Evolvable Self-Reproducing Cells in a Two-Dimensional Artificial Chemistry

Tim J. Hutton

Department of Computer Science, University College London, UK

tim.hutton@gmail.com

January 10, 2007

Abstract

We present a novel unit of evolution: a self-reproducing cell in a two-dimensional artificial chemistry. The cells have a strip of genetic material that is used to produce enzymes, each catalysing a specific reaction that may affect the survival of the cell. The enzymes are kept inside the cell by a loop of membrane, thus ensuring that only the cell that produced them gets their benefit. A set of reaction rules, each simple and local, allows the cells to copy their genetic information and physically divide. The evolutionary possibilities of the cells are explored, and it is suggested that the system provides a useful framework for testing hypotheses about self-driven evolution.

Keywords: artificial chemistry, evolvability, self-replication, cellular automata, mitosis

1 Introduction

The creation of an artificial evolutionary system that displays the open-ended appearance of adaptive innovations is a central goal of artificial life research [8, 7, 34, 29, 53, 55]. The evolutionary process is often seen as creative [9, 54] in the sense that is has produced many sophisticated organs and behaviours that surprise us with their ingenuity, and yet this self-driven process has never been captured in a computer model.

The issue is perhaps best summed up by John Maynard Smith:

> "What features must be present in a system if it is to lead to indefinitely continuing evolutionary change?" [32] (quoted in [54])

Arguably the most progress towards achieving this goal has been in the Avida system [1, 3]. The speed at which self-replicating computer programs can operate means that many generations occur in a short amount

of time, allowing the effects of many different mutations to be explored. If a series of increasingly difficult tasks are rewarded, the complexity needed to perform these tasks soon evolves, giving us an insight into how complex features can arise in nature [25]. However, the same result does not occur without the external fitness function, and thus the evolutionary dynamics [8] are limited. The likely reason for this is that the organisms do not interact in any significant way [54], and thus occupy a fixed fitness landscape, as genetic algorithms do.

Avida has also been used to demonstrate speciation, through the competition for different resources [11]. Here the organisms do interact (albeit in a minor way), by altering the relative levels of the shared resources. Again, however, the evolutionary dynamics are limited because each species' ecological niche becomes fixed and there is no further scope to innovate. It seems that the very design of the Avida system limits evolution by restricting the interactions between organisms, and it is not obvious how the system could be changed to accommodate this.

Taylor [54] suggested a set of features that an evolutionary system should have if we are to expect it to ever demonstrate the open-ended evolutionary growth of complexity. The central issue is the need for rich interactions between the organisms, as mentioned above, and to achieve this it is suggested that the organisms should be fully embedded in the arena of competition and made entirely of material (i.e. conserved) components. This would permit each organism to become a resource for the others, and thus cause interactions between the organisms not only via their shared environment but also directly. Implementing such a virtual world seems to require a different approach to the Avida system, and this is the challenge we are addressing in this paper. The key is that self-replication can be implemented in many different types of system [48].

One class of system that does have rich interactions and embeddedness is cellular automata (CA). Self-replicating structures in CA range from von Neumann's marvellous but still impractical universal constructor [57, 44] through to Langton's Loops [24] and Evoloops [47, 46]. However, CA cannot typically be said to be material systems, since the state of each cell can usually be changed without cost. In the game of life for example [16], a small starting configuration such as the r-pentomino will proliferate into many 'live' cells by creating 'matter' out of nothing.

Artificial chemistries (AChems), by contrast, often have all three properties of rich interactions, embeddedness and materiality, and thus are a promising medium in which to implement an evolutionary system. There are many types of AChem (see [13] for a review), with systems ranging from those that are 'well-stirred' and have no concept of location, to those in which molecules move in a very similar way to molecules in nature. In this paper we will present a two-dimensional AChem system containing genetic operations and a cell membrane, and explore its evolutionary potential.

It should be emphasised that we are not trying to accurately model chemical processes (as [42]), or to

understand what might be needed to implement life in wet chemistry [37, 52, 45], or to understand the chemical origins of life on Earth (see eg. [26]), or even to implement some aspect of what is thought to constitute a living system (eg. autopoeisis [36]), although all of these issues are somewhat related. The similarity of our system to natural cells reflects the similarity of the design constraints of working with molecule-like objects whether in a spatial artificial chemistry or real chemistry.

A similar fretful caution should be made on our use of biological terminology: we use words like membrane, genome, enzyme, etc. because these are convenient labels for the artificial structures we are considering. We could quarantine these terms inside quotation marks but this seems overly fussy, as long as it is borne in mind that any conclusions drawn from experiments in the medium of artificial chemistry should not unthinkingly be assumed to be true of the same structures in nature. On the other hand, evolution as a process can be implemented equally well in different media [12] and this is precisely what we are trying to achieve.

2 Background

Every lifeform on Earth uses DNA molecules to make enzymes that direct the reactions that allow the organism to metabolise and reproduce. Mutations in the genome allow the proliferated population of organisms to adapt to new conditions, and this process has resulted in the richness and sophistication of the biological life we see around us. While we cannot yet simulate the molecular interactions of DNA and its associated machinery directly [5], we can simplify the details of its operation and make a model that hopefully captures the important aspects. Previous AChem models that have followed this approach (eg. [17, 6]) have neglected the physical aspect of the simulation, on the basis that it is the genetic information that is most important. Here we choose to provide molecules with various different properties, and place them in a two-dimensional world with physical interactions, on the assumption that it is the interactions with a complex physical environment that initially drives the growth of complexity.

A simple set of reaction rules for template-directed replication of molecules was presented in [18]. The molecules obtained the material needed to duplicate themselves from a surrounding soup of free atoms, thus leading to competition for these resources. In the absence of any phenotypic effect of the bases on each template molecule, the selection pressure favoured shorter molecules and no other evolutionary activity was observed.

In [19] this system was extended to allow enzymes to be produced, by decoding the sequence of bases. Each enzyme would catalyse a specific reaction, and thus the genomes now had the capacity to affect their environment and potentially give themselves a survival advantage. However, the fact that the population of molecules were mingling with each other freely meant that once an enzyme was produced, it was not only the producer that got any benefit from it but also any molecule that happened to be nearby. Thus

the producers are vulnerable to parasites: shorter molecules that replicate faster and take advantage of the presence of enzymes. In the system described, this would typically lead to a global extinction event, where the parasites would out-compete the enzyme-producers but then find themselves unable to replicate.

The work presented here is an extension of [19] where we also provide a cell membrane, to allow each genome to retain sole use of any enzyme it produces, while still permitting resources to enter. The difficulty of living inside a membrane is that reproduction is more complicated, as the membrane must also be at least partitioned if not divided fully. One solution to this issue is presented here.

Membranes

The membranes of biological cells consist mainly of lipid molecules that are repelled by water molecules in such a way that they naturally form sheets and spheres. Small molecules are able to diffuse through such a membrane, and in real cells this process can be controlled by embedded transport proteins. For the purpose of containing self-replicating molecules these membranes have several relevant properties, including the ability to self-assemble and self-repair, to allow passage of small molecules but not large ones, and to allow different proteins to become embedded in the membrane.

To incorporate the concept of a membrane with these properties into our artificial chemistry system, we need to find a way to model them usefully at minimal computational cost. Membranes can be simulated accurately using molecular dynamics (MD) techniques (see for example, [15]) but this is so costly in terms of computing power that an entire vesicle (membrane forming a spherical shell) has not yet been simulated.

Thus simplifications of membrane dynamics have been explored (see [22] for a review). These range from two-dimensional models of molecular interactions [14, 38] through to lattice models [6, 36, 39, 30, 31, 40, 28] and abstract models [51, 43]. Figure 1 shows one such model, from chapter 3 of [38].

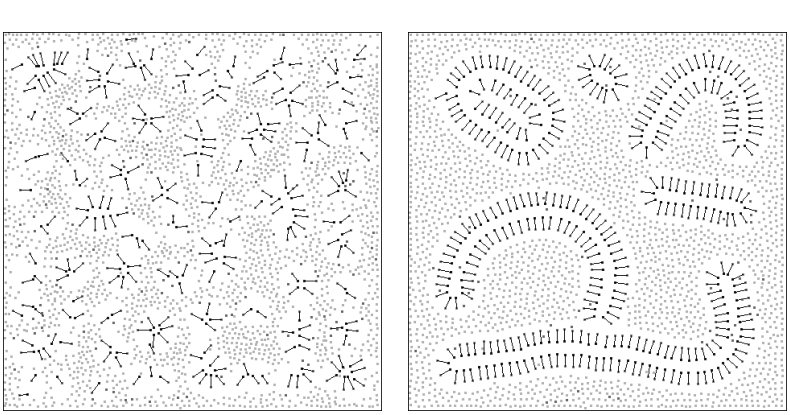


Figure 1: A simulation of the self-assembly (left to right) of phospholipids into bilayer membranes when mixed with water. The repulsion strength is greatest between water molecules and the hydrophobic ends of the amphipathic lipid molecules, thus the lowest energy configuration for the system has the lipids neatly arranged in two-sided layers. Adapted from [38], author's implementation.

If a two-dimensional AChem is used and has suitable rules for its physics, a very simple form of membrane can be made by linking units into a closed chain. In the lattice physics used in [18], where any two 'atoms' that are bonded must remain within each other's Moore neighbourhood, such a loop allows unbonded atoms to pass through but nothing larger. This form of membrane is very efficient, computationally, and is fully compatible with the self-replicating molecules that we wish to enclose. Additionally, this form of membrane permits different atoms to become embedded, allowing the membrane to be pulled about and otherwise manipulated, a necessary constraint for the type of cell division we will present. While such a membrane does not possess the capacity for self-assembly or self-repair, these are losses that can be disregarded for the moment.

In the following section we give details of a system that uses this type of membrane to make simple self-reproducing cells, and illustrate how the process works on an example. In later sections we show how the system can be implemented efficiently and run several experiments on the reproducing cells.

3 System Description

Chemistry

Following [18], the artificial chemistry consists of numerous 'atoms' that move around in a two-dimensional environment. Atoms may be bonded with others to form molecules of arbitrary size. Any two atoms that are bonded must remain near to each other but all atoms are otherwise free to move locally. The type of movement is specified by the physics of the system, for which several possibilities are given in the next section.

The 'atoms' in our AChem have a fixed *type* $\in \{\texttt{a}, \texttt{b}, \texttt{c}, \texttt{d}, \texttt{e}, \texttt{f}\}$ and a variable *state* $\in \{\texttt{0}, \texttt{1}, \texttt{2}\ldots\}$. Reactions can occur when two atoms collide, these can change the states of the atoms involved and may create or break bonds between them. For example, the reaction $\texttt{a0} + \texttt{a0} \rightarrow \texttt{a1a1}$ will cause two atoms with type `a` and state `0` to become bonded with each other and adopt state `1`. Even with such simple rules for chemical interactions, sophisticated behaviour such as template replication can be produced [18]. The exact details of what constitutes a collision are specified by the rules of physics that are chosen, although certain constraints such as that the world is two-dimensional and crossed bonds cannot form, are assumed.

In the system being presented, a string of atoms is kept inside a membrane formed from a loop of other atoms to form a simple cell. With a suitable set of reaction rules, such a cell can duplicate its genetic information and cause the membrane to buckle and fission in such a way as to leave two daughter cells. The form of asexual cell division we are modelling here, in which the gene-string is attached to the cell membrane at division, is commonly seen in bacteria and is one of the simpler forms of cell division that is possible.

The reaction rules that direct the cell division are given in Table 1, with an example starting configuration for the cell shown in Fig. 2. The process of division is followed through on this example in the next few figures.

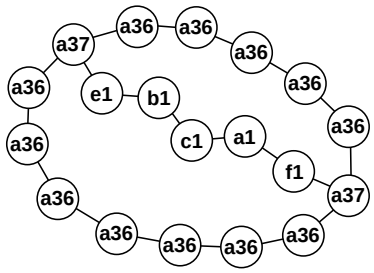


Figure 2: The initial cell configuration, consisting of a linked chain of membrane atoms, plus a string of atoms carrying information, here the string `ebcaf`.

In Fig. 2, the membrane consists of a loop of linked `a36` atoms, with other atoms embedded. The string of atoms that that will form the genetic information each have state `1` and this gene-string is connected at each end to `a37` atoms that are embedded in the membrane. From this figure it can be seen that the 'atoms' in our system approximately correspond to the lipids, proteins and nucleotides of biological cells.

As in [18], the 'start' of the gene-string is marked by an `e` atom. Reactions R1, R6, R2 and R3 are the first to occur, and their effects are illustrated in Fig. 3. A second `e` atom is joined, and attached to the membrane like the first.

R1:	$\texttt{e1a37} \to \texttt{e5a10}$	R18:	$x14y16 \to x27y16$
R2:	$\texttt{a10} + \texttt{e6} \to \texttt{a37e3}$	R19:	$x27\texttt{a11} \to x17 + \texttt{a11}$
R3:	$\texttt{e6e3} \to \texttt{e2e3}$	R20:	$x17y16 \to x17y13$
R4:	$x2y1 \to x7y4$	R21:	$x13\texttt{e8} \to x14\texttt{e15}$
R5:	$x4 + y3 \to x5y7$	R22:	$\texttt{e13a37} \to \texttt{e18a19}$
R6:	$x5 + x0 \to x6x6$	R23:	$\texttt{e13a19} \to \texttt{e18a20}$
R7:	$x6 + y7 \to x3y4$	R24:	$\texttt{a20} + \texttt{a11} \to \texttt{a21a22}$
R8:	$x6y4 \to x1 + y2$	R25:	$\texttt{e18a22} \to \texttt{e32} + \texttt{a23}$
R9:	$x7y1 \to x2y2$	R26:	$\texttt{e18a21} \to \texttt{e32} + \texttt{a24}$
R10:	$\texttt{f2a37} \to \texttt{f9a11}$	R27:	$\texttt{a24} + \texttt{a37} \to \texttt{a26a27}$
R11:	$\texttt{a11} + \texttt{f3} \to \texttt{a11f9}$	R28:	$\texttt{a27a23} \to \texttt{a37} + \texttt{a28}$
R12:	$x2y8 \to x9y1$	R29:	$\texttt{a26a36} \to \texttt{a29a30}$
R13:	$x9y9 \to x8 + y8$	R30:	$\texttt{a29a36} \to \texttt{a31a30}$
R14:	$\texttt{a11a36} \to \texttt{a11a12}$	R31:	$\texttt{a30} + \texttt{a28} \to \texttt{a25a33}$
R15:	$\texttt{f1} + \texttt{a12} \to \texttt{f13a37}$	R32:	$\texttt{a31a25} \to \texttt{a32} + \texttt{a36}$
R16:	$x13y1 \to x14y15$	R33:	$\texttt{a32a30} \to \texttt{a34a36}$
R17:	$\texttt{a11} + x15 \to \texttt{a11}x16$	R34:	$\texttt{a34a33} \to \texttt{a37} + \texttt{a37}$

Table 1: The reaction rules for cellular self-reproduction. Some involve the variables x and y, these stand for any of the six types. For example, R6 says that when two atoms that have states `5` and `0` and the same (but unspecified) type collide, they bond together and adopt state `6`. For clarity, related reactions are grouped together. Reactions are applied asynchronously, in such a way that if two reaction rules can be applied to a given configuration, one will be chosen at random.

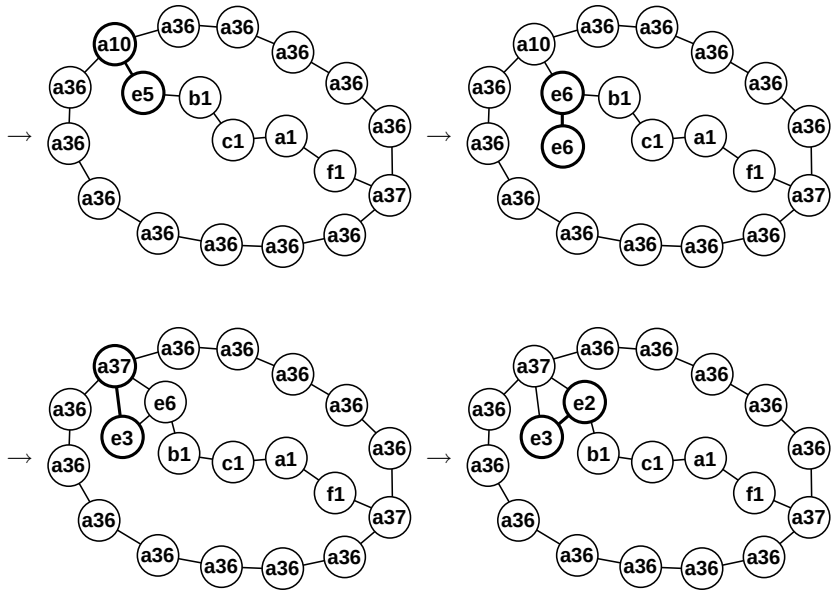


Figure 3: The first stage of cell division, following on from the starting point given in Fig. 2. Firstly, reaction R1 causes a state change at the top of the gene-string, leaving the `e` atom in state `5`. Unbonded atoms are able to diffuse in and out from a surrounding soup of atoms in state `0` and thus an `e0` atom will soon collide with the `e5` atom, allowing reaction R6 to be applied. Reaction R2 causes the new base to be joined to the membrane like the first, and R3 triggers the next stage to proceed when this join has happened. The atoms that have just been involved in a reaction are highlighted in bold.

From this point the main template replication process is triggered, using reactions R4-R9. The process is illustrated in Fig. 4, and works on any string of atoms in state `1`.

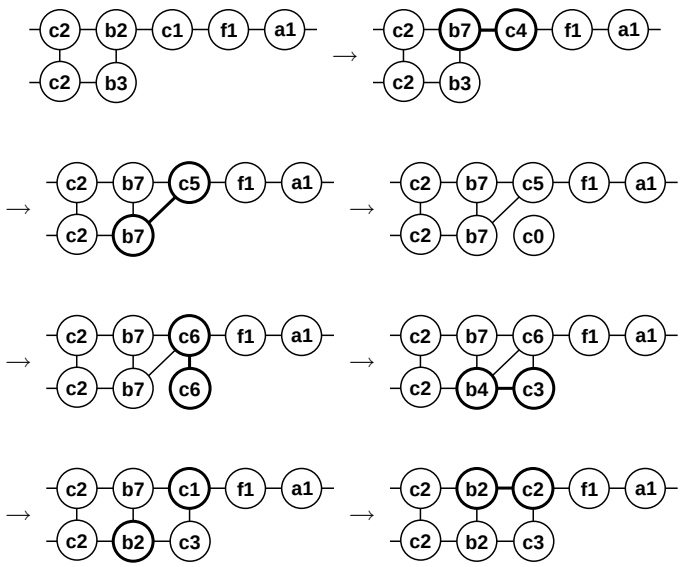


Figure 4: The base-duplication sequence illustrated. Reactions R4-R9 direct this sequence, where atoms from the surrounding soup of atoms in state `0` are bonded to their matching counterparts. Note that the last stage shown here is in a corresponding configuration to the first, allowing the sequence to continue down the molecule. Using this process, a molecule of any length and containing any sequence of bases will be able to duplicate, an important point when we come to consider mutations. Note that the `c0` atom could bond on the other side of the `c5` atom but the sequence would continue in the same way. In the fifth step, the `c6` atom cannot react with the wrong `b7` atom because this would create crossed bonds - this constraint is enforced by the physics used, see next section.

After the duplication sequence has proceeded down the molecule, the cell is left in the state shown in the top-left of Fig. 5. The tail end of the gene-string is then joined to the membrane in a similar way to the head end and the full separation of the cell can begin.

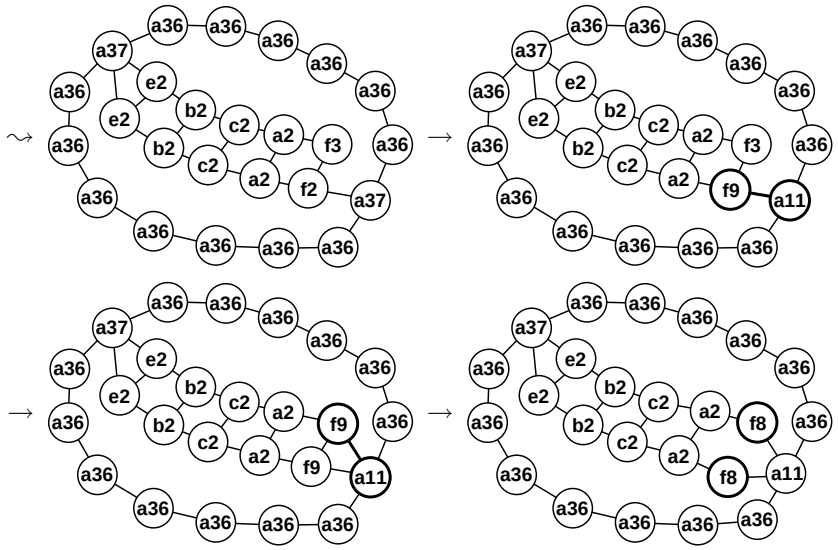


Figure 5: After the duplication of bases, the tail of the completed gene-string (marked with an `f` atom) is attached to the membrane. Reactions R10, R11 and R13 perform the three stages shown here.

Following on from the last stage shown in Fig. 5, a process of unzipping the two copies of the gene-string is triggered. This process is directed by reactions R12 and R13 and is illustrated in Fig. 6.

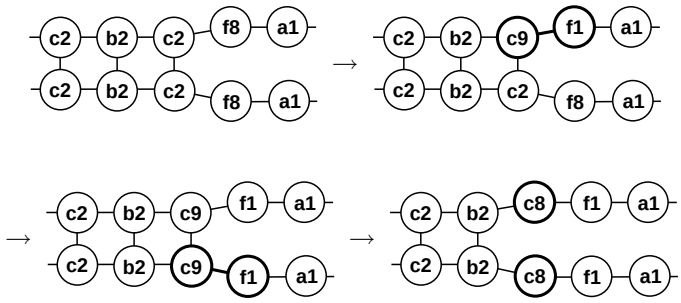


Figure 6: The unzipping sequence, using reactions R12 and R13. Following on from the base-duplication sequence shown in Fig. 4, the two copies of the genetic information are separated. These two processes somewhat resemble (and were inspired by) the replication processes of DNA molecules, although here the genome is single-stranded for simplicity.

After the unzipping sequence has come to a halt, the cell is left in the state shown in the first image of Fig. 7. From this point the cell divides by contracting at the contact points into a dumb-bell shape and then separating completely. To make this happen reliably, a process of 'pulling' is used, to maneouvre the tail-end attachment point up towards the head end. While this may appear an artificial process, more akin to engineering than biology, similar mechanisms are seen in the mechanisms of the mitotic spindle for instance, and in the contractile rings that are used for division in some types of cells [4].

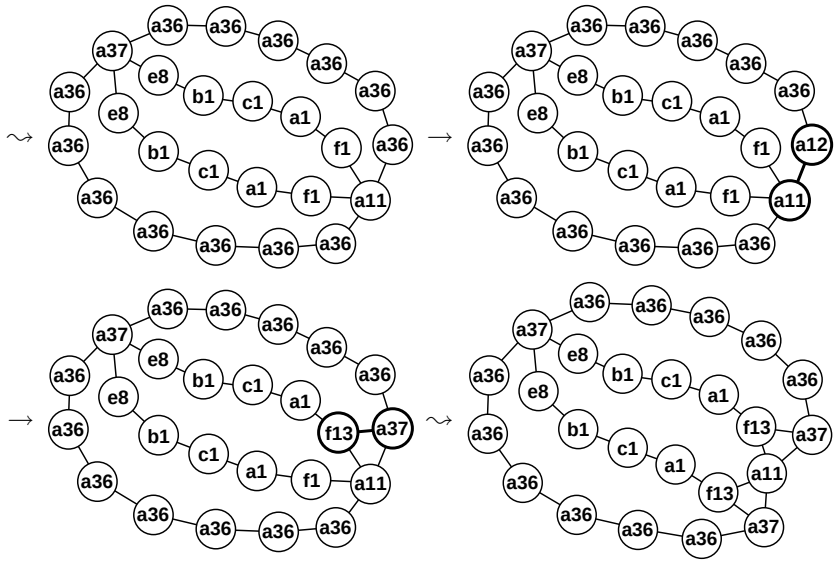


Figure 7: After the gene-strings have unzipped, the pulling sequence is intiated. Firstly the tail-ends of the gene-string (the `f` ends) are attached to new points on the membrane, by reactions R14 and R15. From this point the pulling sequence is triggered.

The pulling sequence is illustrated in Fig. 8. An `a11` atom is sequentially attached and detached from bases along a string. Since the atoms are all moving at random, this results in an inevitable drift along the string. This process causes the tail-end attachment point (now `a11`) to be pulled up between the gene-strings to the head end.

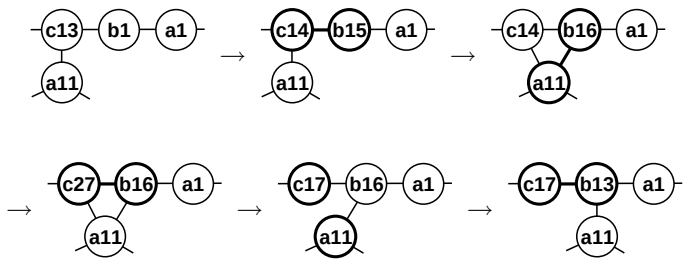


Figure 8: The pulling sequence, using reactions R16-R20.

The effect of this pulling sequence on the cell is shown in Fig. 9. To accommodate the movement of the attachment point, the cell is forced to curl into two lobes, each containing one copy of the genome. When the attachment point reaches the top of the genome (the `e` end) it can fuse with the membrane on the other side, triggering the final separation process that is directed by reactions R25-R34.

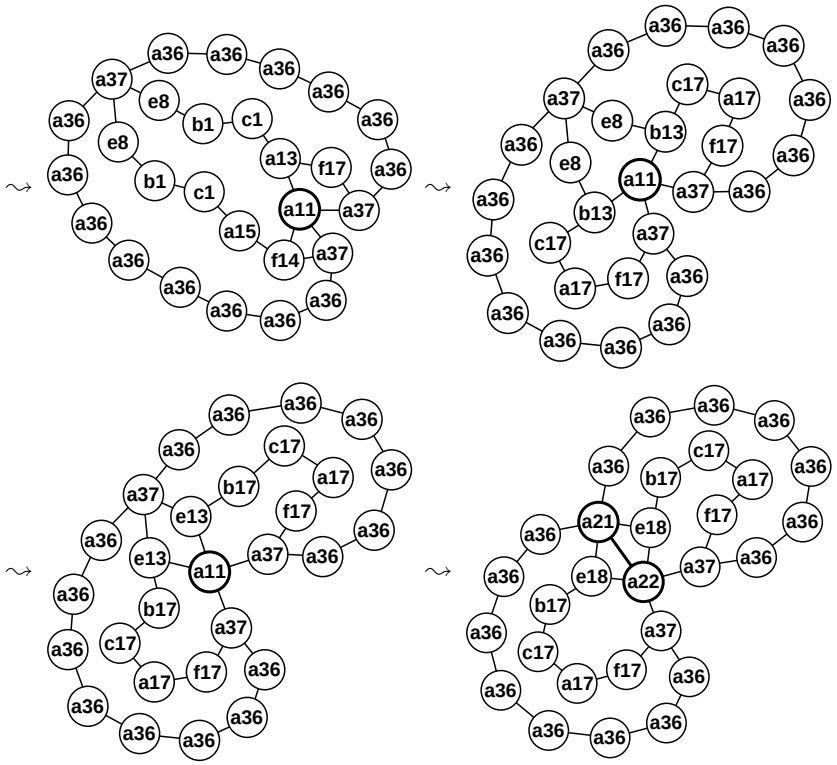


Figure 9: After the attachment point (`a11`, highlighted) has been pulled up to the `e` end of the gene-strings, the two strands are left in separate lobes of the dividing cell. Reaction R21 allows the pulling to proceed past the `e` atom, and reactions R22, R23 and R24 make the two halves of membrane fuse, leaving the two strands in each side.

From this point, reactions R25-R34 cause the two daughter cells to divide completely. This process is shown in Fig. 10.

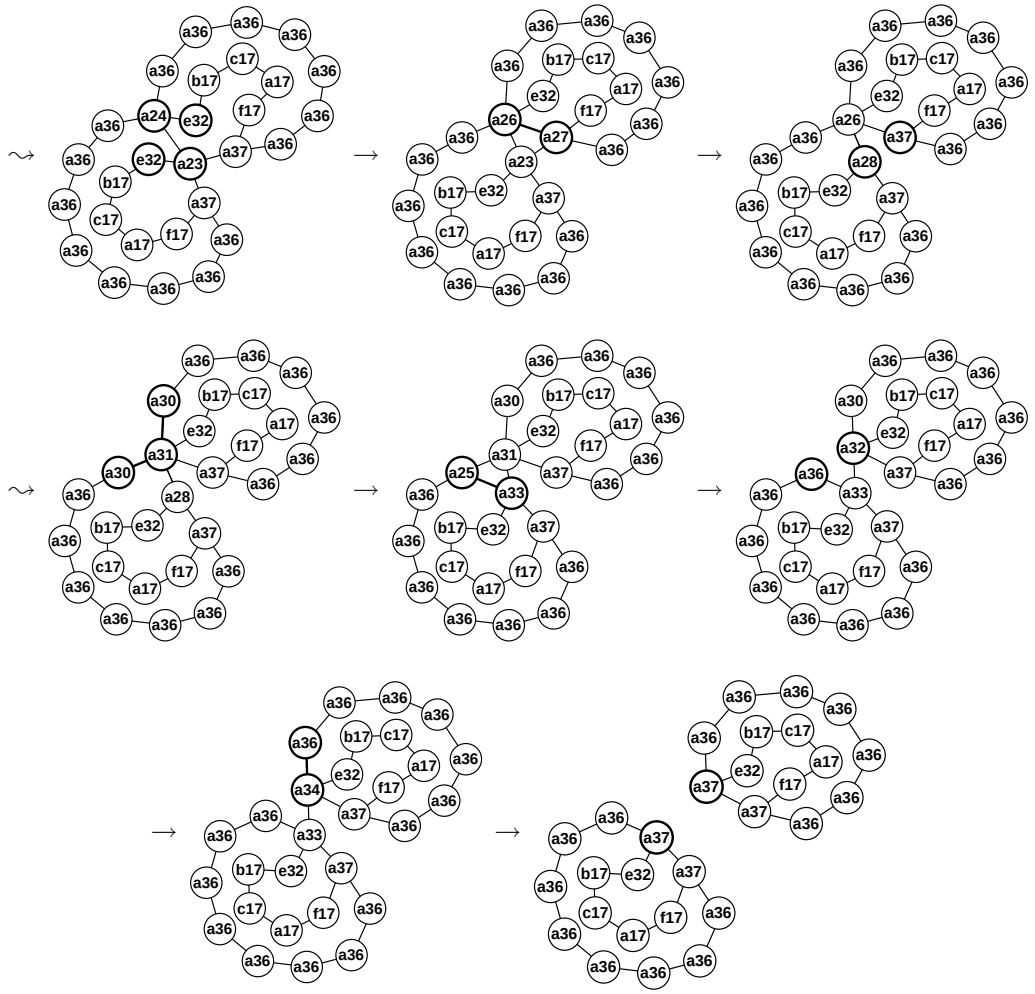


Figure 10: After the membranes have fused, the cells divide in a way that works cleanly. This is directed by reactions R25-R34.

While the reactions described are sufficient for cell division, two issues remain. Firstly, the membrane is now half the size it was to start with, so repeated reproduction would soon be brought to a halt through a lack of room inside the membrane. To deal with this we embody the membranes in our AChem system with an additional property, that they can spontaneously acquire and lose atoms. This property is directed by the reaction shown in Fig. 11. Using this reaction, the cells shown in the last stage of Fig. 10 can grow back to their original size, each becoming as the first image shown in Fig. 13.

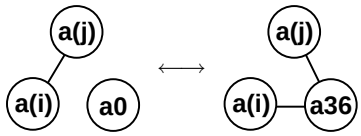
R35 : 

Figure 11: To allow the membrane to expand as required, we allow it to spontaneously acquire or lose atoms from/to the surrounding soup. Reaction R35 (where $i, j \in \{36, 37\}$) directs this, with three reactants being involved. This reaction can operate in either direction at equal rates. Since free atoms may drift away once released, the tendency is for membranes to adopt the smallest possible size around the contents of the cell, while still allowing them to grow and move flexibly.

The second issue is that the atoms comprising the genetic information are not in the states they were initially, so reproduction cannot immediately proceed. This is by design, since we want to use the gene-string for another purpose before allowing reproduction to repeat. This dual use of the information-carriers: for replication in a quiescent state, and function in a more active state, is at the heart of both DNA and von Neumann's universal constructor [57].

Until now, however, these atoms did not technically constitute information because they do not describe anything. To make use of the sequence of types we again draw our inspiration from DNA and map the sequences onto a set of *enzymes* that each catalyse a specific reaction. In DNA this is done via an intermediate medium of mRNA, which is transcribed into a string of amino acids that naturally folds up into the compact three-dimensional shape that drives the reaction. To achieve something similar in our simple AChem, we first map the sequence of base types into an integer value. This process spatially compresses the information contained in a gene-string into a single point in order to make finding the right reactants computationally efficient. It would be possible to model enzyme shape explicitly, even in two dimensions, but this would require the interaction of many atoms at the same time.

The conversion uses a base-4 decoding scheme, with types `abcd` mapping onto digits 0123. Thus the string `bca` (in our worked example) is equivalent to $120_4 = 24$. In an AChem-type setting this is easily achieved by moving an atom down the sequence, changing state as it goes. Figure 12 gives the reactions that are used to do this and Fig. 13 shows a worked example on the cell we have been following.

To produce more than one enzyme, intermediate `f` atoms can be inserted. For example, the string `ebcafbdcaf` would produce enzymes 62 ($= 120_4 + 38$) and 158 ($= 1320_4 + 38$). To continue our (ab)use of biomolecular terminology, we call these sub-sequences genes.

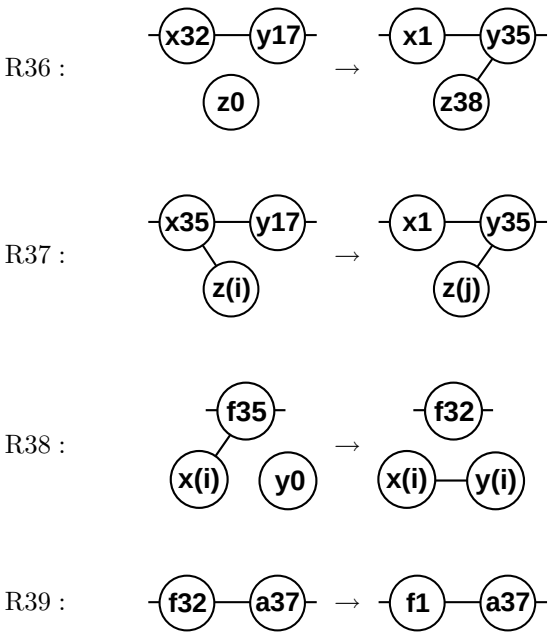


Figure 12: The enzymes are produced in a read-out phase that works down the bases, returning them to state 1. The reactions shown here direct the process. State `32` triggers the production of a new enzyme, the seed for which is attached using R36. This seed then works down the gene-string, changing state as it goes, using R37. In this reaction, $i \geq 38$, $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$ and $j = 4(i-38) + \mathrm{val}(x) + 38$, where $\mathrm{val}(\mathtt{a}) = 0 \ldots \mathrm{val}(\mathtt{d}) = 3$. This formula converts the base-4 string represented by the sequence of bases into an integer value, offset by 38. Reaction R38 releases the completed enzyme as a pair (to prevent it from escaping the membrane) and R39 causes the process to stop when all the enzymes have been produced. A worked example of this sequence is shown in Fig. 13.

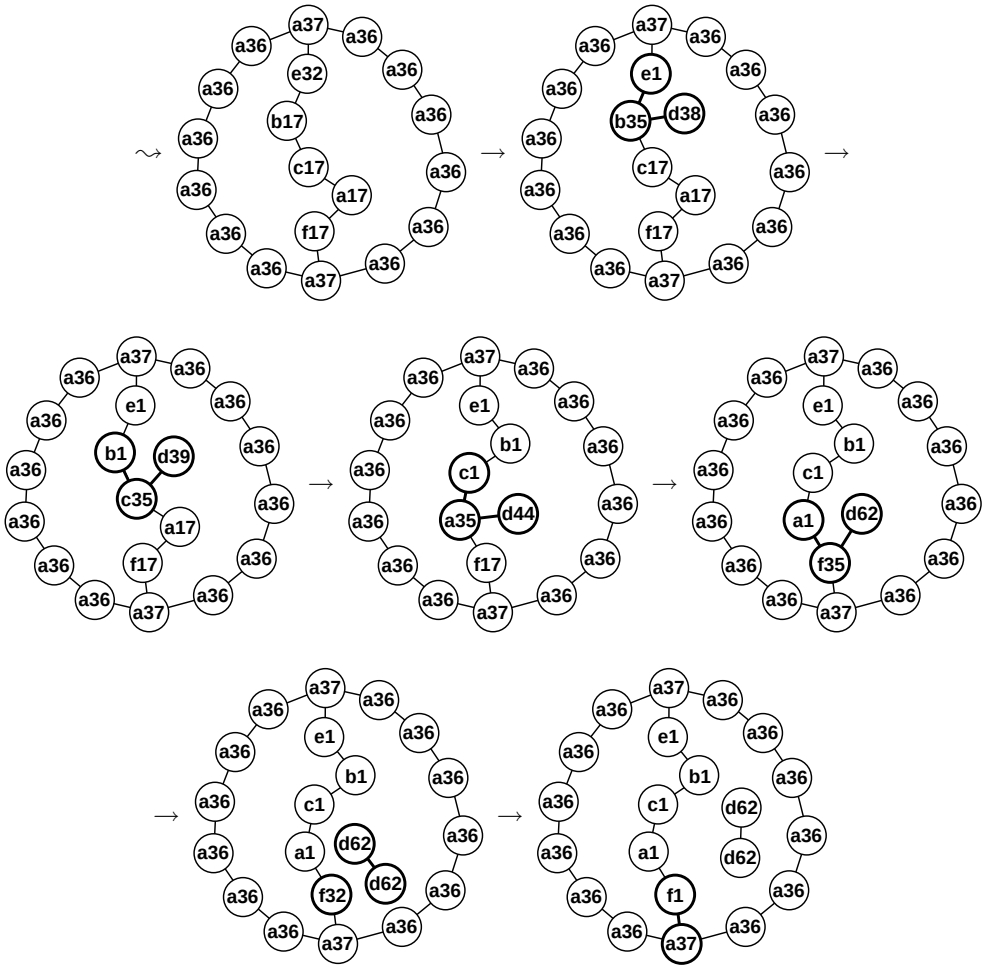


Figure 13: Enzyme production in our example cell. An enzyme seed (here of type `d`) attaches at the top of the gene-string, first acquiring state `38` from R36. It moves down the strand in three steps, taking on states `39`, `44` and `62` as it goes, corresponding to the successive strings: `b` $\equiv 1$, `bc` $\equiv 12_4 = 6$, `bca` $\equiv 120_4 = 24$. When the enzyme is released as a pair, the cell is in an equivalent configuration to when it started, seen in Fig. 2. The number of molecules in the membrane may be different, hence the use of the term 'reproduction' instead of 'replication' which implies exact copying.

After the enzyme-production sequence described above, the cell can reproduce again. Given a supply of raw material (atoms in state 0), the population of cells will grow rapidly. If the resources are finite and there is some process by which cell death occurs, then it can be said that there is competition, since the cells that take up the resources more rapidly will be more likely to survive.

For evolution to occur, two further things are required. Firstly, the sequence of information must have some effect on the environment, in order to create a survival differential between the different cell types. Secondly, mutation must occur, to permit variation to appear.

The first of these requirements is almost complete: we have mapped the gene-string onto a set of enzymes.

All that remains is for each enzyme to catalyse a specific reaction. The details are given in Fig. 14, where any atom in state ≥ 38 will catalyse a given reaction. This mapping allows any reaction between atoms in states 0-37 to occur. Note that this includes all of the reactions in Table 1, allowing the entire self-reproduction sequence to be directed by enzymes (other than the production and effect of enzymes themselves).

R40 :

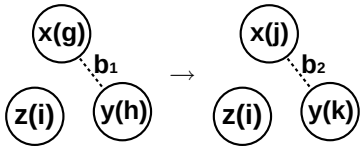


Figure 14: The reaction that determines the action of each enzyme. The atom $z(i)$ is the catalyst that drives a specific reaction between the x and y atoms. The reaction is defined by the value of i, where $i = 2(2(T(T(S(S(Sg + h) + j) + k) + x) + y) + b_1) + b_2 + S$. In this formula, S is the number of states (38) and T is the number of types (6). The values for g, h, j and k are the states of the reactants before and after the reaction, and the values for x and y are the types of the reactants, where `a` $= 0$, `b` $= 1$, ... `f` $= 5$. The bonds, b_1 and b_2, have the value 0 for unbonded, 1 for bonded. Without the presence of the enzyme, the reaction would not occur (unless it is one of the in-built system reactions). When the enzyme is present, the reaction rate is determined solely by how often the correct reactants collide with the enzyme.

The final requirement for evolution is that variation must appear in the population. To do this we introduce one final reaction, given in Fig. 15. Any chain of atoms in state `1` will gain or lose an atom with some low probability, set by the user. The effect of such a mutation on our cells would be to make it produce a different enzyme. Most of the time such an enzyme would likely be detrimental to the fate of the cell, stopping it reproducing by interfering with the division process. Sometimes, however, the enzyme would be beneficial to the cell, and the cells possessing the corresponding gene-string would be selected for. We look at what the nature of such beneficial mutations might be in a later section.

R41 :

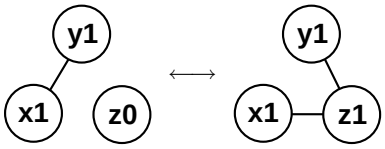


Figure 15: Mutations happen using R41, where a string of atoms in state `1` can gain or lose an atom with some low probability set by the user.

Having given all the details of the cell division process, and shown that the requirements for competitively-selected variation are present, we now proceed to show how the system can be concretely implemented.

Physics

The AChem presented relies on certain features of the underlying physics, such as that the atoms exist in a two-dimensional world and cannot pass through each other - this allows a linked chain of atoms to be an effective membrane. Beyond these considerations there is a wide array of choices for the medium in which the AChem can work. This is a desirable feature because it means that the precise details of the physics are not important; the results will be similar when experiments are run on different implementations.

Here we consider two implementations: a continuous-space world with circular atoms, and a square-lattice based world.

The first is perhaps the simpler to implement. Each atom has floating-point coordinates, and a system-wide *radius* value determines their size. Each atom has a *velocity* that in the absence of interactions will determine its next position. At each update step the velocities are first recomputed, by summing the forces that are acting on the atom, and then the positions of the atoms are updated. Forces come from overlap with other cells or the sides of the area, and from bond distances being greater than twice the radius. In each case the force is either proportional or inversely-proportional to the overstretch/overlap distance, with some user-determined factor k. While this form of motion force integration is unstable when large forces are involved (if two atoms get pushed together), simple measures such as limiting the velocity can prevent numerical problems. To permit small particles to diffuse through the membrane, we simply turn off the cell-overlap force for unbonded atoms. Atoms can react when they are overlapping, since this represents a collision. Figure 16 *(top)* shows a cell in the continuous-space environment described. This form of physics was used in [21] with a different chemistry.

A second method for implementing the required physical constraints is to use a lattice space, as in cellular automata. Here each atom has *integer* coordinates, and at most one atom can occupy a given lattice point. Each atom moves to an empty square in its Moore neighbourhood, as long as such a move does not stretch any bond beyond a certain maximum distance, and the move does not cause any bond to become crossed. Reactions occur when atoms are within each others Moore neighbourhoods with no bond separating them. The atoms are updated in a random order to prevent movement artifacts. For the bond distance, it was found that using a Moore neighbourhood ($|x_1 - x_2| \leq 1$) gave insufficient flexibility to the movement when atoms had multiple bonds, so the extended Moore neighbourhood ($|x_1 - x_2| \leq 2$) is used instead. Figure 16 *(bottom)* shows a cell on a 2D lattice as described. This form of physics was introduced in [22].

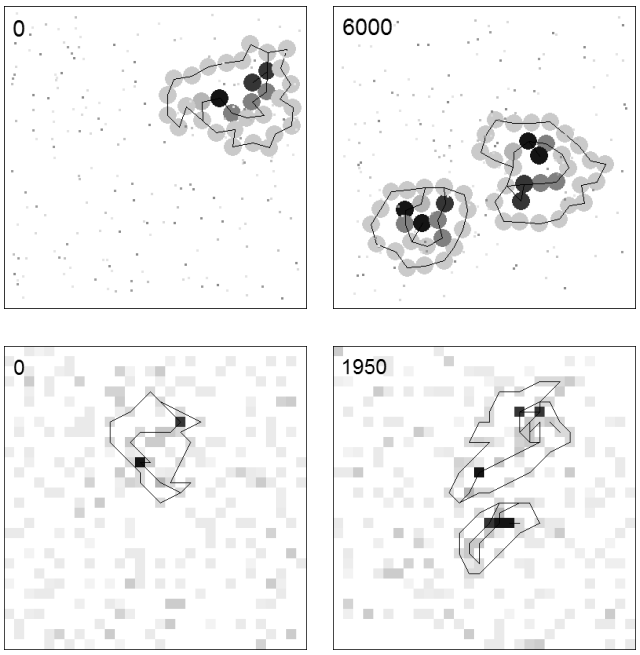


Figure 16: Two concrete implementations of our AChem, with the approximate iteration count shown in each frame. *(top)* A continuous-space model where atoms are represented as circles. Unbonded atoms are shown as dots for clarity. *(bottom)* A square-lattice world where each atom occupies one lattice point. The continuous model took 6,000 iterations at around 600 iterations per second, while the lattice space took only 2,000 iterations at around 2,000 iterations per second.

While the way the cells work is the same in the two implementations, there is a big difference in their speed: the lattice model tends to be much faster. This is hard to demonstrate conclusively since there are many ways of coding each simulation but the point is illustrated in Fig. 16, where the continuous model took around 10 seconds to get to approximately the same point as the lattice model reached in under a second, in a world with a comparable size and number of free atoms (200). One obvious speed-up we are using in the continuous model is space-division to make the search for neighbours a constant-time operation, without this the difference would be even more pronounced.

In the remainder of this paper we will use the lattice model, due to its superior speed.

4 Experiments

The repeated self-reproduction of the cells is shown in Fig. 17, where a whole population of cells grows out of a single ancestor. By 25,000 iterations the population is approaching the point where there are no more free atoms available for use as growth material, and the available space is mostly taken as well.

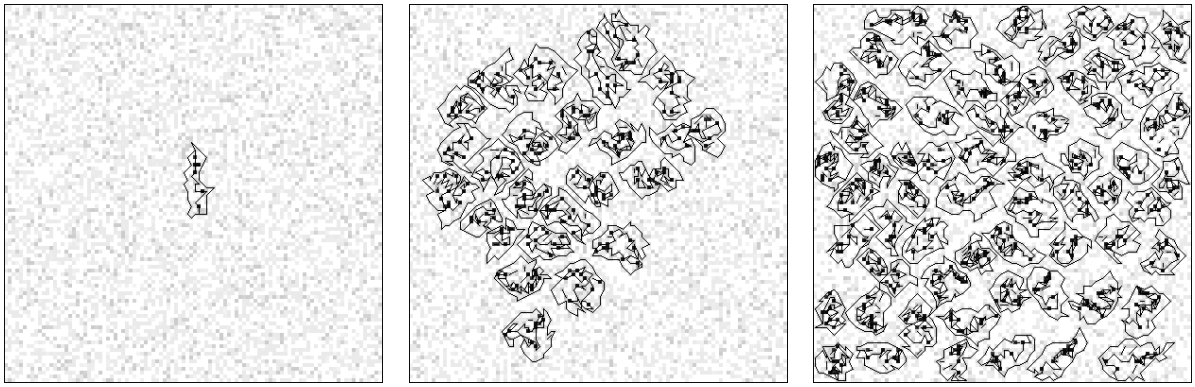


Figure 17: Repeated self-reproduction from a single ancester: snapshots from a simulation run at 9, 13,824 and 25,413 iterations. The world is 100×100 and there are approximately 5,000 free atoms provided to start with.

The cells are not yet capable of attacking each other for food and so to allow reproduction to continue we need some process by which cells are destroyed and fresh material inserted. On an early Earth this might have occurred through the partial mixing of liquid environments - the tidal cycles caused by the moon, for example. Similarly, repeated division and dilution are used in the laboratory to cultivate cell lines for experiments. Here we utilise *floods*, as in [18], where a portion of the world is deleted and replaced by a supply of raw material. This is easily achieved by setting the states of the atoms and those connected to them to `0`, and breaking all their bonds.

Experiment 1 - Information is retained despite mutations

In previous AChem experiments with self-replicating molecules, the effect of mutations was always to cause degeneration, since the gene-string had no phenotypic value [18]. The same effect is seen in vitro, with the Qβ replicase evolving rapidly to the shortest form that allows replication, in an environment where none of its other functions are adaptive [41].

Even if enzymes can be produced, unless the producer gets exclusive benefit of their actions the information is still likely to be lost under the pressure of mutations because of the appearance of molecular parasites [19]. Here, however, with the benefit of a semi-permeable membrane that allows growth but also allows enzymes to be kept inside, we expect to see the information required to produce valuable enzymes being retained.

To test this effect, we first remove one of the in-built reactions, R3. This reaction directs an early step in the cell cycle: `e6e3` $\rightarrow$ `e2e3`, as shown in Fig. 3. Without this reaction or something equivalent, cells cannot reproduce. Instead we provide the genetic code that produces the appropriate enzyme. The enzyme for this reaction (using the formula given in Fig. 14) is $x48044745$, and the type-string that produces this enzyme

(see Fig. 12) is `ecdbdbabccccadf`. Cells with this genome will therefore be able to reproduce despite the absence of R3.

We initialise a world of size 70×70 with some ancestor cells having this genome. We provide approximately 2,500 free atoms in state 0 as raw material (a density of 0.5), and set the mutation probability to 1×10^{-6} per possible occurence (ie. each time R41 *could* occur, there is a 1 in a million chance that it *does* occur). Flooding takes place every 8,000 iterations, deleting a different quarter of the world each time. These parameters are chosen to provide ideal conditions for repeated reproduction, while allowing mutations to appear at a high rate but not so high that natural selection ceases to operate. By monitoring the occurence of R22 and R23 and logging the type-sequence seen at that point, we obtain a reasonable record of the frequency of reproduction of each unique genome - a 'species'. This system runs at around 100 iterations per second, and one typical overnight run produced the data shown in Fig. 18.

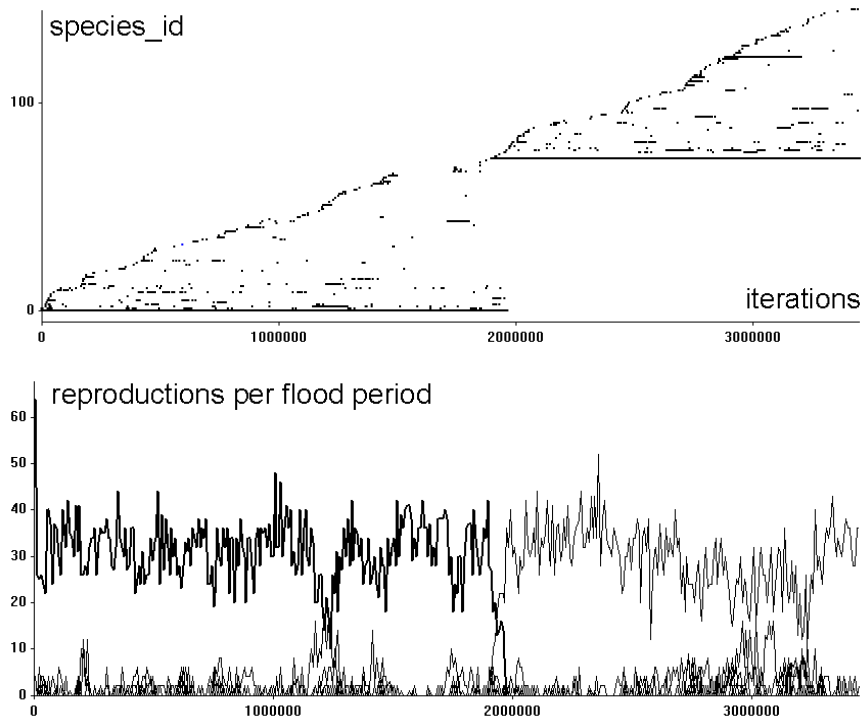


Figure 18: A typical result of running the AChem with an ancestor genome that codes for a necessary enzyme. Running overnight on a single PC, a total of 15,772 reproduction events were logged, and 145 unique genetic species were observed. Each new species seen is assigned a unique species number. *(top)* A scatter plot of species number against time shows the continual appearance of new species as mutations occur. Unlike the ancestor (id 0), most soon vanish because the enzymes they produce do not replace R3, or are even harmful to the cell. At around 2 million iterations a new genome takes over (see main text). *(bottom)* A plot of the reproduction rates for each species against time shows the ancestor (in bold) dominating for the first half of the simulation run before becoming extinct.

In Fig. 18 it can be seen that the ancestor genome we provided is usurped by another. Analysis of the reproduction logs shows this new genome to be `eacdbdbabccccadf` - the same as the ancestor with the addition of an extra base `a` at the front. Because adding a zero to the front of a number does not change its value, this genome actually produces the same enzyme. A cell with this genome will reproduce slightly slower than the original but this difference is minimal - this is a *neutral* mutation [23]. Neutral mutations are important for evolvability because they allow access to parts of the genome-space that would be hard to reach if every non-beneficial mutation were selected against.

It can also been seen in Fig. 18 that many of the unsuccessful species managed to replicate several times before becoming extinct. This is somewhat surprising since these cells are no longer producing the enzyme they need to replace the function of reaction R3. The solution to this puzzle is that the enzymes produced

before the mutation occurred have not been destroyed and in fact have been inherited by and continue to work inside the daughter cells. Indeed the cell cycle as laid out in Figs. 2–13 allows the number of enzymes to build up over time, with a 10th generation cell having as many as 10 copies of the same enzyme. Some process of enzyme depletion could be introduced if this issue became a problem, to mimic the effect in nature where complex molecules cannot exist for extended periods without sustaining damage.

Experiment 2 - Greater complexity can be competitive

The first experiment shows that if a gene is necessary to the continued reproduction of the cell in our AChem then it can be conserved against a mutational load. While this is a gratifying result it only confirms what we were expecting our AChem to be capable of, and does not satisfy any of the aims of this paper. If we wish to recreate the evolutionary growth of complexity in an artificial system we must examine the possible reasons why having more genes might be adaptive, and try to implement them. Then with enough time and a large enough population it might be hoped that the genomes find their way up the adaptive ladder towards sophistication.

We consider more reasons for complexity growth in the next section and for now explore just one of them: extra complexity can be adaptive if it allows the organism to consume a different form of food. Such an organism can then continue to consume and reproduce even while its less complex ancestors are starving around it. A related experiment was run in [20] to show how replicating molecules could adapt to live in different environments.

In this experiment, the free atoms that are provided as raw material are either in state 36 or state 0, in a ratio of 2:1. To allow for this extra state, we increment all states that were 36 or above, including the value of S in Fig. 14. In this environment the ancestor genome from the first experiment can still survive (results not shown) but is only able to reproduce a third as many times as before. We seed the world with ancestor cells as before, but give half of them an extra gene: `daabacaccbddb`. With an additional change that type value 5 maps onto the type variable x instead of `f` in Fig. 14, this gene produces the enzyme for `x5` + `x36` → `x6x6`. This reaction is similar to R6 but allows the new type of raw material to be utilised in addition to the state 0 atoms.

We make our world 80×80 (to allow for the bigger cells) and set the flood period to 30,000 iterations, to allow for the longer reproduction time. All the other settings are the same as in the first experiment. Figures 19 and 20 show the result of a typical run with these conditions.

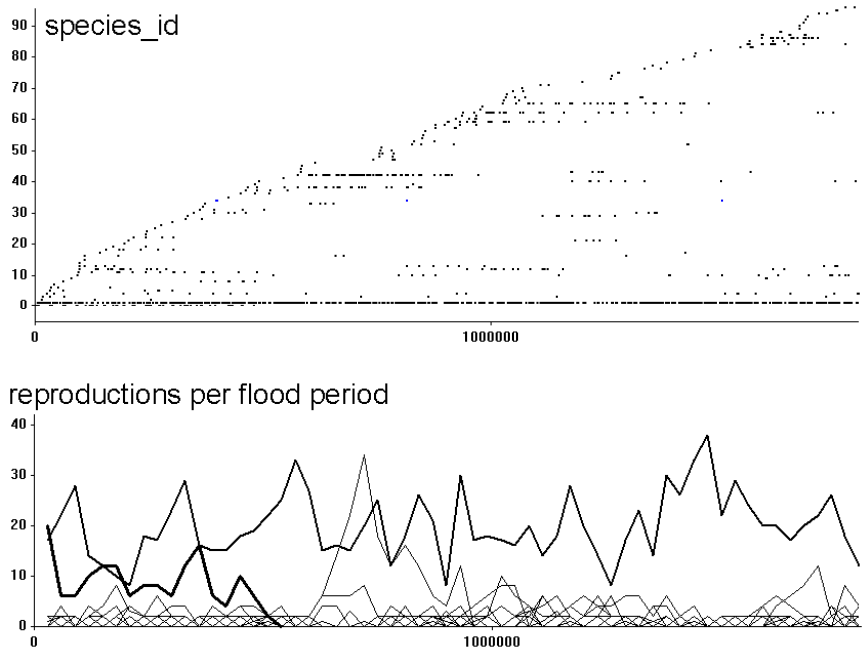


Figure 19: A typical run of experiment 2 shows the species with the two-gene genome `ecdbdbabccccadfdaabacaccbddbf` (id 1, thick line) outcompeting the shorter genome `ecdbdbabccccadf` (id 0, extra thick line). At around 500,000 iterations the smaller genome has become extinct but the longer genome survives. Other species (thin lines) appear as a result of mutations but are not competitive. Note that the concatenation of genes in this fashion could be used to encode all of the reaction rules in Table 1, although this would result in a very large cell.

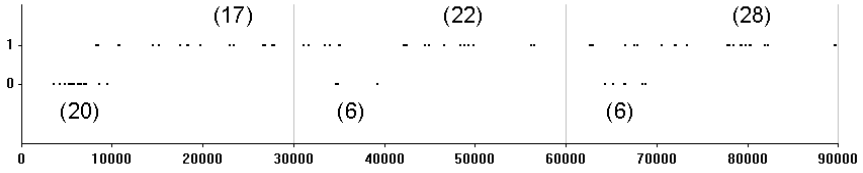


Figure 20: A closeup of the reproduction events at the start of the same run of experiment 2. Before the first flood it can be seen that the smaller genome (id 0) reproduces quickly, using up all of the available state `0` atoms by 10,000 iterations. The longer genome (id 1) reproduces more slowly but is able to keep reproducing because it can also use atoms of state `36` as food. In later flood periods the longer genome outcompetes the shorter one simply because it is able to reproduce more.

This experiment shows that even when a cell takes longer to reproduce and requires more material to do so, under the right environmental conditions this can be adaptive, if the genome codes for the uptake of a new form of food that is abundant. This is one of the ways that greater complexity can be adaptive, and

while this is not a novel conclusion, the fact that such a mechanism can be implemented in a simple artificial chemistry lends encouragement to their investigation. If other mechanisms can be implemented, and the open-ended evolutionary growth of complexity demonstrated, then the system could be used to explore what the *requirements* are for the continual appearance of adaptive innovations.

The difference between the setup seen in this experiment and in experiments in the Avida system where specific tasks are rewarded [25] is that the feature of the environment being adapted to (here the abundance of a new food source) is something that might appear during a simulation run, created by the actions of the reproducing cells. Specifically, if a waste product is produced as a side-effect of the reproduction process then this waste product would quickly become abundant in the environment, and its uptake would present an ecological niche for cells to adapt to. This process of niche construction is thought to be one of the central driving forces of evolution [55].

5 Discussion

The enzymes produced by the cells each catalyse a specific reaction from a large set that includes all of the reactions in Table 1. The important implication of this is that in theory the cells could take control of their own reproduction, directing a specific reaction sequence through the use of their enzymes. Although the set of reaction rules in Table 1 is detailed and specific and hand-designed, these details are not in themelves important because there are certainly many ways of achieving cell division and growth. Out of the set of possible ways, some will tend to operate more quickly, or more reliably, or more accurately, or require less resources, and thus in theory the cells could explore this space of reproduction methods. While the space of reactions is fixed and finite, the space of reaction-sets is combinatorially vast and it is this space that the cells would be exploring. As we have shown in Experiment 2, sometimes a better cell is actually more complex, and thus the mechanisms for complexity growth are in place.

In practice, we do not allow the cells to direct their entire reproduction sequence because the genome required to produce all 34 enzymes is prohibitively large for running evolution experiments - approximately 500 bases. While this is small compared to the genome of *E. coli* for example, at 4 million bases, it currently renders the use of the full genome impractical.

If the cells are to explore the space of reaction-sets, the viable reproduction methods would have to be connected via small mutational jumps. In the system as presented it appears that this is not the case, and thus that the system does not have sufficient evolvability [23, 49]. The appearance of a new enzyme typically requires 14 or so bases, and in the present system unused bases tend to be gradually removed since they have no phenotypic function and slow the reproduction down slightly. To overcome this problem, longer genomes might be the answer, since the selection pressure on having additional junk bases would

be proportionally less and thus the neutral space would be larger. Alternatively, a more compact encoding of enzymes - a different genetic alphabet [27, 58] - would perhaps improve the evolvability by making the appearance of enzymes less unlikely. Another possibility is that using the full genomes of 500 bases and 34 enzymes would actually improve evolvability, since the set of enzymes that could be reached through a simple mutation would be significantly larger. Yet another possibility is that alternatives to single-point mutation such as gene duplication caused by splicing might improve the chances of finding useful adaptations, and some mechanism for allowing these could be considered. The cellular reproduction scheme described here can act as a test-bed for exploring different hypotheses about evolution in a chemical environment.

A more serious obstacle to achieving the evolutionary growth of complexity might be that there are insufficiently many adaptive mutations at each stage. While this cannot easily be quantified since the set of environments in which the cells may find themselves is enormous, a feeling for this can be obtained by hand-designing some innovations and testing their fitness. One such innovation is demonstrated in Experiment 2 as a method for utilising a new food source. If such adaptations are difficult to invent and always require changes to the system then it could be concluded that the cells are unlikely to be able to achieve such innovations themselves, let alone that such innovations will continue to appear. On the other hand, since the cells share and form part of a chemical environment for each other, almost any action they perform changes that environment, potentially presenting a new opportunity for others to adapt to. Immediate work on this system will explore how to encourage these kinds of interactions.

An understanding of all the possible causes for increased complexity to be adaptive is an important longer-term research goal. Evolvability in artificial systems was categorized by Sayama [47] into a) adaptations to the physical environment and b) adaptations to other individuals. In the first category might appear innovations that allow the cells to eat more varied things (as seen), or to endure more environmental damage (UV, temperature, etc.) and thus survive in more environments. In the second category we might expect to see the phenomenon of an arms race [56] between species, and the appearance of simple forms of cooperation and communication. In general, where complexity growth occurs and continues to occur, the genomes are storing information against the flow of entropy (like Maxwell's Demon) because of the reproductive advantage of doing so [2].

One of the most important benefits of using this kind of artificial chemistry is that nowhere is there a definition of what constitutes an *individual*: the reactions are exclusively local, acting only between colliding atoms; the mutation operator merely acts on strings of atoms in state 1 (and thus might be bypassed); the floods merely replenish a certain sector of the environment; the reproduction event observers monitor the occurrence of certain reactions (and thus might fail). Higher-level concepts such as multicellularity and symbiosis are thus not excluded, unlike in many other evolutionary systems, including Avida. Similarly, biogenesis can be modelled since the transition to a 'living' form is nothing more than a rearrangement of

components of the world.

One potential defect of the evolutionary system presented here is that the decoding scheme between base-sequence and phenotypic function is hard-wired into the system, in reactions R36-R40. Several authors have suggested that the translation itself should be evolvable [33, 35, 50] if innovations are to continue to appear. In DNA, for example, the decoding is directed by a ribosome and associated machinery, a complex set of enzymes themselves encoded in and thus produced by the DNA. Similarly von Neumann's self-replicating machine [57] encodes the mechanisms that translate the contents of the tape into actions on the tape itself. Hofstadter tackles this topic [17], using the term 'tangled hierarchies' to describe systems that to some extent encode their own rules. The necessity of this feature for evolvability could be explored in the system presented here, by adding an extra level of complexity to the system: to encode the rules for enzyme production and application in the genome somehow. Cairns-Smith [10] suggests that early life would have started *without* such sophisticated meta-functions, and that later there would have been a transition to adopt them because of the improved speed or accuracy of reproduction.

6 Conclusions

We have presented a two-dimensional artificial chemistry and a set of reaction rules that allows cellular self-reproduction to occur. The behaviour of each cell can be determined by the production of enzymes that catalyse specific reactions, themselves produced by a decoding mechanism from the cell's DNA. The reproduction of the cells is robust enough to operate many times in a shared environment, leading to competition for resources and space. With a mechanism for mutations to appear, the cells are capable in theory of evolving better adapted genomes, thus the system provides a framework for exploring the requirements for the evolutionary growth of complexity and evolvability.

Experiments showed that a sequence of genetic information would be conserved against a mutational load where it was either essential (experiment 1) or beneficial (experiment 2) to the cell. Thus it is hoped that the conditions might be found for continual evolutionary activity, each step being driven by some change to the common environment or to the cells themselves caused by a previous adaptation. The features of an artificial chemistry: conserved components, full embeddedness and rich interactions, support our assertion that this is a suitable medium in which to explore artificial evolution.

Further material

The source code for the systems presented here is available at: `http://www.sq3.org.uk`

Acknowledgements

Many thanks to Dan LaLiberte, Chrisantha Fernando and Katie Bentley for discussion and encouragement.

References

[1] Adami, C., & Brown, C. (1994). Evolutionary learning in the 2D artificial life system "Avida". In R. Brooks & P. Maes (Eds.), *Proc. Artificial Life IV*, (pp. 377–381), MIT Press.

[2] Adami, C., Ofria, C., & Collier, T. (2000). Evolution of biological complexity. *Proc. Nat. Acad. Sci. USA*, *97*(9), 4463–4468.

[3] Adami, C., & Wilke, C. (2004). Experiments in digital evolution (Editor's introduction to the special issue). *Artificial Life*, *10*(2), 117–122.

[4] Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J. (1994). *Molecular Biology of the Cell*. Garland Science.

[5] Allen, F., Almasi, G., Andreoni, W., et al. (2001). Blue gene: A vision for protein science using a petaflop supercomputer. *IBM Systems Journal*, *40*(2), 310–327.

[6] Banzhaf, W., Dittrich, P., & Eller, B. (1999). Selforganization in a system of binary strings with topological interactions. *Physica D*, *125*, 85–104.

[7] Bedau, M., McCaskill, J., Packard, N., Rasmussen, S., Adami, C., Green, D., Ikegami, T., Kaneko, K., & Ray, T. (2000). Open problems in artificial life. *Artificial Life*, *6*(4), 363–376.

[8] Bedau, M., Snyder, E., & Packard, N. (1998). A classification of long-term evolutionary dynamics. In C. Adami, R. Belew, H. Kitano, & C. Taylor (Eds.), *Proc. Artificial Life VI*, (pp. 228–237), MIT Press.

[9] Bentley, P. (1999). Is evolution creative? In P. Bentley & D. Corne (Eds.), *Proceedings of the AISB'99 Symposium on Creative Evolutionary Systems*, (pp. 28–34), The Society for the Study of Artificial Intelligence and Simulation of Behaviour.

[10] Cairns-Smith, A. (1982). *Genetic takeover and the mineral origins of life*. Cambridge: Cambridge University Press.

[11] Chow, S., Wilke, C., Ofria, C., Lenski, R., & Adami, C. (2004). Adaptive radiation from resource competition in digital organisms. *Science*, *305*, 84–86.

[12] Dennett, D. (1995). *Darwin's Dangerous Idea*. New York: Touchstone.

[13] Dittrich, P., Ziegler, J., & Banzhaf, W. (2001). Artificial chemistries - a review. *Artificial Life*, *7*(3), 225–275.

[14] Edwards, L., Peng, Y., & Reggia, J. (1998). Computation models for the formation of protocell structures. *Artificial Life*, *4*(1), 61–77.

[15] Feller, S., Venable, R., & Pastor, R. (1997). Computer simulation of a DPPC phospholipid bilayer: Structural changes as a function of molecular surface area. *Langmuir*, *13*(24), 6555–6561.

[16] Gardner, M. (1970). The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, *223*(4), 120–123.

[17] Hofstadter, D. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books.

[18] Hutton, T. (2002). Evolvable self-replicating molecules in an artificial chemistry. *Artificial Life*, *8*(4), 341–356.

[19] Hutton, T. (2003). Information-replicating molecules with programmable enzymes. In *Proc. Sixth International Conference on Humans and Computers*, (pp. 170–175), University of Aizu, Aizu-Wakamatsu, Japan.

[20] Hutton, T. (2003). Simulating evolution's first steps. In W. Banzhaf, T. Christaller, P. Dittrich, J. Kim, & J. Ziegler (Eds.), *Proc. Seventh European Conference on Artificial Life*, (pp. 51–58), Dortmund, Germany.

[21] Hutton, T. (2004). A functional self-reproducing cell in a two-dimensional artificial chemistry. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, & R. Watson (Eds.), *Proc. Artificial Life IX*, (pp. 444–449).

[22] Hutton, T. (2004). Making membranes in artificial chemistries. In M. Bedau, P. Husbands, T. Hutton, S. Kumar, & H. Suzuki (Eds.), *Workshop and Tutorial Proceedings, Artificial Life IX*, (pp. 71–76).

[23] Kimura, M. (1983). *Neutral theory of molecular evolution*. Cambridge University Press.

[24] Langton, C. (1984). Self-reproduction in cellular automata. *Physica D*, *10*, 135–144.

[25] Lenski, R., Ofria, C., Pennock, R., & Adami, C. (2003). The evolutionary origin of complex features. *Nature*, *423*, 139–144.

[26] Luisi, P., Stano, P., Rasi, S., & Mavelli, F. (2004). A possible route to prebiotic vesicle reproduction. *Artificial Life*, *10*(3), 297–308.

[27] Mac Dónaill, D. (2003). Why nature chose A, C, G and U/T: An error-coding perspective of nucleotide alphabet composition. *Origins of Life and Evolution of Biospheres*, *33*(4–5), 433–455.

[28] Madina, D., Ono, N., & Ikegami, T. (2003). Cellular evolution in a 3D lattice artificial chemistry. In W. Banzhaf, T. Christaller, P. Dittrich, J. Kim, & J. Ziegler (Eds.), *Proc. Seventh European Conference on Artificial Life*, (pp. 59–68), Dortmund, Germany.

[29] Maley, C. (1999). Four steps toward open-ended evolution. In W. Banzhaf, J. Daida, A. Eiben, M. Garzon, V. Honavar, M. Jakiela, & R. Smith (Eds.), *Proc. Genetic and Evolutionary Computation Conference (GECCO99)*, (pp. 1336–1343).

[30] Mayer, B., & Rasmussen, S. (1998). The lattice molecular automaton (LMA): A simulation system for constructive molecular dynamics. *International Journal of Modern Physics C*, *9*(1), 157–177.

[31] Mayer, B., & Rasmussen, S. (2000). Dynamics and simulation of micellar self-reproduction. *International Journal of Modern Physics C*, *11*(4), 809–826.

[32] Maynard Smith, J. (1988). Evolutionary progress and levels of selection. In M. Nitecki (Ed.), *Evolutionary Progress*, (pp. 219–230), University of Chicago Press.

[33] McMullin, B. (1992). *Artificial Knowledge: An Evolutionary Approach*. Ph.D. thesis, Ollscoil na h'Eireann, The National University of Ireland, University College Dublin, Department of Computer Science.

[34] McMullin, B. (2000). John von Neumann and the evolutionary growth of complexity: Looking backwards, looking forwards... *Artificial Life*, *6*(4), 347–361.

[35] McMullin, B., Taylor, T., & von Kamp, A. (2001). Who needs genomes? In *Proc. Atlantic Symposium on Computational Biology, Genome Information Systems & Technology (CBGI)*.

[36] McMullin, B., & Varela, F. (1997). Rediscovering computational autopoiesis. In P. Husbands, & I. Harvey, eds., *Proc. Fourth European Conference on Artificial Life*, (pp. 38–47), MIT Press.

[37] Oberholzer, T., Wick, R., Luisi, P., & Biebricher, C. (1995). Enzymatic RNA replication in self-reproducing vesicles: an approach to a minimal cell. *Biochemical and Biophysical Research Communications*, *207*(1), 250–257.

[38] Ono, N. (2001). *Artificial Chemistry: Computational Models of Self-Maintaining and Self-Reproducing Proto-Cell Structures*. Ph.D. thesis, University of Tokyo.

[39] Ono, N., & Ikegami, T. (2000). Self-maintenance and self-reproduction in an abstract cell model. *Journal of Theoretical Biology*, *206*, 243–253.

[40] Ono, N., & Ikegami, T. (2002). Selection of catalysts through cellular reproduction. In R. Standish, M. Bedau, & H. Abbass (Eds.), *Proc. Artificial Life VIII*, (pp. 57–64), MIT Press.

[41] Orgel, L. (1979). Selection in vitro. *Proceedings of the Royal Society B*, *205*, 435–442.

[42] Ostrovsky, B., Crooks, G., Smith, M., & Bar-Yam, Y. (2001). Cellular automata for polymer simulation with application to polymer melts and polymer collapse including implications for protein folding. *Parallel Computing*, *27*(5), 613–641.

[43] Paun, G. (2002). *Membrane Computing. An Introduction*. Berlin: Springer-Verlag.

[44] Pesavento, U. (1995). An implementation of von Neumann's self-reproducing machine. *Artificial Life*, *2*, 337–354.

[45] Rasmussen, S., Chen, L., Nilsson, M., & Abe, S. (2003). Bridging nonliving and living matter. *Artificial Life*, *9*(3), 269–216.

[46] Salzberg, C., Antony, A., & Sayama, H. (2004). Complex genetic evolution of self-replicating loops. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, & R. Watson, eds., *Ninth International Conference on Artificial Life*, (pp. 262–267).

[47] Sayama, H. (1999). A new structurally dissolvable self-reproducing loop evolving in a simple cellular automata space. *Artificial Life*, *5*(4), 343–365.

[48] Sipper, M. (1998). Fifty years of research on self-replication: an overview. *Artificial Life*, *4*, 237–257.

[49] Smith, T., Husbands, P., Layzell, P., & O'Shea, M. (2002). Fitness landscapes and evolvability. *Evolutionary Computation*, *10*(1), 1–34.

[50] Suzuki, H., Ono, N., & Yuta, K. (2003). Several necessary conditions for the evolution of complex forms of life in an artificial environment. *Artificial Life*, *9*(2), 537–558.

[51] Suzuki, Y., & Tanaka, H. (2000). Chemical evolution among artificial proto-cells. In M. Bedau, J. McCaskill, N. Packard, & S. Rasmussen (Eds.), *Proc. Artificial Life VII*, (pp. 54–64), MIT Press.

[52] Szostak, J., Bartel, D., & Luisi, P. (2001). Synthesizing life. *Nature*, *409*, 387–390.

[53] Taylor, T. (1999). On self-reproduction and evolvability. In D. Floreano, J.-D. Nicoud, & F. Mondala (Eds.), *Proc. Fifth European Conference on Artificial Life*, (pp. 94–103), Springer.

[54] Taylor, T. (2001). Creativity in evolution: Individuals, interactions and environment. In P. Bentley, & D. Corne (Eds.), *Creative Evolutionary Systems*, chapter 1, (pp. 79–108), Morgan Kaufman.

[55] Taylor, T. (2004). Niche construction and the evolution of complexity. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, & R. Watson (Eds.), *Ninth International Conference on the Simulation and Synthesis of Living Systems*, (pp. 375–380), MIT Press.

[56] van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory*, *1*, 1–30.

[57] von Neumann, J. (1966). *Theory of self-reproducing automata*. Urbana: University of Illinois Press, edited and completed by Arthur Burks.

[58] Wills, P. (2004). Stepwise evolution of molecular biological coding. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, & R. Watson (Eds.), *Ninth International Conference on the Simulation and Synthesis of Living Systems*, (pp. 51–56), MIT Press.