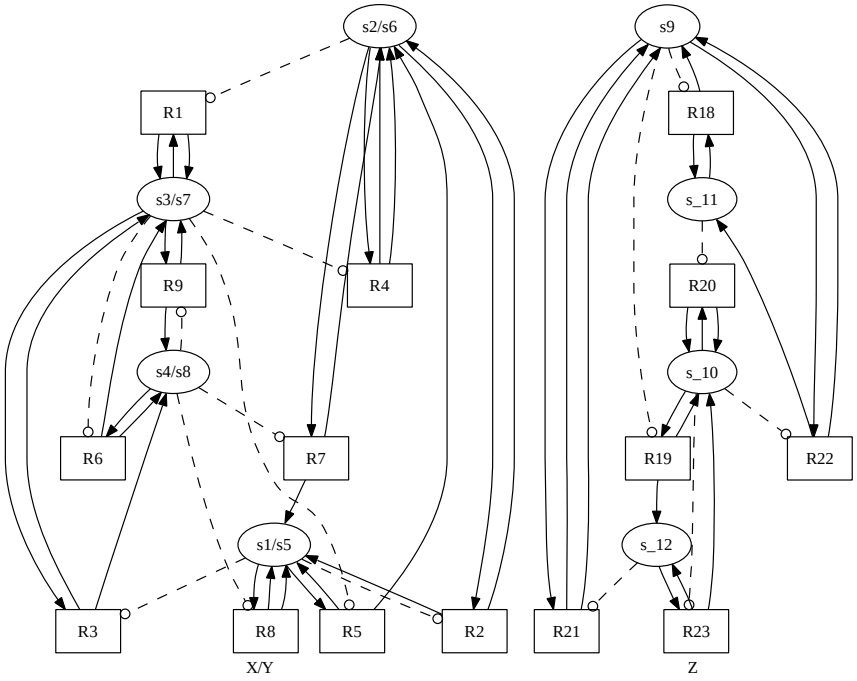

Figure 7.8: Bipartite reaction network graphs of networks X/Y and Z. The topology of molecular interactions of X and Y are equivalent, e.g., the reaction $R4$ would involve the molecular species s_2 and s_3 in X, whereas $R4$ would involve the molecular species s_6 and s_7 in Y.

are employed and initialised with 10 molecules from each species from both X and Y.

As previously mentioned, X and Y were evolved to optimise the production of species s_1 and s_5 respectively. We devise a new cellular division criterion which accounts for both molecular species. The motivation to this criterion is to encourage the maintenance of both networks X and Y. Ultimately we aim at evolving/obtaining a more complex network capable of "multitasking", i.e., a network which is able to carry out the pre-specified tasks of both X and Y. Therefore, a cell c_i divides if $n_1^i \geq 200 \wedge n_5^i \geq 200$. In this experiment, only mutations at the cellular level occur (i.e., mutations at the molecular level are excluded at present and $\alpha = 8$).

Fig. 7.9 depicts the growth of s_1 and s_5 and number of cellular reproductions at the cell population level in a single simulation run. Five additional simulation runs were conducted to explore any significantly differing dynamics. The dynamics

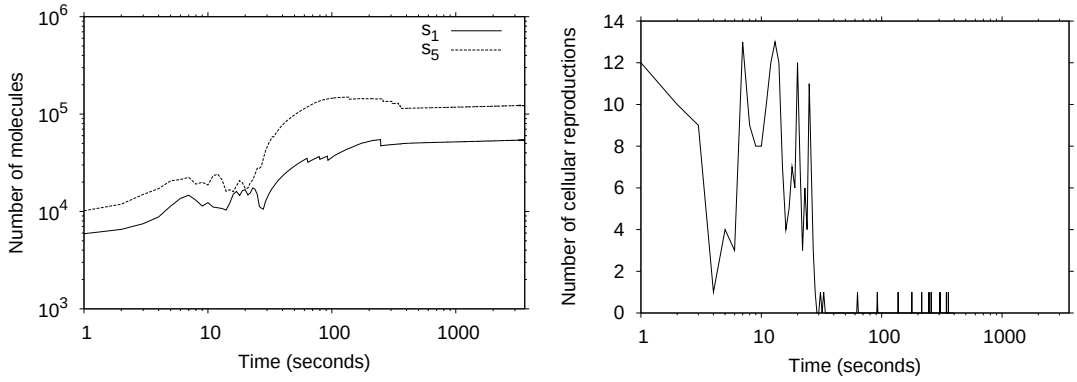

Figure 7.9: Growth of molecular species s_1 and s_5 (left) and number of cellular reproductions (right). Both graphics depict dynamics at the cellular population level in an example simulation run.

described here were found to be exhibited in all of these runs.

The number of cellular reproductions provides an approximate indication of the cells' fitness, i.e., an absence of cellular reproduction would suggest that the cells are not producing both s_1 and s_5 sufficiently to trigger their division. In such cases the cells would thus possess a relatively low fitness.

In Fig.7.9, we first observe an early phase where both the number of s_1 and s_5 molecules vary between 10000 and 11000. Moreover this phase is associated with recurrent cellular reproduction events.

At $t \approx 32$ we note that the number of s_5 is now rapidly increasing, reaching up to 1.0×10^5 when $t \approx 80$ whereas s_1 increases up to 5.0×10^4. From $t > 400$, no further cellular reproductions occur.

Throughout this simulation run, 12 different and unique reaction networks were generated due to cell-level mutations. The growth dynamics of these cellular species are depicted in Fig.7.10.

In this run, we distinguish the following chain of events at the cell population level (Fig.7.10):

- We note that the early phase $0 \leq t \leq 32$ mentioned earlier, where repeated cellular reproductions are observed, is driven only by the cellular species c_1

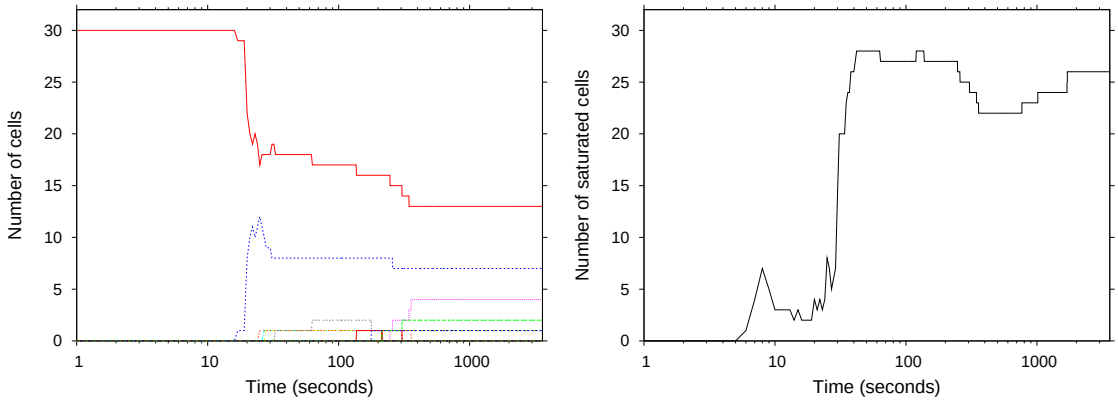

Figure 7.10: Growth dynamics of the 12 cellular species (left) and saturated cells (right). The two dominant cellular species, the red and blue curves, are denoted by c_1 and c_2 respectively. c_1 includes the molecular species from both X and Y, $c_1 = X + Y$. Whereas c_2 is a mutant cellular species of c_1 which does not include the molecular species s_3, $c_2 = \{s_1, s_2, s_4\} + Y$. All reaction networks contained in remaining cellular species are subsets of $X+Y$. A cell c_i is considered as "saturated" if $n_1^i > 2000 \vee n_5^i > 2000$.

(which contains both closed reaction networks X and Y). From $t \geq 32$, we distinguish the emergence of various mutant cellular species which contain reaction networks being subsets of $X + Y$. This emergence of mutant species is associated with the sudden decline in the number of cellular reproductions.

- Moreover we observe that the number of "saturated" cells increases rapidly when $t \approx 32$ which correlates with previous observations reported in Fig. 7.9 where the number of s_1 and s_5 molecules starts to increase rapidly.

- This cellular saturation suggests that although some c_1 cells are still present (which are capable of producing both molecular species s_1 and s_5), these cells are overpopulated/saturated with either species s_1 or s_5. As a result, this cellular saturation and the "survival of the common" dynamics (Section 5.1) occurring here, cause the production rate of s_1 and s_5 to become highly *asymmetric* (as depicted in Fig.7.11).

- When $200 < t < 400$ we note that the number of saturated cells decreases, resulting from the cellular reproduction events occurring sporadically during

this period (Fig.7.9). Nevertheless, from $t > 600$ the number of saturated cells starts to increase again, indicating that either the species s_1 or s_5 are being produced despite no further cellular reproductions occur.

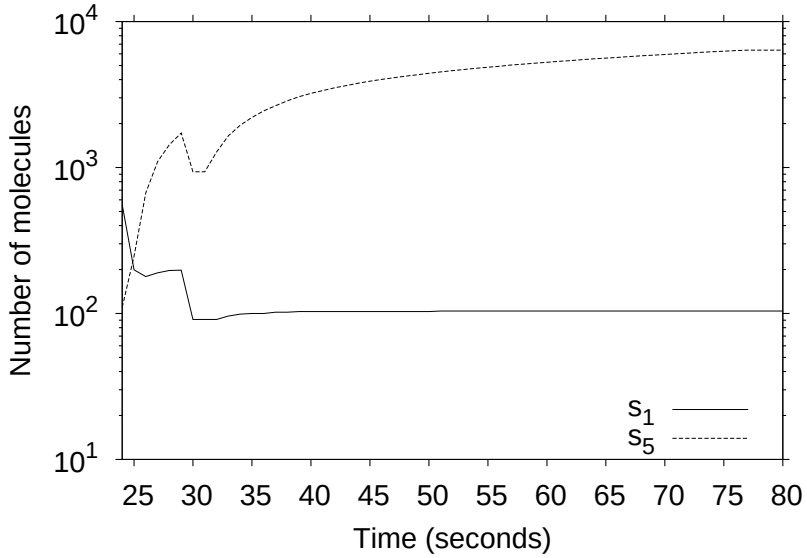

Figure 7.11: Growth of molecular species s_1 and s_5 in an example cell containing the network c_1 when $24 \leq t \leq 80$.

A complementary investigation revealed that some of the non-saturated cells contained reaction networks in which no successful reaction could occur. These reaction networks were essentially composed of target species s_1 and s_5 only. Therefore no chemical interaction could occur within these reaction networks which resulted from cell-level mutations (as defined in Section 6.4.2).

During the successive cellular reproduction events, the numbers of molecules s_1 and s_5 contained in a given cell increasingly deviated from each other until one of these two species started to take over the cell. As a result the production rate of s_1 and s_5 molecules in the reaction networks (including closed ones) was highly asymmetric. In such cells saturated by s_1 or s_5, further divisions (if any) resulted in offspring cells which were likely to contain a majority of, or only, s_1 or s_5 molecules.

Such resulting mutant cellular species did not contain all molecular species necessary to maintain closure for both X and Y. As the closure of X and Y was not maintained in these cells, this consequently penalised the production of further

molecules s_1 or s_5. As a result, the cellular division threshold became increasingly more difficult and ultimately impossible to reach in these cellular lineages.

Although based on a significantly different Artificial Chemistry, these experiments essentially exhibited the same dynamics as in related experiments conducted in Alchemy: the different reactors were rapidly and quasi-deterministically dominated by one of the seed closed reaction networks. In other words, when two non-crosstalking closed reaction networks are mixed together, one displaces the other one.

This above experiment also relates to results obtained with the stochastic corrector model proposed by Szathmary and Demeter (1987). In Szathmary and Demeter's cellular model, the survival of the cells depended on the concentration of two distinct self-replicases having differing growth rates. The difference in growth rate was due to the ability of one the self-replicases to parasite the other one. Cells in which the concentration of the replicases deviated too importantly from each other were selected against.

Although major differences exist between our model and Szathmary and Demeter's one (e.g., we do not employ self-replicase species but collectively autocatalytic reaction networks each of which possesses an initial common growth rate), these models share similarities where the survival of the cells depends on the concentration of two distinct molecular species.

Using the stochastic corrector model, Szathmary and Demeter demonstrated that multi-level selection in such a cellular model was capable of controlling parasitism. However the regulation of the degenerative outcomes due to stochastic variations, occurring during the transmission of molecular species into offspring cells, was possible only when the number of self-replicases was small. This limitation may be involved in the current experiment where: 1) 8 distinct molecular species were necessary to maintain closure. 2) The cellular division threshold required 200 molecular instances of both s_1 and s_5. This may subsequently have affected the survival of the

cells over successive cellular divisions.

Therefore, we may argue that given two distinct non-crosstalking collectively autocatalytic reaction networks containing fewer molecular species and a cellular division threshold requiring less molecules, the current cellular model would theoretically be able to select against cells in which significant deviations occur between both target molecular species. As a result, this model may potentially regulate the degenerative effects due to stochastic variations occurring during successive cellular divisions. However, this hypothesis could not be applicable in Alchemy since the latter lacks the multi-level selectional regime specific to the MCS.bl and stochastic corrector model.

7.3.4 Crosstalking networks

We investigate the effects of crosstalking closed reaction networks upon the system's dynamics. In this experiment, the cells are initialised with molecular species from the crosstalking reaction networks X and Z. A cell c_i divides if $n_1^i \geq 200 \wedge n_9^i \geq 200$. The number of molecular species that may appear in the simulation runs is $\alpha \leq \frac{8^2}{2} = 32$. Any other experimental conditions are identical to those described in the previous section.

Our results showed that the interactions between molecular species from X and Z led to the production of new molecular species s_{13}, s_{14}, s_{15} and s_{16} (which may engage in novel reactions with existing molecular species). This new cellular species, denoted as c_1, contains both networks X and Z, and presents an increased level of complexity (the reaction network now contains 12 molecular species and 55 reactions, see Fig. 7.12). Moreover these c_1 cells were able to self-maintain for a sustained period of time ($\approx$ 400 seconds). This first observation also applied in analogous experiments conducted in Alchemy, in which a meta-reaction network emerged and had the ability to self-sustain and maintain both seed closed reaction networks.

However, an additional phenomenon occurred which was not observed in the

Alchemy system. We distinguish a selective displacement event between c_1 and a new cellular species. In this simulation run, a level of diversity (see Fig. 7.13) was maintained due to cell-level mutations (27 unique reaction networks appeared during this run), a feature specific to the MCS.bl. At $t \approx 380$ we note the emergence of a new cellular species, denoted as c_2 and shown in Fig. 7.14, which later displaced c_1 at $t \approx 400$. During this displacement phase, we note that the cell diversity also increased suggesting that other cellular species may also have contributed to the displacement of c_1 cells.

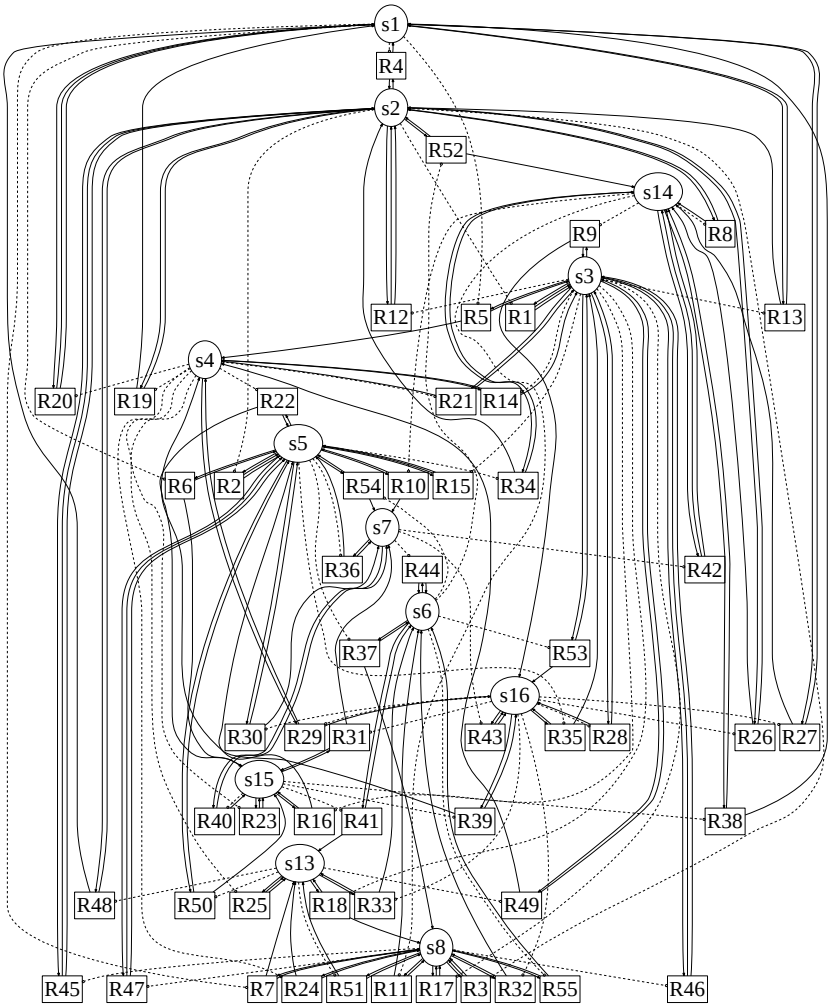

Figure 7.12: Reaction network of cellular species c_1 which contains all molecular species from networks X and Z in addition to new molecular species s_{13}, s_{14}, s_{15} and s_{16}.

In Fig. 7.15 we compare the fitness of reaction networks c_1 and c_2. The fitness of

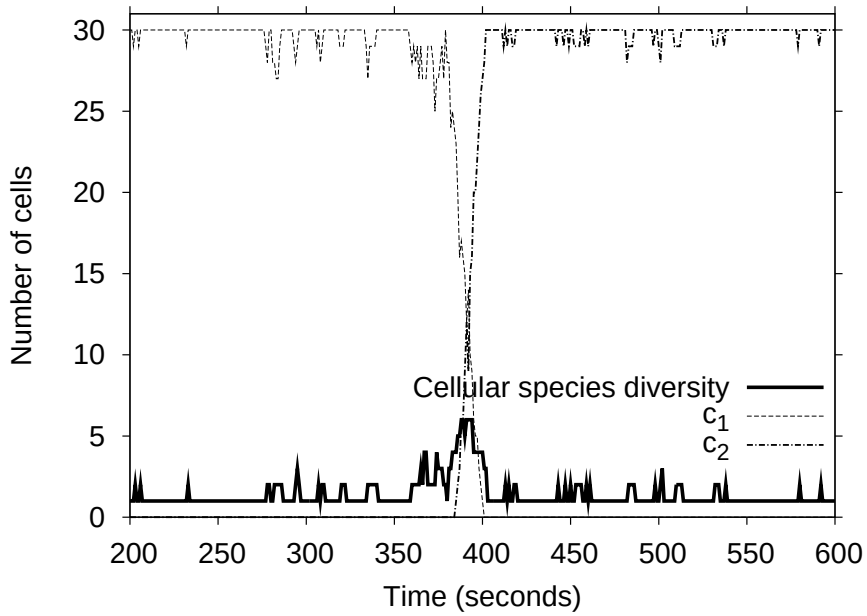

Figure 7.13: Cellular species displacement between c_1 and c_2. The cellular species diversity refers to the number of different (from a qualitative - topological point of view) reaction networks present at a given timestep.

a given cell c_i is reciprocal of time, where both the condition $n_1^i \geq 200 \wedge n_9^i \geq 200$ and the necessary time t_{c_i} to satisfy this condition are accounted for. With the present parallel system, as the speed of production of species s_1 and s_9 increases (augmenting the reproduction rate of the cell), the fitness of the cell increases accordingly.

We note in Fig. 7.15 that c_2 cells produce molecular species s_1 and s_9 at a faster rate than c_1 cells (i.e., $t_{c_2} < t_{c_1}$). According to our definition of fitness, c_2 cells are fitter than c_1 cells. Both the evolved qualitative properties of c_2 and the exploitation of crosstalk led to the maximisation of the production of molecular species s_1 and s_9. We also identify this increase in fitness in Fig. 7.16, in which we distinguish a net increase in the overall cellular reproduction rate following the displacement event. The multitasking c_2 cells were able to self-maintain throughout the entire simulation while cell-level mutations continued to occur. This clear selective displacement event occurred as a direct consequence of both:

- The cell-level mutations which increased cellular diversity.

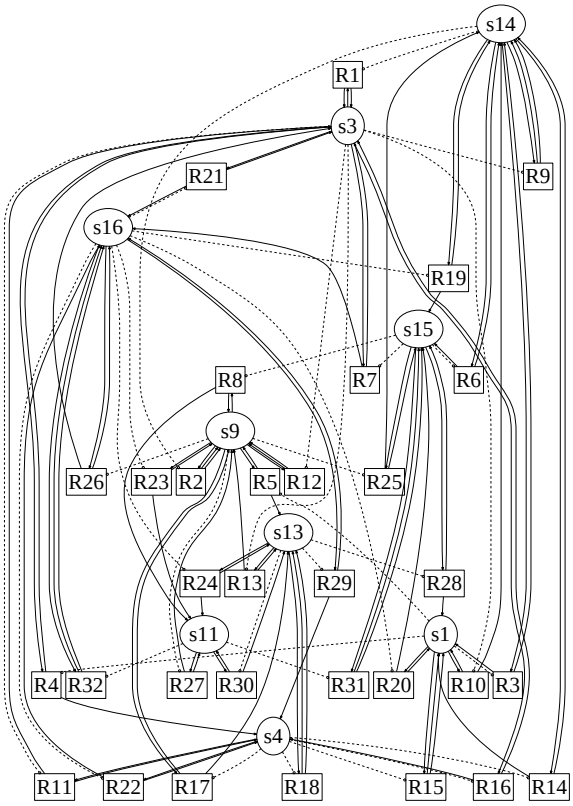

Figure 7.14: Reaction network present in cellular species c_2 in which molecular species s_2, s_{10} and s_{12} from c_1 are absent.

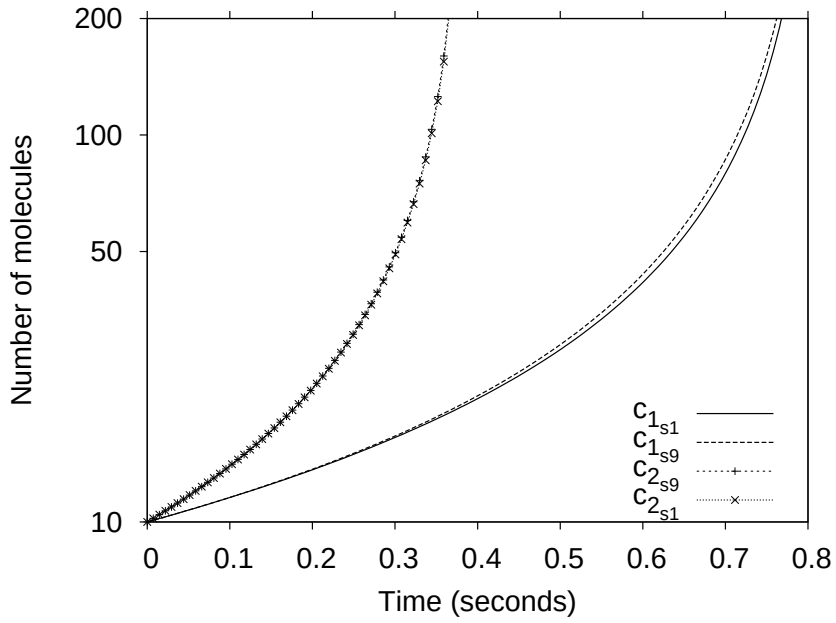

Figure 7.15: Comparison of molecular growth of species s_1 and s_9 in c_1 and c_2. Similarly to Fig.7.2, the time courses of the species' growth were rescaled using experimental data.

- The more elaborate implicit fitness function which affected the fitness landscape. This multi-dimensional fitness landscape allowed for an incremental evolutionary improvement to occur.

The above properties are specific to the MCS.bl and were not present in Alchemy. As a result, comparable evolutionary dynamics described in this section have not previously been reported using Fontana and Buss's system.

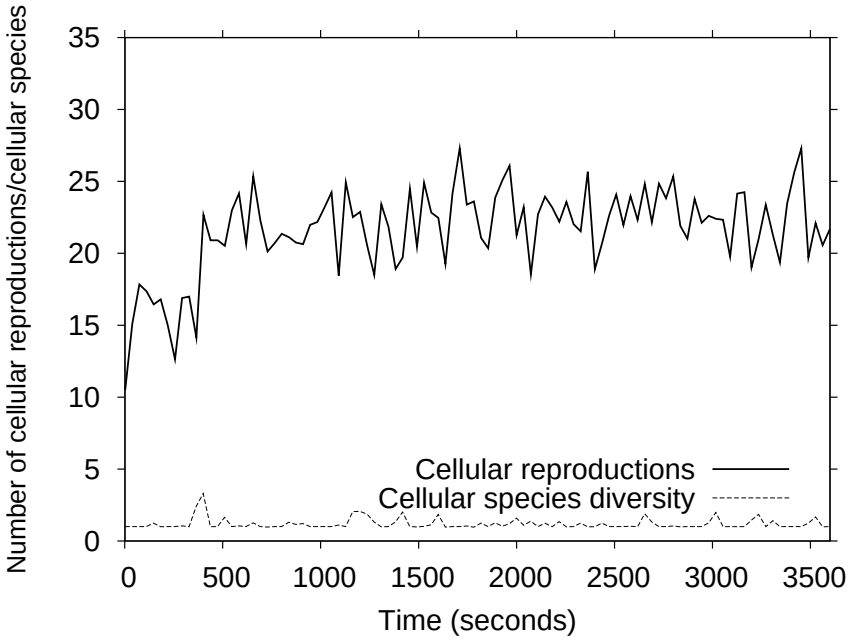

Figure 7.16: Crosstalking networks with no molecular mutation - Dynamics of cellular reproductions and diversity. A spline function was employed to approximate the number of cellular reproductions and cellular species diversity curves.

We also note that the complexity (i.e., the number of molecular species and reactions) actually decreased in c_2 (9 species and 32 reactions) when compared with c_1 (12 species and 55 reactions). This observation suggests that the increased level of complexity of c_1 did not provide any beneficial features, but on the contrary, reduced the speed of reproduction (i.e., fitness) of c_1. The lower level of complexity of c_2 led to a lower computational cost (and consequently a faster reproduction speed) whilst maintaining closure. This ultimately provided c_2 with a selective advantage over c_1.

Finally, as some molecular species from X and Z have been removed in c_2, the latter was thus no longer maintaining the seed original reaction networks X and Z.

As we cannot identify X and Z in c_2, a natural open-question follows: Does c_2 still contain crosstalking closed reaction networks? Such a question could be addressed if we employ an adequate formalism and identify distinct closed reaction networks as subsystems in c_2. This issue is nevertheless beyond the scope of this investigation but the reader may find further details in Wolkenhauer and Hofmeyr (2007), where an abstract cell model is proposed to investigate such issues.

7.3.5 Crosstalking networks with molecular mutations

We finally examine the effects of molecular mutation in systems where the crosstalking reaction networks X and Z are used. Molecular mutations introduce a higher level of both molecular and cellular diversity, which may potentially lead to more complex molecular organisations and richer evolutionary dynamics. Molecular mutations occur with the following probability: $p_{sym} = 5.0 \times 10^{-5}$. Complementary experimental parameters are identical to those presented in Section 7.3.4. Using these conditions, we conduct an experiment in which we identify the following distinctive behaviour.

We first note in Fig. 7.17 that the dynamics of the cellular reproduction rate shares some similarities with analogous dynamics shown in the previous experiment (Fig. 7.16). Indeed we observe a common early phase where the cellular activity is approximately equal to 16 cellular reproductions per second, then at $t \approx 250$ the cellular reproduction rate starts to increase. This common early dynamic is driven by the same cellular species c_1 (i.e., the meta-reaction network containing both seed reaction networks X and Z) which was also able to self-maintain for a period of time. However due to molecular mutations occurring, a significant difference exists in the cellular species diversity. Here a higher average level of cellular diversity per second is observed, being roughly 20 times higher than in the previous experiment, and is maintained throughout the evolutionary simulation. During this run, 37863 unique reaction networks appeared.

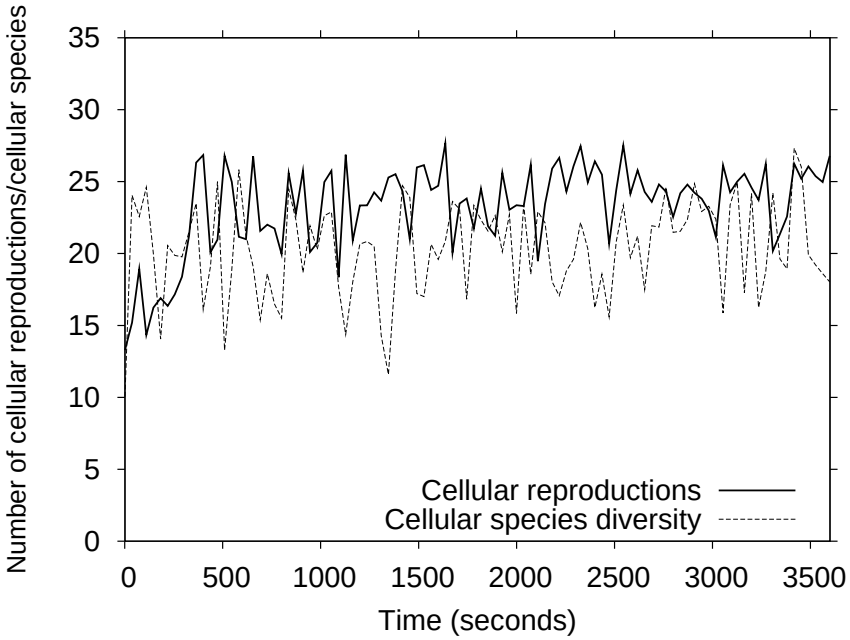

Figure 7.17: Crosstalking networks with molecular mutations - Dynamics of cellular reproduction rate and diversity when molecular mutations occur. A spline function was employed to approximate the cellular reproductions and cellular species diversity curves

In Fig. 7.18, we note that two cellular species displacements occurred at $t \approx 475$ and $t \approx 2500$. The cellular species c_1 is similarly displaced by a mutant cellular species, denoted by c_2 which contains a reaction network that is phenotypically equivalent to c_2 described in previous experiment (Section 7.3.4). The third emergent dominant cellular species is denoted by c_3. The cellular species c_3 shares an equivalent level of complexity (containing 13 molecular species and 66 reactions) with c_1 cells.

In addition, it can be observed that the cellular species' subpopulation rarely exceeded half of the total population. The dominating cellular species have not once succeeded at *fully* displacing the other species for a sustained period of time.

In typical evolutionary simulations it is usually expected to observe incremental improvements in the species' fitness. However when comparing the fitness of the different successive dominant cellular species (Fig. 7.19), we note that this incremental evolutionary improvement did not occur according to our definition of fitness. As

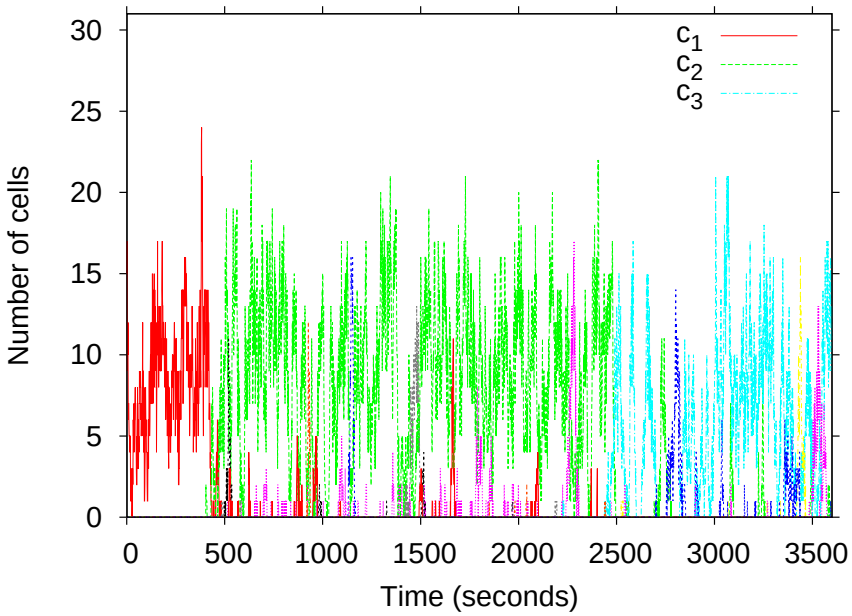

Figure 7.18: Dynamics of the major cellular species. Only the cellular species which invaded, at least once, one third of the cellular population during the simulation run are shown (14 cellular species are plotted).

$t_{c_2} < t_{c_3} < t_{c_1}$, it may be argued that the cellular species c_3 is fitter than c_1 and less fit than c_2. We would thus expect the cellular species c_2 to be the dominant species and not c_3). Therefore our definition of fitness is unsatisfactory here.

When comparing the overall cellular reproduction rate depicted in Fig. 7.16 and Fig. 7.19, we identify a roughly equivalent level of cellular reproduction rate ($\approx$22 cellular reproductions per second). This would thus indicate that although c_2 are fitter (producing molecular species s_1 and s_9 more rapidly) than c_3, the latter (or potentially the cell population as a whole) may have developed other features which maintained a similar cellular reproduction rate.

The details of these particular evolutionary dynamics remain unclear and have not been examined within the timeframe of this thesis. Nevertheless we formulate a number of potential explanations that might merit future investigation:

- Our simplistic view of fitness may not be appropriate in the current experiment. As molecular mutation is now occurring, the cellular species *or* the cellular population as a whole may have developed new features to cope with

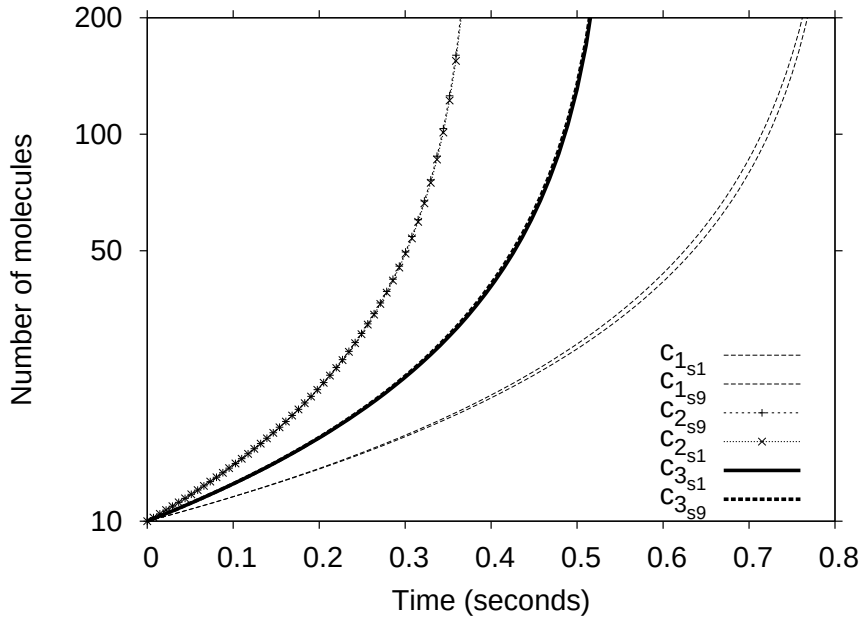

Figure 7.19: Comparison of molecular growth of species s_1 and s_9 in c_1, c_2 and c_3.

negative mutation effects. These features may have enabled the cellular population to maintain a competitive overall cellular reproduction rate while mutations occur. Such features improving the cellular reproduction rate and robustness (Wagner, 2005) should then be accounted for in the cellular species' fitness. This reflects the complex multi-level nature of fitness in chemical reaction networks (Bersini, 2002).

- Our classification of cellular species may not expose the dominant cellular species adequately. A different classification scheme may be defined which would be based on some key properties of the cell's reaction network (and not only on the molecular species being present in the cell).

- The chaotic nature of the dominant cellular species dynamics (Fig. 7.18) may also suggest that the observed displacements might not only be due to selection. This chaotic behaviour may have resulted from the relatively small cellular population size employed here. This parameter may have increased the sensitivity of the cellular population to statistical fluctuations. Note that the

choice of population size was essentially driven the number of CPUs available in the experimental cluster.

- Finally we note that in the different closed networks employed and evolved in Appendix D.2, Section 7.2 and in the current section, the number of molecular species necessary to maintain closure was successively increased. When comparing the dynamics observed in these experiments, we remark that, as this diversity of molecular species increased, the maintenance and domination of cellular species became more chaotic. For example, drift dynamics only appeared in the experiments, conducted in Section 7.2, in which the closed reaction networks were composed of four distinct molecular species.

  We propose that this diversity of molecular species may have been implicated in the different dynamics described in this chapter. Indeed, Szathmary and Demeter (1987) demonstrated, using the stochastic corrector model, that variations due to the stochastic transmission of molecular species during cellular divisions may result in degenerative outcomes (i.e., the cellular species cannot self-maintain over time) when the number of both the molecular species and molecules required for the survival of the cells is too important.

  If this constraint is in effect implicated in the experiments presented in this chapter, then dealing with more molecular species and instances (i.e., more complex information) would increasingly become more difficult using the MCS.bl, averting any significant evolutionary growth of complexity. This would *ultimately* suggest the limitations of the MCS.bl to encode and process more complex information using autocatalytic networks only. To overcome this limit, it may be conjectured that a complementary mechanism enabling the stable storing and subsequently processing of more complex information might be required in the current model. A genetic subsystem could for example address this requirement and potentially lead to the evolution of higher forms

of digital organisms. This would thus naturally relate to a major evolutionary transition as proposed by Smith and Szathmáry (1997).

This experiment presented a range of interesting and unexpected issues which resulted directly from the key differences existing between Alchemy and the MCS.bl system. Further analytical work using adequate tools such as Organisation Theory (Dittrich and Speroni, 2007) may also illuminate these complex evolutionary dynamics.

7.3.6 Conclusion

Inspired by specific experiments related to crosstalk conducted with Alchemy by Fontana, we investigated a potential constructive role of crosstalk: To allow distinct closed reaction networks to cooperate with each other when occurring in the same reaction space. This cooperation would then lead to the emergence of molecular organisations of higher structural and functional complexity. We indicated the similarities and key differences between the Alchemy system and the MCS.bl. Three series of experiments were then detailed:

1. Two non-crosstalking closed reaction networks were employed. Although significant differences exist between the MCS.bl and Alchemy, we essentially identified a similar behaviour: one reaction network would displace the other.

2. Two crosstalking closed reaction networks were utilised. We first noted a phenomenon (which also occurred in the corresponding Alchemy experiments), in which a meta-reaction network emerged and contained both seed closed-reaction networks. This new cellular species was able to self-maintain for a sustained period of time. However a second phenomenon occurred (which was not observed in Alchemy), in which a selective displacement took place. A mutant cellular species emerged and displaced the meta-reaction network. This mutant cellular species was no longer maintaining the seed reaction networks

but was in fact fitter at performing the pre-specified tasks. This resulted in a faster cellular reproduction rate which provided the mutant species with a selective advantage.

3. Two crosstalking closed reaction networks were used and molecular mutations were applied. We identified a common selective displacement as reported in previous experiment. However an additional cellular species displacement was observed and presented evolutionary dynamics which are not fully understood. We discussed some of the possible pitfalls of our analysis and outlined potential explanations.

These experiments demonstrated the constructive role of crosstalk in enabling co-operation to occur between closed reaction networks. The evolutionary process was also able to optimise the reaction networks and their crosstalk properties to carry out the pre-defined multitask function.

The resulting evolved networks presented a higher level of functional and structural complexity which supports our initial hypothesis: Crosstalk is a key mechanism enabling the evolutionary growth of complexity in biochemical networks. More precisely, crosstalk enabled the symbiogenesis of separate closed reaction networks to occur, leading to the emergence of novel closed reaction networks of higher complexity. However future work remains necessary as the final series of experiments presented very complicated and difficult to interpret evolutionary dynamics.

7.4 Summary

Using the extended version of the MCS.bl where a cellular model is employed, we conducted a series of experiments focusing on the evolution of closed reaction networks to carry out pre-specified information processing tasks. We first presented a simple experiment in which a closed reaction network was evolved to perform a signal amplification function. Following this, we extended this work and examined

networks of higher complexity. We hypothesised the constructive role of crosstalk to allow the evolution of more complex closed reaction networks. This work was inspired by the symbiogenesis theory and preliminary experiments conducted by Fontana and Buss. We demonstrated that crosstalk was in fact necessary for the cooperation of distinct closed reaction networks. This cooperation subsequently permitted the evolutionary growth of complexity of crosstalking closed reaction networks. This chapter demonstrated the possibility of evolving closed reaction network which are capable of performing information processing tasks.

Chapter 8

Conclusion

This final chapter first summarises and discusses the research contributions of this thesis. Following this, future work that has been identified to extend the work reported in this thesis is outlined.

8.1 Research contributions

The research contributions of this thesis are summarised and discussed as follows:

- *Modelling chemical reaction networks:* A state of the art review on computational techniques applied to the modelling of chemical networks was proposed. Several families of modelling techniques were distinguished: deterministic, stochastic, probabilistic, algebraic and agent-based techniques. For each of these modelling approaches, specific techniques were individually presented and evaluated. A comparison table was provided to highlight the strengths and weaknesses of the different approaches. This evaluation showed that algebraic and agent-based techniques, both families originating from the field of computer science, are the most flexible modelling techniques. This flexibility is essentially due to the high descriptive power of these techniques which allows one to model the hierarchical and intricate nature of chemical networks and molecular species. Moreover, transformation techniques and standardised formats (i.e., SBML, CellML) permit the partial translation of

algebraic/agent-based models into deterministic ones. The latter complements the range of analysis that can be conducted given an original algebraic/agent-based model. Although deterministic and stochastic approaches remain the most employed techniques within the modelling community, the rapidly growing field of algebraic/agent-based modelling suggests the limitations of *traditional* techniques and the need for flexible and more accurate modelling tools.

- *Evolving Cellular Information Processing Networks:* A review on evolutionary techniques applied to the evolution of organisationally closed Cellular Information Processing Networks (CIPNs) was presented. Although no methods directly applied to evolving closed CIPNs were identified, two complementary and indirectly related families of evolutionary methods were distinguished: top-down/Evolutionary Computation (EC) techniques and bottom-up/Artificial Chemistry (AC) approaches. Individual techniques from both evolutionary approaches were selected, presented and evaluated. In terms of realising and evolving computational functions using reaction networks, this evaluation showed that EC techniques have successfully demonstrated the feasibility of evolving chemical networks to perform computational functions. Nevertheless no EC techniques have to date addressed closure and self-organisation dynamics in chemical networks. On the other hand, we showed that ACs have extensively been used to examine the emergence, self-maintenance and evolution of closed reaction networks with little focus on signal-processing capabilities.

  A comparison of both complementary evolutionary approaches was conducted. This comparison identified the explicit definition of fitness functions as the major drawback of EC techniques. Explicit fitness functions constrain the evolutionary process and prevent, by design, an open-ended evolution from occurring. Moreover such explicit fitness functions determine the genotype-

/phenotype mapping which ultimately defines the complexity/hardness of the EC algorithm. As a result the quality of evolved solutions is dependent on the specific design of the employed EC algorithm.

Our evaluation suggested that AC approaches are adequate to evolve CIPNs as these methods rely on implicit fitness functions. In ACs, agents determine by themselves their own fitness and collectively determine the system's fitness as a whole. Although no ACs have so far demonstrated an open-ended evolutionary growth of complexity, as occurring in the biosphere, we concluded that agent-based AC methods are a suitable technique to study the evolution of organisationally closed CIPNs.

- *Evolutionary simulation platform:* A novel simulation platform capable of evolving organisationally closed reaction networks was implemented. This stochastic agent-based system termed the MCS.bl employs the Holland broadcast language to specify the molecular reaction and species. The novelty of this string-based Artificial Chemistry relies on the use of the broadcast language which addresses the reflexive nature of molecular species regarded here as condition/action rules. Since Holland's original proposal in the 1970s, no studies on the broadcast language formalism have been reported in the literature. The work presented in this thesis and related publications constitute the first published evaluation of the broadcast language.

  Prior to the development of the MCS.bl, an implementation of the original broadcast language was also conducted. This system is the first publicly available implementation of the broadcast language and may assist in the evaluation of the broadcast language in allied fields (e.g., Genetic Programming and Genetic Algorithms).

- *Emergence and self-maintenance of closed reaction networks:* A first series of experiments was conducted focusing on the spontaneous emergence and self-

maintenance of closed reaction networks in the MCS.bl. These experiments first demonstrated the role of binding specificity in the dynamics of replicase species. This binding specificity affected the capability of the replicases to displace other molecular species. Although this property may have been previously implicated in the dynamics of a variety of ACs, it has not been explicitly exposed in the manner presented in these experiments.

Additional evolutionary experiments suggested that the spontaneous emergence of autocatalytic species/organisations, being able to self-maintain in the MCS.bl, was unlikely to occur. We suggested a number of factors which may have contributed to this phenomenon which has also been reported in other ACs such as Tierra. These experiments provided supplementary insights on the potential conditions necessary for the spontaneous emergence and self-maintenance of closed reaction networks in ACs.

Following the Tierra system, further evolutionary experiments were conducted in which an ancestor species was employed. These experiments presented unexpected evolutionary dynamics in which a degenerative elongation catastrophe phenomenon was identified. This phenomenon was due to a form of parasitism which prevented replicase species from self-maintaining over time. These results indicated counter-intuitive outcomes when compared with the evolutionary dynamics reported in other ACs.

These results are potentially artefacts of the broadcast language which could have been avoided by utilising a different AC. Nevertheless this remains hypothetical as these degenerative evolutionary dynamics may be due to specific system properties which were desired in this project (e.g., variable molecular length, reflexive structure, pattern matching based reaction scheme).

- *Evolutionary capability in multi-level selectional models:* A parallel implementation of the MCS.bl using distributed computing facilities was conducted.

This extended MCS.bl addressed the concurrent nature of chemical processes and introduced compartmentalisation and multi-level selection. Using this novel version of the MCS.bl, two model variants were evaluated to improve the system's evolutionary capability and prevent the degenerative evolutionary phenomena from occurring.

The first model explored the effects of molecular diffusion between static reactors. This model was inspired by the analytical study conducted by McCaskill et al. which demonstrated that such a model, according to a range of parameters, can stabilise the self-maintenance of closed reaction networks when subjected to disruptive parasitic effects. A series of evolutionary simulations was conducted using a limited range of parameters where the molecular diffusion coefficient was varied. Our results indicated that the elongation catastrophe phenomenon could not be averted in any of these simulations. Although these results suggest that molecular diffusion cannot control the degenerative effects due to parasitism, we cannot infer that there exists (or not) a range of parameters that could provide the autocatalytic species with resistance against parasites in the MCS.bl. These results highlighted the limitations of agent-based systems that are not analytically tractable.

Following this, a cellular model where compartments/cells can grow and divide was evaluated. Similarly to previous experiments, only a limited range of parameters was examined in which the mutation rate was varied. These evolutionary experiments demonstrated that a mutation rate threshold exists where the parasitic effects can be controlled. Moreover, the introduction of chemical kinetics (due to the parallel nature of the model) improved the system's robustness and evolutionary capability in particular cases.

Both experimental investigations provide complementary insights on the potential effects of compartmentalisation over evolutionary capability in Artificial

Chemistries. As no analytically tractable method is currently available for the examination of complex ACs, such empirical investigations are necessary and may provide guidance on the construction and analysis of future evolutionary systems.

- *Evolution of closed reaction networks:* Using the cellular model of the MCS.bl, further evolutionary experiments were conducted. To drive the evolution of the closed reaction networks, a novel and simple cellular division criterion was introduced. A series of evolutionary experiments was performed in which self-replication reactions were disabled. The target objective of the reaction networks was to promote/amplify a designated molecular species/signal. The modification of the cellular division criterion affected the fitness landscape which became multi-dimensional. The reaction networks' fitness was determined by their ability to both maintain closure and to amplify the production of target species. In these experiments, we observed the evolution of reaction networks in which the ability to grow the target molecular species was enhanced whilst maintaining closure. These results indicated the feasibility of evolving closed reaction network to perform pre-specified information processing tasks. To our knowledge, this work is the first attempt to evolve closed reaction networks capable of distinct information processing functions to be reported in the literature.

- *Crosstalk and the evolution of complexity:* The role of crosstalk in closed reaction networks was investigated. Inspired by the symbiogenesis theory and preliminary experiments conducted by Fontana and Buss, three series of evolutionary experiments were conducted to explore the effects of crosstalk on the evolutionary growth of complexity in closed reaction networks. In these evolutionary experiments, the cellular division criterion was modified to account for the pre-specified functions of both seed reaction networks. These experiments

thus aimed at evolving multitasking reaction networks. In the first two series of experiments, only cell-level mutations occurred.

We first examined the dynamics of non-crosstalking closed reaction networks when mixed in the same reaction space. Although significant differences between the MCS.bl and Alchemy were identified, a common phenomenon was observed: Non-crosstalking closed reaction networks cannot cooperate and would quasi-deterministically displace each other.

In the second experiment, crosstalking closed reaction networks were examined. First observed was a phenomenon which also occurred in the analogous Alchemy experiments. A meta-reaction network emerged which contained both seed closed-reaction networks. This cellular species was able to self-maintain for a sustained period of time. However a second phenomenon occurred (which was not reported in any Alchemy-based studies), in which a selective displacement took place. A mutant cell emerged and displaced the meta-reaction network species. This mutant cellular species was no longer maintaining both seed reaction networks but was in fact fitter at performing the pre-specified tasks. The emergence of this mutant cell was due to variations introduced by cell-level mutations.

Finally, the third experiment extended the previous one by introducing mutations at the molecular level. A common selective displacement reported in the previous experiment was observed. However additional cellular species displacements were also observed and presented evolutionary dynamics which have not been fully understood. We discussed some of the possible pitfalls of our analysis and outlined potential explanations which may be addressed in future work.

These experiments demonstrated that crosstalk was necessary in enabling cooperation to occur between distinct closed reaction networks. The evolution-

ary process was also able to optimise the reaction networks and their crosstalk properties to carry out the objective multitask function. These evolved reaction networks presented a higher level of functional and structural complexity. These experiments suggest the constructive role of crosstalk to contribute to the evolutionary growth of complexity in CIPNs.

- *The open-ended evolutionary growth of complexity:* Our evolutionary experiments demonstrated the feasibility to evolve organisationally closed CIPNs to achieve pre-specified information processing tasks. Through this evolutionary process, we observed a relative growth of complexity in CIPNs. More particularly, the crosstalk based experiments suggested an interesting avenue of research in investigating the evolutionary growth of complexity in biochemical networks.

  Nevertheless, when compared with the evolutionary dynamics reported in other ACs such as Tierra or Avida, a significant difference exists in terms of the evolutionary growth of complexity. Where Tierra exhibited numerous complex emerging phenomena (Section 3.3.3), our MCS.bl based experiments hardly presented any comparable evolutionary dynamics at both the molecular and population/system levels.

  As discussed earlier, a potential reason for this limitation is the specification of the broadcast language which may not provide a robust method to support evolvability. Future work may illuminate on these system properties which are critical to the realisation of an open-ended evolutionary system.

In the following section, we outline and discuss future research directions that may further develop some of the above contributions.

8.2 Future work

To extend the work reported in this thesis, two axes of research have been identified and are developed as follows:

1. **Building a better analytical and theoretical framework**: The experimental work conducted in this thesis presented unexpected and complex evolutionary dynamics that have not been fully understood (e.g., Section 5.5 and 7.3.5). These observations highlight the lack of appropriate analytical and theoretical tools which limits the study of Artificial Chemistries and, more generally, Complex Adaptive Systems (CAS). Several directions are suggested to address these limitations.

   (a) *A formal Artificial Life:* Future work would benefit from further theoretical research such as the development of a formal method for the study of ACs. Employing a common/unified formal framework may enable one to compare differing ACs or to map a given AC into another one. Exposing the common properties or differences may suggest potential research directions for examining and understanding the effects of specific AC features (e.g., molecular folding, space, etc.) upon the system's evolutionary dynamics. Such a formal approach could be addressed by extending the AC formalism to account for further molecular and environmental properties.

   (b) *Developing analytical frameworks:* Developing and extending analytically tractable models such as McCaskill et al.'s may provide a valuable tool for evaluating ACs. As discussed in Section 6.3, major differences existed between the MCS.bl and McCaskill et al.'s model. Extending the latter with a richer repertoire of molecular species and reactions may offer a critical tool for predicting evolutionary dynamics in ACs. Although

this extension may constitute an intricate and tedious enterprise, such a mathematical approach remains a key tool to examine complex chemical systems. Chemical organisation theory is another promising research avenue which was initiated by Fontana and Buss (1994a) and later enriched by Dittrich and Speroni (2007). This algebraic approach proposes a theoretical foundation to describe closed chemical organisations and their dynamics. Organisation theory could thus be naturally employed to examine the MCS.bl's evolutionary dynamics. Such "conventional" mathematical techniques may assist in the analysis and understanding of complex chemical organisations using ACs.

2. **Examining the conditions for the evolutionary growth of complexity in CIPNs:** We propose a number of system modifications that could lead to the emergence and evolution of CIPNs of higher complexity. This proposed research direction aims at understanding the conditions for the evolutionary growth of complexity in CIPNs.

   (a) *Cellular division criteria:* In Chapter 7, two simple cellular division criteria were devised in which the objective was to promote the production of specific molecular species. Further cellular division criteria could be designed to investigate the emergence of more complex information processing functions. For example, the conditions to trigger the cellular division could be dynamic. In this proposal, the cellular division criterion may vary according to the states of several molecular species. Similarly to experiments conducted in Chapter 7, a cell divides if n_{target} molecules of species s_T are produced. However this condition is here modulated by the presence of an additional species s_{sw} acting as a switch operator. An additional target species s_U is identified. When s_{sw} is present in a given cell, the latter has to generate n_{target} s_U molecules to trigger the cellu-

lar division. If s_{sw} is not present, the cellular division criterion remains the production of n_{target} molecules s_T. The insertion and removal of s_{sw} species are carried out manually over time. The evolutionary process may encourage the emergence of cellular species which are able to rapidly process this switching condition and promote the growth of appropriate target species. A cellular division probability can also be introduced to specify further constraints, e.g., to penalise cellular species which simultaneously promote the growth of both target species s_T and s_U regardless of s_{sw} being present or not.

(b) *Detectors and effectors:* Following Holland's LCS/agent-based approach (Section 2.2.8), introducing a set of detectors and effectors is proposed to encourage the emergence of a chemotactic behaviour. In this extended MCS.bl model, cells are situated in a two-dimensional space in which detectors may probe the cell's surrounding environment for chemicals. Detectors and effectors are implemented as broadcast devices that, similarly to molecular species, may be subjected to evolution. The environment is populated with gradients of food molecules (again specified as broadcast devices) that are necessary for the cells to grow and divide; this growth condition is addressed by the cellular division criterion. Upon detecting the required food species, detectors generate signalling molecular species within the cell. In contrast to detectors, effectors do not produce further molecular species upon binding to signalling species. In this chemotactic model, effectors may activate "flagella" which affect the cell's movement in space. The flagella's actions vary according to the nature of the effectors's action statement (a coding scheme is devised to specify this function). Such an extended MCS.bl model may potentially give rise to the emergence of regulatory/control feedback which is distinctive of the bacterial chemotaxis signalling pathway. In this approach,

a clear input/output signal demarcation is introduced by the detectors and effectors.

Additional more "realistic" properties such as mass conservation, molecular folding, a genetic subsystem or advanced chemical kinetics could be introduced. These complementary properties would certainly broaden the complexity of an already difficult investigation. However there would be no guarantee of improved results, i.e., exhibiting a more interesting evolutionary growth of complexity. A first reason for this assertion is that the impact of environmental constraints on the evolution of complexity still remains to date an open question (Gershenson and Lenaerts, 2008).

Moreover, developing a unified theoretical framework may simply not be feasible using mathematical methods that are currently available. As a result we believe that further empirical investigations need to be performed to assemble a set of key observations. By integrating these observations we may be able to formulate further theories with regards to the evolution of complexity in CAS. Nevertheless this development will only be feasible if the employed models are not burdened with *unnecessary* complex features which may distract and prevent the thorough analysis of CAS.

Therefore we suggest that minimalist approaches to CAS, where the system is still analytically tractable and examined using available mathematical methods, should be adopted. In keeping with this final suggestion, our thesis contributed, to some extent, to the understanding of the evolutionary growth of complexity of CAS using ACs.

8.3 Summary

In this concluding chapter, we summarised and discussed the main research contributions of this thesis with regards to autocatalytic closure and the evolution of CIPNs. Following this, we enumerated a number of future directions that may di-

rectly extend the work presented in this thesis. A series of system modifications was proposed to encourage the evolutionary growth of complexity in the MCS.bl. Additional proposed research investigations advocated the development of further analytical and theoretical frameworks to study Complex Adaptive Systems. However we also mentioned that the development of such a theoretical foundation may not be feasible given the mathematical methods currently available and the complexity of studied systems. We consequently and finally argued that supplementary empirical investigations should be conducted using minimalist and analytically tractable approaches to examine CAS. Such minimalist CAS approaches, as the one presented in this thesis, may facilitate the comparison and analysis of differing systems and lead to the formulation of critical theories in the field of complex systems.

Appendix A

List of Publications

A number of publications resulted from the work carried out throughout the course of this thesis and are listed below:

- **Decraene J.**, Mitchell G.G., McMullin B., *Crosstalk and the Cooperation of Collectively Autocatalytic Reaction Networks*, To appear in Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2009), Trondheim, Norway.

- **Decraene J.**, *Closure in Artificial Cell Signalling Networks: Investigating the Origins of Cognition in Collectively Autocatalytic Reaction Networks*, In Proceedings of the Second International Conference on Bio-inspired Systems and Signal Processing (BioSignals 2009), Porto, Portugal.

- **Decraene J.**, Mitchell G.G., McMullin B., *Unexpected Evolutionary Dynamics in a String-based Artificial Chemistry* . In S. Bullock, J. Noble, R. A. Watson, and M. A. Bedau (Eds.) In Proceedings of the Eleventh International Conference on Artificial Life (Alife 2008), pp.158-165, MIT Press, Cambridge, MA. Winchester, UK.

- **Decraene J.**, Mitchell G.G., McMullin B., *Modeling and Evolving Biochemical Networks: Insights into Communication and Computation from the Biolog-*

ical Domain, In Proceedings of the China-Ireland International Conference on Information and Communications Technologies (CIICT 2008), Beijing, People's Republic of China.

- **Decraene J.**, Mitchell G.G., McMullin B., *Exploring Evolutionary Stability in a Concurrent Artificial Chemistry.* In Proceedings of the European Conference on Complex Systems (ECCS 2008), Jerusalem, Israel.

- **Decraene J.**, Mitchell G.G., McMullin B., *Evolving Artificial Cell Signaling Networks: Perspectives and Methods.* In F. Dressler, I. Carreras, editors, Advances in Biologically Inspired Information Systems: Models, Methods, and Tools. Series Studies in Computational Intelligence, Vol. 69, pp 165-184, Springer Verlag, 2007.

- **Decraene J.**, Mitchell G.G., McMullin B., *Studying Complex Adaptive Systems using Molecular Classifier Systems.* In Proceedings of the European Conference on Complex Systems (ECCS 2007), Dresden, Germany.

- **Decraene J.**, Mitchell G.G., McMullin B., *A Molecular Approach to Complex Adaptive Systems.* In Proceedings of the IEEE (UK&RI chapter) Conference on Cybernetic Systems (SMC 2007), Dublin, Ireland.

- **Decraene J.**, *On the Emergence and Evolution of Artificial Cell Signalling Networks*, In Proceeding of the European Graduate Student Workshop on Evolutionary Computation (EvoPhD 2007), Valencia, Spain.

- **Decraene J.**, Mitchell G.G., McMullin B., *The Holland Broadcast Language and the Modeling of Biochemical Networks.* In Proceedings of the 10th European Conference on Genetic Programming (EuroGP 2007), Lecture Notes in Computer Science, Vol. 4445, Springer, Valencia, Spain.

- **Decraene J.**, Mitchell G.G., McMullin B., *Evolving Artificial Cell Signaling Networks using Molecular Classifier Systems.* In Proceedings of the 1st international conference on Bio inspired models of network, information and computing systems (Bionetics 2006), ACM Press, Cavalese, Italy.

- **Decraene J.**, Dittrich P., Hinze T., Lenser T., McMullin B., Mitchell G.G., *A Multidisciplinary Survey of Modeling Techniques for Biochemical Networks.* In Proceedings of the Integrative Post-Genomics Conference (IPG 2006), Lyon, France.

- **Decraene J.**, Mitchell G.G., McMullin B., *Evolving Artificial Cell Signaling Networks.* In Proceedings of the European Conference on Complex Systems (ECCS 2006), Oxford, UK.

- **Decraene J.**, Mitchell G.G., Kelly C., McMullin B., *An Approach to Evolving Cell Signalling Networks in silico.* In Proceedings of the International Workshop on Systems Biology (IWSC 2006), Maynooth, Ireland.

- **Decraene J.**, *The Holland Broadcast Language*, Technical report ALL-06-01, ALife Lab, RINCE, School of Electronic Engineering, Dublin City University, 2006.

Appendix B

Simplifying the broadcast language

The original broadcast language, as devised by Holland (1975, 1992a), differs from the MCS.bl by including the following and additional system properties. The broadcast language alphabet Λ contains ten *symbols*, Λ^* is the set of strings over Λ. The symbols constitute the atomic elements of the language.

$$\Lambda = \{0,\ 1,\ *,\ :,\ \Diamond,\ \triangledown,\ \blacktriangledown,\ \triangle,\ p,\ '\}$$

Broadcast units

Four types of broadcast unit can be distinguished, any other broadcast units that do not follow one of those four schemes are *null units*. An arbitrary string from Λ^* which contains neither unquoted $*$ nor unquoted : is denoted by I_n, with $n = \{1, 2, 3\}$. Broadcast units may engage in the following interactions based on discrete timesteps:

1. $*I_1 : I_2$

   If a signal of type I_1 is detected at time t then the signal I_2 is broadcast at time $t + 1$.

2. $* : I_1 : I_2$

   If there is no signal of type I_1 present at time t then the signal I_2 is broadcast at time $t + 1$.

3. $*I_1 :: I_2$

   If a signal of type I_1 is detected at time t then a *persistent* string (preceded by the p symbol) of type I_2 (if any) is removed from the environment at the end of time t.

4. $*I_1 : I_2 : I_3$

   If a signal of type I_1 and a signal of type I_2 are both present at time t then the signal I_3 is broadcast *at same time* t unless the string I_3 contains unquoted symbols $\{\triangledown, \blacktriangledown, \triangle\}$ or singly quoted occurrence of $*$, in which case the string I_3 is broadcast at time $t + 1$.

For broadcast units of type 1 and 2, the string I_2 refers to the output signal. Whereas I_1 is said to be a broadcast unit argument, and this applies to any types of broadcast unit. Nevertheless, we also have additional broadcast unit arguments I_2 for broadcast units of type 3 and 4. Finally, in the case of type 4 broadcast unit, I_3 corresponds to the output signal.

When a broadcast unit of type 2 is fired at time t, this implies the deletion of a persistent signal. Persistent signals include signals starting with an unquoted occurrence of p but also active broadcast devices.

Also when an output signal is interpreted for broadcast, one quote is removed from each quoted symbol. This allows one to use the quote symbol to "protect" special symbols to be passed into the output signal. A broadcast unit may broadcast only once at each time step.

The symbols

The interpretation of the symbols $\{\blacktriangledown,\ p\}$ is now presented.

$\blacktriangledown$ This symbol is similar to $\triangledown$ but can concatenate different input signals.

For example, with $S(t) = \{*10\triangledown : 11\blacktriangledown : 000\triangledown\blacktriangledown, 10111, 1100\}$ we obtain at $t+1$: $S(t+1) = \{*10\triangledown : 11\blacktriangledown : 000\triangledown\blacktriangledown, 10111, 1100, 00011100\}$. In this case $\triangledown$

designates the suffix 111 occurring in the input signal 10111 and ▼ designates the suffix 00 found in the detected signal 1100. The format of the broadcast signal is 000▽▼, therefore we replace and concatenate ▽ and ▼ accordingly and we obtain the output signal 00011100.

p When this symbol occurs at the first position of a string, it designates a *persistent* string which persists over time until deleted even if the string is not an active broadcast unit. A null device occurring at time t which is not persistent exists only for a single timestep and is removed at the end of time t.

The modifications

It was demonstrated that the broadcast language can model Genetic Regulatory Networks (Decraene et al., 2007b). This was due to the ability of the broadcast language to mirror Boolean networks which illustrates the wide ranging processing power that broadcast systems are capable of. Nevertheless, it was also highlighted that the broadcast language is limited regarding the representation and simulation of biochemical networks. To address this issue, we propose to combine the Molcular Classifier System concept with the broadcast language in a new system termed MCS.bl. The MCS.bl complements the broadcast language and extends it by including the following refinements:

- Instead of processing all broadcast devices sequentially and deterministically during a time step, the MCS.bl processes as follows: At each time step t, we pick a pair of broadcast devices at random. For each pair of devices, one of the broadcast devices is designated (at random) as the *enzyme device* and the second one as the *substrate device*. If the conditional statement of the enzyme device is satisfied by the informational string of the substrate device, then the action statement of the enzyme is executed upon the substrate.

- In the broadcast language specification given by Holland, additional rules were required to resolve some ambiguities raised by the interpretation of broadcast devices. To facilitate this, the MCS.bl simplifies the interpretation of broadcast units by preserving broadcast units of type 1 only (i.e., we remove broadcast units of type 2, 3 and 4).

- The "black reversed triangle" ▼ is removed. This symbol was used to concatenate matched strings occurring in type 4 broadcast units. As type 4 broadcast units are removed, this symbol may no longer function.

- Similarly the notion of non-persistent devices is removed: By default all broadcast devices are considered as persistent molecules. Therefore the p symbol is removed.

- As type 3 broadcast units and non-persistent devices no longer exist in this proposal, no molecule can be deleted from the population. However the deletion/dilution of molecules is needed to obtain a selective mechanism at the molecular level. Our suggestion is as follows, each time a successful reaction occurs, we pick a molecule at random and delete it from the population.

Appendix C

Static Reactors with Molecular Diffusion - Results

We describe three series of experiment (where 5 simulations were run in each series) in which the static reactors model with molecular diffusion is employed (Section 6.3). The following diffusion coefficients are employed: $m_1 = 0.01$, $m_2 = 0.05$ and $m_3 = 0.1$ in respective series.

The following set of fixed parameters is utilised in all simulations:

- 30 compartments are utilised and executed in parallel using 30 AMD Opteron 270 (2.0 GHZ) CPUs.
- Experiments are run for 3600 seconds.
- The maximal compartment carrying capacity is $n_{max} = 1000$.
- The diffusion probability is set to $p_m = 0.05$.
- The spontaneous decay probability is set to $d = 0.1$.
- Similarly to experiences conducted in Chapter 5, the maximal species string length is set to $BD_{Lmax} = 500$.
- The per-symbol mutation is set to $p_{sym} = 1.0 \times 10^{-5}$.
- The global spontaneous mutation rate is set to $r_{mut} = 0$.

- As in the evolutionary experiments presented in Section 5.5, each compartment is seeded/initialised with $s_{R_4} = \nabla 0101 : \nabla 0101$ molecular species. However the initial amount differs, we seed/fill the compartments with 1000 instances of s_{R_4} species.

We present an overview of the dynamics of all simulation runs where the diffusion coefficient is set to $m_1 = 0.01$, $m_2 = 0.05$ and $m_3 = 0.1$. The following chain of events were observed in all runs:

1. We note an initial phase where the system is stable with an average species length of 12 symbols (i.e., the length of s_{R_4} species) and the average population size stagnating at nearly 1000 molecules. Most compartments are thus full during this phase (i.e., the molecular production rate is higher than the decay rate).

2. However at some stage, we note that the average length of species starts to increase rapidly. This behaviour suggests that the elongation phenomenon is occurring.

3. Shortly after, we observe a rapid decrease in the average population size throughout the 30 compartments. This indicates that the production rate has now become smaller than the decay rate, i.e., the compartments are depleting.

4. Nevertheless this decrease does not apply to the species length which continues to increase for a period of time, where the species reach a peak in string length. During this period, successive species having an increasing length emerged and displaced each others. As a result we observe a linear increase in the average species length throughout all compartments.

5. Following the observed peak in the species length, the average species length is then rapidly decreasing (similarly to the average population size) until the system becomes extinct, i.e., all molecules have decayed.

Figures C.2 and C.3 complement the current analysis by providing more detailed information about the dynamics of each compartment when $m_1 = 0.01$:

1. During the early stable phase, we note that the molecular string length is relatively homogeneous throughout the 30 compartments with little variance occurring. This assertion applies for both the average population size and species string length.

2. Following this, we observe a divergence in the composition of the compartments (i.e., compartments with different population sizes exist). This variance is maintained until close to the end of the extinction phase.

3. Although the composition of compartments start diverging at some point in time, the average species length (which is globally increasing) is more or less homogeneous throughout the compartments for a period of time. Thus during this phase, the mutant species (here classified by their string length) are well diffused throughout the compartments. However as the average molecular population size is already decreasing, it indicates that these mutant species have a production rate lower than the decay rate.

4. Finally, a significant range of variances is observed in the species string length. This phenomenon suggests that only few reactions leading to the creation of much longer species are succeeding in some compartments.

Figures C.5 and C.6 detail the extinction phases of the simulation runs with $m_2 = 0.05$ whereas the extinction phases of the runs with $m_3 = 0.1$ are depicted in Figures C.8 and C.9. In these figures, we distinguish that the range of variances in the average molecular population size actually decreases in constrast to the increase trend reported earlier in the first experiment where $m_1 = 0.01$. Little variance of average molecular string length is observed over the whole simulation runs excepting during a few seconds where the systems collapse. These results suggest that

higher diffusion coefficients allowed for the diffusion equilibrium to be achieved more rapidly. Consequently we identify a homogeneous molecular composition throughout all compartments over time.

C.1 5 example simulations with diffusion coefficient $m_1 = 0.01$

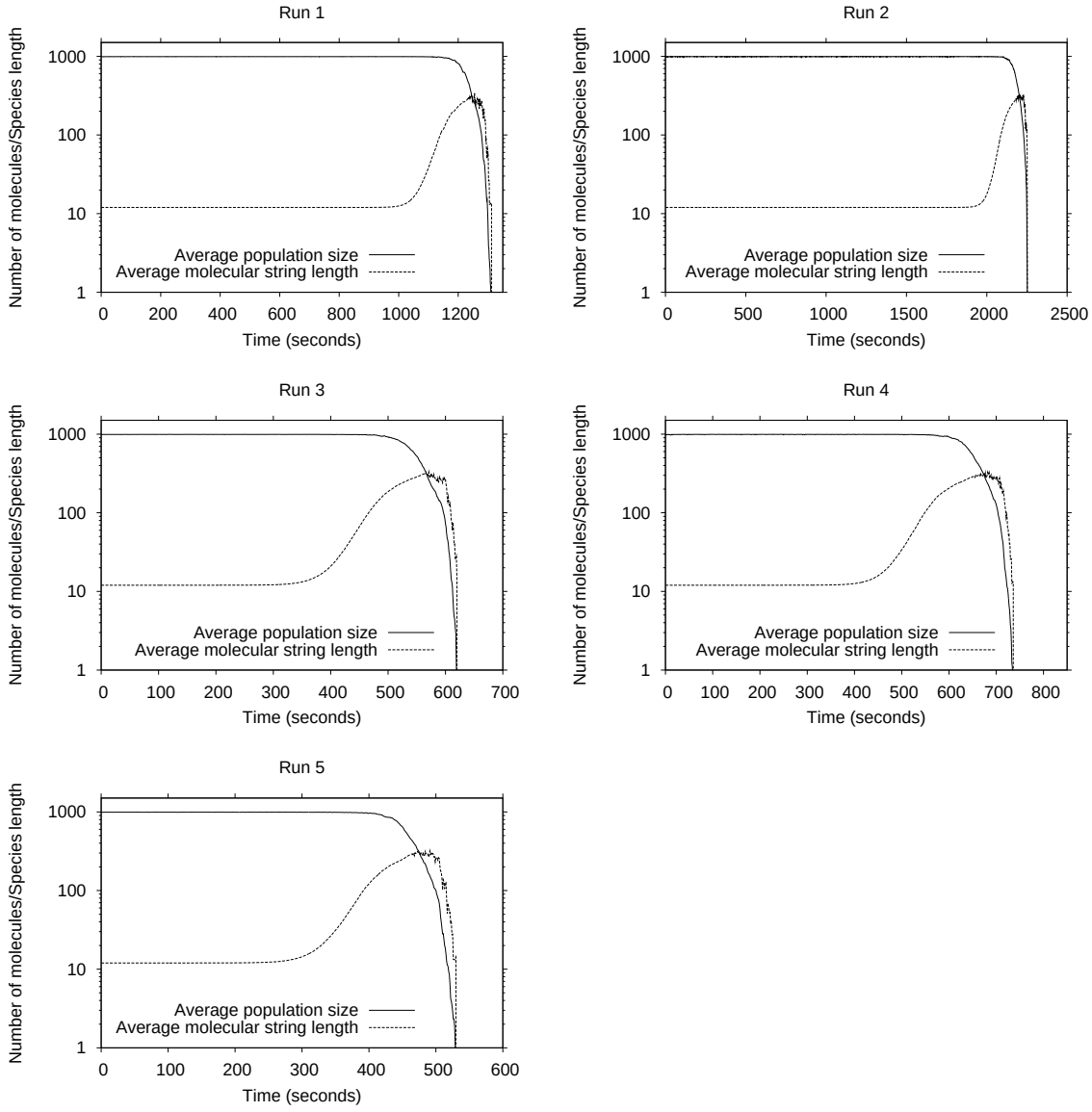

Figure C.1: 5 example simulations with $m_1 = 0.01$.

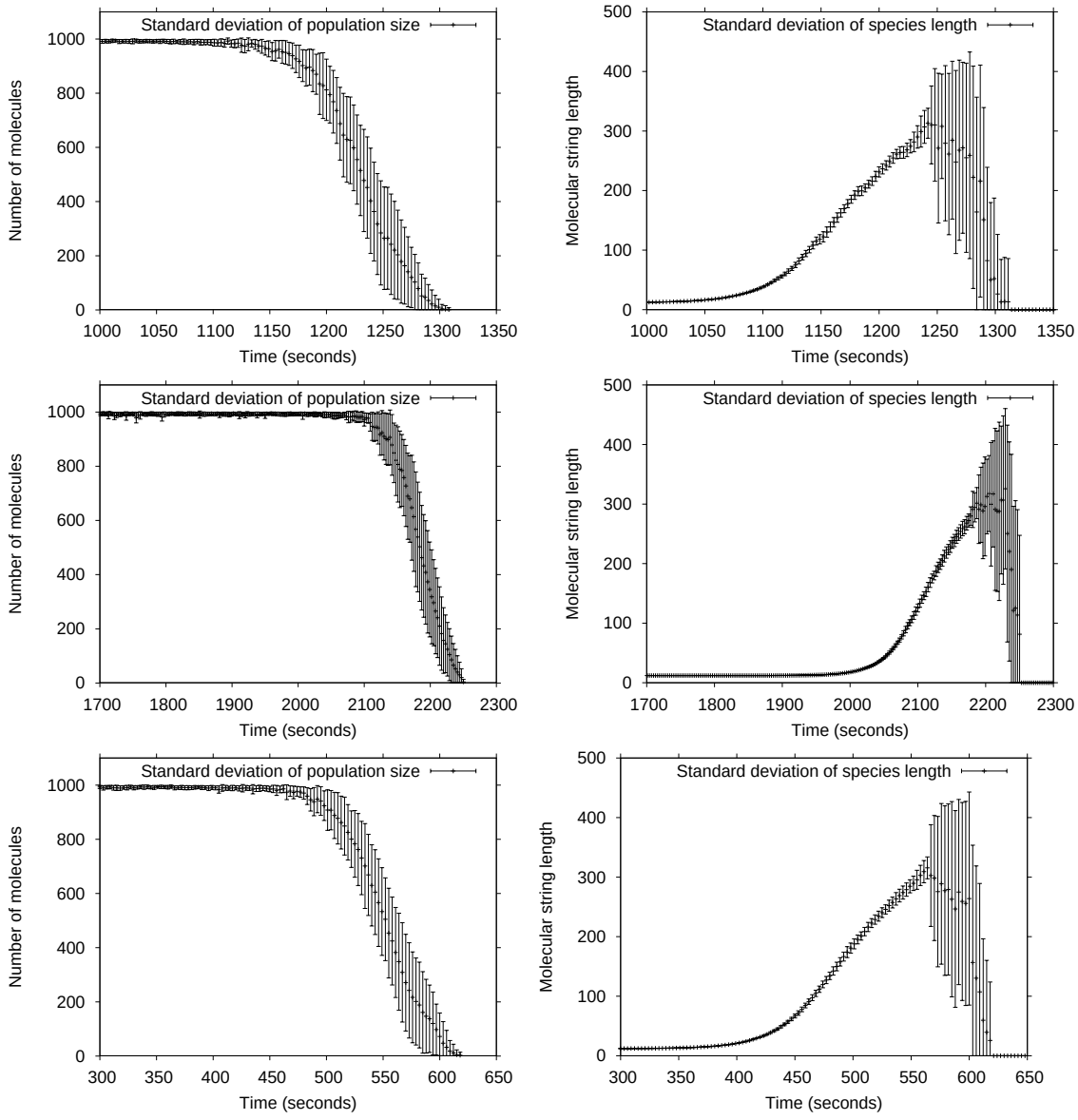

Figure C.2: Extinction phases of simulation runs 1, 2 and 3 - from top to bottom, with diffusion coefficient $m_1 = 0.01$.

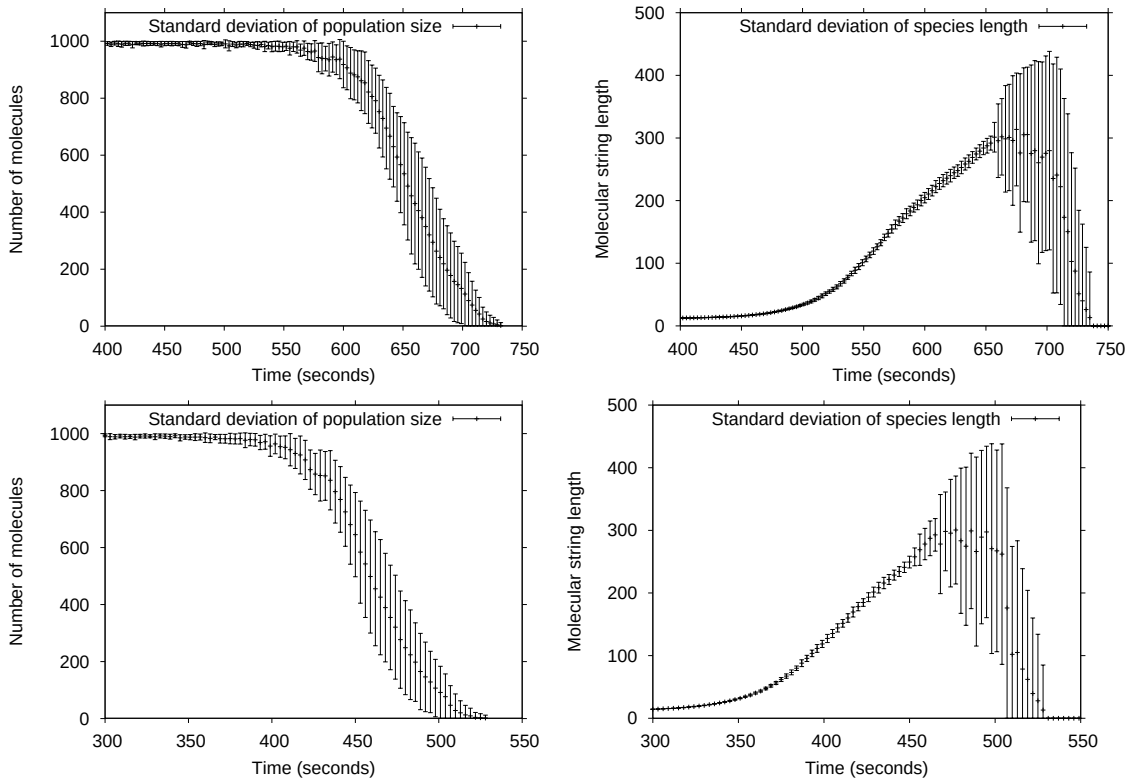

Figure C.3: Extinction phases of simulation runs 4 and 5 with diffusion coefficient $m_1 = 0.01$.

C.2 5 example simulations with diffusion coefficient $m_2 = 0.05$

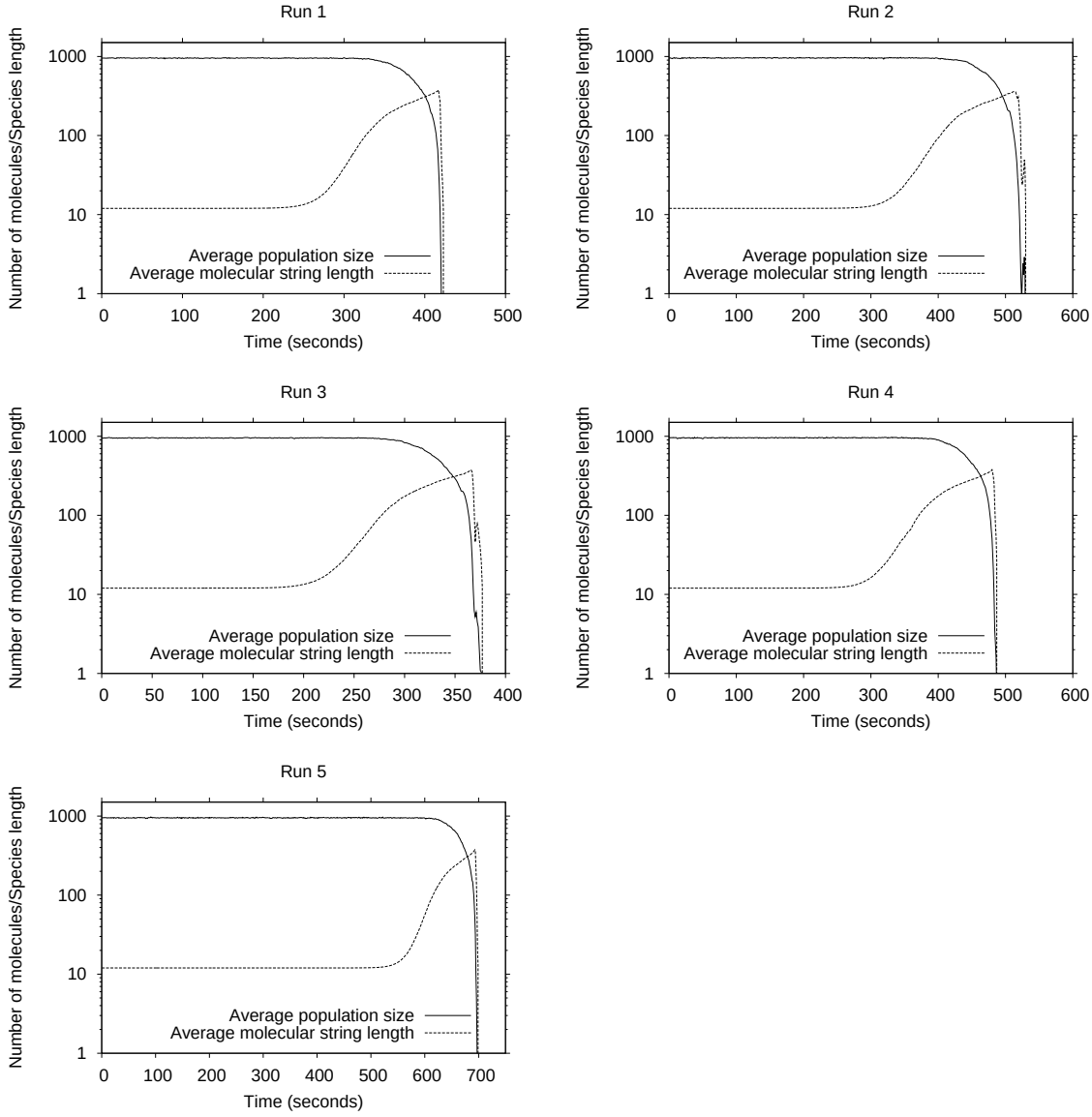

Figure C.4: 5 example simulations with $m_2 = 0.01$.

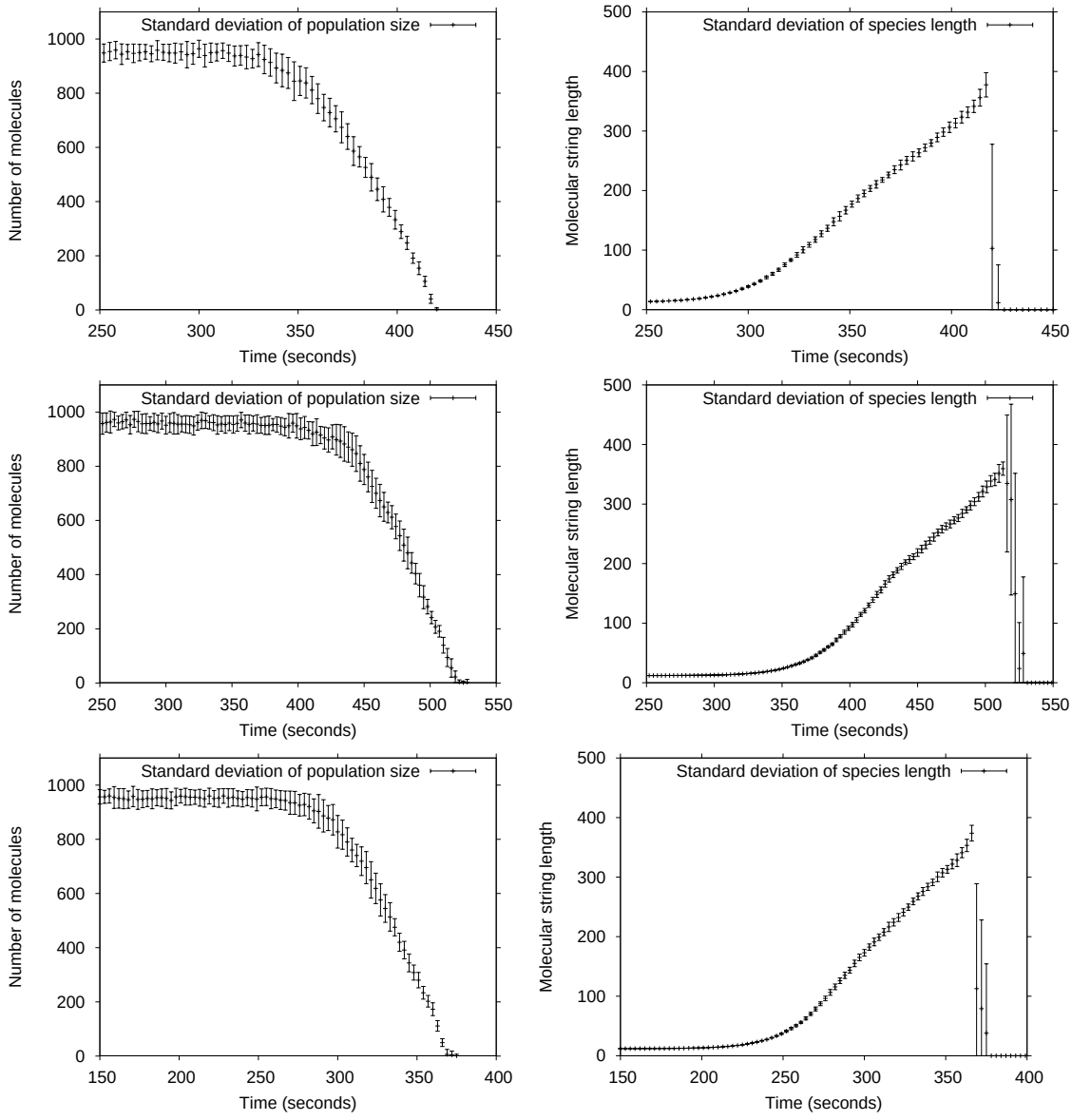

Figure C.5: Extinction phases of simulation runs 1,2 and 3 with diffusion coefficient $m_2 = 0.05$.

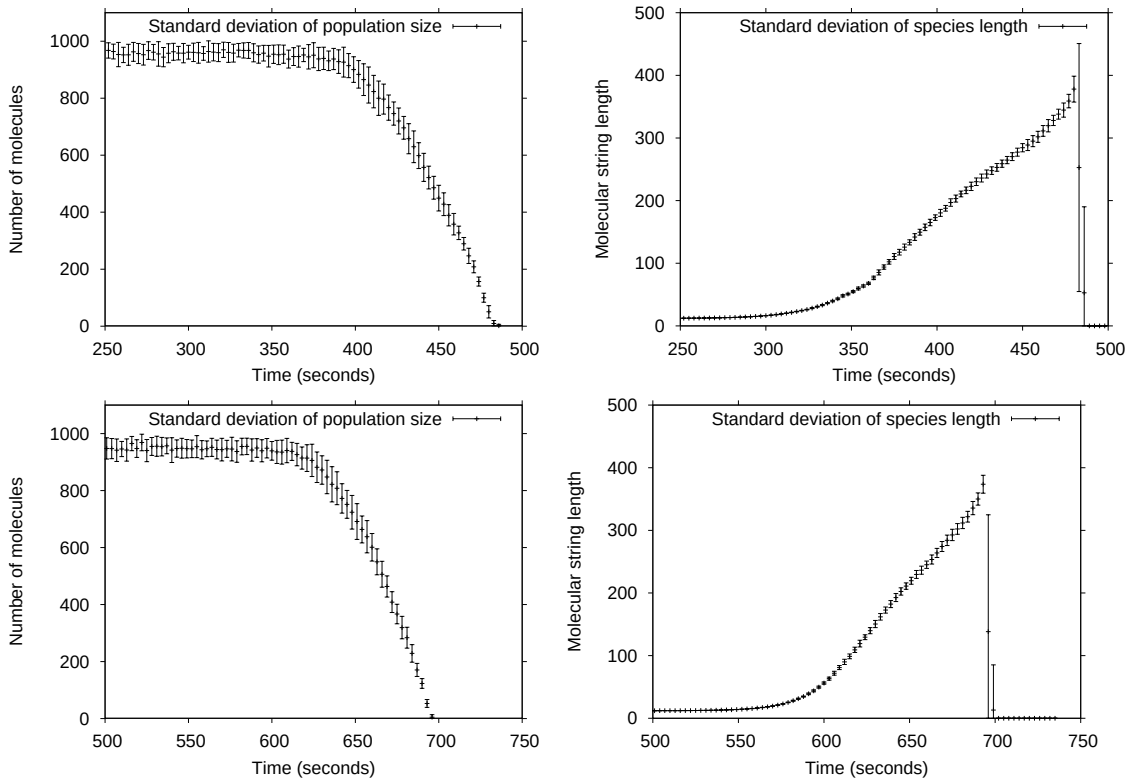

Figure C.6: Extinction phases of simulation runs 4 and 5 with diffusion coefficient $m_2 = 0.05$.

C.3 5 example simulations with diffusion coefficient $m_3 = 0.1$

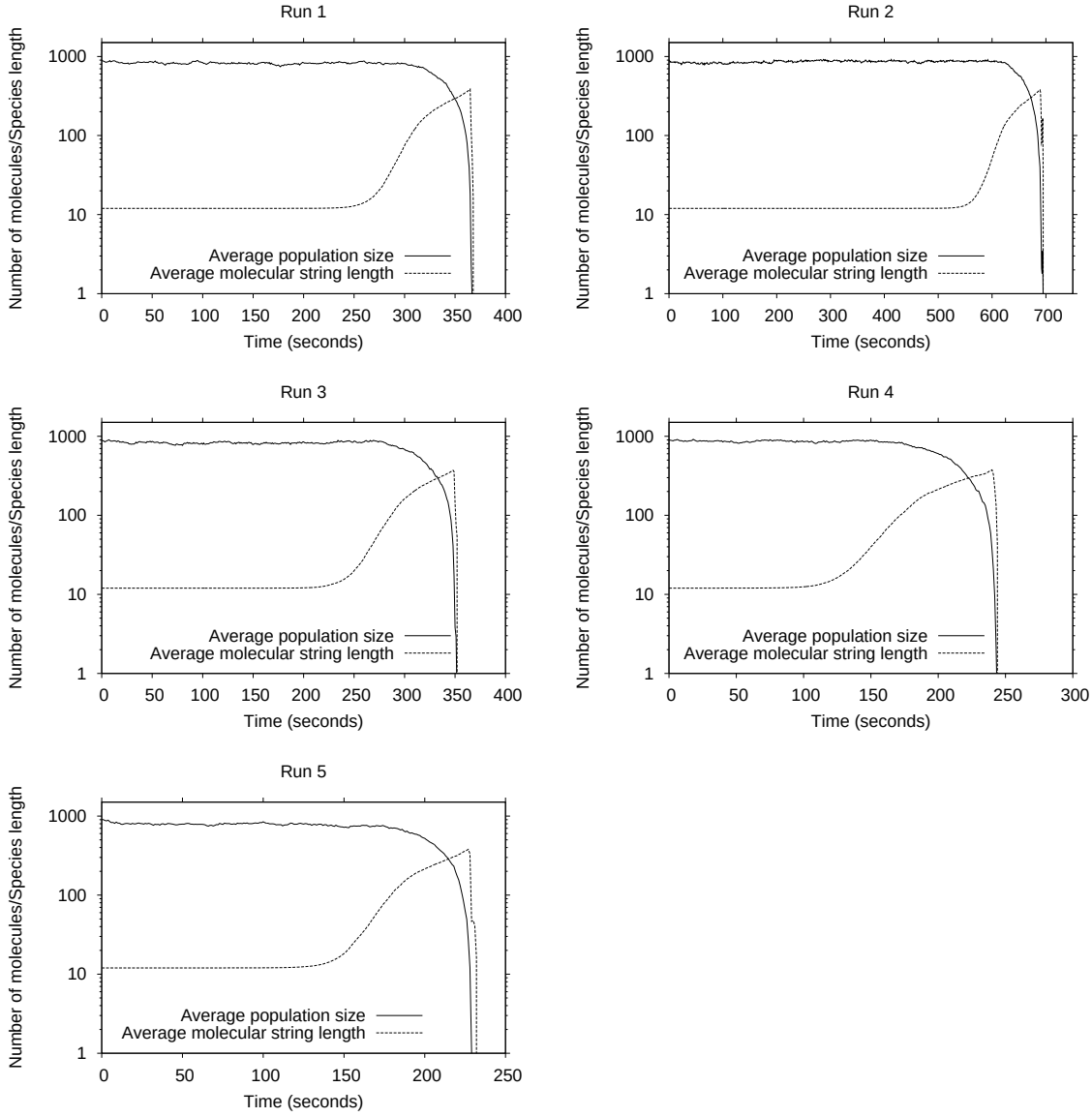

Figure C.7: 5 example simulations with $m_3 = 0.1$.

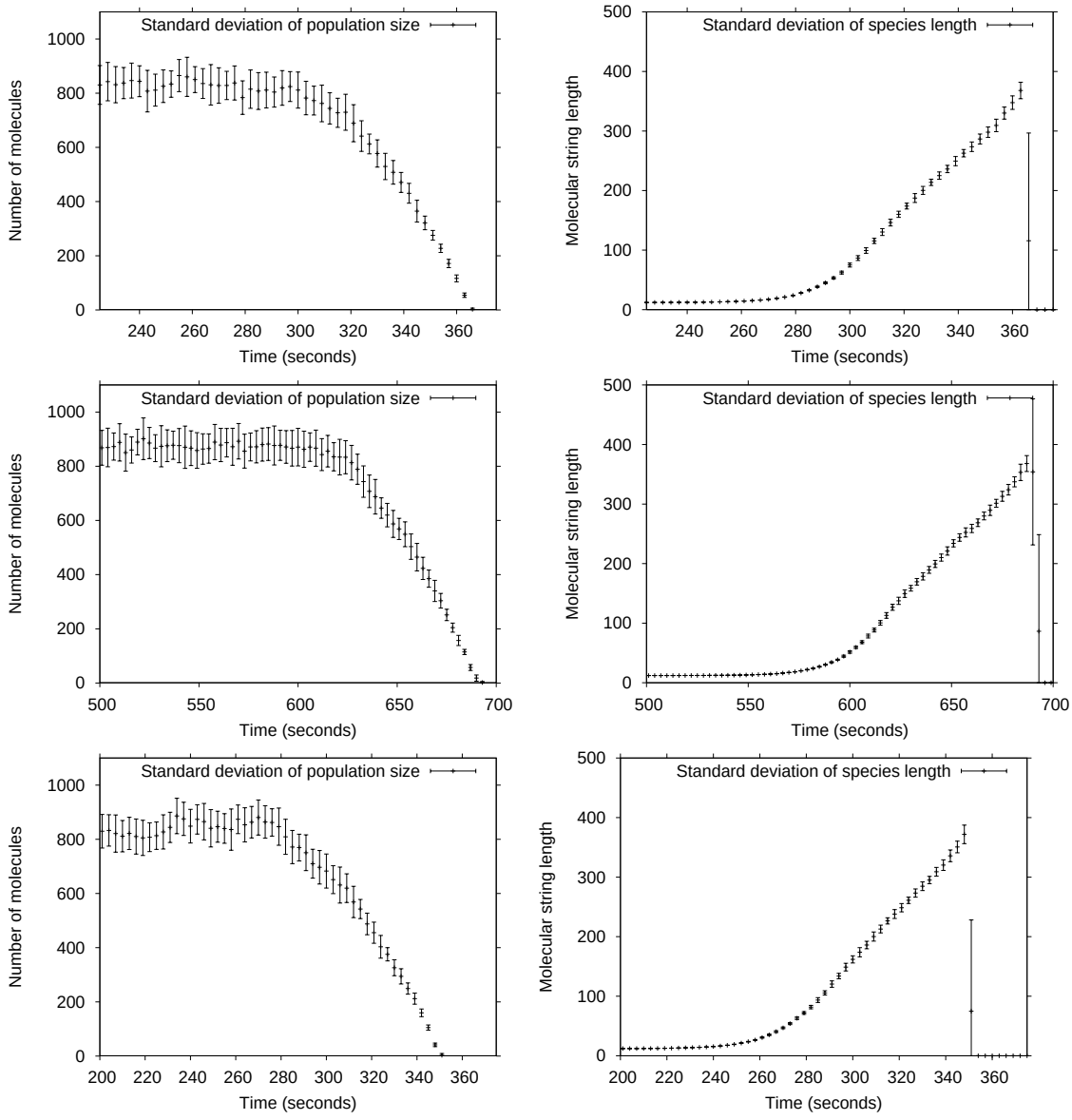

Figure C.8: Extinction phases of simulation runs 1,2 and 3 with diffusion coefficient $m_3 = 0.1$.

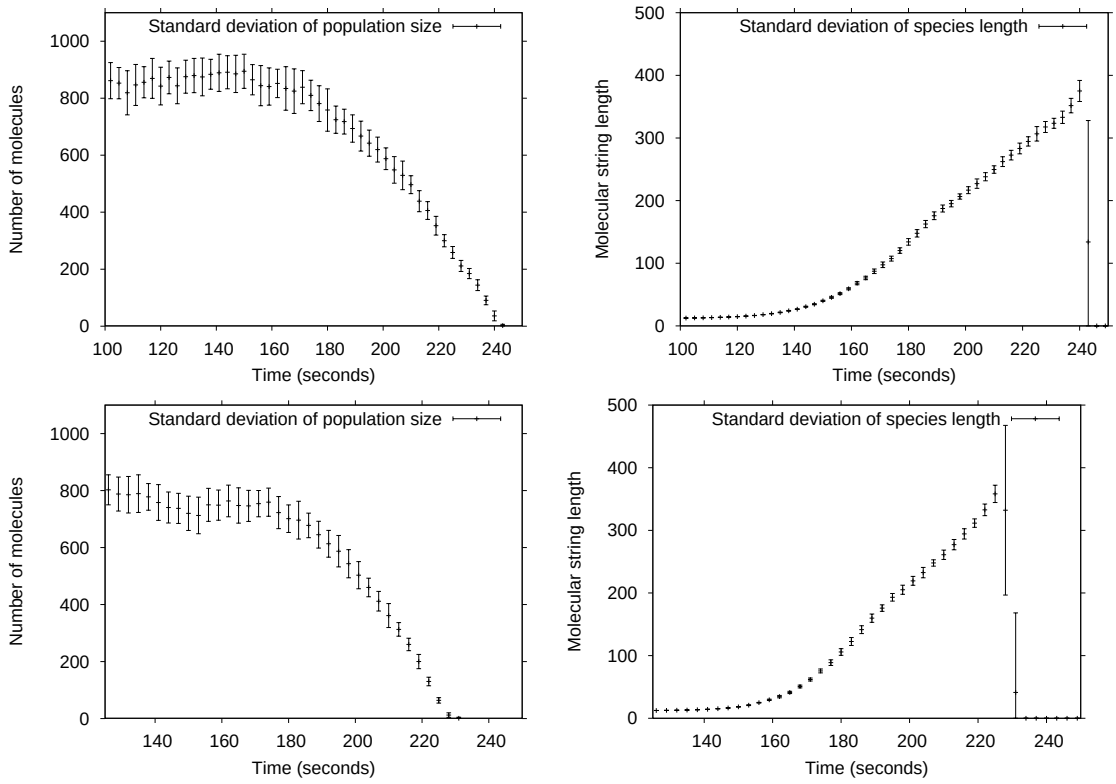

Figure C.9: Extinction phases of simulation runs 4 and 5 with diffusion coefficient $m_3 = 0.1$.

Appendix D

Evolving Closed Reaction Networks - Results

A series of experiments are reported in which closed reaction networks are evolved to promote the growth of a target species s_t.

D.1 Two example simulation runs with self-replication reactions enabled

Two evolutionary simulations are conducted and utilise the following set of parameters:

- 30 cells are utilised and executed in parallel using 31 AMD Opteron 270 (2.0 GHZ) CPUs.
- self-replications reactions are allowed.
- Simulations are run for 3600 seconds.
- The maximal compartment carrying capacity is $n_{max} = \infty$.
- The target molecular species division threshold is set to $n_{target} = 200$.
- The maximal species string length is set to $BD_{Lmax} = 500$.
- The global spontaneous mutation rate is set to $r_{mut} = 0$.

- The per-symbol mutation probability is set to $p_{sym} = 0.00005$.
- Each compartment is seeded/initialised with 10 instances of the species s_1.
- The target molecular species is $s_t = s_1$.

All reactors are seeded with a common closed reaction network which contains the species s_1 only (Fig.D.1).

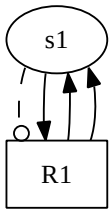

Figure D.1: Bipartite reaction graph of seed closed reaction network containing the self-replicase species s_1 only.

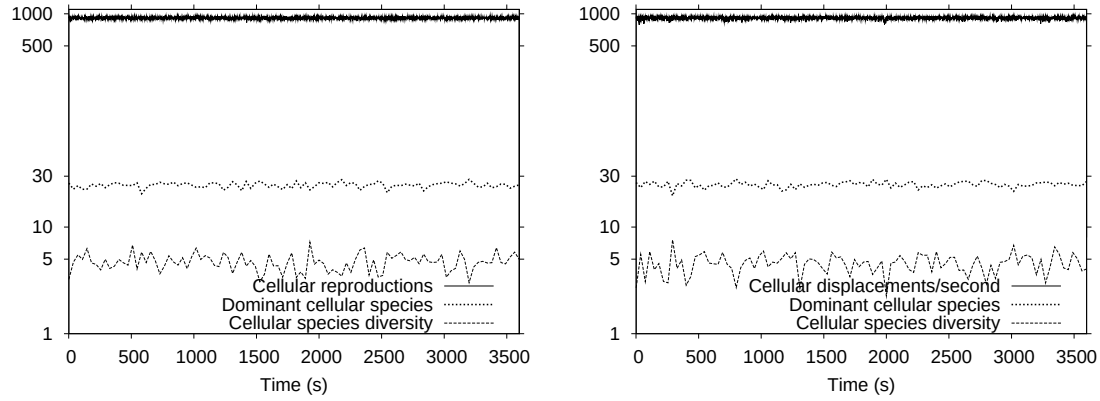

Figure D.2: Example simulation runs 1 (left) and 2 (right)

In Fig. D.2, it is first observed that no displacement occurred in both simulations. A level of cellular species diversity is observed averaging at ≈ 5 throughout the simulation runs. The seed closed reaction network remained the dominant cellular species throughout the simulations.

D.2 Constructing a minimalist collectively autocatalytic reaction network

In this section we present the construction of a minimalist collectively autocatalytic reaction network. We aim at realising the simplest collectively reaction network in terms of complexity at both the molecular (i.e., using the shortest molecular species) and network level (i.e., involving the least number of species and interactions between the latter). In this reaction network, no self-replication reaction may occur.

We showed that the simplest autocatalytic species that can be built in the MCS.bl is $s_{R_0} = *\nabla : \nabla$. Indeed the length of s_{R_0} is equal to BD_{Lmin} and ∇ is the only single enzymatic operator that can both match and output multiple characters (as required to produce an enzymatic molecule having at least 4 symbols).

As self-replication reactions are now disabled then the minimalist closed reaction network would thus contain at least two distinct molecular species. Both of these molecular species would mutually contribute to the maintenance of each other. We first attempt to build such a closed network composed of two distinct molecular species denoted by s_1 and s_2.

In this first attempt, s_1=s_{R_0}. The second molecular species s_2 is required to be as well a replicase species being able to replicate s_1. Therefore, in this minimalist approach, s_2 is to be equivalent to s_1 from a enzymatic function point of view with $s_2 = *\nabla : \nabla$ as a *starting* point. Nevertheless as we cannot decrease the length of s_2, an additional symbol is to be inserted to differ s_2 from s_1.

If we insert an *executable* symbol in the binding condition or action statement of s_2, excluding the structural symbols * or : and the quote symbol as this would inhibit the s_2's enzymatic function, then we would obtain an elongator or "reductor" species which by definition cannot replicate s_1.

For example:

$$*' \bigtriangledown 0 : \bigtriangledown \ + \ * \bigtriangledown : \bigtriangledown \ \rightarrow \ \emptyset \tag{D.1}$$

$$* \bigtriangledown 0 :' \bigtriangledown \ + \ * \bigtriangledown : \bigtriangledown \ \rightarrow \ \emptyset \tag{D.2}$$

$$* \diamond \bigtriangledown 0 : \bigtriangledown \ + \ * \bigtriangledown : \bigtriangledown \ \rightarrow \ \bigtriangledown : \bigtriangledown \tag{D.3}$$

$$* \bigtriangledown 0 : \diamond \bigtriangledown \ + \ * \bigtriangledown : \bigtriangledown \ \rightarrow \ * \bigtriangledown : \diamond \bigtriangledown \tag{D.4}$$

As a result we are left with the option to insert a *non-executable* symbol in s_2. In this case we may obtain a closed network containing two distinct species. However this also implies that both molecular species are equivalent from a phenotypic point of view. Let $s_2 = * * \bigtriangledown : \bigtriangledown$. We present an example simulation run in which this reaction network is employed and where the cellular division criterion is to promote the growth of $s_t = s_1$.

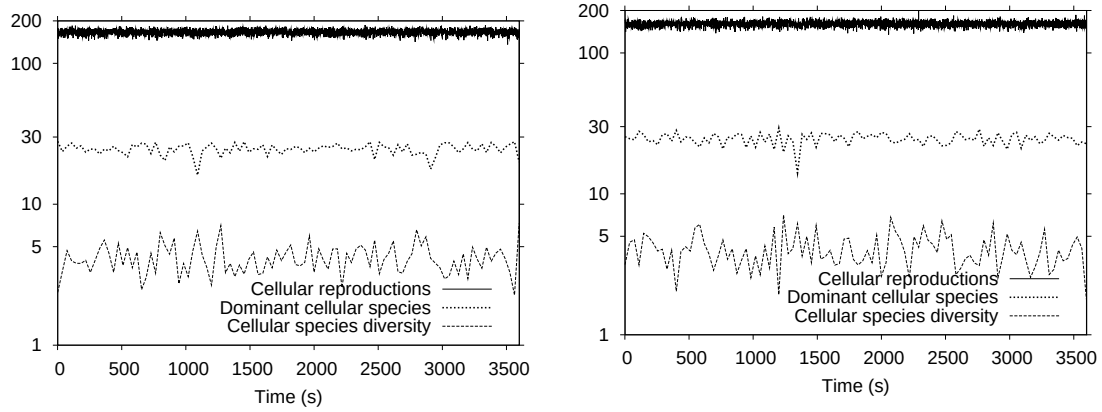

Figure D.3: Two example simulation runs where the seed reaction contained 10 instances of both $s_1 = * \bigtriangledown : \bigtriangledown$ and $s2 = * * \bigtriangledown : \bigtriangledown$ with $p_{sym} = 0.00005$. and $r_{mut} = 0$

Fig.D.3 depicts two simulation runs where the two replicase species s_1 and s_2 are employed. Although a level of cellular diversity was maintained throughout the simulation runs, no cellular species with a significant difference in fitness emerged. As a result no displacements between cellular species occurred. These results suggest that the employed seed reaction network is already the optimal catalytic network to produce s_t.

In contrast with the above proposal, we now attempt to build a minimalist closed network where the molecular species are phenotypically different from each other and no universal replicases are employed.

If we were to employ 5 executable symbols to specifiy the molecular species, we would, again, be confronted with that asymmetric relationship between the binding and action statements leading to the elongation/trimming enzymatic functions. We thus propose to relax the minimal molecular string length to 6. The following two molecular species are proposed:

$$\begin{aligned} s_1 &= \nabla 0 : \nabla 1 \\ s_2 &= \nabla 1 : \nabla 0 \end{aligned}$$

However reactions between the molecular species s_1 and s_2 lead to the production of further molecular species as follows:

$$\begin{aligned} s_1 + s_2 &\rightarrow s_3 \\ s_2 + s_1 &\rightarrow s_4 \end{aligned}$$

with $s_3 = \nabla 1 : \nabla$ and $s_4 = \nabla 0 : \nabla 0$.

We enumarate all possible reactions (with $\alpha = 4^2$) that may occur between s_1, s_2, s_3 and s_4:

$$
\begin{aligned}
s_1 + s_1 &\rightarrow \emptyset \\
s_1 + s_2 &\rightarrow s_3 \\
s_1 + s_3 &\rightarrow \emptyset \\
s_1 + s_4 &\rightarrow s_1 \\
s_2 + s_1 &\rightarrow s_4 \\
s_2 + s_2 &\rightarrow \emptyset \\
s_2 + s_3 &\rightarrow s_2 \\
s_2 + s_4 &\rightarrow \emptyset \\
s_3 + s_1 &\rightarrow s_1 \\
s_3 + s_2 &\rightarrow \emptyset \\
s_3 + s_3 &\rightarrow \emptyset \\
s_3 + s_4 &\rightarrow \emptyset \\
s_4 + s_1 &\rightarrow \emptyset \\
s_4 + s_2 &\rightarrow s_4 \\
s_4 + s_3 &\rightarrow \emptyset \\
s_4 + s_4 &\rightarrow \emptyset
\end{aligned}
$$

No further molecular species have been produced in the above reactions.

The above informal investigation suggests that the above set of molecular species $c_0 = \{s_1, s_2, s_3, s_4\}$ is potentially a minimalist collectively autocatalytic reaction network that can be constructed in the MCS.bl given the following condition:

- The reaction network is not, by design, the optimal catalytic network to realise the target task. Evolutionary experiments using this reaction network as the seed networks are presented in the next section.

- Molecular species which can perpetually generate new species (such as elongator species) are not allowed.

D.3 10 simulation runs with self-replication reactions disabled

10 evolutionary simuations are executed using c_0 and the following set of parameters:

- 31 cells are utilised and executed in parallel using 31 AMD Opteron 270 (2.0 GHZ) CPUs.
- Simulations are run for 3600 seconds.
- The maximal compartment carrying capacity is $n_{max} = \infty$.
- The target molecular species division threshold is set to $n_{target} = 200$.
- The maximal species string length is set to $BD_{Lmax} = 500$.
- The global spontaneous mutation rate is set to $r_{mut} = 0$.
- The per-symbol mutation probability is set to $p_{sym} = 1.0 \times 10^{-5}$.
- Each compartment is seeded/initialised with 10 instances of each species s_1, s_2, s_3 and s_4 (as presented in the previous section).
- The target molecular species is $s_t = s_1$.

In four of these runs, we observed the emergence and domination of reaction networks which are phenotipically equavalent to $c_1 = \{s_1, s_2, s_5, s_6\}$ with $s_5 = *\nabla\triangle : \nabla 0$ and $s_6 = *\nabla\triangle : \nabla 1$.

In four other runs, the emergence of reaction networks containing the molecular species of either c_1 or c_3 in addition to some molecular species were noted. However these additional molecular species did not improve the fitness of the cellular species. It is thus conjectured that, given enough time, these reaction networks would displace these extra molecular species and collapse to c_1 or c_3. In the remaining two runs, the emergence of c_0 mutants with no phenotypic differences was observed.

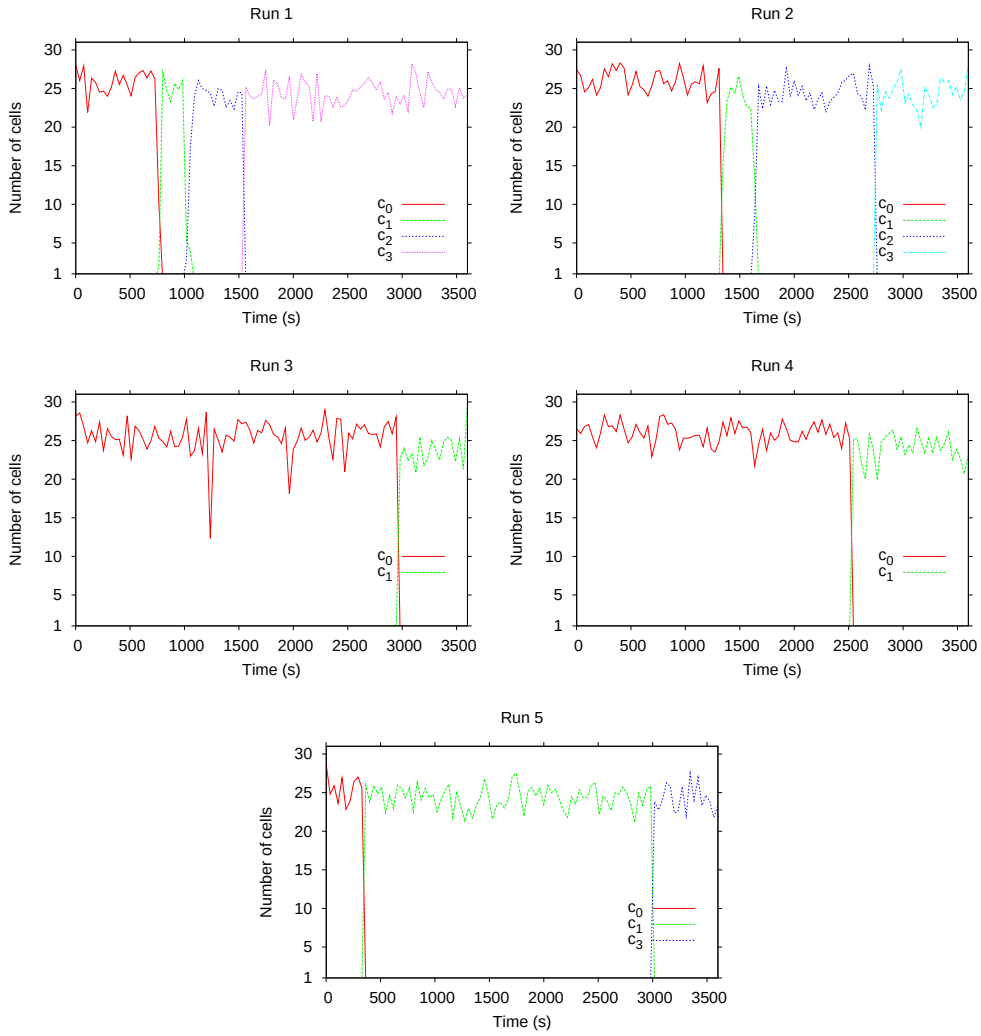

Figure D.4: Simulation runs 1 to 5 with self-replication reactions disabled

c_1	c_2	c_3
$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$
$*\nabla 1 : \nabla 1$	$*\nabla 1 : \nabla 1$	$*\nabla 1 : \nabla 1$
$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$
$*\nabla 1 : \nabla 0$	$*\nabla 1 : \nabla 0$	$*\nabla 1 : \nabla 0$
$1 * \nabla 0 : \nabla 1$	$* * \nabla 0 : \nabla 0$	$*0 : \nabla 0$
		$*0 : \nabla 1$

Table D.1: Molecular species contained in successive dominant closed reaction networks in simulation run 1.

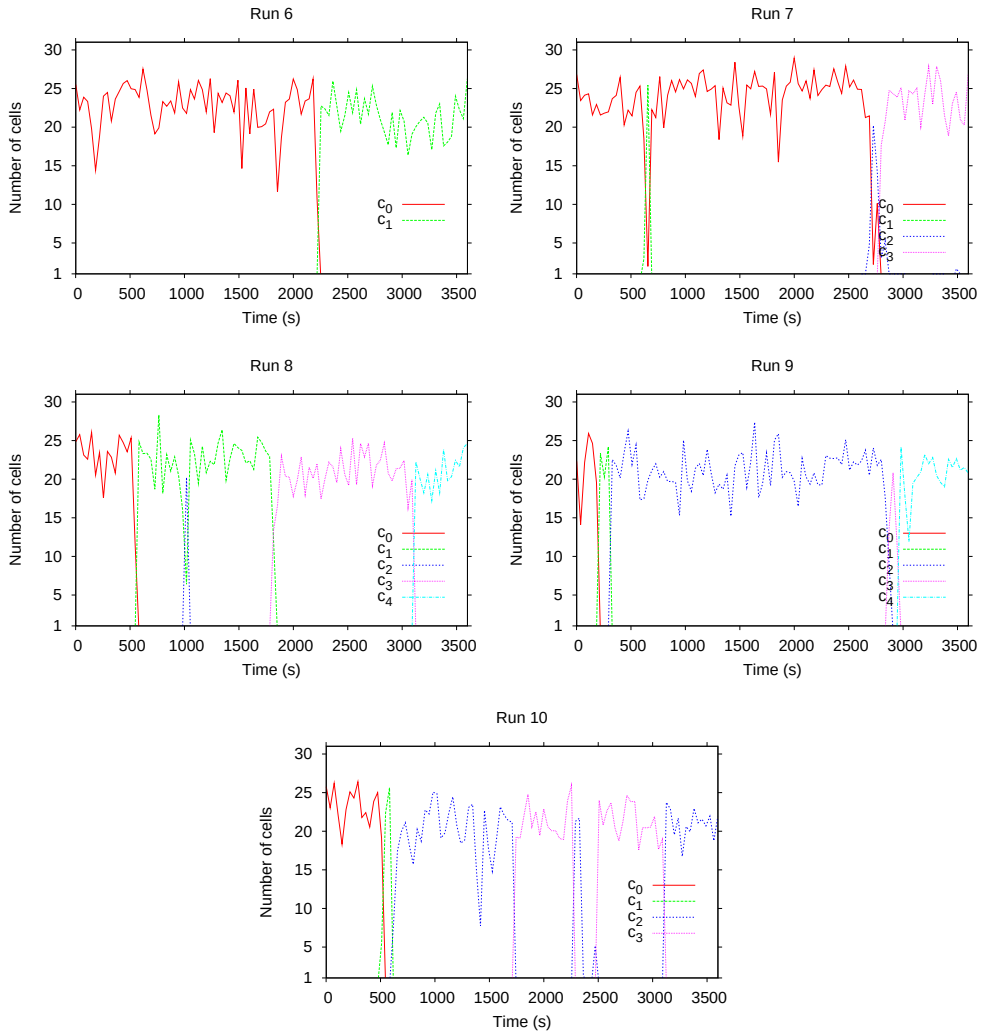

Figure D.5: Simulation runs 6 to 10 with self-replication reactions disabled

c_1	c_2	c_3
$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$
$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$
$*\nabla 1 : \nabla 0$	$*\nabla 1 : \nabla 0$	$*\nabla \triangle : \nabla 1$
$*\nabla 1 : \nabla 1$	$*\nabla 1 : \nabla 1$	$*\nabla \triangle : \nabla 0$
$*\nabla 1 \nabla 1$	$*\nabla 0 : \nabla 00$	$*\nabla \nabla 0 : \nabla 0$
		$*\nabla \nabla 0 : \nabla 1$

Table D.2: Molecular species contained in successive dominant closed reaction networks in simulation run 2.

c_1
$*\nabla 0 : \nabla 1$
$*\nabla 0 : \nabla 0$
$*\nabla\triangle : \nabla 1$
$*\nabla\triangle : \nabla 0$
$*0 : \nabla 1$
$*0 : \nabla 0$

Table D.3: Molecular species contained in successive dominant closed reaction networks in simulation run 3.

c_1
$*\nabla 0 : \nabla 1$
$*\nabla 0 : \nabla 0$
$*\nabla\triangle : \nabla 1$
$*\nabla\triangle : \nabla 0$

Table D.4: Molecular species contained in successive dominant closed reaction networks in simulation run 4.

c_1	c_2
$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$
$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$
$*\nabla 1 : \nabla 1$	1
$*\nabla 1 : \nabla 0$	0
$**\nabla 0 : \nabla 1$	$*\nabla\triangle : \nabla 0$
	$*\nabla\triangle : \nabla 1$
	$*\nabla\triangle : 1$

Table D.5: Molecular species contained in successive dominant closed reaction networks in simulation run 5.

c_1
$*\nabla 0 : \nabla 1$
$*\nabla 0 : \nabla 0$
$*\nabla\triangle : \nabla 1$
$*\nabla\triangle : \nabla 0$

Table D.6: Molecular species contained in successive dominant closed reaction networks in simulation run 6.

c_1	c_2	c_3
$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$
$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$
$*\nabla 1 : \nabla 1$	$*\nabla\triangle : \nabla 0$	$*10 : \nabla 1$
$*\nabla 1 : \nabla 0$	$*\nabla\triangle : \nabla 1$	$*10 : \nabla 0$
$*\nabla 01 : \nabla 1$	$*\nabla 0 : \nabla\triangle 0$	$*\nabla\triangle : \nabla 1$
$*\nabla 01 : \nabla 0$	$*\nabla 0 : \nabla\triangle 1$	$*\nabla\triangle : \nabla 1$

Table D.7: Molecular species contained in successive dominant closed reaction networks in simulation run 7.

c_1	c_2	c_3	c_4
$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$
$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$
$*\nabla 1 : \nabla\triangle 1$	$*\nabla\triangle : \nabla 1$	$0 * \nabla\triangle : \nabla 1$	$0 * \nabla\triangle\nabla : \nabla 1$
$*\nabla 1 : \nabla\triangle 0$	$*\nabla\triangle : \nabla 0$	$0 * \nabla\triangle : \nabla 0$	$0 * \nabla\triangle\nabla : \nabla 0$

Table D.8: Molecular species contained in successive dominant closed reaction networks in simulation run 8.

c_1	c_2	c_3
$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$	$*\nabla 0 : \nabla 1$
$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$
$: *\nabla 1 : \nabla 0$	$*\nabla\triangle : \nabla 0$	$*\nabla\triangle\triangle : \nabla 1$
$: *\nabla 1 : \nabla 1$	$: *\nabla\triangle : \nabla 1$	$*\nabla\triangle\triangle : \nabla 0$

Table D.9: Molecular species contained in successive dominant closed reaction networks in simulation run 10.

Bibliography

Adami, C. and Brown, C. (1994). Evolutionary Learning in the 2D Artificial Life System Avida. In *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, pages 377–381. MIT Press.

Alon, U. (2007). *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC.

Ausk, B., Gross, T., and Srinivasan, S. (2006). An Agent-based Model for Real-Time Signaling induced in Osteocytic Networks by Mechanical Stimuli. *Journal of biomechanics*, 39(14):2638–2646.

Barricelli, N. (1957). Symbiogenetic Evolution Processes Realized by Artificial Methods. *Methodos*, 9(35-36):143–182.

Barricelli, N. (1963). Numerical Testing of Evolution Theories. Part 11. Preliminary Tests of Performance. Symbiogenesis and Terrestrial Life. *Acta Biotheoretica*, 16(3/4).

Bedau, M., McCaskill, J., Packard, N., Rasmussen, S., Adami, C., Green, D., Ikegami, T., Kaneko, K., and Ray, T. (2000). Open Problems in Artificial Life. *Artificial Life*, 6(4):363–376.

Bernstein, D. (2005). Simulating Mesoscopic Reaction-Diffusion Systems using the Gillespie Algorithm. *Physical Review E*, 71(4):41103.

Bersini, H. (1999). Design Patterns for an Object-Oriented Computational Chemistry–In proceedings of the 5 th European Conference on Artificial Life (ECAL99)–Eds: Floreano. *Springer*, pages 389–398.

Bersini, H. (2000). Reaction Mechanisms in the OO Chemistry. In *Artificial Life VII: Proceedings of the Seventh International Conference on Artificial Life*, page 39. MIT Press.

Bersini, H. (2002). And the Winner is...Not the Fittest. In *Proceedings of the 2002 IEEE Congress on Evolutionary Computation, 2002. CEC '02.*, pages 1510–1515.

Bersini, H. (2008). *La Programmation Orientée Objet. Cours et Exercices en UML 2 avec Java 5, C++, Python, PHP 5 et LINQ (4 Éd.).* Eyrolles.

Bersini, H. and Carneiro, J., editors (2006). *Artificial Immune Systems, 5th International Conference, ICARIS 2006, Oeiras, Portugal, September 4-6, 2006, Proceedings*, volume 4163 of *Lecture Notes in Computer Science*. Springer.

Bersini, H. and Varela, F. (1990). Hints for Adaptive Problem Solving Gleaned from Immune Networks. *Parallel problem solving from nature*, pages 343–355.

Binder, B. and Heinrich, R. (2004). Interrelations between Dynamical Properties and Structural Characteristics of Signal Transduction Networks. *Genome Inform Ser Workshop Genome Inform*, 15(1):13–23.

Blossey, R., Cardelli, L., and Phillips, A. (2008). Compositionality, Stochasticity, and Cooperativity in Dynamic Models of Gene Regulation. *HFSP Journal*, 2(1):17–28.

Bray, D. (1995). Protein Molecules as Computational Elements in Living Cells. *Nature*, 376(6538):307–312.

Bray, D. and Lay, S. (1994). Computer Simulated Evolution of a Network of Cell-Signaling Molecules. *Biophysical Journal*, 66(4):972–977.

Bray, T., Paoli, J., Sperberg-McQueen, C., Maler, E., and Yergeau, F. (2000). Extensible Markup Language (XML) 1.0. *W3C recommendation*, 6.

Bull, L. and Kovacs, T. (2005). *Foundations of Learning Classifier Systems*. Springer.

Chatterjee, A., Vlachos, D. G., and Katsoulakis, M. A. (2005). Binomial Distribution based Tau-Leap Accelerated Stochastic Simulation. *Journal of Chemical Physics*, 122(2).

Chavali, A., Gianchandani, E., Tung, K., Lawrence, M., Peirce, S., and Papin, J. (2008). Characterizing Emergent Properties of Immunological Systems with Multi-cellular Rule-based Computational Modeling. *Trends in Immunology*, 29(12):589–599.

Chickering, D. M. (1996). Learning Bayesian Networks is NP-complete. In Fisher, D. and Lenz, H.-J., editors, *Learning from data: AI and statistics V*, chapter 12, pages 121–130. Springer.

Church, A. (1932). A Set of Postulates for the Foundation of Logic. *The Annals of Mathematics*, 33(2):346–366.

Cramer, N. (1985). A Representation for the Adaptive Generation of Simple Sequential Programs. In *Proceedings of the 1st International Conference on Genetic Algorithms*, pages 183–187. L. Erlbaum Associates Inc. Hillsdale, NJ, USA.

Cronhjort, M. and Blomberg, C. (1997). Cluster Compartmentalization may provide Resistance to Parasites for Catalytic Networks. *Physica D: Nonlinear Phenomena*, 101(3-4):289–298.

de Jong, H. (2002). Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *Journal of Computational Biology*, 9(1):67–103.

Deckard, A. and Sauro, H. (2004). Preliminary Studies on the in silico Evolution of Biochemical Networks. *Chembiochem*, 5(10):1423–1431.

Decraene, J. (2006). The Holland Broadcast Language. Technical Report ALL-06-01, RINCE, Dublin City University. http://elm.eeng.dcu.ie/~alife/jd/ALL-06-01/.

Decraene, J. (2009). Closure in Artificial Cell Signalling Networks - Investigating the Emergence of Cognition in Collectively Autocatalytic Reaction Networks. In Encarnação, P. and Veloso, A., editors, *BIOSIGNALS*, pages 107–114. INSTICC Press.

Decraene, J., Dittrich, P., Hinze, T., Lenser, T., McMullin, B., and Mitchell, G. (2006). A Multidisciplinary Survey of Modeling Techniques for Biochemical Networks. *Proceedings of the Integrative Post-Genomics (IPG06), Lyon, France.*

Decraene, J., Mitchell, G., and McMullin, B. (2008a). Modeling and Evolving Biochemical Networks: Insights into Communication and Computation from the Biological Domain. *IET Conference Publications*, 2008(CP544):235–239.

Decraene, J., Mitchell, G., and Mcmullin, B. (2008b). Unexpected Evolutionary Dynamics in a String-based Artificial Chemistry. In Bullock, S., Noble, J., Watson, R., and Bedau, M. A., editors, *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*, pages 158–165. MIT Press, Cambridge, MA.

Decraene, J., Mitchell, G. G., and McMullin, B. (2007a). Evolving Artificial Cell Signaling Networks: Perspectives and Methods. In Dressler, F. and Carreras, I., editors, *Advances in Biologically Inspired Information Systems*, volume 69 of *Studies in Computational Intelligence*, pages 165–184. Springer.

Decraene, J., Mitchell, G. G., and McMullin, B. (2009). Crosstalk and the Coop-

eration of Collectively Autocatalytic Reaction Networks. In *IEEE Congress on Evolutionary Computation*. To appear.

Decraene, J., Mitchell, G. G., McMullin, B., and Kelly, C. (2007b). The Holland Broadcast Language and the Modeling of Biochemical Networks. In Ebner, M., O'Neill, M., Ekárt, A., Vanneschi, L., and Esparcia-Alcázar, A., editors, *EuroGP*, volume 4445 of *Lecture Notes in Computer Science*, pages 361–370. Springer.

Dittrich, P. (2004). Chemical Computing. In Banâtre, J.-P., Fradet, P., Giavitto, J.-L., and Michel, O., editors, *Unconventional Programming Paradigms*, volume 3566 of *Lecture Notes in Computer Science*, pages 19–32. Springer.

Dittrich, P. and Speroni, P. (2007). Chemical Organisation Theory. *Bulletin of Mathematical Biology*, 69(4):1199–1231.

Dittrich, P., Ziegler, J., and Banzhaf, W. (2001). Artificial Chemistries - A Review. *Artificial Life*, 7(3):225–275.

Eigen, M. (1971). Self-organization of Matter and the Evolution of Biological Macromolecules. *Naturwissenschaften*, 58(10):465–523.

Eigen, M. and Schuster, P. (1977). A Principle of Natural Self-Organization. *Naturwissenschaften*, 64(11):541–565.

Emonet, T., Macal, C., North, M., Wickersham, C., and Cluzel, P. (2005). AgentCell: a Digital Single-cell Assay for Bacterial Chemotaxis. *Bioinformatics*, 21(11):2714–2721.

ESIGNET (2006a). Deliverable number 3.1 Cell Signalling Networks: State of the Art Report. Eu technical report, Dublin City University. http://esignet.net/dokumente/WP3-D31/.

ESIGNET (2006b). Deliverable number 4.1 Specification of the Properties of Cell Signalling Systems Document Describing Formats for Phenotypic Representation.

Eu technical report, University of Jena. http://esignet.net/dokumente/WP4-D41/.

Eungdamrong, N. J. and Iyengar, R. (2004). Modeling Cell Signaling Networks. *Biol Cell*, 96(5):355–362.

Farmer, J., Kauffman, S., and Packard, N. (1986a). Autocatalytic Replication of Polymers. *Physica D*, 2(1-3):50–67.

Farmer, J., Packard, N., and Perelson, A. (1986b). The Immune System, Adaptation, and Machine Learning. *Physica*, 22(2):187–204.

Fisher, M. J., Paton, R. C., and Matsuno, K. (1999). Intracellular Signalling Proteins as Smart' Agents in Parallel Distributed Processes. *Biosystems*, 50(3):159–171.

Fogel, D. (2007). *Introduction to Evolutionary Computation*. Wiley-IEEE Press.

Fontana, W. and Buss, L. (1994a). The Arrival of the Fittest: Toward a Theory of Biological Organization. *Bulletin of Mathematical Biology*, 56(1):1–64.

Fontana, W. and Buss, L. (1994b). What Would be Conserved if "the Tape were Played Twice"? *Proceedings of the National Academy of Sciences*, 91(2):757–761.

Fontana, W. and Buss, L. (1996). The Barrier of Objects: From Dynamical Systems to Bounded Organizations. In Casti, J. and A., K., editors, *Boundaries and Barriers*, pages 56–116. Redwood City, MA. Addison-Wesley.

Fritz, J. (1982). *Partial Differential Equations*. Applied Mathematical Sciences. Springer-Verlag, New York, NY, 4th edition.

Gabora, L. (2008). The Cultural Evolution of Socially Situated Cognition. *Cognitive Systems Research*, 9(1-2):104–114.

Genoud, T. and Metraux, J.-P. (1999). Crosstalk in Plant Cell Signaling: Structure and Function of the Genetic Network. *Trends in plant science*, 4(9):503–507.

Gershenson, C. and Lenaerts, T. (2008). Evolution of Complexity. *Artificial Life*, 14(3):241–243.

Gesteland, R., Atkins, J., and Storchova, H. (2005). *The RNA World*. Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY.

Gibson, M. and Bruck, J. (2000). Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *J. Phys. Chem. A*, 104:1876–1889.

Gilbert, W. (1986). The RNA World. *Nature*, 319(6055):618.

Gillespie, D. (1976). A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comput. Phys*, 22(4):403–434.

Gillespie, D. et al. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions. *The journal of physical chemistry*, 81(25):2340–2361.

Gillespie, D. T. (2001). Approximate Accelerated Stochastic Simulation of Chemically Reacting Systems. *Journal of Chemical Physics*, 115(4):1716–1733.

Gomez, S., Lo, S., and Rzhetsky, A. (2001). Probabilistic Prediction of Unknown Metabolic and Signal-Transduction Networks. *Genetics*, 159(3):1291–1298.

Gonzalez, P., Cardenas, M., Camacho, D., Franyuti, A., Rosas, O., and Lagunez-Otero, J. (2003). Cellulat: an Agent-based Intracellular Signalling Model. *BioSystems*, 68(2-3):171–185.

Goto, H., Kiyono, T., Tomono, Y., Kawajiri, A., Urano, T., Furukawa, K., Nigg, E., and Inagaki, M. (2005). Complex Formation of Plk1 and INCENP required for MetaphaseAnaphase Transition. *Nature Cell Biology*, 8(2):180–187.

Gould, S. (2002). *The Structure of Evolutionary Theory*. Belknap Press.

Goutsias, J. (2006). A Hidden Markov Model for Transcriptional Regulation in Single Cells. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 3(1).

Groß D. and McMullin, B. (2002). The Creation of Novelty in Artificial Chemistries. In *In Proceedings of The 8th International Conference on the Simulation and Synthesis of Living Systems*, pages 400–409. MIT Press.

Gwak, C. and Wee, K. (2007). Construction of Hypercycles in Typogenetics with Evolutionary Algorithms. *Lecture Notes in Computer Science*, 4648:1060.

Heinrich, R. and Schuster, S. (1996). *The Regulation of Cellular Systems*. Chapman & Hall.

Helmreich, E. J. M. (2001). *The Biochemistry of Cell Signalling*. Oxford University Press, USA.

Hoare, C. A. R. (1983). Communicating Sequential Processes. *Commun. ACM*, 26(1):100–106.

Hofstadter, D. (1979). *Godel, Escher, Bach: an Eternal Golden Braid*. Basic Books, Inc. New York, NY, USA.

Hogeweg, P. and Takeuchi, N. (2003). Multilevel Selection in Models of Prebiotic Evolution: Compartments and Spatial Self-Organization. *Origins of Life and Evolution of the Biosphere*, 33(4-5):375–403.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor MI: University of Michigan Press, 1st edition.

Holland, J. (1976). Studies of the Spontaneous Emergence of Self-replicating Systems using Cellular Automata and Formal Grammars. *Automata, Languages, Development*, pages 385–404.

Holland, J. (1990). ECHO: Explorations of Evolution in a Miniature World.(eds) Farmer JD, Doyne J. In *Proceedings of the second conference on artificial life, Addison-Wesely*.

Holland, J. (1992a). *Adaptation in Natural and Artificial Systems*. MIT Press Cambridge, MA, USA, 2nd edition. See also Holland(1975).

Holland, J. (1992b). Complex Adaptive Systems. *Daedalus*, 121(1):17–30.

Holland, J. (1994). Echoing Emergence: Objectives, Rough Definitions, and Speculations for Echo-class Models. *Complexity: Metaphors, Models, and Reality*, pages 309–342.

Holland, J. (1996). *Hidden order: How Adaptation Builds Complexity*. Addison Wesley Publishing Company.

Holland, J. (2001). Exploring the Evolution of Complexity in Signaling Networks. *Complexity*, 7(2):34–45.

Holland, J. (2006). Studying Complex Adaptive Systems. *Journal of Systems Science and Complexity*, 19(1):1–8.

Huang, C. Y. and Ferrell, J. E. J. (1996). Ultrasensitivity in the Mitogen-activated Protein Kinase Cascade. *Proc Natl Acad Sci U S A*, 93(19):10078–10083.

Hucka, M., Finney, A., Bornstein, B., Keating, S., Shapiro, B., Matthews, J., Kovitz, B., Schilstra, M., Funahashi, A., Doyle, J., et al. (2004). Evolving a Lingua Franca and Associated Software Infrastructure for Computational Systems Biology: the Systems Biology Markup Language (SBML) project. *IEE Proceedings Systems Biology*, 1(1):41–53.

Jerne, N. K. (1974). Towards a Network Theory of the Immune System. *Annales d'immunologie*, 125C(1-2):373–389.

Kauffman, S. (1986). Autocatalytic Sets of Proteins. *Journal of Theoretical Biology,(119)*, pages 1–24.

Kauffman, S. (1995). *At Home in the Universe: The Search for Laws of Self-organization and Complexity*. Oxford University Press, USA.

Kauffman, S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press.

Kim, S., Imoto, S., and Miyano, S. (2003). Inferring Gene Networks from Time Series Microarray Data using Dynamic Bayesian Networks. *Briefings in Bioinformatics*, 4(3):228–235.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press.

Korn, G. and Korn, T. (1956). *Electronic Analog Computers (DC Analog Computers)*. McGraw-Hill.

Krauss, G. (2003). *Biochemistry of Signal Transduction and Regulation*. John Wiley & Sons.

Kvasnicka, V. and Pospichal, J. (2001). Autoreplicators and Hypercycles in Typogenetics. *Journal of Molecular Structure: THEOCHEM*, 547(1-3):119–138.

Lanzi, P. L., Stolzmann, W., and Wilson, S. W., editors (2002). *Advances in Learning Classifier Systems, 4th International Workshop, IWLCS 2001, San Francisco, CA, USA, July 7-8, 2001, Revised Papers*, volume 2321 of *Lecture Notes in Computer Science*. Springer.

Le Novère, N. and Shimizu, T. S. (2001). StochSim: Modelling of Stochastic Biomolecular Processes. *Bioinformatics*, 17:575–576.

Lenaerts, T. and Bersini, H. (2009). A Synthon Approach to Artificial Chemistry. *Artificial Life*, 15(1):89–103.

Lenser, T., Hinze, T., Ibrahim, B., and Dittrich, P. (2007). Towards Evolutionary Network Reconstruction Tools for Systems Biology. *Lecture Notes in Computer Science*, 4447:132–142.

Lenski, R., Ofria, C., Collier, T., and Adami, C. (1999). Genome Complexity, Robustness and Genetic Interactions in Digital Organisms. *Nature*, 400(6745):661.

Lenski, R., Ofria, C., Pennock, R., and Adami, C. (2003). The Evolutionary Origin of Complex Features. *Nature*, 423:139–144.

Li, L. M. and Lu, H. H.-S. (2005). Explore Biological Pathways from Noisy Array Data by Directed Acyclic Boolean Networks. *Journal of Computational Biology*, 12(2):170–185.

Lloyd, C., Halstead, M., and Nielsen, P. (2004). CellML: its Future, Present and Past. *Progress in Biophysics and Molecular Biology*, 85(2-3):433–450.

Luck, M., Ashri, R., and d'Inverno, M. (2004). *Agent-Based Software Development*. Artech House, Inc., Norwood, MA, USA.

Machné, R., Finney, A., Müller, S., Lu, J., Widder, S., and Flamm, C. (2006). The SBML ODE Solver Library: a Native API for Symbolic and Fast Numerical Analysis of Reaction Networks. *Bioinformatics*, 22(11):1406–1407.

Mangan, S. and Alon, U. (2003). Structure and Function of the Feed-forward Loop Network Motif. *Proceedings of the National Academy of Sciences*, 100(21):11980–11985.

Margulis, L. (1981). *Symbiosis in Cell Evolution*. Freeman San Francisco.

Margulis, L. and Sagan, D. (2002). *Acquiring Genomes: A Theory of the Origins of Species*. New York: Basic books.

Matsumoto, M. and Nishimura, T. (1998). Mersenne Twister: a 623-dimensionally Equidistributed Uniform Pseudo-random Number Generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30.

McCaskill, J., Füchslin, R., and Altmeyer, S. (2001). The Stochastic Evolution of Catalysts in Spatially Resolved Molecular Systems. *Biological Chemistry*, 382(9):1343–1363.

McMullin, B. (1992). The Holland α-universes Revisited. In *Proceedings of the First European Conference on Artificial Life, MIT Press, Cambridge*, pages 317–326.

McMullin, B. (2000). Remarks on Autocatalysis and Autopoiesis. *Annals of the New York Academy of Sciences*, 901(1):163–174.

McMullin, B., Kelly, C., OSBrien, D., Mitchell, G., and Decraene, J. (2007). Preliminary Steps toward Artificial Protocell Computation. In *Proceedings of the 2007 International Conference on Morphological Computation*.

Meng, T. (2004). Modeling and Simulation of Biological Systems with Stochasticity. *In Silico Biology*, 4(3):293–309.

Mereschkowsky, C. (1910). Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen. *Biol. Centralbl*, 30:278–288.

Milner, R. (1999). *Communicating and Mobile Systems: the π-calculus*. Cambridge University Press, New York, NY, USA.

Morris, H. C. (1987). Typogenetics: A logic for artificial life. In Langton, C., editor, *ALIFE*, volume 6 of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 369–396. Addison-Wesley.

Mukai, M., Kusama, T., Hamanaka, Y., Koga, T., Endo, H., Masaharu, M. T., and Inoue, M. (2005). Crosstalk between Apoptosis and Invasion Signaling in Cancer Cells through Caspase-3 Activation. *Cancer Res*, 65(20):9121–9125.

Needham, C. J., Bradford, J. R., Bulpitt, A. J., and Westhead, D. R. (2006). Inference in Bayesian Networks. *Nature Biotechnology*, 24(1):51–53.

Nii, H. P. (1986a). Blackboard Systems, Part One: The Blackboard Model of Problem Solving and the Evolution of Blackboard Architectures. *AI Magazine*, 7(2):38–53.

Nii, H. P. (1986b). Blackboard Systems, Part Two: Blackboard Application Systems, Blackboard Systems from a Knowledge Engineering Perspective. *AI Magazine*, 7(3):82–106.

Palsson, B. (2006). *Systems biology: properties of reconstructed networks*. Cambridge University Press New York, NY, USA.

Pargellis, A. N. (2001). Digital Life Bahavior in the Amoeba World. *Artificial Life*, 7(1):63–75.

Pattee, H. (1973). *Hierarchy Theory: The Challenge of Complex Systems*. George Braziller New York.

Paun, A., Perez-Jimenez, M., and Romero-Campero, F. (2006). Modeling Signal Transduction using P systems. *Lecture Notes in Computer Science*, 4361:100.

Paun, G. and Rozenberg, G. (2002). A Guide to Membrane Computing. *Theor. Comput. Sci.*, 287(1):73–100.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Pe'er, D. (2005). Bayesian Network Analysis of Signaling Networks: a Primer. *Sci STKE*, 2005(281).

Pérez, P. P. G., Gershenson, C., Cárdenas-García, M., and Lagunez-Otero, J. (2002). Modelling Intracellular Signalling Networks using Behaviour-based Systems and the Blackboard Architecture. *CoRR*, cs.MA/0211029.

Peterson, J. L. (1981). *Petri Net Theory and the Modeling of Systems*. Prentice Hall PTR, Upper Saddle River, NJ, USA.

Petri, C. (1962). *Kommunikation mit Automaten*. Auch im Handel als: Schriften d. Rheinisch-Westfälischen Instituts f. instrumentelle Mathematik an d. Universität Bonn. Nr 2.

Pinney, J. W., Westhead, D. R., and McConkey, G. A. (2003). Petri Net Representations in Systems Biology. *Biochem Soc Trans*, 31(Pt 6):1513–1515.

Polyanin, A. D. and Zaitsev, V. F. (2002). *Handbook of Exact Solutions for Ordinary Differential Equations, Second Edition*. Chapman & Hall/CRC, 2 edition.

Ray, T. (1991). An Approach to the Synthesis of Life. *Artificial Life II*, 10:371–408.

Ray, T. (1992). Evolution, Ecology and Optimization of Digital Organisms. *Santa Fe Institute working paper 92-08-042*.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution*. Frommann-Holzboog.

Reddy, V., Mavrovouniotis, M., and Liebman, M. (1993). Petri Net Representations in Metabolic Pathways. *Proc Int Conf Intell Syst Mol Biol*, 1:328–36.

Regev, A. and Shapiro, E. (2004). The π-calculus as an Abstraction for Biomolecular Systems. *Modelling in Molecular Biology*, page 219.

Regev, A., Silverman, W., and Shapiro, E. Y. (2001). Representation and Simulation of Biochemical Processes Using the π-Calculus Process Algebra. In *Pacific Symposium on Biocomputing*, pages 459–470.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., and Lorensen, W. (1991). *Object-Oriented Modeling and Design*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Sachs, K., Gifford, D., Jaakkola, T., Sorger, P., and Lauffenburger, D. A. (2002). Bayesian Network Approach to Cell Signaling Pathway Modeling. *Sci STKE*, 148(EG10).

Sauro, H. (2004). The Computational Versatility of Proteomic Signaling Networks. *Current Proteomics*, 1:67–81.

Schwab, E. D. and Pienta, K. J. (1997). Modeling Signal Transduction in Normal and Cancer Cells using Complex Adaptive Systems. *Med Hypotheses*, 48(2):111–123.

Smith, J. and Szathmáry, E. (1997). *The Major Transitions in Evolution*. Oxford University Press, USA.

Stadler, P. F., Fontana, W., and Miller, J. H. (1993). Random Catalytic Reaction Networks. *Phys. D*, 63(3-4):378–392.

Stock, J., Surette, M., McCleary, W., and Stock, A. (1992). Signal Transduction in Bacterial Chemotaxis. *Journal of Biological Chemistry*, 267(28):19753–19756.

Szathmary, E. and Demeter, L. (1987). Group Selection of Early Replicators and the Origin of Life. *Journal of Theoretical Biology*, 128(4):463–486.

Szathmáry, E. and Maynard Smith, J. (1997). From Replicators to Reproducers: the First Major Transitions Leading to Life. *Journal of Theoretical Biology*, 187(4):555–571.

Tominaga, K., Suzuki, Y., Kobayashi, K., Watanabe, T., Koizumi, K., and Kishi, K. (2009). Modeling Biochemical Pathways using an Artificial Chemistry. *Artificial Life*, 15(1):115–129.

Tominaga, K., Watanabe, T., Kobayashi, K., Nakamura, M., Kishi, K., and Kazuno, M. (2007). Modeling Molecular Computing Systems by an Artificial Chemistry—Its Expressive Power and Application. *Artif. Life*, 13(3):223–247.

Van Kampen, N. (2007). *Stochastic Processes in Physics and Chemistry*. North-Holland.

Vera, J., Balsa-Canto, E., Wellstead, P., Banga, J., and Wolkenhauer, O. (2007). Power-law Models of Signal Transduction Pathways. *Cellular signalling*, 19(7):1531–1541.

Volfson, D., Marciniak, J., Blake, W., Ostroff, N., Tsimring, L., and Hasty, J. (2005). Origins of Extrinsic Variability in Eukaryotic Gene Expression. *Nature*, 439(7078):861–864.

von Neumann, J. (1949). Theory and Organization of Complicated Automata. *Burks (1966)*, pages 29–87.

Wagner, A. (2005). *Robustness and evolvability in living systems*. Princeton University Press Princeton, NJ.

Wee, K. and Lee, W. (2005). Extensions and Variations on Construction of Autoreplicators in Typogenetics. *Lecture notes in computer science*, 3630:685.

Winikoff, M. and Padgham, L. (2004). *Developing Intelligent Agent Systems: A Practical Guide*. Halsted Press, New York, NY, USA.

Wolkenhauer, O. and Hofmeyr, J. (2007). An Abstract Cell Model that describes the Self-organization of Cell Function in Living Systems. *Journal of Theoretical Biology*, 246(3):461–476.

Woolf, J., Wendy, P., Laurence, D., Daley, Q., and Lauffenburger, A. (2005). Bayesian Analysis of Signaling Networks Governing Embryonic Stem Cell Fate Decisions. *Bioinformatics*, 21(6):741–753.

Xiong, W. and Ferrell, J. (2003). A Positive-feedback-based Bistable Memory Module that Governs a Cell Fate Decision. *Nature*, 426(6965):460–465.

Yee, D. and Lee, A. (2000). Crosstalk between the Insulin-like Growth Factors and Estrogens in Breast Cancer. *J Mammary Gland Biol Neoplasia*, 5(1):107–115.

Zwillinger, D. (1992). *Handbook of Differential Equations*. Academic Press, San Diego, CA, 2nd edition.