Unexpected Evolutionary Dynamics in a String Based Artificial Chemistry

James Decraene, George Mitchell, Barry McMullin

Artificial Life Laboratory - Research Institute for Networks and Communications Engineering
School of Electronic Engineering - Dublin City University, Ireland
Email: james.decraene@eeng.dcu.ie

Abstract

This work investigates closure in Cell Signaling Networks, which is one research area within the ESIGNET project[1]. We employ a string-based Artificial Chemistry based on Holland's broadcast language (*Molecular Classifier System, Broadcast Language*, or MCS.b). We present a series of experiments focusing on the emergence and evolution of self-maintaining molecular organizations. Such experiments naturally relate to similar studies conducted in artificial chemistries such as Tierra, Alchemy and Alpha-Universes. However, our results demonstrate some counter-intuitive outcomes, not indicated in previous literature. Each of these "unexpected" evolutionary dynamics (including an elongation catastrophe phenomenon) are examined and explained both informally and formally. We also demonstrate how the elongation catastrophe can be prevented using a multi-level selectional model of the MCS.b (which acts both at the molecular and cellular level). This work provides complementary insights into the understanding of evolutionary dynamics in minimal artificial chemistries.

Introduction

Cell Signaling Networks (CSNs) are complex biochemical networks of interacting molecules (proteins, ions, secondary messengers, etc.) occurring in living cells. Through complex molecular interactions (e.g., signal transduction), CSNs are able to coordinate critical cellular activities (e.g., cell differentiation, apoptosis) in response to internal and external stimuli.

As CSNs occur in cells, these networks have to replicate themselves prior to the cellular division. This allows the replicated CSNs to be "distributed" to the offspring cells. Errors may occur during this replication process, e.g., an offspring cell may inherit only a partial CSN. Thus resulting in potentially defective cells which would lead to a variety of undesired effects (e.g., premature cell death). As a result, the "fitness" of a cell is implicitly represented by the *survival* and *performance* of a cell in achieving self-maintenance *and* cell-level replication.

[1]ESIGNET: Evolving Cell Signaling Networks *in silico*, an EU FP6 project, contract no. 12789, http://www.esignet.net

Based on the above assumption, we hypothesize that CSNs may be regarded as subsets of closed (and thus self-maintaining) systems. The latter would have the additional ability to replicate themselves as a whole (cellular division). The signal processing ability of CSNs would emerge from the closure properties of these systems.

Examining such phenomena relates closely to other studies which have been conducted on Holland's Alpha-Universes (Holland, 1976), Tierra (Ray, 1991) and Alchemy (Fontana and Buss, 1994). Although these Artificial Chemistries (ACs) were developed for different purposes and were implemented differently, these systems exhibited common evolutionary phenomena such as the emergence of (collectively) autocatalytic reaction networks (Dittrich et al., 2001; McMullin, 2000). In this investigation, such classes of network are of interest as they would allow CSNs to self-maintain and replicate themselves. Moreover, as demonstrated in several ACs, it is commonly accepted that the emergence and maintenance of such collectively autocatalytic reaction networks is relatively trivial.

We introduce the *Molecular Classifier System, Broadcast Language System*, or MCS.b (J.Decraene et al., 2007). This addresses the reflexive nature of molecular species and automatically gives rise to an implicit molecular fitness function represented by the "replication" ability of the individual molecular species. We present a series of experiments focusing on the emergence of self-maintaining organizations and finally we examine the outcomes of these experiments together with possible modifications for further work.

Molecular Classifier Systems

Molecular Classifier Systems are a class of string-rewriting based AC inspired by the broadcast language (BL; see Holland, 1992). As opposed to more traditional string-rewriting systems, operations are stochastic and reflexive (no distinction made between operands and operators). The behavior of the condition (binding) properties and action (enzymatic functions) is defined by a language specified within the MCS. This "chemical" language defines and constrains the complexity of the chemical reactions that may be mod-

eled and simulated. In this AC, all reactants are catalytic in the sense that they are not consumed during reactions. These reactions result from successful molecular interactions which occur at random. When a reaction occurs, a product molecule is inserted in the reactor whereas another molecule, selected at random, may be removed from the reactor space (designating the system outflow).

A molecule may contain several condition/action rules which define the binding and enzymatic properties. A reaction between molecules occurs if at least one conditional part from any rules in a molecule A matches a target molecule B. A is regarded as an enzyme whereas B is regarded as a substrate molecule. When a reaction occurs, the action part from the satisfied rule in A is utilized to perform the enzymatic operations upon the bound substrate molecule B. This operation results in the production of another offspring (product). If several rules in A are satisfied by B, then one of these rules is picked at random and employed to carry out the enzymatic function.

We proposed a simplification of the BL (J.Decraene et al., 2007) which is used as the MCS chemical language resulting in the MCS.b system. MCS.b has some similarity with the Learning Classifier Systems, also pioneered by John Holland (Holland and Reitman, 1978); however there are also a number of differences. For example, the LCS strings are fixed length on an alphabet of $\lambda = \{\ 1,\ 0,\ \#\}$; whereas the BL strings are of variable length using a significantly larger alphabet of $\Lambda = \{1,\ 0,\ *,\ :,\ \Diamond,\ \triangledown,\ \triangle,\ '\}$. BL strings are referred to as *broadcast devices*. A broadcast device is parsed into zero, one or more *broadcast units*, where each unit represents a single condition/action rule. The symbol $*$ separates broadcast units within a broadcast device. The symbol : separates a condition from an action within a single broadcast unit. $\{\Diamond,\ \triangledown,\ \triangle\}$ are single/multiple character wildcards that may also copy matched (sub-)strings into output strings. A detailed description is omitted in this paper, see (J.Decraene, 2006) for full specification of our BL implementation.

Autocatalytic organizations

A series of experiments using the MCS.b is now outlined. These experiments first examine both the self-maintenance and the spontaneous emergence of autocatalytic molecules (i.e., molecules that can self-replicate). Both spontaneous emergence and self-maintenance were reported as easily obtained in Alchemy. Spontaneous emergence was not expected or reported for the original Tierra system; however, it did arise in the related Amoeba system, specifically devised for this purpose (Pargellis, 2001).

No selective advantages for universal replicases

An artifact 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 ($R_0 = *\triangledown : \triangledown$) is autocatalytic. Although the probability of spontaneously obtaining such autocatalytic molecules is therefore quite low in MCS.b, the intuition was that, *once* such a molecule does appear, it should be able to rapidly fill the reaction space. This phenomenon was indeed observed in Alchemy and was expected to occur in MCS.b. We present here a series of experiments which explore and test this conjecture.

The behavior of the minimal self-replicase, R_0, is as follows. The matching condition is defined by a single symbol, $\triangledown$, which designates a multiple character wildcard. This indicates that R_0 may bind to any molecule. In addition when a reaction occurs between R_0 and a substrate molecule I_0, $\triangledown$ is assigned a value, being the matched substring of I_0. In this case, this will be the complete string I_0. A unique symbol $\triangledown$ also constitutes the action part of R_0. This specifies that the output string of R_0 is exactly the string bound by the $\triangledown$ in the condition part, i.e., a copy of I_0. Therefore the broadcast device 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., $I_0 = R_0$). The "specificity" of R_0 is said to be *null*.

Fig. 1 presents a first experiment examining the behavior of R_0 averaged over 30 simulation runs. The broadcast "universe" (reaction space) is configured as follows:

- The system is seeded with 900 randomly generated molecules, each of length 10 symbols.
- In addition, 100 instances of 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 A and B are picked at random, A is considered as an enzyme and B as a substrate. If A can bind and react with B then a molecule C is produced. If the current size of the population, n, is less than n_{max} then C is simply added to the population (and n increases by 1); otherwise a molecule is picked at random and is replaced with C (and the population size remains unchanged at n).
- No mutation may occur in these experiments.
- A single "timestep" is arbitrarily defined as 50 molecular interactions.

A high concentration (0.1) of R_0 was chosen to minimise early extinction due simply to stochastic fluctuation.

From Fig. 1 it is clear that the species R_0 never grows to take over the population; on the contrary, it consistently diminishes, contrary to the original, informal, prediction. A

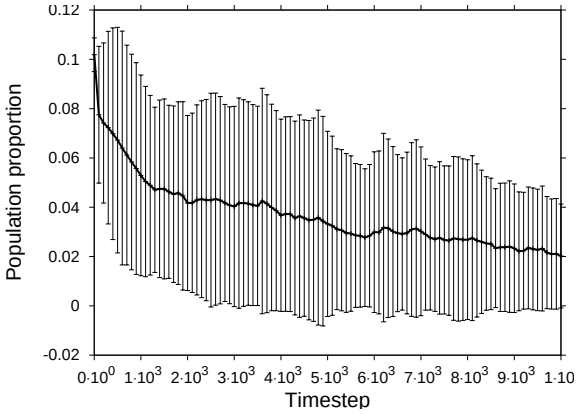

Figure 1: Relative population growth of replicators R_0 averaged over 30 simulation runs. Solid line is average concentration; error bars denote standard deviation.

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:

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

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:

$$\alpha_{ij}^{k} = \begin{cases} 1 \text{ if } s_i + s_j \rightarrow s_i + s_j + s_k \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

For simplicity, consider the simple case where only universal replicases (R_0) and non-enzymatic molecules (NE) (that may only act as substrates) are present. This is clearly the *most* favourable case for the growth of R_0. Denote the molecular concentrations of R_0 and NE by x_1 and x_2 respectively. Then $\alpha_{ij}^{1} = 1$ if $i = 1, j = 1$; otherwise $\alpha_{ij}^{1} = 0$. Similarly, $\alpha_{ij}^{2} = 1$ if $i = 1, j = 2$; otherwise $\alpha_{ij}^{2} = 0$. Inserting into Eq. 1, we obtain:

$$\dot{x_1} = x_1^2 - x_1(x_1^2 + x_1 x_2) \tag{3}$$

But given that $x_2 = 1 - x_1$:

$$\begin{aligned} \dot{x_1} &= x_1^2 - x_1^3 - x_1^2 + x_1^3 \\ \dot{x_1} &= 0 \end{aligned} \tag{4}$$

whereas the growth rate of molecules NE is:

$$\begin{aligned} \dot{x_2} &= x_1(1 - x_1) - (1 - x_1)[x_1^2 + x_1(1 - x_1)] \quad (5) \\ \dot{x_2} &= x_1 - x_1^2 - (1 - x_1)(x_1^2 + x_1 - x_1^2) \\ \dot{x_2} &= x_1 - x_1^2 - x_1 + x_1^2 \\ \dot{x_2} &= 0 \quad (6) \end{aligned}$$

Thus, both molecular species R_0 and 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 R_0 species. Qualitatively this is due to the fact that any (self-)replicase having low or zero specificity, such as 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 R_0, where none of the other molecules had any enzymatic activity. In the practical case of Fig. 1 the collection of such additional side reactions will give a nett negative growth rate for R_0, which therefore, quasi-deterministically, decays.

Specificity and domination of the replicases

To confirm the importance of specificity, we proceeded to a series of experiments in which we incrementally increased the specificity of the (self-)replicases. Table 1 shows the different replicases employed in these experiments. R_1 designates a molecule that would only react with molecules whose strings end with the symbol "1". As the latter occurs at the rightmost position of R_1, it may react with itself, producing another instance of R_1. Similarly, R_2 only binds to molecular strings containing the suffix 01. This "signature" forms a constraint on the replicases, allowing them to react only with a progressively more restricted set of substrate molecules. This impacts directly on these molecules' binding specificity.

Replicase	Informational string
R_0	$*\nabla : \nabla$
R_1	$*\nabla 1 : \nabla 1$
R_2	$*\nabla 01 : \nabla 01$
R_3	$*\nabla 101 : \nabla 101$
R_4	$*\nabla 0101 : \nabla 0101$

Table 1: (self-)replicases with increasing specificity

The results depicted in Fig. 2 confirm the importance of specificity upon the system dynamics. The ability of a (self-)replicase to dominate and sustain itself, against a random initial population of molecules, increases progressively with its binding specificity. As in the previous section, we can explain and demonstrate this behavior through the use of a simple ODE model.

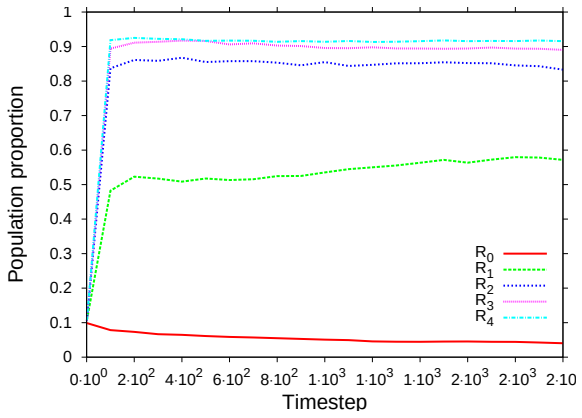

Figure 2: Population growth of replicators R_0, R_1, R_2, R_3 and R_4. Each line represents the average concentration of corresponding replicase over 30 simulation runs.

In this case, we consider a reactor containing only the following molecular species:

- Replicases R_1 which only replicate molecules terminating with the symbol "1" (which includes R_1 molecules themselves).
- A variety of non-enzymatic molecules NE which are randomly generated. $NE_1 \subseteq NE$ is the subset of molecules whose strings terminate with the designated symbol. These molecules contained in NE_1 can be replicated by molecules 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 R_1 and x_2 is the sum of concentrations of molecules in NE_1. The growth rate of the different molecular species in this reactor are as follows:

$$\dot{x_1} = x_1^2 - x_1(x_1^2 + x_1x_2) \quad (7)$$
$$\dot{x_1} = x_1^2 - x_1^3 - x_1^2x_2$$
$$\dot{x_1} = x_1^2(1 - x_1 - x_2) \quad (8)$$

The growth rate of molecules NE_1 is:

$$\dot{x_2} = x_1x_2 - x_2(x_1^2 + x_1x_2) \quad (9)$$
$$\dot{x_2} = x_1x_2 - x_1^2x_2 - x_1x_2^2$$
$$\dot{x_2} = x_1x_2(1 - x_1 - x_2) \quad (10)$$

Since $x_1 + x_2 + \ldots + x_n = 1$, we have $x_1 + x_2 < 1$ and therefore $\dot{x_1} > 0$ and $\dot{x_2} > 0$. Whereas the growth rate of any other molecules (that may be not replicated by R_1) in the reactor space is given by:

$$\dot{x} = 0 - x_i(x_1^2 + x_1x_2) \quad (11)$$
$$\text{with } 2 < i \leq n$$

In Eq. 12, we note that any given molecules $s = (s_3, \ldots, s_n)$ possess a negative growth rate which indicate that these molecules would be displaced by molecules R_1 and NE_1.

In this model, only NE_1 molecules are able to parasite the replicases R_1. By increasing the specificity of replicases, we decrease the range of molecule that may parasite the replicases. This explains the behavior observed in Fig. 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 sustain themselves and/or to dominate the molecular population, a significant binding specificity is required. We conjecture that this underlying phenomenon may have been implicated in the dynamics of a variety of previously reported artificial chemistries; but, to our knowledge, it has not previously been explicitly isolated in the manner presented here.

Spontaneous emergence of replicases

In the previous set of experiments, mutation was turned off in order to facilitate our investigation on replicases, which were hand-designed and inserted into the initial population. This led to a limited diversity in the population. To examine the spontaneous emergence of autocatalytic molecules, we performed a second series of experiments in which no replicases are specified and molecular mutation could occur. The latter is implemented as follows:

- When a new molecule is produced, a mutation with probability $p_{sym} = 0.001$ may be applied to each of its symbols. Therefore, the longer the molecule, the higher the probability of mutation occurring.
- Three types of mutation are distinguished and are applied with equal probabilities:
  - *Symbol flipping:* The current symbol is replaced with a symbol picked uniformly at random from Λ.
  - *Symbol insertion:* A symbol is picked uniformly at random from Λ and inserted after the current symbol.
  - *Symbol deletion:* The current symbol is removed.
- To maintain diversity in the event of low ongoing reaction activity, a global mutation technique occurring every 100 timesteps is also available. A subset ($r_{mut} = 0.01$) 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.

As mutation now occurs, diversity is maintained during long term evolution. The spontaneous appearance of replicators was expected. Results indicated that (self-)replicases do emerge, however they never manage to self-sustain.

This is explained as follows:

- As already noted, the BL syntax does not strongly facilitate the spontaneous emergence of replicators. This syntactical constraint may discourage the spontaneous emergence of self-replicators. The BL syntax may also have an impact on the robustness of these self-replicators against mutation effects.

- Secondly if self-replicators do emerge, they would be required to possess a specificity higher than null to sustain themselves.

- Finally, replicators are likely to possess a low molecular concentration when emerging. This low concentration diminishes the capacity of these molecular species to persist against side reactions and mutation events.

These three factors, when combined, significantly lower the probability of having a replicator spontaneously emerge and self-sustain in the MCS.b.

We examined the nature of the (self-)replicases that may emerge during evolution. An additional set of experiments was specified as follows:

- Each simulation run was initialised with 100 randomly generated, 10-symbol long, molecules.

- $n_{max} = 1000$ (i.e., the population initially grew without any displacement; but once the total number of molecules reached 1000 it was limited to this value, by displacing one random molecule for each new molecule generated, as previously described).

- 30 simulation runs were performed, each for 100000 timesteps.

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 experiments are listed in Table 2. This shows that 15 distinct self-replicases appeared. However, 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 original universal self-replicase, $R_0 = *\nabla : \nabla$, discussed earlier. 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 R_1, having just the minimal specificity of one symbol.

In the 30 experimental runs, the highest concentration achieved by any of these spontaneously occurring self-replicases was 0.001—i.e., just a single isolated molecule.

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 2: Spontaneously emergent self-replicases in MCS.b

This is consistent with the comments earlier in this section, and the results of the previous section. 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); but self-replicases of very low specificity (which do spontaneously occur) cannot grow to significant concentrations.

The spontaneous emergence of a "sustainable" self-replicase (i.e., of sufficient specificity to establish itself) remains theoretically possible in MCS.b. However, both the experimental results and the informal 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.b therefore shares this property with the Tierra system.

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.b system.

However, the results indicate that MCS.b does *not* exhibit an evolutionary dynamic at all comparable to Tierra in this case. Fig. 3 presents an example of such an experiment. The "ancestor" self-replicators do, at first, quickly fill the reaction space ($n_{max} = 1000$), 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 were arbitrarily limited to a maximum length of $BD_{lmax} = 500$. Other experiments, without such a limit, indicated that the growth in molecular length appeared to continue indefinitely, subject only to available physical (computer) resources.

As with the experiments discussed earlier, these results were not expected. In fact, certain mutants of the original au-

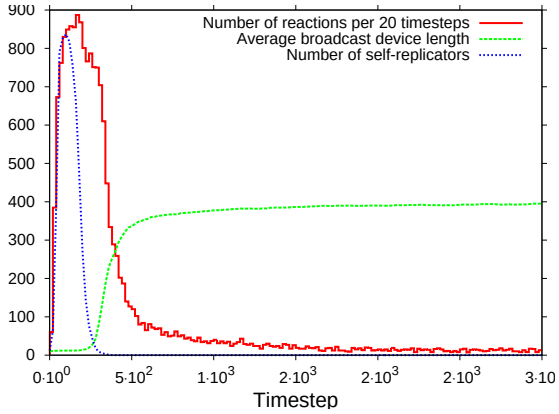

Figure 3: Effects of molecules length growth upon overall system reactions rates. In this experiment, an ancestor ($R_4 = \nabla 0101 : \nabla 0101$) is inserted (with initial concentration $[R_4] = 0.1$) in addition to randomly generated molecules. Moreover mutation per molecule and per symbol is turned on.

tocatalytic 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 signature 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 which we define two molecules: $R_4 = *\nabla 0101 : *\nabla 0101$ and $R'_4 = *\nabla 0101 : *\nabla 00101$. The latter is a readily accessible mutant of R_4. Once it appears, the mutant R'_4 allows for a runaway degenerative scenario to occur. The possible reactions are as follows:

$$
\begin{aligned}
R_4 + R_4 + X &\longrightarrow 3R_4 \\
R_4 + R'_4 + X &\longrightarrow R_4 + 2R'_4 \\
R'_4 + R_4 + X &\longrightarrow R_4 + 2R'_4 \\
R'_4 + R'_4 + X &\longrightarrow 2R'_4 + R''_4
\end{aligned}
$$

X is a molecule picked at random and removed from the population. The product R''_4 is of the form $*\nabla 0101 : *\nabla 000101$ and similarly has a selective advantage over both R_4 *and* R'_4. The reaction $R''_4 + R''_4 + X$ would result in the production of a molecule 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 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 : will commonly "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 $1s$s. As a result, although some molecules may still possess an active site capable of some enzymatic function, their high specificity decrease 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). Fig. 4 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.

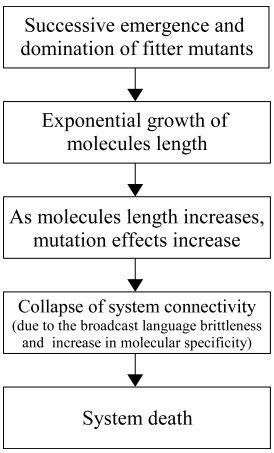

Figure 4: Elongation catastrophe in MCS.b

Fixing the elongation catastrophe: 1

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

1. In the system presented earlier, reactions leading to the production of molecules that were longer than BD_{lmax} were simply not permitted. An initial modification was to permit such reactions to proceed, but to truncate the product molecules at length BD_{lmax}. The system could

then remain active (with ongoing reactions) even though the molecules have reached a critical size.

2. The multiple symbols 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 form of "parametric" mutation, where its value would change randomly in $[1, c_{max}]$ over time.

3. Similarly to (1), a finite total number, A, of "free" symbol objects (atoms) available in the broadcast universe was defined. This reservoir of (untyped) atoms is reduced when new molecules are produced, and increased when molecules are destroyed. If insufficient atoms are available to complete a reaction, the reaction fails. This should favor smaller molecules over longer ones suffering from elongation catastrophe.

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

5. Another extension to (3) was to vary the probability of a reaction to occur according to the product's length and A. 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 activity was also quasi null.
- 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 occur.

Thus, although a range of modifications were implemented, the different outcomes do not differ substantially from the degenerative cases presented above (section *Rise and fall of the fittest*).

Fixing the elongation catastrophe: 2

In this section we present an alternative approach to the MCS.b elongation catastrophe, based on *multi-level* selection. This has previously be demonstrated to be an effective means to provide resistance against parasites for catalytic networks Hogeweg and Takeuchi (2003). In such systems parasitized cells decay and may be displaced by neighboring healthy cells.

In the single-level selectional MCS.b model, competing molecules were contained in a single reactor, which we refer to as the molecular level of selection. In the multi-level selectional model, we introduce multiple reactors, each containing a population of molecules. These reactors ("cells") may be subjected to cellular division, which results in the replacement of the parent cell and creation of two offspring cells. However, the number of cells in the broadcast universe is fixed. As a result such a cellular division also triggers the removal of another cell selected at random. In a similar manner to molecules, cells are competing with each other which is regarded as the second level of selection.

In contrast to the single level model, successful reactions do not lead to the removal of a random molecule in the reaction space. Thus the number of molecules contained in a cell may increase until it reaches a finite limit l. When a cell reaches this size, a spontaneous division occurs. Half of the molecules are selected at random. These are removed from the "parent" cell and inserted into a newly created "daughter" cell. This is then inserted in the population of cells. Finally, a cell is picked at random (other than the parent and daughter cell) and removed from the population.

For time efficiency our multi-level model was implemented on a distributed, symmetrical, computer cluster where each cell was run on a single CPU. In this concurrent model, the fittest cells would not only be the cells that exhibit a high molecular growth rate, but cells that also contain molecules that are fast to compute (in real time). In other words, if we consider two cells which present an equal overall molecular growth rate, but contains molecules with different computational complexities, the cell which possesses a smaller overall molecular computational complexity will have the selective advantage.

We conducted a series of experiments as follows:

- 32 cells are employed.
- $l = 1000$ is the cell capacity.
- Mutation is turned on.
- Each cell is seeded with 250 replicases $R_4 = *\triangledown 0101 : \triangledown 0101$ and 250 randomly generated molecules of length 10.
- 5 simulation runs were conducted for at least 50 million molecular interactions per "cell object" (i.e., the run terminates when every concurrent cell object, one per CPU, has run for at least 50 million interactions each).

Results indicated that none of the evolved cells resulting from the simulation suffered from elongation catastrophe.

During an evolutionary run, we may observe the elongation catastrophe phenomenon to occur as expected. However we know that if the parasitic mutants appear in a cell, the cell would degenerate and not produce sufficient molecules to trigger cellular division and ultimately the displacement of another cell. As a result those cells would not have any selective advantages over the other "healthy" cells. On the contrary, healthy cells may still possess a high molecular reactivity and would consequently displace the infected cells.

Moreover, our results indicate that infected cells do not get displaced only when their connectivity is quasi null. In fact, these cells can get displaced at an early stage, when they would still present a high molecular activity although still being considered as being infected. As mentioned earlier, another fitness aspect to be considered is the computational complexity of molecules contained in a cell. Infected cells would rapidly produce molecules which have an increasing length, and this elongation has the effect of increasing their computational cost. As a consequence, although such cells may contain molecules with a richly connected reaction network, and therefore with a high continuing molecular replication rate, the overall cell growth rate is now penalized for having a higher molecular computational cost; as opposed to the healthy cells which generally still exhibit relatively short molecules and thus have lower computationally cost. This ultimately leads to a rapid displacement of infected cells whenever they would appear.

This multi-level selectional model successfully prevented the elongation catastrophe phenomenon from occurring. The nature of the evolved populations resulting from the simulation runs were at least somewhat comparable to those expected from systems such as Alchemy. Specifically, we have observed the rapid domination of molecular organizations which involve a range of replicases, capable of self-sustaining over time.

Conclusion

We conducted a series of experiments using the MCS.b system. These focused on the emergence and evolution of self-maintaining molecular organizations. Our results indicated counter-intuitive outcomes when compared with a variety of other AC systems in the literature. Each of these unexpected evolutionary dynamics was described and explained in detail. We also demonstrated how the elongation catastrophe can be prevented using a multi-level selectional model, which allowed for the evolution of organizations that were capable of self-sustaining over time. We propose to extend this multi-level selectional model by introducing new cellular division criteria, which would constrain and drive the evolution of the molecular networks. This may ultimately give rise to the emergence of proto-CSNs, being subsets of closed molecular systems, capable of some distinct CSN control-like features.

Acknowledgments

This work was funded by the EU FP 6 ESIGNET (Evolving Cell Signaling Networks in Silico)contract no. 12789. The computing facilities of ESIGNET and the Irish Center for High-End Computing (project code: dccom011c) are acknowledged. We are grateful to Ciarán Kelly for valuable comments.

References

Dittrich, P., Ziegler, J., and Banzhaf, W. (2001). Artificial Chemistries-A Review. *Artificial Life*, 7(3):225–275.

Fontana, W. and Buss, L. (1994). What Would be Conserved if "the Tape were Played Twice"? *Proc. of the National Academy of Sciences*, 91(2):757–761.

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. (1976). Studies of the spontaneous emergence of self-replicating systems using cellular automata and formal grammars. *Automata, Languages, Development*, pages 385–404.

Holland, J. (1992). *Adaptation in natural and artificial systems*. MIT Press Cambridge, MA, USA.

Holland, J. H. and Reitman, J. S. (1978). Cognitive systems based on adaptive algorithms. In Waterman, D. A. and Roth, H. F., editors, *Pattern-Directed Inference Systems*. Academic Press, New York.

J.Decraene (2006). The Holland Broadcast Language. Technical Report ALL-06-01, RINCE, Dublin City University. `http://elm.eeng.dcu.ie/~alife/jd/ALL-06-01/`.

J.Decraene, Mitchell, G., and McMullin, B. (2007). Evolving Artificial Cell Signaling Networks: Perspective and Methods. *Advances in Biologically Inspired Information Systems: Models, Methods, and Tools*, 69.

McMullin, B. (2000). Remarks on Autocatalysis and Autopoiesis. *Annals of the New York Academy of Sciences*, 901(1):163–174.

Pargellis, A. N. (2001). Digital life bahavior in the amoeba world. *Artificial Life*, 7(1):63–75.

Ray, T. (1991). An approach to the synthesis of life. *Artificial Life II*, 10:371–408.

Stadler, P. F., Fontana, W., and Miller, J. H. (1993). Random catalytic reaction networks. *Phys. D*, 63(3-4):378–392.