This document is downloaded from DR-NTU (https://dr.ntu.edu.sg)
Nanyang Technological University, Singapore.

The evolution of complexity in self-maintaining cellular information processing networks

Decraene, James; McMullin, Barry

2011

Decraene, J., & Mcmullin, B. (2011). The Evolution of Complexity in Self-maintaining Cellular Information Processing Networks. Advances in Complex Systems, 14(1), 55-75.

https://hdl.handle.net/10356/93823

https://doi.org/10.1142/S0219525911002913

© 2011 World Scientific Publishing. This is the author created version of a work that has been peer reviewed and accepted for publication by Advances in Complex Systems, World Scientific Publishing. It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: http://dx.doi.org/10.1142/S0219525911002913.

Downloaded on 21 Apr 2025 06:15:50 SGT

Advances in Complex Systems
© World Scientific Publishing Company

The Evolution of Complexity in Self-Maintaining Cellular Information Processing Networks

JAMES DECRAENE
Parallel and Distributed Computing Center
School of Computer Engineering
Nanyang Technological University, Singapore
jdecraene@ntu.edu.sg

BARRY MCMULLIN
Artificial Life Laboratory
The Rince Institute
Dublin City University, Dublin, Ireland
barry.mcmullin@dcu.ie

Received (received date)
Revised (revised date)

We examine the role of self-maintenance (collective autocatalysis) in the evolution of computational biochemical networks. In primitive proto-cells (lacking separate genetic machinery) self-maintenance is a necessary condition for the direct reproduction and inheritance of what we here term Cellular Information Processing Networks (CIPNs). Indeed, partially reproduced or defective CIPNs may generally lead to malfunctioning or premature death of affected cells. We explore the interaction of this self-maintenance property with the evolution and adaptation of CIPNs capable of distinct information processing abilities. We present an evolutionary simulation platform capable of evolving artificial CIPNs from a bottom-up perspective. This system is an agent-based multi-level selectional 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. Using this system, we successfully evolve an artificial CIPN to improve performance on a simple pre-specified information processing task whilst subject to the constraint of continuous self-maintenance. We also describe the evolution of self-maintaining, crosstalking and multitasking, CIPNs exhibiting a higher level of topological and functional complexity. This proof of concept aims at contributing to the understanding of the open-ended evolutionary growth of complexity in artificial systems.

Keywords: Self-maintaining chemistry; collective autocatalysis; evolutionary growth of complexity; agent-based artificial chemistry.

1. Introduction

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

stimuli. Examples of CIPNs include cell signalling networks[a] [14] (pp. 1–154) such as the chemotaxis signalling pathway [22] which may 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.

As signal processing systems, CIPNs can be regarded as special purpose computers [5]. 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. We consider the computational nature of individual molecules. A single enzyme molecule can be regarded as carrying out pattern matching to identify and bind target substrate(s), 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).

The concept of *collective autocatalysis* was formulated by Kauffman [9]. This was proposed to help explain the emergence and early evolution of life through a process of spontaneous self-organisation. A collectively autocatalytic set is a collection of molecular species where each of them is the product of at least one reaction catalysed by at least one other species of the set. It is argued that given a critical diversity of molecular species, the spontaneous emergence and self-organisation of an autocatalytic set may occur. The emergence of autocatalytic sets was also examined by Fontana and Buss [11] using the Alchemy artificial chemistry, where it corresponds to a more general formal concept of (collective) *self-maintenance*. This has been further elaborated and refined in the *Chemical Organization Theory* of Dittrich and Speroni [8].

This formal property of collective autocatalysis or self-maintenance ensures that reaction networks can reconstitute themselves (self-repair) when subjected to internal and external perturbations and during cellular divisions (given a continuous inflow of "food molecules"). Self-maintenance may thus potentially mediate between the conflicting properties of robustness and evolvability in reaction networks.

In contrast to modern living cells, the cellular model considered here does not incorporate a distinct genetic translation system. The latter can be regarded as a "programmable self-maintaining core" from which relatively arbitrary reaction networks may be regenerated. That is, in modern cells, most CIPNs are presumably inherited via a coded genetic representation; i.e., they are not required to be self-

[a]The work presented in this paper 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.

maintaining in themselves. By contrast, the model presented here may address the evolution of information processing in (proto-)cells *prior* to the emergence of the genetic architecture.

The paper is organised as follows: We first present our evolutionary simulation platform which is an agent-based multi-level selectional Artificial Chemistry called the Molecular Classifier System (MCS.bl). We then describe two evolutionary experiments where self-maintaining reaction networks are evolved to achieve pre-specified information processing functions. The aim of these experiments is to investigate the evolution of complexity in reaction networks through the emergence and evolution of computational capabilities under the constraint of preserving collective chemical self-maintenance. Finally we discuss the limitations and impact of this work.

2. The Artificial Chemistry

We first present the MCS.bl metaphor and outline the Holland broadcast language (BL) which is employed to specify the molecular species and reactions. We then describe the reactor algorithm which was implemented on a distributed system (using a cluster of computers).

2.1. *The Molecular Classifier System*

We employ an agent/string-based Artificial Chemistry (AC) called the MCS.bl which is based on Holland's broadcast language [15] (pp. 143-152). In this AC, the chemical operations are stochastic and reflexive (no distinction made between operands and operators). All reactants are catalytic in the sense that they are not consumed during reactions. A molecule may contain several condition/action rules which define the binding and enzymatic properties.

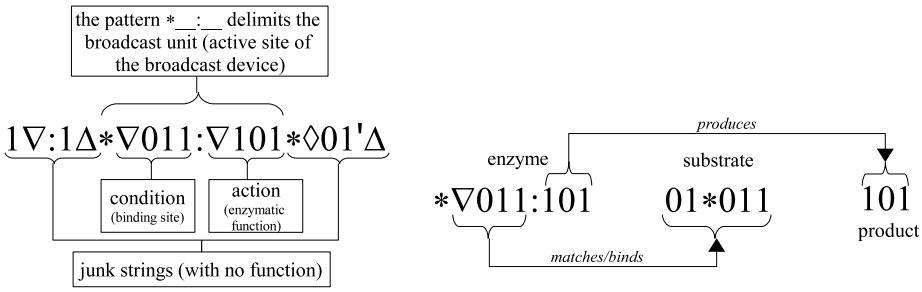
The basic elements of the Broadcast Language (BL) are strings called *broadcast devices* (i.e., molecular species), which are strings over $\Lambda = \{1,\ 0,\ *,\ :,\ \Diamond,\ \triangle,\ ',\ \triangledown\}$, see Fig. 1.

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. 0s and 1s are basic informational symbols.

When occurring in the condition statement, $\{\Diamond,\ \triangle\}$ act as single character wildcards. Whereas $\triangledown$ is a multiple character wildcard. If $\triangle$ or $\triangledown$ occurs in both the condition and action statements, then these symbols also transpose the matched string (either a single or multiple characters) into the output string.

Quoted symbols (preceded by $'$) are prevented from interpretation. Within broadcast devices, we may also 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).

Table 1 presents a comparison between the biological and the broadcast system terminology. Table 2 presents a number of example reactions that can be realised



(a) (b)

Fig. 1. The broadcast language. (a) An example broadcast device: the active site designates the condition/action rule defining the binding/enzymatic properties of the molecule. (b) An example reaction: ∇ is acting as a multiple character wildcard enabling the expression $\nabla 011$ to match 01*011. When ∇ occurs in the action expression, it may transpose the matched string into the product, see Table 2 for examples.

Biology	Broadcast language
sequence of amino acids from $\{A, R, N, D, C, E, \ldots\}$	string of symbols from Λ
molecule	broadcast device
molecule with no enzymatic function	null broadcast device
molecule with enzymatic function	active broadcast device
non-coding DNA string	string of ignored symbols from Λ
enzyme's active site	broadcast unit
enzyme molecule	broadcast device
cellular milieu	list of strings from Λ

Table 1. Comparison of biological and broadcast language terminology

with the MCS.bl.

Enzyme	Substrate	Product	Reaction
$*\nabla 1 : \nabla 0$	$1 : 0$	$\emptyset$	elastic collision (no product)
$*\nabla 1 : {}' * \nabla$	$0 : 1$	$*0 : 1$	activation (product is an active broadcast device)
$* {}' * 0\nabla : 0\nabla$	$*0 : 1$	$0 : 1$	inhibition (product is a null broadcast device)
$*\nabla : \nabla$	$*00 : 11$	$*00 : 11$	replication (∇ matches and transposes the string $*00 : 11$ into the product which is identical to the substrate, i.e., $*\nabla : \nabla$ is a copier)
$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$	$*\nabla 0 : \nabla 0$	self-replication (product is identical to both the enzyme and substrate)
$*\nabla 1 : \nabla 10$	$*0 : 1$	$*0 : 10$	concatenation
$*\nabla 1 : \nabla$	$*0 : 1$	$*0 :$	cleavage

Table 2. Example reactions realised with the MCS.bl

The full specification of our broadcast language implementation is available at `http://alife.rince.ie/jd/ALL-06-01/`.

2.2. *Molecular mutation*

We define the different operators which allow molecular variations to occur in the MCS.bl. These variations or "molecular mutations" are regarded as *inheritable* molecular processing "errors" that may occur during molecular reactions. These mutations affect the structure of molecules (the sequence of monomers) and may result in some changes of the molecular functions. DNA and RNA replications are example reactions where such mutations may occur.

In the MCS.bl, we devise molecular mutations as follows. 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 types 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.

2.3. *Multi-level selectional and concurrent model*

We implemented the MCS.bl[b] as a multi-level selectional model, i.e., two distinct levels of selection dynamics exist, one at the molecular level and one at the "cellular level". Multiple reactors or *cells* were introduced where each of them contains a population of molecules.

These cells may be subjected to cellular division, which results in the replacement of the parent cell and creation of two offspring cells. However, the total number of cells is fixed: a cellular division triggers the removal of another cell selected at random. In a similar manner to molecules, cells are competing with each other (to remain in the population) which is regarded as the second level of selection. The latter was introduced as a possible means of improving the evolutionary capability of the MCS.bl which has been previously implemented as a single-level selectional model [6].

The number of molecules (each molecule having a distinct presence) contained in a cell may increase until the cell matches a specified division criterion (for example, when the total number of molecules, or the number of some individual molecular species, reaches a specified threshold). When this criterion is satisfied, division occurs as follows: half of the molecules contained in the cell are selected at random, then these molecules are removed from this cell and are inserted into a newly created offspring cell. This offspring is then inserted into the cellular population. Finally,

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

a cell is picked at random (other than the parent and offspring cells) and removed from the cell population, see Fig. 2.

```
INIT molecular species
WHILE simulation termination criterion not satisfied
   GET two molecules at random
   IF selected molecules can react with each other THEN
      CREATE product molecule
      IF cell division criterion is satisfied THEN
          Divide
          DELETE another cell selected at random
      ENDIF
   ENDIF
ENDWHILE
```

Fig. 2. Pseudocode of the multi-level selectional model algorithm, each single cell/CPU runs this algorithm simultaneously and asynchronously. The Message Passing Interface (MPI) is employed to handle the communications between the different CPU nodes/cells. A simplex topology is utilised to condition the interactions between cells, i.e., the "distance" between any two-cells is equal.

Cellular divisions are stochastic processes during which individual molecules are randomly selected and distributed between the parent and the (new) offspring cell. Thus the number of molecules of each molecular species may vary significantly between the two cells after division. In particular, if the absolute number of molecules of some molecular species is small in the parent, then that species may not be present at all in one or other of the two cells after division. Such a cell would then present a reaction network whose dynamics may well diverge significantly from the parent cell. We refer to this potential loss of molecular species due to the stochastic assortment of molecules during cell division as *cell-level/cellular mutation*. Such cell-level mutation introduces further perturbations, separate from and in addition to molecular mutations, which may dramatically disrupt the catalytic activities of reaction networks. In contrast with molecular mutations, cellular mutations have no direct analog in modern organisms. We here refer to cell-level inheritable errors that may occur in some hypothetical protocells during the cellular division processes.

Furthermore this multi-level model was implemented as a concurrent system where each cell is run on a single CPU. In this concurrent model, cell-level fitness (cell reproduction rate) is the *reciprocal of gestation time*. This gestation time or reproduction rate is dependent on the real-time rate of catalytic reactions occurring in the cell, and on the specific criteria in effect for cell division. Note that, in general, the real-time required for individual molecular interactions varies with the specific detailed structures of the molecules involved.

3. Experiments

We present two experiments in which we evolve self-maintaining reaction networks to perform pre-specified information processing functions. The aim of these exper-

iments is to investigate the evolution of complexity in self-maintaining reaction network through the emergence and evolution of computational capabilities.

3.1. *Evolving a self-maintaining reaction network for signal amplification*

We describe an experiment in which collectively self-maintaining reaction networks (in which individual autocatalysis, or self-catalysed replication reactions, are explicitly disallowed) are evolved to carry out a simple target information processing task: signal amplification, i.e., the reaction networks are evolved to promote the growth of a target molecular species. This is motivated by conceptually similar *in vivo* investigations reported in the literature, where experiments were conducted to maximise the production of a target molecular species (lactic acid) using the *E.coli* bacterium [10].

To drive the evolution of self-maintaining reaction networks towards the achievement of a pre-specified task, we modify the conditions triggering the cellular divisions. The latter indirectly or *implicitly* affects the fitness of the cellular species. In contrast to fitness functions *explicitly* imposed in top-down evolutionary approaches[c], implicit fitness functions are not externally defined and do not stipulate and impose the differential fitness (birth/death-rate) of the candidate species. Here, the difference in fitness arises intrinsically from differences in the internal dynamics/behavior of the cells. 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. Note that an early discussion of such an evolutionary approach addressing intrinsic fitness in ACs was proposed by Adami in 1995 [1].

Here, cells divide when a specific target molecular species, denoted by s_T, reaches n_{target} instances. The cellular reproduction rate therefore depends on the molecular growth rate of s_T. The ability of the cell's reaction network to promote the growth of s_T while preserving overall collective self-maintenance of all molecular species in the network now determines the cell's fitness. Ongoing reproduction (fitness) therefore relies on continuing maintenance of the network with respect to whatever molecular species are necessary to continue production of the target species s_T. The pre-specified task assigned to these reaction networks is therefore to amplify the "signal" s_T. As a motivation for this particular mechanism, we may also interpret s_T as a necessary molecular species to allow the cellular division process to occur, e.g., a membrane-related species.

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. An evolutionary experiment is presented in which c_0 (Fig. 3)

[c]An example top-down evolutionary approach is the genetic algorithm where a fitness function is externally devised and employed to evaluate the fitness of candidate solutions. According to these fitness evaluations, candidate solutions are selected for reproduction or removed from the population. Moreover such fitness functions and their specifications, which depend on the coding/representation of the problem, ultimately determine the effectiveness of the algorithm.

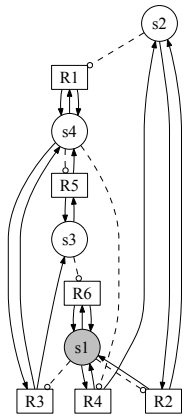


Fig. 3. Bipartite reaction network graph of the collectively self-maintaining network of cellular species c_0. This network was hand designed in a minimalist manner with regard to the complexity at both the molecular (i.e., using simplest/shortest molecular species) and network level (i.e., involving the least number of both molecular species and reactions). The molecular species are listed in Table 3.

is employed as the reaction network of the seed cellular species and is evolved to preferentially promote the growth of the target molecular species $s_T \equiv s_1$. A fixed total population of 31 cells is utilised and executed in parallel. All simulations are run for a pre-defined amount of time $t_{max} = 3600$ (seconds in real wallclock time) using 31 AMD Opteron 270 (2.0 GHZ) CPUs. The maximum molecular carrying capacity of any single cell is $n_{max} = 1.0 \times 10^6$. The target molecular species division threshold is set to $n_{target} = 200$. Finally the per-symbol (per-monomer) mutation probability in each molecular reaction is set to $p_{sym} = 1.0 \times 10^{-5}$.

Fig. 4 presents a typical run in which several successive dominant cellular species successfully preserved collective molecular self-maintenance. Moreover 1235 different and unique cellular species were generated in total during this run, due to molecular and cellular mutations.[d] It was also noted that both molecular and cellular mutations did regularly give rise to non-self-maintaining reaction networks. Nevertheless, we observed that any molecular species lying outside the collectively self-maintaining set of the reaction networks were progressively diluted through cell divisions and eventually lost. Thus, the only (sub-)reaction networks that had long term evolutionary influence, in this experiment, are those that are self-maintaining.

Further repetitions of this experiment were conducted: these presented equivalent dynamics and involved similar emerging cellular species.

We observe that three displacements occurred in the cell population during this

[d]Cellular species are here classified simply by the sets of distinct molecular species present in each cell. Such sets are not necessarily self-maintaining. If the set contained within a cell is not self-maintaining then such a cell will necessarily have a finite maximum number of progeny, and that cellular species will ultimately be displaced.

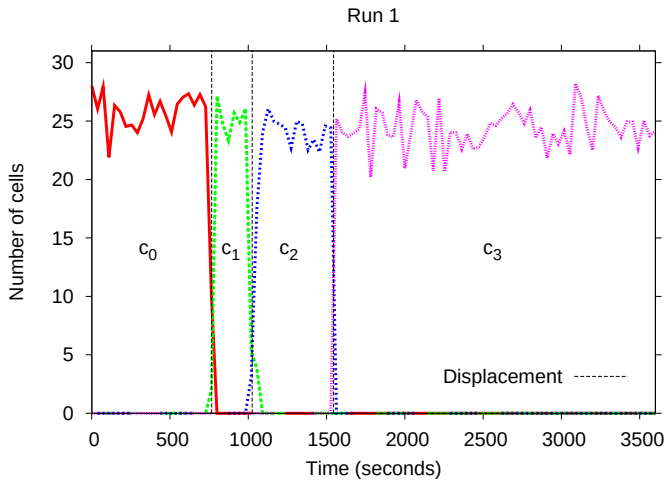


Fig. 4. Dynamics of dominant cellular species in simulation run 1. A spline function was employed to approximate the different curves. The reaction network contained in c_1 is depicted in Fig. 6. The dotted vertical lines indicate the recognizable displacement events.

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 3. Molecular species contained in successive dominant collectively self-maintaining reaction networks denoted by c_1, c_2 and c_3.

run. We note the relative *abruptness* (over the timescale of the complete experiment) of the successive transitions which is suggestive of selectional displacement. However, we should also consider that these displacements may be due to neutral drift dynamics. To address this hypothesis, let us consider the Moran process [19], a well known analytical model which can be employed to study drift and selection dynamics in finite populations. The Moran process model is employed here as it best reflects the MCS.bl system.

Using the Moran process[e] approach, it is possible to predict the expected fixation time of a species (i.e., when an emerging species eventually dominates the population) under neutral drift. Thus, according to the duration of the displacement

[e]The Moran process is a stochastic process describing the dynamics of finite populations in which two reproducing species compete with each other. When a reproduction occurs, an individual is randomly selected and removed from the population, keeping the population size fixed. Neutral drift occurs when a neutral mutation (i.e., not affecting the fitness of a species) manages to spread out throughout the population, replacing the original species. This displacement phenomenon may thus occur due to randomness only.

events, we may judge whether the displacements should be interpreted as primarily due to drift or selection, where the latter is expected to occur at a faster pace than neutral drift. The average expected fixation time T_{c_i} (in number of reproduction events) of a species c_i is given by Eq. 1 (detailed in [24]):

$$T_{c_i} = -N^2(ln1/p - 1)(ln(1-p)) \tag{1}$$

Where N is the fixed population size, c_i^n is the initial number of c_i instances and $p = c_i^n/N$. When $N = 31$ and $c_i^n = 1$, we obtain an average number of reproduction events of 945. Complementary numerical analysis (not detailed here) indicated a standard deviation of 506 and an overall spread of approximately 2 standards deviations.

The displacements between c_0, c_1, c_2 and c_3 took between 4 and 9 seconds to occur with an average aggregate number of 350 reproductions per second (i.e., between 1400 and 3150 reproductions occurred during the distinct displacements). These results are compatible with the expected fixation time under neutral drift obtained with Eq. 1. As a result, if we only consider the durations of each displacement event, these measurements provide no strong basis for discriminating whether any of the displacements are due to drift or selection.

We now examine and analyse the successive dominant cellular species, and their contained reaction networks, to identify whether there exist any significant differences in fitness between them. Table. 4 compares the gestation times of the different cellular species, i.e., the time (seconds in real wallclock time) required for a newborn cell of a given cellular species to reach the cellular division threshold.

	Cellular species			
Gestation time (seconds)	c_0	c_1	c_2	c_3
Mean	3.94×10^{-2}	3.65×10^{-2}	3.52×10^{-2}	3.65×10^{-2}
Standard deviation	1.02×10^{-2}	0.75×10^{-2}	0.63×10^{-2}	0.74×10^{-2}

Table 4. Gestation time of the dominant cellular species. The measurements were obtained as follows. For each cellular species, 150 distinct "newborn" cells having differing initial molecular amount distributions were picked at random from the experimental run (thus, the statistical variations in the initial amounts of the molecular species, resulting from random assortment at cell division, are accounted for). Each of these newborn cells was then incubated, with no molecular mutation occurring, until the cellular division threshold was satisfied. This was repeated 30 times using different random number seeds for each randomly picked newborn cell. Overall, 4500 independent cellular "incubations" were conducted for each cellular species. The gestation time (seconds in real wallclock time) is averaged over the 4500 distinct incubations. Finally, note that these experiments were conducted on a different hardware and may not directly reflect the gestation times reported in the evolutionary experiments, nevertheless this does not affect the relative performances of the cellular species for our comparison purposes.

In Table 4, we note that the average gestation time of c_1 cells decreased to 3.65×10^{-2} seconds from 3.94×10^{-2} seconds for c_0. In other words c_1 possesses a

faster reproduction rate than c_0. This led c_1 to gain a *selective advantage* over c_0. The emergence of the molecular species s_5 and s_6 in c_1 had the effect of promoting the growth of s_1 whilst still preserving collective network self-maintenance. The network properties evolved and permitted the network to preferentially promote the growth of molecular species s_1. Moreover, c_1 presents a higher level of complexity, in some reasonable sense, in that it involves three additional reactions compared to c_0.

The cellular species c_1, c_2 and c_3 are now examined. Although the molecular species contained in these cellular species are different from a genotypic point of view, their phenotypes seem similar as suggested by their very comparable gestation times (Table 4). The symbols $\Diamond$ and $\triangle$ act in the same manner (as single character wildcards) when occurring in the condition statement, see Section 2.1. The symbols $\triangle$ occurring in the action statement of c_2 are *ignored* (Section 2.1). 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 cell-level fitness.

We may thus affirm that the first displacement between c_0 and c_1 was primarily selective whereas the other displacements were the result of neutral drift (with c_1, c_2 and c_3 being more or less equally fit, and all three fitter than c_0).

Finally, we note in Fig. 5 that these displacements are associated with a relative increase in the cellular species diversity during the displacement events themselves. The observed displacements are thus more intricate than straight competition between two "pure" cell lines (as in the simplistic Moran process involving two species only). Indeed, when examining the composition of the additional mutant cellular species, we noticed that these mutant cells were actually in the near "mutational neighborhood" of c_1. The emergence of these cellular mutants precisely initiated at $t = 760$ with the appearance of a new molecular species which, through interactions with the other molecules, generated a set of additional molecular species. Then, through successive cellular mutations, a set (which includes c_1) of cellular mutants rapidly appeared (as observed in the increase of the cellular species diversity, Fig. 5). These mutants eventually fully displaced the c_0 ancestors at $t = 768$. The displacement event completed with c_1 displacing related cellular mutants. This phenomenon was also observed in the second and third displacement events.

Finally with regard to the evolutionary growth of complexity, we note that in this evolutionary experiment, the number of molecular reactions was increased from 6 to 9 when comparing the reaction networks in the seed and evolved cellular species. On the other hand, the complexity of the individual molecular species remains comparable, with an average string length of 6. Moreover both the seed and evolved reaction networks contain the same number of distinct molecular species. In summary, although the seed reaction network was successfully evolved and improved (somewhat) in relation to the pre-specified task, only a very limited growth of

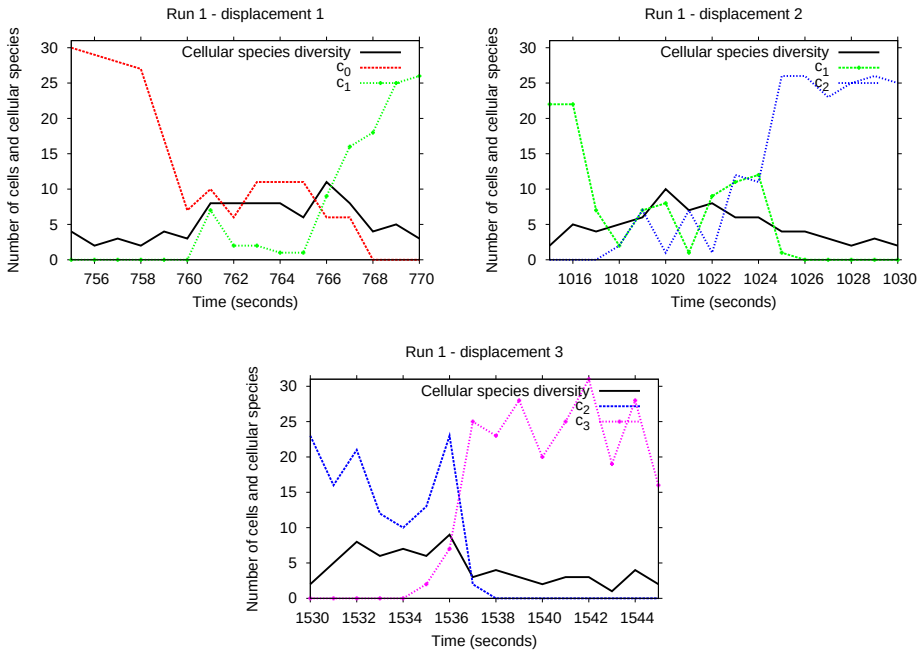


Fig. 5. Successive displacements that occurred in simulation run 1. The cellular diversity indicates the number of distinct cellular species (discriminated by their reaction networks) occurring at a given timestep.

complexity was observed.

Ten additional repetitions of the above experiment were conducted. 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, there would be further drift in the mutational space of the cellular species, such that the reaction networks could lose these extra molecular species and collapse again to c_1 or c_3. In the remaining two runs, the emergence of c_0 mutants with no phenotypic differences was observed.

3.2. *Evolving self-maintaining reaction networks with crosstalk*

The previous experiment demonstrated only very limited evolutionary growth of complexity. In this section, we extend this preliminary work on the evolution of self-maintaining reaction networks, with the explicit aim of evolving networks of higher complexity. To assist this research on the evolutionary growth of complexity, we examine a phenomenon occurring in real biochemical information processing 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 previous work, we demonstrated that crosstalk was a key property enabling the *merging* and *cooperation* of distinct self-maintaining reaction networks when mixed together [7]. It was found out that if no crosstalking molecular species exist between two distinct self-maintaining networks mixed into a common reaction space (with limited size), then, both networks would compete against each other. Such one-versus-one competitions ultimately resulted in the displacement of one of the networks.

We now present a further investigation suggesting the evolutionary growth of complexity through crosstalk. This work is naturally related to the symbiogenesis theory which was originally postulated by Mereschkowsky [18], and already explored computationally by Barricelli, on the first stored program digital computers, in the 1950's [3, 4]. According to this theory, separate (proto-)organisms may merge with each other to form new organisms of higher complexity [17] (pp. 3–24).

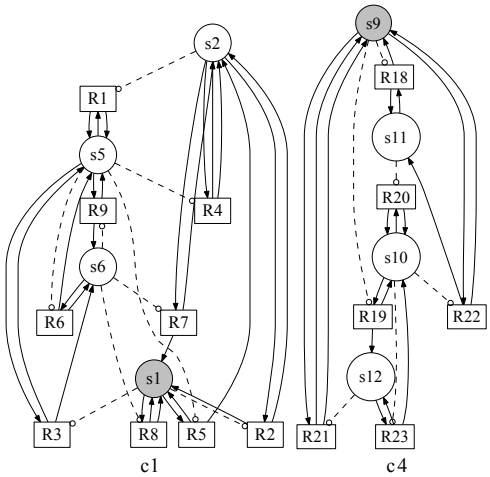


Fig. 6. Bipartite reaction network graphs of the reaction networks contained in cellular species c_1 and c_4. c_1 and c_4 were obtained from separate precursor experiments in which they were evolved to promote the production of target molecular species s_1 and s_9 respectively (coloured in grey). The evolution of c_1 was already presented in Section 3.1 above; c_4 derives from a similar, but separate experiment series, which is not reported here in detail. The reaction networks of c_1 and c_4 are classified as (potentially) *crosstalking* because the molecular species contained in the two networks may react with others; specifically, s_1, s_2, s_5 and s_6 can all bind to s_9 and s_{12}, whereas s_{10} and s_{12} can both bind to s_2 and s_5. Thus, if the two networks are introduced into the same reaction space, that will immediately introduce additional reactions which cross-connect between these previously isolated graphs. These new reactions may, in turn, give rise to a cascade of further changes, affecting both the molecular species present and the composite reaction network which they ultimately give rise to.

The collectively self-maintaining, and potentially crosstalking, reaction networks of cellular species c_1 and c_4 (Fig. 6) are utilised as the seed networks in this experiment. We applied a new cellular division criterion which accounts for both target molecular species s_1 and s_9. The motivation for this criterion is to require the retention of the evolved functions of *both* cellular species c_1 and c_4. Ultimately we aim at evolving/obtaining a cellular species with a more complex self-maintaining network specifically adapted or optimised in this "multitasking" role, i.e., a network which is better able to carry out the pre-specified tasks of both cellular species c_1 and c_4. Therefore, in this scenario, a cell c_i divides if $(n_1^i \geq 200 \wedge n_9^i \geq 200)$. An evolutionary experiment is conducted using the complementary experimental conditions presented in the previous section.

We present an example of a typical run in which the following distinctive behaviour is observed. We observe an 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 (Fig. 7). This early dynamic is driven by the cellular species c_5 (Fig. 9). c_5 is characterised by a "meta-reaction network" containing the complete reaction networks of both c_1 and c_4, in addition to four new molecular species, which was able to self-maintain for a period of time. During this run as a whole, an overall total of 37863 unique reaction networks appeared.

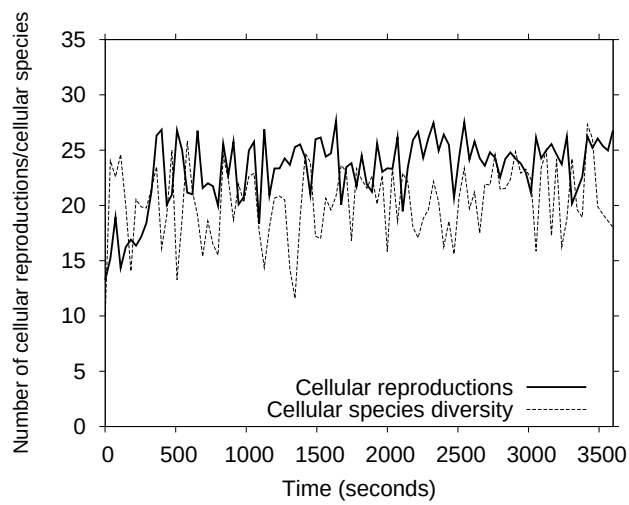


Fig. 7. Crosstalking networks with molecular mutations. Dynamics of cellular reproduction rate and diversity with molecular mutations occurring. A spline function was employed to approximate the cellular reproductions and cellular species diversity curves.

In Fig. 8, we note that two cellular species displacements are recognizable at $t \approx 475$ and $t \approx 2500$. The cellular species c_5 is displaced by a mutant cellular species, denoted by c_6. The third emergent dominant cellular species is denoted by c_7. These cellular lines are associated with CIPNs of higher complexity, which are capable of

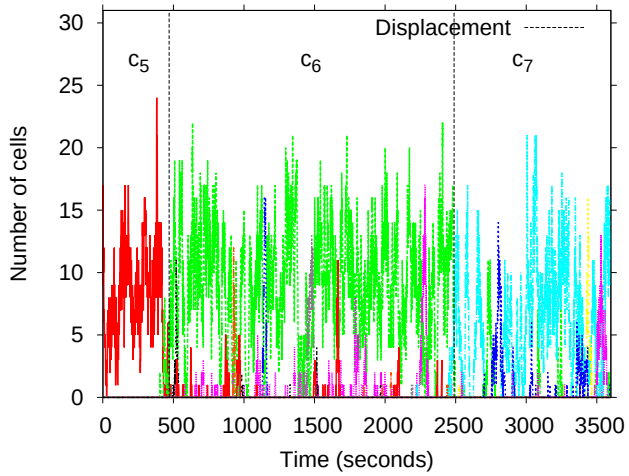


Fig. 8. Dynamics of the major cellular species. Only the cellular species which reached a proportion of at least one third of the cellular population, at least once during the simulation run, are shown (14 cellular species in total). The dotted vertical lines indicate the recognizable displacement events.

producing both target molecular species s_1 and s_9 whilst preserving their collective self-maintenance. These networks can thus perform the functions of the seed cellular species c_1 and c_4 simultaneously and can be regarded as "multitasking". Through exploiting crosstalk properties, we successfully evolved CIPNs of significantly higher complexity (from a topological and functional point of view) which were able to self-maintain and reproduce for a sustained period of time.

In contrast with the previous experiment, it can be observed that the largest cellular species' subpopulation rarely exceeded half of the total population. The dominating cellular species have not once succeeded at *fully* displacing the other species. This phenomenon suggests that more complex dynamics are occurring at the cellular population level.

	Cellular species		
Gestation time (seconds)	c_5	c_6	c_7
Mean	1.03	0.80	0.74
Standard deviation	0.57	0.43	0.32
Number of molecular species	12	9	12
Number of reactions	55	32	66

Table 5. Gestation times of the dominant cellular species. The measurements were obtained using the methodology previously described in Table 4.

In Table 5, we note that the gestation time of the dominant cellular species suc-

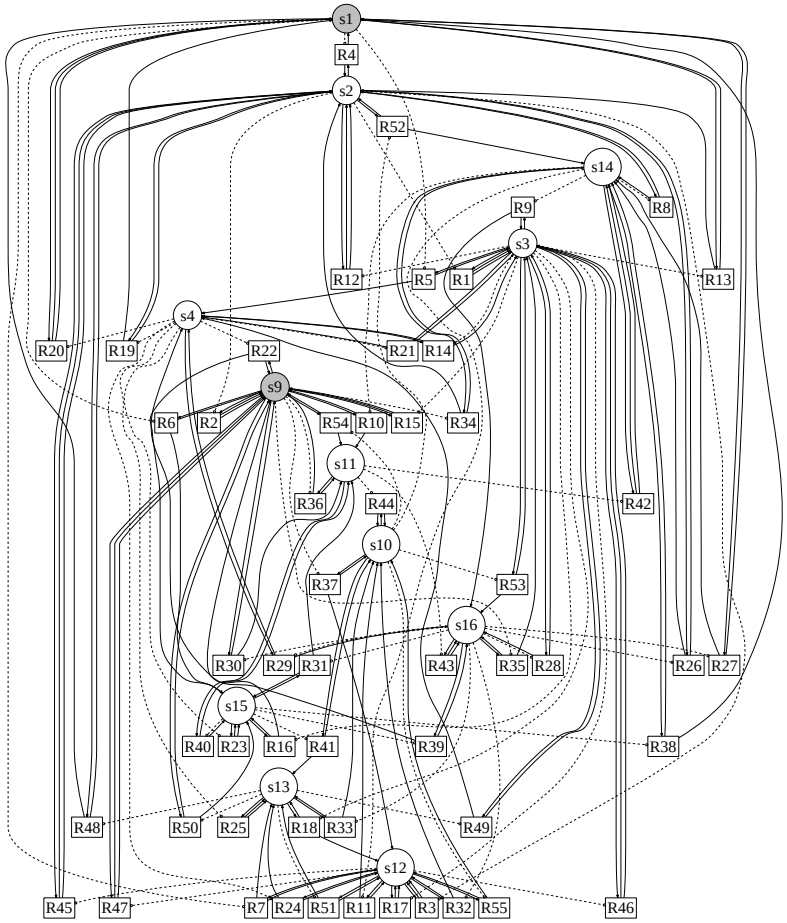


Fig. 9. Reaction network of cellular species c_5 which contains all molecular species from networks c_1 and c_4 in addition to new molecular species s_{13}, s_{14}, s_{15} and s_{16}.

cessively decreased. This suggests that the evolutionary process has successfully led to the emergence of fitter cellular species with improving reproduction rates. When examining the molecular species composing the self-maintaining reaction network of cellular species c_6, it was observed that this was a subset of c_5. In other words, three molecular species were removed from c_5 to constitute c_6. This "simplification" effectively hastened the reproduction rate of c_6 leading to the first selective displacement event. It may be thus be argued that these (removed) molecular species and associated reactions were penalising the production of target molecular species.

In contrast with c_6, three novel molecular species emerged in cellular species c_7 which significantly increased the number of reactions to 66. However, these additional molecular species and reactions (i.e., "complexification") resulted in promoting the production of the target molecular species, as indicated by the improved

reproduction rate of c_7.

In the next section, we discuss the achievements of the above experiments and potential limitations of the MCS.bl to study the evolutionary growth of complexity in CIPNs.

4. Discussion

Our evolutionary experiments demonstrated the feasibility of evolving self-maintaining CIPNs to achieve pre-specified information processing tasks. Through this evolutionary process, we observed a (limited) relative growth of complexity in CIPNs. More particularly, the crosstalk based experiments suggested an interesting avenue of further 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 [20] or Avida [2], a significant difference exists in terms of the evolutionary growth of complexity. Where Tierra exhibited numerous complex emerging phenomena (e.g., emergence of ecologies, exhibition of punctuated equilibrium dynamics [13], pp. 745–1025), our MCS.bl based experiments hardly presented any comparable evolutionary dynamics at either the molecular or cell population/system levels. This suggests that the MCS.bl may not be suitable to provide a *robust* method to support evolvability of complex systems. A candidate explanation, relating to Szathmáry and Demeter's work based on the stochastic corrector model [23], is here considered.

In the experiments reported earlier, we noted that in the different evolved networks, the number of molecular species 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 in the cell population became more variable.

We propose that this increased diversity of molecular species may have been implicated in the different dynamics described here. Indeed, Szathmáry and Demeter 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., some offspring 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 paper, then dealing with more molecular species and instances (i.e., more complex information) would increasingly become more difficult using the MCS.bl, pre-empting any significant evolutionary growth of complexity. This would *ultimately* suggest the limitations of the MCS.bl to encode and process more complex information using self-maintaining 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 naturally relate to a major evolutionary transition as proposed by Smith and Szathmáry [21].

In the next section we propose some future research directions to further investigate the evolution of complexity in self-maintaining CIPNs using the MCS.bl.

5. Future work

We propose a number of system modifications that could lead to the emergence and evolution of CIPNs of higher complexity. These proposed research directions aim at complementing the understanding of the evolutionary growth of complexity in CIPNs using the MCS.bl, without (yet) having to introduce a complementary genetic subsystem.

(1) *Cellular division criteria:* In the above experimental sections, 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 the experiments presented in this paper, 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 cellular 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.

(2) *Detectors and effectors:* Following Holland's classifier systems/agent-based approach proposed in [16], 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 [12].

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 CIPNs. 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 CIPNs.

Therefore we suggest that minimalist approaches, where the system is still analytically tractable and examined using available mathematical methods, should continue to be explored.

6. Conclusion

To address the grand-challenge problem of achieving the open-ended evolutionary growth of complexity in artificial, computational, systems, we investigated the significance of collective self-maintenance or auto-catalysis to the evolution of complexity in information processing reaction networks. We first presented our hypothesis where CIPNs are regarded as self-maintaining sets which may grow in complexity to achieve computational functions through evolution. To assist this research, we built an agent-based multi-level selectional Artificial Chemistry called MCS.bl. Using this system, we conducted a series of evolutionary experiments. To drive the evolution of the CIPNs, we devised novel cellular division criteria to encourage the emergence of relatively more complex "computational" signalling pathways from a bottom-up perspective. This method was applied to both the evolution of a single and multiple/crosstalking self-maintaining reaction networks. In these experiments, the networks were successfully evolved to achieve the pre-specified in-

formation processing functions more effectively and exhibited a relatively higher level of complexity (by at least some reasonable measures). By evolving CIPNs to achieve computational capabilities, we successfully demonstrated the growth of complexity in self-maintaining reaction networks. Nevertheless, we also discussed some potential limitations of the MCS.bl which may have prevented a more significant evolutionary growth of complexity as observed in ACs such as Tierra or Avida. This proof of concept should contribute, to some extent, to understanding of the open-ended evolutionary growth of complexity using ACs.

References

[1] Adami, C., Learning and Complexity in Genetic Auto-adaptive Systems, *Physica D: Nonlinear Phenomena* **80** (1995) 154–170.

[2] Adami, C. and Brown, C., 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* (MIT Press, 1994), pp. 377–381.

[3] Barricelli, N., Symbiogenetic Evolution Processes Realized by Artificial Methods, *Methodos* **9** (1957) 143–182.

[4] Barricelli, N., Numerical testing of evolution theories, *Acta Biotheoretica* **16** (1963) 99–126.

[5] Bray, D., Protein Molecules as Computational Elements in Living Cells, *Nature* **376** (1995) 307–312.

[6] Decraene, J., Mitchell, G., and McMullin, B., Unexpected Evolutionary Dynamics in a String-based Artificial Chemistry, in *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems* (MIT Press, Cambridge, MA, 2008), pp. 158–165.

[7] Decraene, J., Mitchell, G. G., and McMullin, B., Crosstalk and the Cooperation of Collectively Autocatalytic Reaction Networks, in *2009 IEEE Congress on Evolutionary Computation*, ed. Tyrrell, A., IEEE Computational Intelligence Society (IEEE Press, Trondheim, Norway, 2009), pp. 2249–2256.

[8] Dittrich, P. and Speroni, P., Chemical Organisation Theory, *Bulletin of Mathematical Biology* **69** (2007) 1199–1231.

[9] Farmer, J., Kauffman, S., and Packard, N., Autocatalytic Replication of Polymers, *Physica D* **22** (1986) 50–67.

[10] Fong, S., Burgard, A., Herring, C., Knight, E., Blattner, F., Maranas, C., and Palsson, B., In Silico Design and Adaptive Evolution of Escherichia coli for Production of Lactic Acid, *Biotechnology and bioengineering* **91** (2005) 643–648.

[11] Fontana, W. and Buss, L., The Arrival of the Fittest: Toward a Theory of Biological Organization, *Bulletin of Mathematical Biology* **56** (1994) 1–64.

[12] Gershenson, C. and Lenaerts, T., Evolution of Complexity, *Artificial Life* **14** (2008) 241–243.

[13] Gould, S., *The Structure of Evolutionary Theory* (Belknap Press of Harvard University Press; Sixth Printing edition, 2002).

[14] Helmreich, E. J. M., *The Biochemistry of Cell Signalling* (Oxford University Press, USA, 2001).

[15] Holland, J., *Adaptation in Natural and Artificial Systems*, 2nd edn. (MIT Press Cambridge, MA, USA, 1992), see also Holland(1975).

[16] Holland, J., Exploring the Evolution of Complexity in Signaling Networks, *Complexity* **7** (2001) 34–45.

[17] Margulis, L. and Sagan, D., *Acquiring Genomes: A Theory of the Origins of Species* (New York: Basic books, 2002).

[18] Mereschkowsky, C., Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen, *Biol. Centralbl* **30** (1910) 278–288.

[19] Moran, P., *The Statistical Processes of Evolutionary Theory* (Clarendon Press, 1962).

[20] Ray, T., Evolution, Ecology and Optimization of Digital Organisms, *Santa Fe Institute working paper 92-08-042.* (1992).

[21] Smith, J. and Szathmáry, E., *The Major Transitions in Evolution* (Oxford University Press, USA, 1997).

[22] Stock, J., Surette, M., McCleary, W., and Stock, A., Signal Transduction in Bacterial Chemotaxis, *Journal of Biological Chemistry* **267** (1992) 19753–19756.

[23] Szathmary, E. and Demeter, L., Group Selection of Early Replicators and the Origin of Life, *Journal of Theoretical Biology* **128** (1987) 463–486.

[24] Traulsen, A., Pacheco, J., and Imhof, L., Stochasticity and Evolutionary Stability, *Physical Review E* **74** (2006) 021905–1–021905–6.