Simulating Evolution's First Steps

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.

Abstract. We demonstrate a simple artificial chemistry environment in which two small evolutionary transitions from the simplest self-replicators to larger ones are observed. The replicators adapt to increasingly harsh environments, where they must synthesise the components they need for replication. The evolution of a biosynthetic pathway of increasing length is thus achieved, through the use of simple chemical rules for catalytic action.

1 Introduction

The evolutionary growth of biological complexity [11] is an intriguing process but not one we are able to experiment with directly because of the time-scales involved. The creation of a virtual organism, one that replicates, competes for survival in a simulated environment and undergoes largescale evolutionary development is a primary goal of ALife research. Being able to tinker with such a creature inside a computer simulation would be a useful tool for exploring the requirements for life and for the growth of complexity to occur naturally.

Ideally, the constituent components of the creature would be closely related to those in nature, in order that any conclusions drawn about the simulated biology might be expected to hold for our actual biology. One way to achieve this is through direct simulation of our physics and chemistry but to accurately simulate even a single molecule stretches current computing capabilities. The IBM Blue Gene project aims to combine one million CPUs into a supercomputer in order to simulate the folding of proteins suspended in water. Even with an awesome 10^{15} floating-point operations per second it is estimated that simulating a system of 32,000 atoms for 100 microseconds would take three solid years of computing time [20].

Thus for the moment we are forced to abstract out some of the details, leaving a simulation that contains the features that we think are important but that requires far less computing power. There have been many different approaches taken to this problem over the years, with cellular automata (see [15]) and machine code systems (see [21]) perhaps the most popular. Biological processes can also be modelled using artificial chemistries [3], with a division between abstract chemistries that have no representation of the physical location of the components (eg. AlChemy [4], P-systems and membrane computing [14] and ARMS/ACS [16]) and more concrete artificial chemistries that do (eg. [12, 10, 6]).

Modern Darwinian theory tells us that there are three fundamental requirements for the evolutionary growth of complexity. Firstly, there must exist entities that duplicate information (replicators). Secondly, there must exist an evolutionary path from the simplest replicator to the most complex, with only minimal changes at each step, achievable through random mutation. Thirdly, at every step, the minimally more complicated replicator must outperform (or at least survive in the presence of) all the surviving replicators. Without this third requirement there would exist an evolutionary path but there would be no drive to follow it. Of course this does not mean that *every* minimally more complicated replicator has to outperform all surviving replicators, merely that for an evolutionary path to have been followed this must have been true at every step.

In this paper we present a novel artificial chemistry (AC) environment that meets all three requirements and does indeed exhibit evolutionary growth, though in the current system only two small steps are demonstrated. Previous systems in which the evolutionary growth of complexity has been observed include Geb [2], an agent-based simulation in which neural networks evolve, and Avida [8], a machine code system in which replicators that perform certain pre-specified operations are rewarded with the energy needed to execute their code. One advantage that concrete ACs have over such systems is that their representation is much closer to the substrate of our own biology - this should make the organisms more recognisable in their design solutions and any experimental findings should be more directly applicable to an understanding of natural evolution. Additionally, unlike the two systems mentioned, some ACs have the features that have been suggested as necessary for creative, open-ended evolution [19]: implicit reproduction, embeddedness of individuals, materiality and rich interactions. Of these four features, Geb arguably has two (implicit reproduction and rich interactions) while Avida, interestingly, has none.

The AC presented in [6] showed a strong survival pressure for smaller replicators, since these were able to copy themselves more rapidly. The same effect is observed in vitro with the Qβ replicase [13]. A fascinating question, and one we must answer if we are to understand why complex life evolved on Earth, is this: what are the features of a system that drive the evolutionary growth of complexity? Towards this question, though falling short of answering it, we can explore ways of extending the AC given in [6] to produce a growth in complexity. Note that achieving the equivalent of this in vitro is very difficult because we do not have the luxury of being able to change the rules of chemistry and cannot easily monitor the dynamics of the system but perhaps the same result might be achieved in other ways.

Satisfying the third requirement for evolution (that at least some more complex replicators should be naturally selected for) was found to be difficult in our AC. Without a membrane surrounding each replicator, frequently it was found that any phenotypic benefit of having certain bases in the molecule would be shared with other replicators nearby, thus conferring no survival advantage. One solution that was found was that the molecules should be able to catalyse the units needed for their replication ('food') from the surrounding primordial soup

but that these food particles would soon be converted back into non-food particles if not utilised. The second part of this provision ensures that it is likely to be only the replicator that catalysed the food that gets the benefit. It was found that with limited flow between a sequence of environments in which it was possible for the replicators to catalyse the components they needed but increasingly difficult to do so, the replicators would adapt, increasing in length as they acquired more catalytic bases.

The process of synthesising required components is a major part of the activity of cells, and the development of this ability was an important step in evolution:

> "The mechanism for the evolution of long biosynthetic pathways was probably that envisioned by N. H. Horowitz [5]. Any organism that could convert some available compound to a compound required for cell reproduction could then survive in the absence of the formerly required compound. Organisms that developed pathways for the synthesis of required cell compounds had selective advantages over others." [9] (p. 8)

Firstly, in the next section we specify the chemical rules of the system and its starting configuration. In section 3 we show that by introducing mutation and occasional mixing between different compartments the molecules evolve upwards in length as they adapt to survive in the different environments.

2 System Description

As before (see [6]), we use a simple diffusion algorithm on a square grid in 2D, giving us the movement that brings our simulated atoms into contact with each other. (It should be noted that certain features of our atoms are not shared with actual atoms, and perhaps a closer analogy is with a small molecule such an amino acid.) Each atom moves at random to an empty square in its Moore neighbourhood, if there are any, with the additional constraint that it is also not allowed to move outside of the Moore neighbourhood of any atom it is bonded to.

Bonds between atoms are made and broken by reactions, which can also change the state of atoms (0,1,...) but not their type (a-f). One problem with the reactions in [6] was that the molecules had a tendency to become tangled with each other when replicating. While this gave them the ability to mutate we found that the rate of mutation was too high for the experiments here. A different set of reactions was found that gave much more robust replication, these are shown in Fig. 1a. By using three-way reactions we can ensure that in the normal replication sequence bonds only form between atoms in the same molecule, something which is very difficult to enforce when using only two-way reactions. Additionally we were able to reduce the number of states required from 10 to 7.

To introduce catalytic effects we also introduce reactions R9-R13 as shown in Fig. 2. R9 is responsible for converting the 'food' atoms (state 0) to non-food

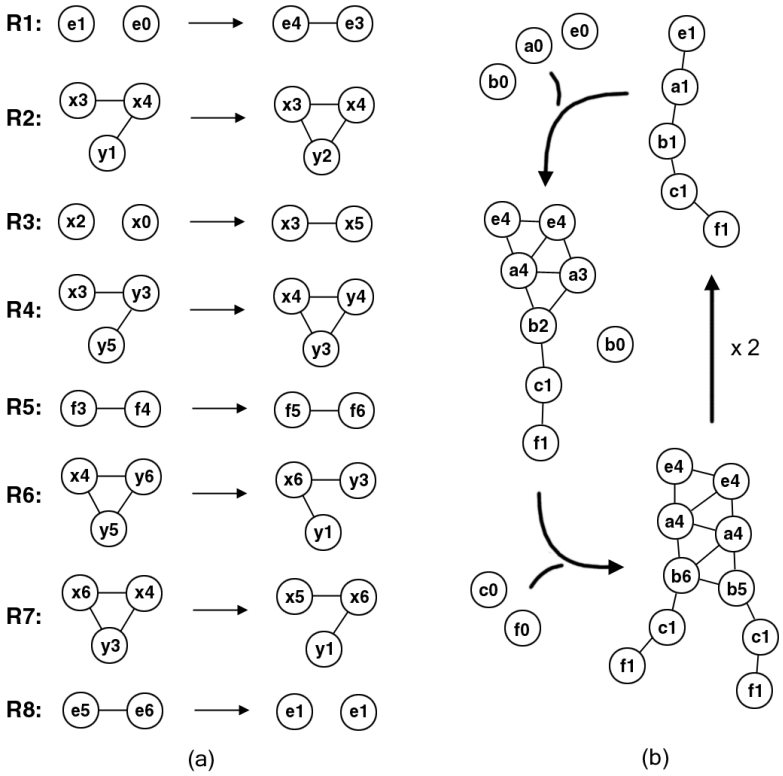

Fig. 1. *(a)* The reactions required for molecular replication in our artificial chemistry. Some act between just two atoms, while others involve three atoms. x and y are variables standing for any type (a-f). These reactions only specify the atoms that are directly involved, the atoms may be bonded to other atoms not shown. *(b)* Three stages in the replication sequence of an eabcf molecule. Any chain of atoms a-d in state 1 with an e1 at one end and an f1 at the other will replicate repeatedly using these reactions when immersed in a soup of free atoms in state 0.

atoms (state 7). One atom in state 7 in an environment containing atoms in state 0 will very rapidly convert them all to state 7 through a cascade of reactions. Reactions R10 and R11 introduce another non-food atom state, 8. Atoms in state 8 convert both state 0 and state 7 to state 8.

Reactions R12 and R13 provide the catalytic countermeasures to this removal of food. R12 says that an atom of type a with state i (where $i \in \{1, 2 \ldots 6\}$) will convert an atom in state 7 to state 0, thus rendering it available for use in replication. A molecule in the form e1a1f1 would therefore be able to replicate when only atoms in state 7 were available by first converting them to state 0. Similarly, atoms b1-b6 convert atoms in state 8 to state 7. Molecules with both a and b atoms in them would be able to replicate (albeit slowly) if only atoms in state 8 were available. These statements are verified in the next section.

Additionally, reaction R14 allows the spontaneous formation of the smallest possible replicator (e1f1), while R15 allows replicators to add or lose an atom (mutation). Unlike R1-R13, these reactions do not take place every time the right atoms come together, instead only happening with some low probability.

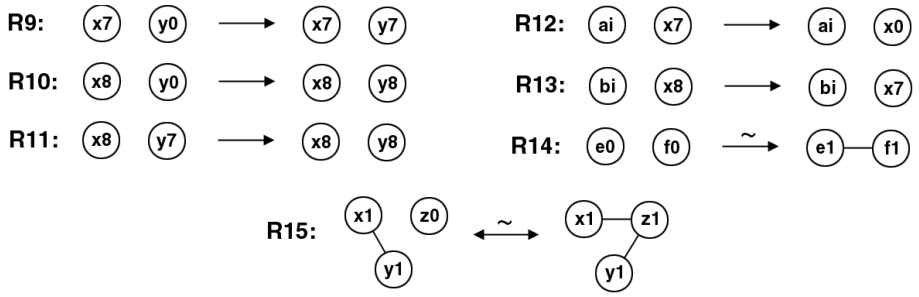

Fig. 2. R9-R13: Additional reactions that give replicating molecules some catalytic properties. x and y are variables standing for any type (a-f), i is a variable standing for any state in the range 1-6. R14-R15: Two additional reactions that allow evolutionary change. R14 permits the minimal replicator (e1f1) to appear spontaneously when an e0 and an f0 come into contact. R15 permits replicators to change length occasionally, either adding or losing an atom. Reactions R14 and R15 only happen with a low probability, P = 0.00001.

3 Observations

We initialize a virtual world with the above physics and chemistry and the following environment. The world is 152×50 and is divided into three zones by two vertical walls of static atoms in a non-reacting state (-1) at columns 51 and 102. The three zones are initialised with atoms in state 0, 7 and 8 respectively, at a density of 1 atom to every 6 squares.

As before [6], periodically the zones are flooded by removing all the atoms in one half and filling that half with atoms in the raw material state (in this case 0, 7 or 8 depending on the zone). This happens every $T_{flood} = 10,000$ timesteps. This repeated dilution ensures that only replicators can persist[1].

Every $10 \times T_{flood} = 100,000$ iterations we cause a degree of mixing between the zones. First we delete half of zone 3 and move half of zone 2 into it. Then, we move half of zone 1 into the empty half of zone 2 and refill the empty half of zone 1 with atoms of random type with state 0. Any replicators that get carried between zones will either be able to survive in the new chemical environment or will be removed with successive floods.

In this experiment we do not seed any of the zones with replicators but instead allow them to form and evolve naturally. We keep a record of each replication event (by waiting for R8 and then analysing the molecules that were involved) so that we can track the frequency of replication of each molecule in each zone.

Figure 3 shows the results of a typical run. In zone 1 (top) the molecule ef dominates, although various mutants are seen. In zone 2 there is no replication until the appearance of eaf at 90,000 iterations. The first eaf replicator could either have mutated in zone 1 and been transported over, or could have mutated from an ef molecule in zone 2 that had been transported previously. In zone 3 (bottom), there is no lasting replication until 290,000 iterations when ebaf molecules appear. In each zone there are occurances of mutated replicators that perform less well than the dominant replicators but these are quickly eradicated, indicating the selection pressures at work.

In zone 3, the ebaf molecule does not dominate forever. In this run, at approximately 2,200,000 iterations (not shown) the equivalent molecule eabf takes over, after a brief struggle.

4 Conclusions

We have demonstrated a simple artificial chemistry environment in which the evolutionary growth in replicator length from 2 to 4 can be observed. While this is only a modest increase, the experiments show that adding simple catalytic properties to replicating molecules can create a survival differential. It has been suggested that the earliest evolution on Earth from the smallest replicating molecules to more complex ones must have involved some simple phenotypic properties of the molecule, since an explicit decoding mechanism could not yet have evolved [7, 1, 17]. While we may never know exactly which phenotypic properties were first acquired, the experiments in this paper confirm that evolutionary growth is possible with simple catalytic effects, and without an enclosing membrane.

Clearly a membrane is a desirable thing from a replicator's point of view. Being able to protect the genomic information from harmful reactions with unknown external agents, and additionally to keep the products of catalytic re-

[1] Likewise, homeopaths should be careful about bacteria lest their dilutions inadvertently have a physical effect on their customers...

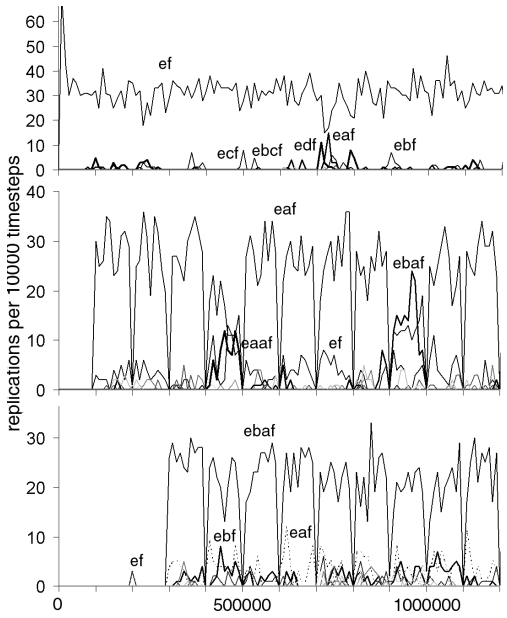

Fig. 3. The average replication rates of different replicators for the three different zones 1, 2 and 3 (top to bottom). While molcules in the form ef appear early on in zone 1, zones 2 and 3 remain without life until molecules eaf and ebaf respectively are evolved.

actions to oneself are strong incentives to develop and maintain the necessary information sequence. It would be instructive to try to evolve this mechanism in an AC, perhaps by incorporating ideas from other ACs where membranes have been created [10, 12]. It is also conceivable that we may one day be able to combine replicators and lipids in the laboratory to make synthetic proto-cells [18], this would also be exciting.

Perhaps the biggest drawback in the current system is that we are only getting as much out as we put in, since the evolutionary changes are in direct response to the environmental conditions and catalytic reactions that we concocted. A major goal of evolutionary systems is to find a way to provide the minimum of features after which evolution takes-off and develops complexity on its own. The only system to date that seems to have achieved this (other than biology on Earth) is Geb [2]. In our AC this would mean finding a set of reactions that allows the sequence of bases to affect the survival chances of the replicator. If such reaction-sets can be found then their common features should enable general conclusions to be drawn about the requirements for the evolutionary growth of complexity.

We have chosen to implement the artificial chemistry described on a 2D grid but it would work equally well with other physics, including continuous space in 2D or 3D. Some demonstrations of this and of the experiments in the paper are available at: `http://www.eastman.ucl.ac.uk/~thutton/Evolution/Squirm3`

References

1. A.G. Cairns-Smith. *Seven clues to the origin of life*. Cambridge University Press, Cambridge, 1985.
2. 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.
3. P. Dittrich, J. Ziegler, and W. Banzhaf. Artificial chemistries - a review. *Artificial Life*, 7(3):225–275, 2001.
4. W. Fontana and L.W. Buss. What would be conserved if "the tape were played twice"? *Proc. Nat. Acad. Sci. USA*, 91:757–761, 1994.
5. N.H. Horowitz. On the evolution of biochemical synthesis. *Proc. Nat. Acad. Sci. USA*, 31:153–157, 1945.
6. T.J. Hutton. Evolvable self-replicating molecules in an artificial chemistry. *Artificial Life*, 8(4):341–356, 2002.
7. 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.
8. R.E. Lenski, C. Ofria, R.T. Pennock, and C. Adami. The evolutionary origin of complex features. *Nature*, 423:139–144, 2003.
9. L. Margulis. *Symbiosis in Cell Evolution*. Freeman, New York, 1981.
10. B. Mayer and S. Rasmussen. Dynamics and simulation of micellar self-reproduction. *International Journal of Modern Physics C*, 11(4):809–826, 2000.
11. B. McMullin. John von Neumann and the evolutionary growth of complexity: Looking backwards, looking forwards... *Artificial Life*, 6(4):347–361, 2000.
12. N. Ono and T. Ikegami. Artificial chemistry: Computational studies on the emergence of self-reproducing units. In J. Kelemen and P. Sosík, editors, *Proc. European Conference on Artificial Life*, pages 186–195. Springer, 2001.
13. L.E. Orgel. Selection in vitro. *Proceedings of the Royal Society B*, 205:435–442, 1979.
14. Gh. Paun. *Membrane Computing. An Introduction*. Springer-Verlag, Berlin, 2002.
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. Y. Suzuki and H. Tanaka. Chemical evolution among artificial proto-cells. In M.A. Bedau, J.S. McCaskill, N.H. Packard, and S. Rasmussen, editors, *Proc. Artificial Life VII*, pages 54–64. MIT Press, 2000.
17. E. Szathmáry and L. Demeter. Group selection of early replicators and the origin of life. *Journal of Theoretical Biology*, 128:463–486, 1987.
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, *Proceedings of the AISB'99 Symposium on Creative Evolutionary Systems, The Society for the Study of Artificial Intelligence and Simulation of Behaviour*. Morgan Kaufman, 1999.
20. The IBM Blue Gene team. Blue gene: A vision for protein science using a petaflop supercomputer. *IBM Systems Journal*, 40(2), 2001.
21. C.O. Wilke and C. Adami. The biology of digital organisms. *Trends in Ecology and Evolution*, 17(11):528–532, 2002.