Diversity From a Monoculture: Effects of Mutation-On-Copy in a String-Based Artificial Chemistry

Simon Hickinbotham[1], Edward Clark[1], Susan Stepney[1], Tim Clarke[2],
Adam Nellis[1], Mungo Pay[2], Peter Young[3]

[1]Department of Computer Science, [2]Department of Electronics, [3]Department of Biology
York Centre for Complex Systems Analysis, University of York, UK
sjh@cs.york.ac.uk www.yccsa.org

Abstract

We have developed an artificial chemistry that allows self-maintaining molecular systems to mutate and exhibit innovative behaviour. The molecular species in the chemistry are defined by strings of symbols that specify both the binding affinity and the reaction. We define a replicase molecule that can copy any other molecule that binds at a particular region on the replicase. Molecules are copied on a symbol-by-symbol basis. Occasional mis-copying of an individual symbol forms our mutation scheme. This paper describes the characteristics of the resulting evolutionary system. We ran 1,000 open-ended trials and observed an unexpectedly wide range of emergent phenomena, with many parallels to biological systems. We report these phenomena in qualitative terms, and give details of one of the most interesting among them: the emergence of co-dependent replicase hypercycles.

Introduction

Early-earth molecular systems are of interest due to their relatively simple replication mechanisms, gene multiplicity, and the blurring of the genotype-phenotype boundary. The simplicity of these systems make them a good target for models of chemical evolution. We have been working on an artificial chemistry called Stringmol [4, 3], which combines a stochastic chemistry, variable binding rates and a simple sequence-based programming language.

Stringmol is a rich intra-cellular RNA-world analogue in which there is no distinction between molecular template and molecular machine. We have recently been experimenting with a unimolecular system, where the molecule is capable of self-copying. We call this molecule a replicase. The sequence of symbols that specify a particular molecular species can be interpreted both as a template (a sequence of symbols) and as a program, which can be executed to carry out the reaction between molecules. If two molecules bind to each other by having a sufficiently "strong" match in their sequences, a handshaking process determines where the program that specifies the reaction starts. In our replicase example, this handshaking determines which molecule is copied and which molecule carries out the copying. In earlier work [5] we found that the function of simple molecular simulations is heavily influenced by bind affinity between molecules, so it is important that the representation of the molecules allows bind affinity to be specified on the genome.

String- or tape-based evolutionary simulations have been reported frequently in the literature, and there are many parallels between biology and computer science in the area. Turing machines make use of a tape and read-write heads [13]. They preceded von Neumann's self-reproducing automata [15]. Both of these architectures have interdependence of data and program, and use self-copying as key demonstrators of the function of the system. These are very simple state machines, with only a loose analogue to the concept of the organism. More recently, Ray's Tierra [11] and the AVIDA architecture [7] have expanded on the paradigm of organism-as-tape, with interesting emergent phenomena that mirror biology. A less well-known but related theme is that of expressing the organism as a container for a large *set* of strings, each of which contribute to the metabolism (and hence fitness) of the organism. Examples include Laing's kinematic machines from the 1970s [8], Hofstader's Typogenetics [6, 14], and Suzuki's string rewriting system [10]. The concept of mutation is realised only in Tierra and AVIDA. These two systems have a single tape per individual, mirroring the function of DNA in the organism. We believe that string systems have the potential to encode more than the genome of the system - the phenotypic machinery of gene expression can also be encoded on string-like agents and so lead to the evolution of effective machinery for genome organisation.

This paper concerns our early experiments with mutation in our replicase system. We believe that there should only be one form of "spontaneous" mutation in the system, and that this should occur when a symbol is copied from one sequence to another. We call this process "mutation-on-copy". In biology, mutation-on-copy certainly happens, especially when resources are running low; i.e. while the cell is under stress [16]. We believe that other forms of genome change should be effected by mechanisms intrinsic to the chemical model. For example it should be possible to construct

a transposon in the Stringmol language, which would allow macromutations whilst itself being a candidate for genomic control. Biological genomes are highly organised, and are responsible for their own expression. In other words, *the phenotype includes the genotype-reading structures, and is completely encoded in the genotype*. In yet other words, the genotype in its purest form is a sequence of symbols, and this encodes *everything else* that is manufactured in the cell, including the machinery for curating the genotype. We have preserved this property in our Stringmol model, and detail here a control experiment that attempts to determine the effects of single point mutations on such a system.

What might be expected of a single-container system that contains mutating molecular replicators? Our experiments confirm the prediction that a series of stable states would emerge, with eventual collapse of the system due to emergent selfish parasites. However, the observed range of reactive behaviour and the interesting dynamics were not expected to occur so rapidly in such a simple system. Analogues of parasitism, hypercycles, random drift, gene repression and co-evolution are reported. Unlike real biology, we are in a position to fully examine the system, and can detail the key events that led to the observed dynamics.

In an RNA-world analogue, such as the chemistry we present here, a molecule can act as both template and machine. Initially, two identical molecules come together, with one acting as the machine which makes a copy of the other. Mutants that are better templates subsequently sweep through the population, replacing the initial molecular species. More interestingly, we repeatedly observe the emergence of a molecular species that does not self-replicate but drives evolution to a state where the system is dominated for a long period by two co-dependent replicase species that are not self-maintaining. This is a catalytic hypercycle as defined by Eigen [2, fig.7].

It is interesting to consider the role of the container in these experiments. Many explanations for the origin of life include the use of membranes to keep the molecular template in close association with the machinery it specifies [9, 1], allowing selective advantage to operate on the machine-template complex as an entity. In early living systems, where mutation was rampant and much less tightly controlled, we observe that containers have a more time-critical role of preventing the rampant spread of emergent pathogens.

System overview

We give here a brief overview of our molecular system, which is described fully in [3] and [4]. A summary of the container metabolism is presented below, followed by a description and discussion of molecular structure. We pay particular attention to the role of sequence alignments and the mutation scheme in our chemistry.

Metabolism

A simulation can be considered as a set of reacting molecules whose movements inside a container are governed by a stochastic mixing function. All molecules are subject to *decay* (spontaneous destruction), which places a requirement upon the system to act in order to maintain itself in the face of entropy. Should molecules come sufficiently close to one another, then they can *bind* and *react*. The system has a clock. At each time step, all the molecules in the system are processed. Actions only occur if energy is available. Energy is consumed via binding and executing each instruction in a reaction. The likelihood of binding and the nature of the reaction is encoded in the string of each molecule in the encounter. Binding and reacting have an energy cost. At one particular time step, we specify that 25 energy units are available. Selection of which events consume the energy is stochastic. The balance between energy availability and the decay rate of the molecule maintains a population of around 350 molecules. We currently specify that only two molecules can ever participate in a single reaction, and that raw materials for the assembly of new molecules are available in saturation. These assumptions will be addressed in future work.

Molecular representation

Our molecular representation is a string of symbols. Each unique string is considered to be a unique molecular species. There are 33 symbols, most of which are non-functional. Maximum string length is 2000 symbols (to accommodate longer molecules with richer functionality), so there exists $n = \sum_{i=1}^{2000} 33^i \approx 10^{3037}$ potential molecular species. An important feature of the molecular representation is that it allows the possibility of several complementary subsequence alignments. Complementary alignments are necessary in order to prevent two identical molecules from binding to each other perfectly. Alignments have two key roles: firstly, they specify binding regions on molecules such that the more precise the alignment, the stronger the binding affinity; secondly they specify program flow in the functional region, commonly acting as placemarkers in "goto" statements. An important property of the representation is that the location of functional and binding regions is solely specified by the subsequences themselves, and different molecular species can bind at different sites on the sequence, so triggering different functions of the molecule. The sequence of the molecule is used to determine how likely a bind between molecules is via a process of Smith-Waterman alignment [12] of complementary symbols. Once a bind occurs, the sequence is treated like a program, commencing at the beginning of whichever aligned subsequence is furthest from the beginning of the string. There are 7 functional symbols, shown as non-alphabetical characters '`$`', '`>`', '`^`', '`?`', '`=`', '`%`', and '}'. Stringmol uses functional symbols to specify the manipulation of a set of point-

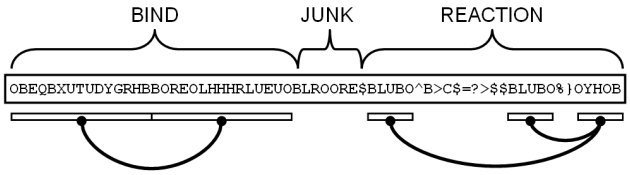

Figure 1: The seed replicase. The top line indicates the regions of the sequence. The sequence itself is shown in the centre box. Complementary alignments are indicated by black connecting lines at the bottom of the figure

ers which indicate positions on the molecular strings, and the symbols that the pointers index.

Mutation Scheme

One of the functional symbols is the copy operator '**=**'. This operator reads the symbol at the read pointer, and writes a copy of that symbol at the write pointer. To implement mutation-on-copy, we specify that a copy operation occasionally writes a different symbol to that being read with a probability $p_s = 0.00001$. More rarely still, insertion of an extra random symbol, or deletion of the symbol, take place with a much smaller probability $p_i = p_s/(10n)$, where n is the number of different symbol codes.

Experimental framework

We ran 1,000 simulations of a replicase environment under the mutation scheme described above. The goal was to evaluate whether the system would be robust to mutation, and if so, what effects it had on the molecular ecosystem. Each of the 1,000 trials had the potential to run indefinitely and only terminated when there were no molecules remaining in the system. This occurs when the replication mechanism deteriorates in some way so that the replicating molecules cannot copy themselves sufficiently quickly to counter the process of decay. In particular, we sought to identify emergent behaviours in the system that were not part of the original specification and arose by mutation.

The "seed replicase"

Here we describe the molecule used as the seed for the trial. It is one of many possible replicase molecules and is shown in figure 1. There are several features to note:

1. Two binding regions. Two are needed to allow a replicase to bind to a copy of itself because binding is *complementary*: a symbol is a perfect match to a different symbol in the set.

2. A junk region. Mutations here have no effect on the binding or reaction-program, allowing us to explore the effects of neutral mutation *drift*.

3. A functional region. This program specifies that the reaction involves creating a copy of the partner molecule in the reaction.

The seed replicase is 65 instructions long. The reactions takes 240 time steps to construct a new replicase molecule. All of the template codes in the seed replicase are more than one mutation away from a function code. Alignments in the functional region specify program flow. The two binding sites in our seed molecule do not align perfectly, which enables us to evaluate the evolutionary pressure on binding.

Analysis

As part of our evaluation, we developed several ways of representing the simulation data. Each molecule has a sequence of symbols. A particular sequence of symbols denotes a particular molecular species, which has an associated species number. The seed replicase is always species number 1. When a mutation occurs, a molecule with a novel sequence is generated, and this is assigned a new species number. In this way, we can record all new molecular species as they arise. We must also record the dynamics that ensue. Occasionally a new species increases in number and rises to dominance of the system, driving the previous dominant species to extinction. This is known in biology as a *sweep* event. We can capture these events by monitoring when the species number of the most abundant species changes (examples are shown in figure 4). We can record the reactions that exist between all species present in a system at any one time (see figure 6). Finally, we can record the *ancestry* of a molecular species: a new molecule is the product of a reaction between two other molecules, which belong to either one or two species types (see figure 7). These figures are described in more detail later.

With these tools to hand, we are able to demonstrate that our system is capable of producing innovative behaviour even from very simple starting conditions and with no external selection pressure. Essentially, the molecular community acts as a co-evolutionary system, in which the fitness of a particular molecular species is largely determined by the cohort of molecular species with which it shares the container. To demonstrate this, we present results on three levels. The first level gives summary observations and statistics from the 1,000 trials. Secondly, we offer a qualitative analysis of these trials, in which a range of emergent phenomena are qualitatively described. The third analysis gives details of a single trial with emergent phenomena and shows how a series of single-point mutations change the seed replicase system to a mutually-dependent "hypercycle" in which two molecular species cannot self-maintain, but maintain a population by copying each other.

General observations

The mutation rate delivers a mean time of 18,700 time steps for the creation of new molecular species. The majority of

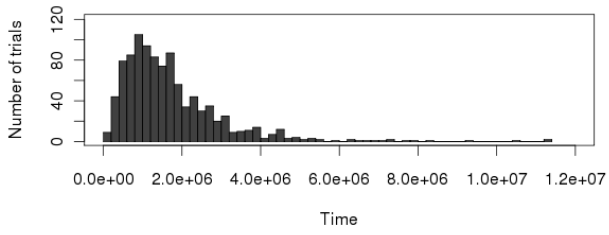

Figure 2: Distribution of extinction times for 1,000 trials

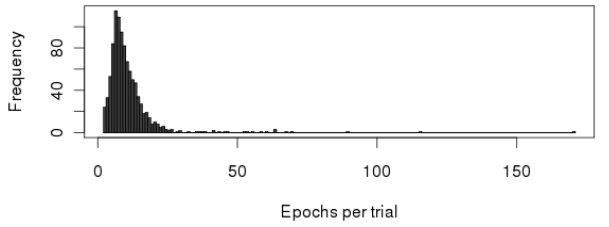

Figure 3: Histogram of number of epochs per trial

these new mutations are not "fixed" in the population and go extinct very quickly. Occasionally a new species arises that has some advantage over the current dominant species.

None of the 1,000 trials self-maintains indefinitely. The nature of extinction follows a uniform pattern as described below, but the timing of the extinction varies. Figure 2 shows the distribution of time to extinction for the molecular populations. The modal extinction time is 750000 time steps. In this time an average of 40 new species are produced.

Mutations occasionally produce molecules that rapidly multiply to become the dominant species in the system via the phenomenon of *invasion when rare*. We use the term *epoch* to describe the period over which a particular molecular species is dominant in the system; *sweep* describes a change in epoch. A histogram of the number of epochs per trial is shown in figure 3. The long tail on the histogram is a caused by runs where periods with co-dominant species that should be labelled as a single epoch are recorded by the analysis as a high number of very short epochs due to small fluctuations in abundance of the two species. This definition of the epoch is not particularly useful in situations where two species are co-dominant, but this behaviour was not predicted. Epochs for a single trial can be seen in figure 4.

A classification of emergent phenomena

In this section we give brief descriptions of the key phenomena we have observed in the 1,000 trials. These were identified by visual inspection of the plots of changes in the populations of molecular species, e.g. figures 4 and 5.

Extinction

All trials end when no molecules exist in the system. This occurs when there is a catastrophic decline in replicating molecules. The common cause of this is when a new 'parasitic' molecule arises that is 1) incapable of replicating itself, and 2) copied by the incumbent replicase at a higher rate than the replicase. Note that in order to be copied, a parasite must bind to the replicase sufficiently frequently. This tends to make the system more robust to molecular "junk" and explains why some of the trials continued for so long. A characteristic spike may be observed at the end of each run, which shows this new parasitic molecule as it rapidly increases and then declines when the last replicase molecules decay. Occasionally a parasite begins to overrun the replicase population, but it is unable to bind to a new replicase mutant that is created as the parasitic molecule is increasing. This is rare, occurring in only two of the trials.

Dynamics

Characteristic sweep. The majority of sweeps in our system take a constant form, as shown in figure 4. These are the the main cause of epoch change, and take less than 50,000 time steps for a new mutant to drive the previous dominant species to extinction.

Drift. Drift is observed when a neutral mutation of a dominant individual builds in numbers due to a random walk. Drift is common, occurring in 92 trials. It is plausible that sub-populations and slow sweeps (described below) are both commonly caused by drift. Species exhibiting drift tend to have mutations in the junk region, but can also show mutations in binding regions that do not change the bind affinity.

Sub-populations. These are species which persist in the community in fairly large numbers (more than 50 molecules of approximately 350 in the system). These are very common, occurring in nearly all runs. These sub-populations are nearly always wiped out when a new epoch begins, demonstrating the biological phenomenon of selective sweeps. *Enduring Sub-Populations*, that persist across more than one epoch, occur in 26 trials. This indicates that sub-populations tend to depend on some property of the dominant species in the system, essentially acting as non-lethal parasites. Co-dependence between dominant and sub-populations cannot be determined by examination of population numbers alone. In 2 trials we observed a sweep in a subpopulation whilst the dominant population remained stable.

Slow sweeps. A sweep can occasionally take much longer than the 50,000 time steps of a typical sweep. These are called "slow sweeps" and may be due to drift alone. An example can be seen in one of the hypercycle partners in figure 4 at around $t = 2,600,000$. Slow sweeps occurred in 52 trials.

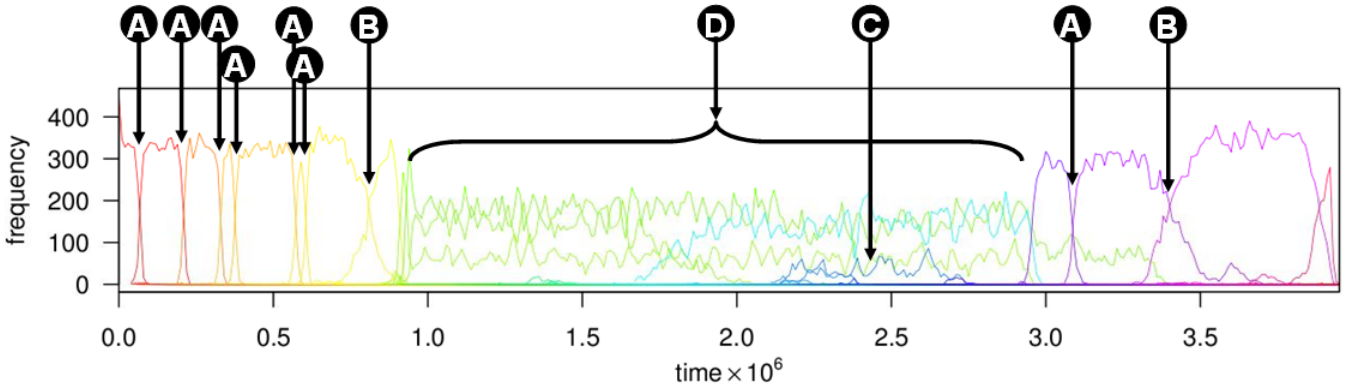

Figure 4: Dominant species in run 112. This trial exhibits (A) characteristic sweeps, (B) slow sweeps, (C) subpopulations, and (D) multispecies hypercycles.

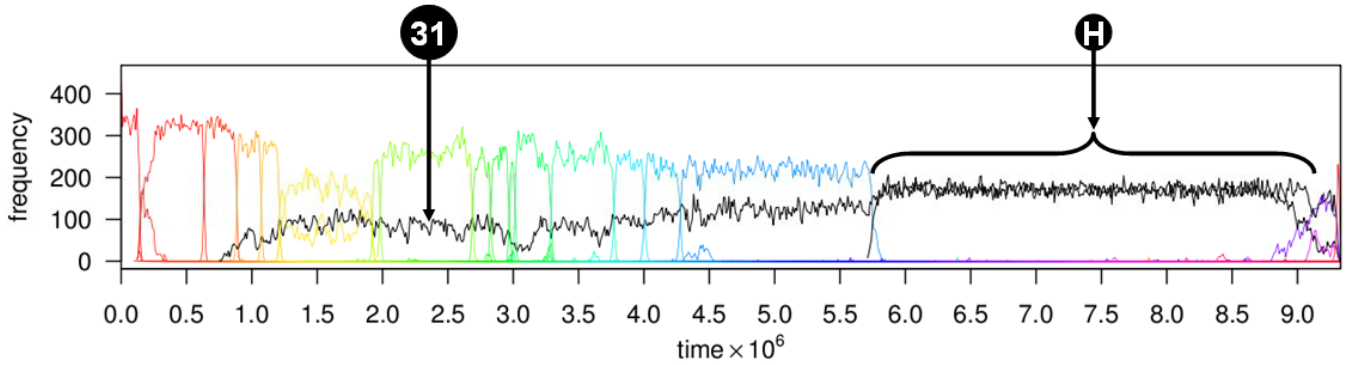

Figure 5: Dominant species in run 277. The short replicase (species 31) emerges at $t = 748,199$ and forms a hypercycle (H) at $t = 5,750,000$.

Rapid sweep sequences. Occasionally a mutant causes a "cascade" of new molecules by triggering a sequence of new unseen molecules that quickly dominate the population. The most common mechanism for this is a mutation that gives rise to a series of molecules that bind to a replicase such that less than their entire sequence is copied. This occurs in 31 trials.

Complex behaviour

Emergent hypercycles. A hypercycle occurs when an enduring sub-population increases in number until it becomes co-dominant with a dominant species. The species forming the enduring sub-population is not self-maintaining, but acts as a copier for the dominant species. The dominant species then repeatedly loses self-self affinity until it loses the ability to self-maintain altogether. The hypercycle occurs when the ability of the dominant population to self-maintain is lost, and the two species become co-dependent. This occurs in 8 trials. Hypercycles end with a sweep, but occasionally one of the partner molecules is still able to maintain a sub-population. A series of sweeps ensues, in which the sub-population declines slightly following each sweep. This occurs in 6 trials.

Spontaneous hypercycles. are the same as the emergent hypercycle, but forms from species that both arise in the immediately preceding epoch. The mechanism is under investigation. This occurs in 15 trials.

Multispecies hypercycles. occur in 14 trials, when there appears to be a mutual dependence among more than two chemical species, as shown in figure 4.

Detailed evaluation of a single trial

We present here details of one of the more interesting sequences of mutation that leads to a hypercycle of co-dependent molecular species. This was observed in trial 277 (figure 5), but hypercycles of one form or another occurred in 30 trials.

We classify this trial as an "emergent hypercycle". At $t = 748,199$ one of the eventual partners (species 31) is first produced via a mutation. This molecule exists as a sub-population for around $5,750,000$ time-steps before forming one partner in a co-dominant pair of molecular species. The

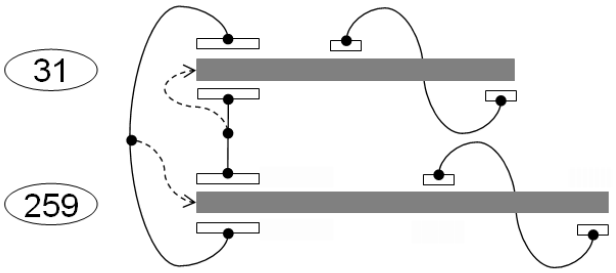

Figure 6: Reactions in the hypercycle. Molecules are represented by grey bars. Binding sites are shown as white boxes, with active binds shown above and passive binds shown below the molecule. Bind alignments are shown as black lines between molecules. Dashed lines show the product of the reaction (where one occurs).

partnership runs for approximately 3 million time steps before a parasitic molecule emerges to end the trial.

The molecular species in a hypercycle

The two molecular species (31 and 259) in the hypercycle are shown in figure 6. The bindings that occur between them are shown as black lines. The assignment of roles in the reaction (i.e. whether the molecule is passive (acts as the template) or active (acts as the program) occurs with equal probability for both molecules, meaning that for 50% of the time species 31 is produced and for the other 50% of the time species 259 is produced. Also note that species 31 is shorter than species 259 - it has lost one of the binding regions required for the reaction-program to initialise such that a copy of the replicase is created. This means it tends to be copied more quickly. Neither molecule is able to self-copy.

This phenomenon was neither foreseen in the original design nor expected to form without further design effort. It is particularly surprising that both partners in our hypercycle have no ability to self-copy. How could this have happened, and what is the evolutionary advantage of it?

Origin of the short partner

We need to explain how species 31, that is missing a key functional component, can rise to co-dominance in our system. We can trace the ancestry of the molecular species, and examine the reaction networks at key stages in any trial (figure 7). A white box indicates that a new species is synthesised *de novo* in the reaction, whereas a grey box indicates that the new species arises by modification of one of the reactants. Replicase molecules should act as catalysts, remaining unchanged when they emerge from a reaction. We can conclude that there is something in the reaction with molecules of species 29 that has produced species 30, which then reacts with species 9 to form species 31. The single

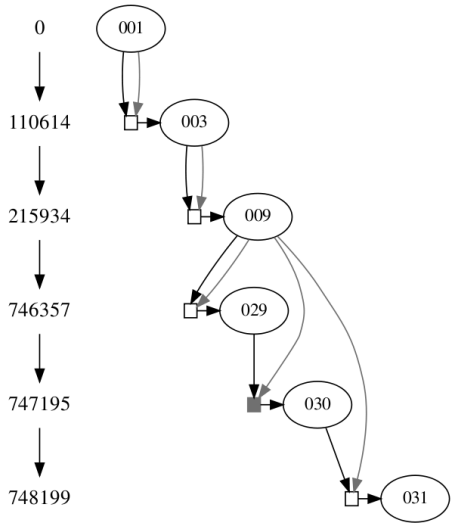

Figure 7: Ancestry of species 31. Numbers on the left indicate the time of reaction. Black arrows indicate the active partner. Grey arrows indicate the passive partner

point mutation of species 9 to create species 29 is shown below by a vertical line:

```
009 OBEQBX...LHHHRLUEUOBLROORE$BLUBOˆB>C$=?>$$BLUBO%}OYHOB
                                |
029 OBEQBX...LHHHRLUEUOBLROORE$BLUBPˆB>C$=?>$$BLUBO%}OYHOB
```

The subsequence `$BLUBO` has mutated to `$BLUBP`. The `$` symbol is a code for "seek", and (in this situation) positions the molecule's flow pointer at the end of the best complementary alignment for the sequence `BLUBO`, which is the sequence `OYHOB`. With the mutation in species 29, the alignment spans only the first four letters of `$BLUBO`, so the copy of the molecule is constructed one symbol in from the end of the molecule. When the construction is complete, the newly-created string must be cleaved from the active molecule's sequence. The pointers are arranged to achieve this via a second "seek" command with the same target (`OYHOB`). However, since the target has been overwritten in the original molecule, the seek command positions the pointer at the *end* of the newly copied molecule instead. The "cleave" command is applied to the far end of the string and is thus ineffective. The reaction-program terminates, and the new molecule (species 31) is created from most of a molecule of species 29 with a copy of species 9 pasted over the penultimate symbol.

In this manner, the reaction between species 29 and 9 creates species 30, which is nearly twice as long as the seed replicase, as shown in figure 8. Note there is only ever a single molecule of species 29, which is immediately transformed into species 30 when it reacts with a molecule from species 9. When species 9 binds to species 31, the bind site is shifted to a new position, as shown in figure 8. This changes the action of the replicase program such that the first 14 characters of the string are not copied. In this way,

```
030 OBEQBXUUUDYGRHBBOSEOLHHHRLUEUOBLROORE$BLUBPˆB>C$=?>$$BLUBO%}OYHOOBEQBXUUUDYGRHBBOSEOLHHHRLUEUOBLROORE$BLUBOˆB>C$=?>$$BLUBO%}OYHOB
Bind site:                                                  |--------------|
009                                                 OBEQBXUUUDYGRHBBOSEOLHHHRLUEUOBLROORE$BLUBOˆB>C$=?>$$BLUBO%}OYHOB
Product:                                                    |-----------------------------------------------|
031                                                         BBOSEOLHHHRLUEUOBLROORE$BLUBOˆB>C$=?>$$BLUBO%}OYHOB
```

Figure 8: Origin of species 31

the single instance of species 30 can create many molecules of species 31 until it decays. Species 31 is then copied by dominant species in the system in 50% of reactions with it. Note that this cascade of reactions all occurs as a result of the single-point mutation on species 9.

Evolutionary pressure towards a hypercycle

Having established how a shorter molecule can arise via single-point mutations, we need to investigate how the molecule persists in the system, and what evolutionary pressure there is towards the formation of a hypercycle. It is important to note that in our replicase system a molecule that ensures it will always act as the template in a reaction is likely to sweep the population, as it will increase in numbers whenever it binds to another molecule. This is often achieved by *reducing* the bind probability for self-self reactions: as long as a bind is sufficiently likely, all the energy available in the system can be consumed. Binds stronger than this critical value have no advantage, whereas increasing any bias towards becoming the template in a reaction is clearly advantageous. For single-replicase systems, this is straightforward to understand, but with the introduction of species 31, the dynamics get more interesting.

Once present in the system, species 31 becomes a resource for other molecules. In all of the reactions with species 31, the chances of acting as a template are 50-50 (since the position of the alignment is the same on each string). This means that new species that bind to 31 can use it as a resource for increasing their number, even though half the time they will be exploited by species 31 to maintain its own population. Through a series of sweeps, each new dominant species binds increasingly strongly to species 31, thus flushing the previous incumbent from the system. Any new species that binds *less* strongly to species 31 than the previous dominant species is unsuccessful: it loses in the competition to exploit a valuable resource. Once bind affinity to species 31 is maximised, the old strategy of weakening self-self binds to guarantee template status in a reaction takes over again.

These processes are illustrated in figure 9, which plots binding rates for new dominant species in trial 277. The plots show the changes in bind probabilities with each successive sweep of the population as illustrated in figure 5. The line labelled "Bind to self" shows the probability of self-self binding for each new dominant species. The line labelled "Bind to 31" shows the bind probability between the new dominant species and species 31. There are three phases.

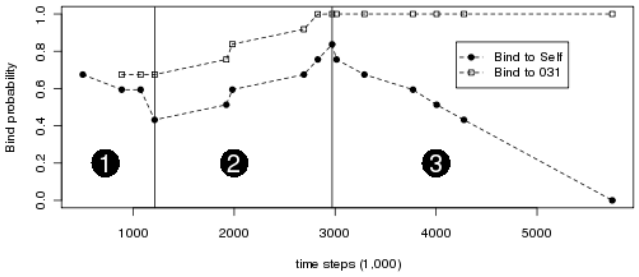

Figure 9: Change in binding rates as a precursor to hypercycle emergence

The first phase shows a decrease in self-binding probability between successive dominant species. We then see a second phase in which new species have an increasing affinity for binding to molecule 31. Once this is maximised, the third phase begins, in which successive dominant species sacrifice their self-bind probability to ensure they act as templates when reacting with the previous dominant species. In this way, dependence upon species 31 increases, until self-replication disappears altogether, and a hypercycle emerges.

The single-point mutations between dominant species are shown in figure 10. It shows that all mutations that confer an advantage occur in the binding regions of the molecule. Phases 1 and 3 of the run show changes in the second bind region, whereas phase 2 shows mutations in the first bind region. This corresponds with the change in phase noted for figure 9. The functional region of the molecule, which occupies the last half of the string, is preserved throughout. This is far from a random walk: the critical function of the replicase is preserved throughout, whilst a continual turnover of the binding site sequences illustrates the evolutionary pressure on the molecular species to act as a template for the molecule that the replicase builds.

Conclusions

We have presented an evaluation of the effect of mutation on an open-ended chemical system. The richness of behaviour we have shown is striking; indeed it was unexpectedly rich given that the only form of mutation is single-point. The need for such richness in complex systems was one of our main considerations during the design of this system. In addition, our chemistry reveals something of the dynamics of replicase systems that is very difficult to observe in biology. The decrease in binding affinity was not predicted, and the mechanism by which the hypercycle emerged was the result

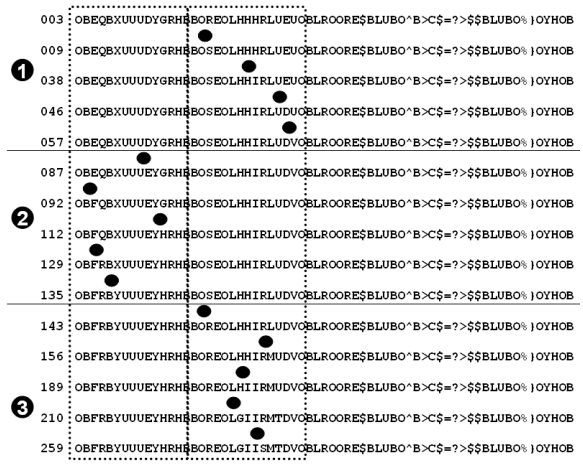

Figure 10: Mutations for the dominant species in run 277. Bind sites are indicated with dashed lines.

of a macromutation that was not "designed in" to the system.

Our replicase molecules are "imperfect replicators": they have a small chance of making an error when copying anything that binds to a certain region on the molecule. The imperfections in the copy process are not currently encoded on the genome; they are preset in the microcode of the copy instruction and thus unavailable for manipulation on the genome. In future work, we could represent the copy instruction at a finer level of granularity and use template codes to specify the accuracy of each sub operation, possibly including some cost for an increased accuracy of copy. We observed macro-mutations arising as a result of single-point changes that delivered emergent phenomena due to the wide heritable range of the system.

Finally, we must emphasise that these trials form a control experiment in which the effects of single-point mutation were evaluated. Future work will examine the effects of running a "population" of these trials, such that when a population of molecules collapses in an individual container, it can be replenished by a neighbour. This gives us a full model of early life, in which replicating templates and machinery self-maintain within membrane-bounded containers that can be replenished by neighbours.

Acknowledgements

This work was supported by EPSRC grant EP/F031033/1.

References

[1] Eigen, M. (1992). *Steps Towards Life: A Perspective on Evolution*. Oxford University Press.

[2] Eigen, M. and Schuster, P. (1977). The hypercycle, a principal of natural self-organization. part a: Emergence of the hypercycle. *Die Naturwissenschaften*, 64:541–565.

[3] Hickinbotham, S., Clark, E., Stepney, S., Clarke, T., Nellis, A., Pay, M., and Young, P. (2010). Specification of the stringmol chemical programming language version 0.2. Technical Report YCS-2010-458, Univ. of York.

[4] Hickinbotham, S., Clark, E., Stepney, S., Clarke, T., Nellis, A., Pay, M., and Young, P. (in press). Molecular microprograms. In *ECAL '09, LNCS 5777*.

[5] Hickinbotham, S., Clark, E., Stepney, S., Clarke, T., and Young, P. (2009). Gene regulation in a particle metabolome. In *CEC '09*, pages 3024–3031. IEEE Press.

[6] Hofstadter, D. R. (1979). *Godel, Escher, Bach: An Eternal Golden Braid*. Penguin.

[7] Johnson, T. J. and Wilke, C. O. (2004). Evolution of resource competition between mutually dependent digital organisms. *Artif. Life*, 10(2):145–156.

[8] Laing, R. (1977). Automaton models of reproduction by self-inspection. *J. of Theor. Biol.*, 66(3):437 – 456.

[9] Maynard-Smith, J. and Szathmary, E. (1997). *The Major Transitions in Evolution*. Oxford University Press.

[10] Ono, N. and Suzuki, H. (2002). Universal replication in a string-based artificial chemistry system. *Journal of Three Dimensional Images*, 16(4):154–159.

[11] Ray, T. (1991). An approach to the synthesis of life. In *Artificial Life II*, pages 371–408. Addison-Wesley.

[12] Smith, T. F. and Waterman, M. S. (1981). Identification of common molecular subsequences. *J Mol Biol*, 147(1):195–197.

[13] Turing, A. M. (1937). On computable numbers. *Proc. Lond. Math. Soc.*, 2:230–265.

[14] Varetto, L. (1998). Studying artificial life with a molecular automaton. *J Theor Biol*, 193(2):257–85.

[15] von Neumann, J. (1966). *Theory of self-reproducing automata*. University of Illinois Press.

[16] Wright, B. E. (2000). A biochemical mechanism for nonrandom mutations and evolution. *J Bacteriol*, 182(11):2993–3001.