Information-Replicating Molecules with Programmable Enzymes

Tim J. Hutton
Biomedical Informatics Unit,
Eastman Dental Institute for Oral Health Care Sciences,
University College London, 256 Gray's Inn Road, London WC1X 8LD, UK
T.Hutton@eastman.ucl.ac.uk
http://www.eastman.ucl.ac.uk/~thutton/Evolution/

Abstract

We present an artificial chemistry that supports molecules that make copies of themselves and also produce specific enzymes. As part of the replication process the sequence of bases is first transcribed onto a non-replicating molecule which then transforms itself into one or more enzymes using a base 3 encoding. The aim of the system is to give replicators control over their environment with the intention that complexity might evolve, however this is not yet achieved. Simulation runs show that the naked replicators are highly vulnerable to parasites, and thus obtain no survival benefit from their enzymes. We speculate that membranes together with an enzyme production mechanism might be necessary for replicating molecules to evolve towards greater complexity.

Keywords: artificial life, artificial chemistry, evolution, complexity

1 Introduction

In the search for an artificial system that recreates the *creativity* of biological evolution, many different approaches have been investigated. The two key concepts of information-replication and competition have been implemented in machine-code systems [14, 13, 22, 10], cellular automata (CA) [15], artificial chemistries [7] and individual-based models [16, 20, 4].

While there is not a clear consensus on what creativity means in the context of evolution [3], a working definition is that the system should pass the test proposed in [2]. This 'ALife test' was refined in [4] using Geb, which remains the only artificial system that has actually passed the test. However, the nature of Geb as currently implemented means that any innovation is hard to appreciate other than numerically. If evolutionary creativity could be recreated in a system that was closer to our own biology then not only might it be more comprehensible but it may also stand a better chance of allowing biologically relevant conclusions to be drawn. Towards this goal, spatially-explicit artificial chemistries seem to be a good medium in which to work.

Achieving the evolutionary growth of complexity in an artificial system has so far proved to be a most difficult challenge. Various suggestions have been made as to the features that a system must have in order to pass the ALife test [6, 19, 17] and this is an important open question [1].

In von Neumann's venerable self-replicating CA machine [21] evolvability is assured through the use of a universal constructor, a device capable of transcribing any configuration of cells from a sequence of instructions stored on a linear tape. When the instructions for the constructor itself are encoded on the tape, the machine will proceed to make a functional copy of itself, and with a large enough array of cells this could continue indefinitely. It is clear however [11] that von Neumann's intention was not just self-replication but in fact evolution, since any mutation to the tape would be inherited, if it were not deleterious. The use of a universal constructor is a powerful guarantee of future evolutionary capability since it allows phenotypes of any complexity to be produced. However, von Neumann's machine is currently of only theoretical interest because without a massively-parallel architecture on which to implement the CA the computational demands of its operation

are too great to run even a single cycle of self-replication.

While DNA does not have the same kind of explicitly universal constructor as von Neumann's machine, it does have a similarly powerful decoding mechanism for converting stored information into functional structure. A section of DNA (a gene, delimited by start and stop markers) is copied onto a free-floating length of messenger RNA, which is then converted into a chain of amino acids (a protein) by a ribosome. This protein then folds itself into a complex three-dimensional shape that allows it to catalyse specific chemical reactions; it becomes an *enzyme*. Together, such enzymes direct the chemical reactions that keep the organism alive, such as biochemical synthesis and transportation. Additionally, the enzymes can interact with the process by which enzymes are produced, either by triggering other genes to transcribe, or by inhibiting them. Clearly then, enzymes play a central role in giving modern information-replicators the power to survive in complex and varied environments.

In this paper we explore some possible ways of implementing these ideas in a simple artificial chemistry, where they can be studied with regard to the question of evolvability of information-replicators.

2 Background

In [7] we introduced Squirm3, a simple artificial chemistry environment in which molecules that could make copies of themselves were possible. These molecules could exist in different forms and this led to a rapid evolution from longer molecules to shorter ones, since the shorter ones could replicate faster and thus use up the available atoms. Regarding creativity, however, this system was a failure since only degeneration was observed. (The same result is seen *in vitro* with simple replicating molecules [12].)

In [8] the artificial chemistry was extended to give the bases on the replicating molecules some simple catalytic properties, and it was found that these were sufficient to give longer replicators a survival advantage over shorter ones in certain chemical environments. However, the catalytic possibilities were severely limited and so no evolution beyond length 4 was observed.

In this paper we present a further extension, inspired by modern DNA and the mechanism by which enzymes are produced. We consider *universal enzymes* in the hope that this might guarantee evolvability. Our conclusions are only speculative, however, since significant evolution is not yet observed.

Simple information-carrying molecules are thought to have been the earliest forms of life, at least according to the 'RNA World' hypothesis [5, 9]. It is not clear, however, how the RNA molecules evolved in complexity, since 'naked' replicators — those without a surrounding membrane — have no choice but to share anything they produce (through catalysis) with their neighbours. Thus, the survival benefit to naked replicators of producing enzymes (or of being enzymes themselves) is minimal [18]. Our experiments *in silico* confirm this and seem to indicate that membranes are necessary if replicating molecules are to increase in complexity through evolution.

3 System Description

Our artificial chemistry (AC) consists of a set of *atoms*, each with a *type* $\in \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}, \mathtt{f}\}$ that cannot change, a *state* $\in \{\mathtt{0}, \mathtt{1}, \mathtt{2} \ldots\}$ that can, and an exclusive position on a 2D grid. *Bonds* between two atoms are created and destroyed by *reactions*. A set of bonded atoms is called a *molecule*. Atoms move to neighbouring (unoccupied) squares at random but must remain near to any atoms with which they have bonds.

In addition to the reactions necessary for replication, we also introduce some extra ones to implement a reasonably close analogy of the biological process of firstly transcribing the bases and then producing an enzyme. For the first part of this we can take advantage of the existing 'zippering' mechanism that moves up and down a sequence first copying the bases and then separating the two copies. The modified reactions for the main replication loop are shown in Fig. 1.

For example, with these reactions and a supply of atoms in state `0`, the molecule `e1-a1-b1-c1-d1-f1` will first duplicate its bases using reactions R1–R4. The two copies will then separate (using reactions R5–R8), leaving two molecules in the form `e3-a1-b1-c1-d1-f1`. Each of these will then duplicate their bases

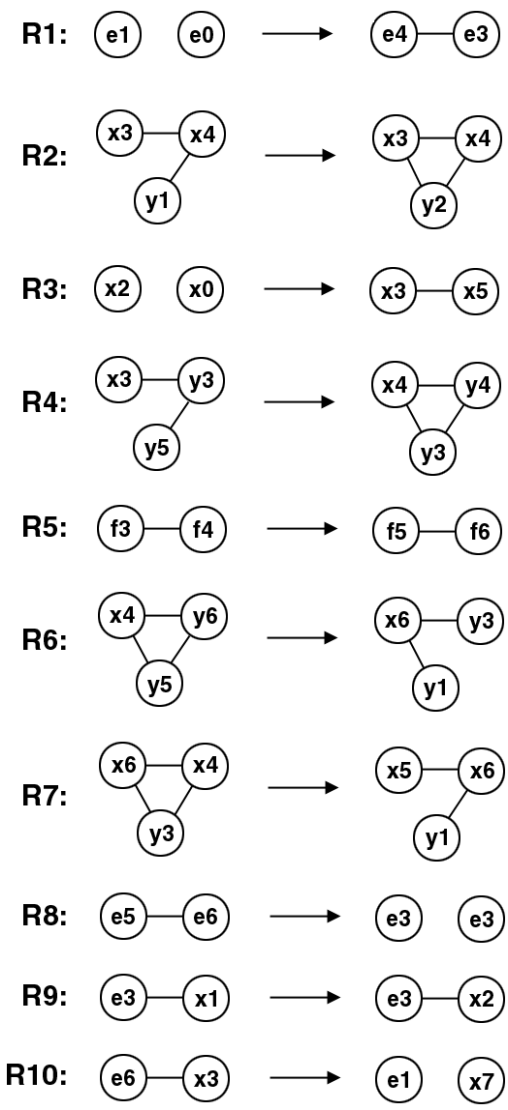
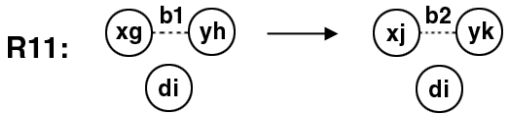


Figure 1: The main reactions for molecular replication and gene-sequence transcription. The symbols x and y stand for any type (`a`–`f`).

again (starting with R9) but without the leading `e` atom. When each of these duplex molecules separates (ending with R10), it leaves one `e1-a1-b1-c1-d1-f1` (which is ready to begin the whole process from the beginning) plus one `a7-b1-c1-d1-f1`, which will become our enzyme. The process is illustrated in Fig. 5.

For the second part, the production and operation of enzymes, we consider how something approaching universality might be achieved. Enzymes in nature work by having a specific shape that fits only the target molecules. When the target molecules drift into place the enzyme catalyses a reaction between them. Since simulating shape is computationally demanding, especially in 3D, we instead utilize a more amenable representation in our AC: the state of each atom. Figure 2 shows how every two-atom reaction (involving atoms with states `0`–`6`) can be encoded into a unique number. Reaction R11 says that an enzyme with such a number as its state and `d` as its type catalyses a single specific reaction. For example, the enzyme `d8` catalyses the reaction `a0+a0`→`a0a0`.



where g, h, j, $k \in \{\texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}\}$ and b_1 and b_2 represent whether the atoms are bonded (=1) or not (=0). The values of g, h, j, k, x, y, b_1 and b_2 are specified by the value of i:

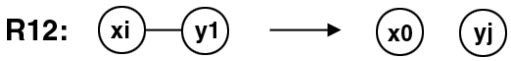
$$i = 2(2(6(6(7(7(7g + h) + j) + k) + x) + y) + b_1) + b_2 + 7$$

and for x and y: `a`=0, `b`=1, ... `f`=5.

Figure 2: Reaction R11 specifies how an enzyme `d`i catalyses a specific two-atom reaction, whose properties are encoded in the enzyme's state, i, where $i \geq 7$.

To create these enzymes, a decoding scheme is required that transforms a sequence of bases into an enzyme. The idea we have adopted here is to use the three bases `a`, `b` and `c` to represent a base 3

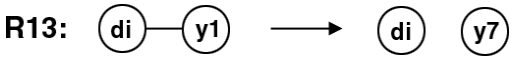
number, with the mapping a=0, b=1 and c=2. Figure 3 shows the reaction that implements this scheme. For example, the sequence c7-b1-a1-d1 would reduce first to b9-a1-d1, then to a14-d1, and then to d28 (as shown in Fig. 5). With enough bases a, b and c, any number ≥ 7 could be encoded.



where : $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$
$i \geq 7$
$j = 3(i-7) + \text{value}(x) + 7$
$\text{value}(\mathtt{a}) = 0$
$\text{value}(\mathtt{b}) = 1$
$\text{value}(\mathtt{c}) = 2$

Figure 3: Reaction R12 decodes a sequence of bases into a single number which is assigned to an atom as a state value.

The final reaction is shown in Fig. 4. When the decoding process reaches a d atom a new sequence is started and the d atom is left in its programmed state. This allows multiple enzymes to be encoded on a string of bases, giving the replicators the power to control different aspects of their environment.



where $i \geq 7$

Figure 4: Reaction R13 separates a prepared enzyme di from the rest of the sequence and also triggers the sequence to continue decoding. This allows multiple enzymes to be encoded on a single string of bases.

An illustration of the replication process including the production of an enzyme is shown in Fig. 5.

4 Experiments

Having created a system in which a degree of controlled action is possible, we now explore whether in a competitive situation the replicators are able to make use of it. A similar situation was faced

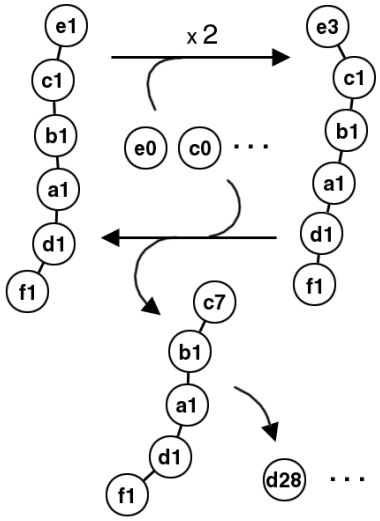


Figure 5: An illustration of the main replication loop, showing how enzymes are produced from a transcription of the bases of the replicating molecules. The molecule e1-c1-b1-a1-d1-f1 is used as an example.

in [8], albeit with a much simpler level of enzymatic action. The solution there was to have different chemical environments in which different configurations of replicator were optimal. While obviously a great deal of novelty can be attained through adaptation to different but fixed environments in this way, true creativity can surely only be attained by replicators finding new ways of getting a competitive edge over the others in a given environment (although partly this may consist of replicators altering that environment).

We seeded a world with one molecule: ebaaacbacbaccdf (where all the atoms are in state 1). When this molecule replicates using the reactions given above, it produces the enzyme d182454, which is the catalyst for the reaction: f3f4→f5f6 (see Figs. 2 and 3). Note that this is reaction R5, one of the ones necessary in the replication sequence. We disable R5 (as is possible when working in an artificial chemistry), thus making the enzyme d182454 the only way that molecules can replicate.

The simulated world was 70x70 in size, and was additionally initialised with six copies of the necessary enzyme to get things started (one would be sufficient but the simulation run would take longer in the initial stages). We used a mutation reaction (see [8]) to allow occasional (P=0.000001)

base insertion or deletion. Additionally, since it was found that longer molecules had a greater propensity to get stuck against each other (an effect imposed by our use of a square grid) we found it necessary to increase the permissible bond length from the Moore neighbourhood to the 5x5 neighbourhood less the corners. This gave the molecules some ability to move while replicating.

The results of a typical run are shown in Fig. 6. As the `ebaaacbacbaccdf` molecule replicates it produces more copies of the enzyme `d182454`. However, at 50000 iterations or so mutations start appearing, and the shorter ones come to dominate because they can replicate faster, using the `d182454` enzymes that are floating around. However, as they drive the population of `ebaaacbacbaccdf` molecules down, the number of enzymes present also drops as a result of the periodic *floods* [7] with $T_{\text{flood}} = 20000$. The eventual outcome is a global extinction event. This has been the outcome of every run that we observed.

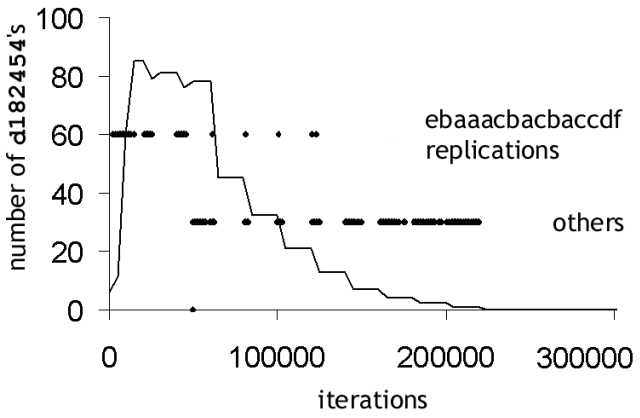


Figure 6: A typical simulation run, showing how many of the enzyme `d182454` are present against time. Overlaid on the graph are the timesteps at which different molecules managed to replicate, with `ebaaacbacbaccdf` molecules in the top row and all others in the bottom row. As the other molecules came to dominate the world this led to a gradual drop in the number of enzymes since they were no longer being produced. This in turn led to a global extinction since the remaining molecules were no longer able to replicate without the necessary enzyme.

These results are not surprising, since the `ebaaacbacbaccdf` molecules are very vulnerable to *parasites* - replicators that hijack other organisms (in this case the `d182454` enzyme) to benefit themselves. Enclosing a replicator in a membrane would allow it to retain sole use of any enzymes that it produces but if parasites could somehow pass through the membrane then they would again be able to take advantage of the replication machinery, as viruses do in nature.

5 Conclusions

We have presented an artificial chemistry containing molecules that replicate information and can program enzymes to catalyse specific reactions. With the correct sequence of bases, any two-atom reaction (involving states `0`–`6`) can be catalysed in this way. The eventual goal is that the replicators should be able to use these enzymes to control their environment and improve their own survival chances but this is not yet achieved. A likely reason for this is that naked replicators are unable to derive any survival advantage from the enzymes they produce since these enzymes and their reaction-products diffuse away.

It must be noted that in one sense the enzymes are not universal since they cannot catalyse reactions between themselves, for example. However, note that this is also true in nature, where only a limited number of reactions have catalysts. Compare also von Neumann's 'universal' constructor which is only capable of writing an array of cells in a non-active state. It seems likely that this is not a limiting factor with regards to evolvability, however, and that control over some aspects of the environment is sufficient.

Our system differs from genes in DNA that have promoter regions that can make them transcribe individually rather than all at once. It seems inevitable that this sophistication is a highly evolved ability; an adaptation to the need to respond to chemical stimuli in different ways. It might be interesting to see if this kind of ability can be evolved in our AC from some simpler mechanism such as the one presented in this paper.

Another interesting observation is that the decoder in DNA (the ribosome and associated machinery) is a highly complex construction that can only be the result of evolution — some simpler mechanism for converting stored information into functional structure must have existed. In this paper we have supplied a ready-made decoder in the form of purpose-built reactions but this was

in order to show that decoders are possible in this form of AC. If we could find the requirements for different decoders to evolve then this might shed some light on how complex life evolved on Earth.

If the replicators could exist inside a membrane then this would make enzyme-production much more useful to them. Not only would the replicators get the sole benefit of the enzymes but also there would be a greater concentration of them, allowing reaction-sequences to proceed faster [18]. However, successful living inside a membrane is not a trivial task, requiring both membrane growth and membrane splitting to be controlled. Showing how naked replicators can make the transition to living inside membranes is another important question that might be studied using this form of artificial chemistry.

References

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

[2] M.A. Bedau, E. Snyder, and N.H. Packard. A classification of long-term evolutionary dynamics. In C. Adami, R.K. Belew, H. Kitano, and C.E. Taylor, editors, *Proc. Artificial Life VI*, pages 228–237, Cambridge, MA, 1998. MIT Press.

[3] P.J. Bentley. Is evolution creative? In P.J. Bentley and D. Corne, editors, *Proceedings of the AISB'99 Symposium on Creative Evolutionary Systems*, pages 28–34. The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 1999.

[4] A. Channon. Improving and still passing the ALife test: Component-normalised activity statistics classify evolution in Geb as unbounded. In R. Standish, M.A. Bedau, and H.A. Abbass, editors, *Proc. Artificial Life VIII*, pages 173–181. MIT Press, 2002.

[5] W. Gilbert. The RNA world. *Nature*, 319:618, 1986.

[6] F. Heylighen, J. Bollen, and A. Riegler. *The Evolution of Complexity*. Kluwer Academic, Dordrecht, 1999.

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

[8] T.J. Hutton. Simulating evolution's first steps. In *Proc. European Conference on Artificial Life*, 2003. (to appear).

[9] G.F. Joyce and L. Orgel. Prospects for understanding the origin of the RNA world. In R.F. Gesteland, T.R. Cech, and J.F. Atkins, editors, *The RNA World*, pages 49–77, New York, 1999. Cold Spring Harbor Laboratory Press.

[10] R.E. Lenski, C. Ofria, R.T. Pennock, and C. Adami. The evolutionary origin of complex features. *Nature*, 423:139–144, 2003.

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

[12] D.R. Mills, R.L. Peterson, and S. Spiegelman. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Nat. Acad. Sci. USA*, 58:217–224, 1967.

[13] A.N. Pargellis. The evolution of self-replicating computer organisms. *Physica D*, 98:111–127, 1996.

[14] T.S. Ray. An approach to the synthesis of life. In C.G. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, editors, *Artificial Life II: Proceedings of the Workshop on Artificial Life (Vol. X of SFI Studies in the Sciences of Complexity)*, pages 371–408, Redwood City, CA, 1991. Addison-Wesley.

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

[16] K. Sims. Evolving virtual creatures. In *Proc. SIGGRAPH94*, pages 15–22, 1994.

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

[18] J.W. Szostak, D.P. Bartel, and P.L. Luisi. Synthesizing life. *Nature*, 409:387–390, 2001.

[19] T. Taylor. Creativity in evolution: Individuals, interactions and environment. In P. Bentley and D. Corne, editors, *Creative Evolutionary Systems*, chapter 1. Morgan Kaufman, 2001.

[20] T. Taylor and C. Massey. Recent developments in the evolution of morphologies and controllers for physically simulated creatures. *Artificial Life*, 7(1):77–87, 2001.

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

[22] C.O. Wilke and C. Adami. The biology of digital organisms. *Trends in Ecology and Evolution*, 17(11):528–532, 2002.