The Origin of Informational Replicators by Serial Dilution of a Primordial Soup

Chrisantha Fernando[1,2,3]

[1]Center for Computational Neuroscience and Robotics, University of Sussex, Falmer, Brighton, BN1 9RH
[2]Collegium Budapest, 1014 Budapest, Szentháromság utca 2, Hungary
[3]The National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, UK

Abstract

How can informational replicators (Zachar and Szathmáry 2010) such as template replicators, arise from non-informational autocatalysts (Szathmáry and Maynard Smith 1997; Szathmary 2000)? Variants of an informational replicator have a high probability of being autocatalytic, thus allowing potentially unlimited heritable variants to be replicated, for example, mutants of a DNA sequence have this property. Variants of non-informational replicators such as glycolaldehyde in the Formose cycle are *not* in general autocatalytic; therefore, there is little capacity for hereditary variation (Szathmáry 2006). This paper asks; what are the necessary and sufficient conditions for an increase in the probability that a variant of an autocatalyst will itself be capable of autocatalysis? Given some well-defined assumptions, serial dilution in a rich generative chemistry such as that found in the Miller experiment should result in the emergence of informational replicators, i.e. autocatalysts whose variants have a high probability of themselves being capable of autocatalysis.

Introduction

A reactor such as that of Millar's famous experiment (Miller 1953) contains reactions that are simple autocatalytic cycles (and probably more complex kinds of autocatalytic structure, e.g. reflexive autocatalytic sets (Farmer, Kauffman et al. 1986; Kauffman 1986)). An example of a simple autocatalytic cycle is the Formose reaction (Fernando, Santos et al. 2005), see Figure 1. It is known that this autocatalytic cycle is notoriously subject to side-reactions, the reaction of molecules external to the cycle with the intermediates of the cycle to produce new molecules. Some of these new molecules will themselves be autocatalytic with some probability p that we assume is a property of the parental autocatalytic cycle. The same fate of side-reactions befalls these newly produced autocatalysts.

Real chemistry is very complicated, but it is possible to get some idea of the dynamics of a growing chemical network of reactions by using simplified artificial chemistries. A typical abstraction is to use linear binary strings as molecules and allow ligation and cleavage reactions between these strings. This paper will use an even simpler artificial chemistry where a chemical is described by only two parameters. What is the motivation for this? In simulations carried out previously using a artificial chemistry (Fernando and Rowe 2007; Fernando and Rowe 2008) it was observed that the probability of an autocatalytic molecule producing another autocatalytic molecule in a side-reaction decreased with the size of the molecule. This is an inevitable consequence in a random chemistry of linear strings because longer strings are less likely to produce two copies of one reactant by chance, than are shorter strings, given random rearrangement of the monomers of in a bimolecular rearrangement reaction (the type used in the simulation). The reality for real organic molecules is of course much more complicated. Some classes of autocatalytic molecule will inevitably be more likely to produce autocatalysts than others (i.e. have different p values). The complexity of the chemical models that would be needed to determine these probabilities for various classes of molecule are bewildering and possibly beyond that which is currently feasible. Therefore, a model is presented that abstracts certain properties of this generative chemical process. The model assumes simply that an autocatalyst can be described by a small number of parameters. Firstly, a probability p that a side-reaction to the autocatalytic cycle produces an autocatalyst. Secondly, a structural parameter $\bar{p}$ that describes the mean of a lognormal distributed set of values from which is drawn the probability p' that an autocatalyst produced by a side-reaction will be capable of itself producing autocatalysts, see Figure 1. Thirdly, for some variants of the model it is assumed that each autocatalyst has some observable property f. f is drawn randomly for each autocatalyst from a normal distribution with mean 0 and s.d. = 1. In the models, there is no correlation between f of a parent and f of an offspring molecule. This f is intended to be some function that may contribute to fitness at a higher level.

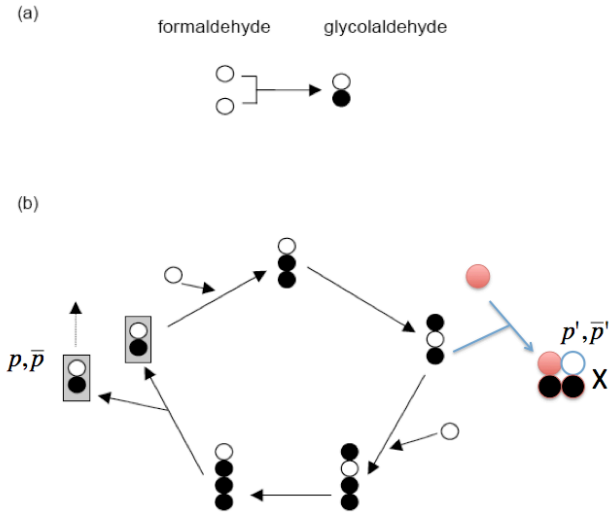

Figure 1 (Top) The molecule glycolaldehyde is autocatalytic using Formaldehyde as food, and making copies of itself, and growing in concentration exponentially. **(Bottom)** The intermediates of the glycolaldehyde autocatalytic cycle can undergo side-reactions with other species (red) to produce autocatalysts with a low probability *p*. Let these new autocatalysts have a probability *p'* of producing autocatalysts themselves by side-reactions. In the model, a structure parameter of the parental autocatalyst $\overline{p}$ determines stochastically the actual value of *p'* that an offspring autocatalyst will have.

Methods

A reactor is initialized with one core autocatalyst that has *p* drawn from a lognormal distribution with mean $\overline{p} = e^{-10}$. This is a small value, e.g. 0.001. The production of novel autocatalysts is simulated using a discrete time simulation. At each time-step, each existing autocatalyst has a probability *p* of producing another autocatalyst. If it does produce an autocatalyst, then this new autocatalyst has its *p'* value assigned by choosing a random number from the lognormal distribution defined by the $\overline{p}$ value of the "parent" autocatalyst. The new autocatalyst then has its $\overline{p}$' value defined based on the original $\overline{p}$ value of its parent. The crucial question in any realistic chemistry is whether there is a correlation between the $\overline{p}$ value of a parental autocatalyst and the $\overline{p}$' value of the autocatalyst produced from it. In other words, is the probability of producing an autocatalyst in a side-reaction a heritable parameter; is $\overline{p}$ heretable? It is clearly the case that there is no such simple correlation for all classes of molecule, although for some molecules there clearly is, for example, polymer template replicators. Such molecules have a very high probability that a variant will also be capable of replication. Several functions that relate the heretability $\overline{p}$ of parent and heritability $\overline{p}$' of offspring are examined in this paper. The simplest function assumes correlated $\overline{p}$ values where the $\overline{p}' = Norm(1,\sigma)\overline{p}_p$, where *Norm* is a Gaussian random number with mean 1 and standard deviation σ. An uncorrelated function is one in which $\overline{p}_c = e^{rand(-10,-9.5)}$, where rand(-10,-9.5) is a uniform random number between -10 and -9.5, the typical values evolved in the previous experiments when $\overline{p}$ was an evolvable parameter.

The reactor produces autocatalysts for a fixed time period T after which *M* random samples (containing autocatalysts) are taken from the reactor. Each autocatalyst has some probability *q* of being chosen for each sample, and let this value be fixed throughout a simulation. Let the chance of choosing an autocatalyst be low, e.g. 5%. In reality this probability *q* will depend on abundance, but here we have no model of chemical kinetics. Also, we do not allow the number of autocatalysts chosen to exceed some maximum *C* e.g. *50.* The sample will also inevitably contain many non-autocatalytic molecular species that are not modeled here.

One of the *M* samples are chosen based on maximizing the linear sum of *f* values of the autocatalyst species present in the reactor. Another valid option is just to choose a random sample. Both options are modeled here. The chosen sample then is used to reinitialize a new reactor. All autocatalysts not present in this sample are discarded. This is the serial dilution phase of the experiment.

Results

Correlated $\overline{p}$

Figure 2 shows the results obtained for a run in which selection is for highest summed *f*. The initial value of $\overline{p} = e^{-10}$, *q = 0.05, C = 50, M = 10.* In the function $\overline{p}' = Norm(1,\sigma)\overline{p}_p$, σ *= 0.001,* i.e. there are small correlated changes to the potential to produce autocatalysts (P1 in the diagram).

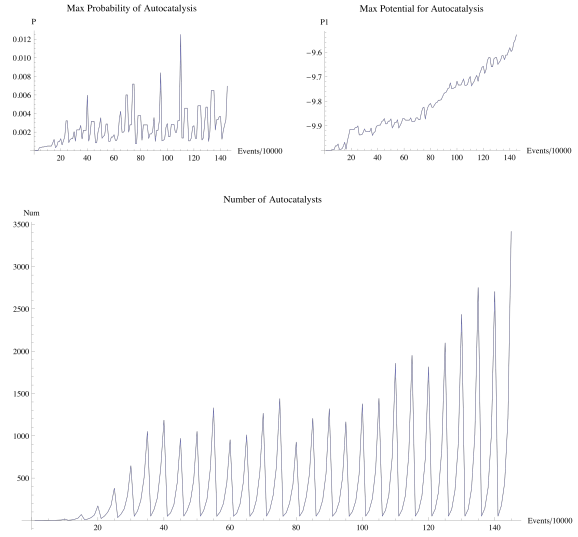

Figure 2. 28 serial dilutions, with selection for highest *f* sample. **(Top Left)** Maximum *p* value obtained. **(Top Right)** Maximum $\overline{p}$ value obtained. **(Bottom)** Total number of autocatalysts in the reactor.

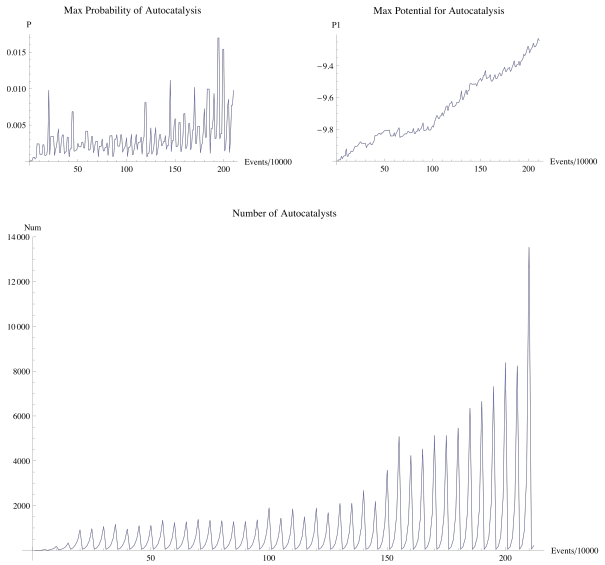

Figure 3. 34 serial dilutions, with random selection of a compartment. **(Top Left)** Maximum p value obtained. **(Top Right)** Maximum $\bar{p}$ value obtained. **(Bottom)** Number of autocatalysts.

After 28 serial dilutions of the system, the maximum value of $\bar{p}$ has increased significantly, and more autocatalysts are being produced in each round of network growth. Random compartment selection has a similar effect, see Figure 3. Selection for the compartment with the largest number of autocatalysts also has a similar effect (not shown). Next we consider the effect of making $\bar{p}$ a non-heritable structural parameter.

Uncorrelated $\bar{p}$

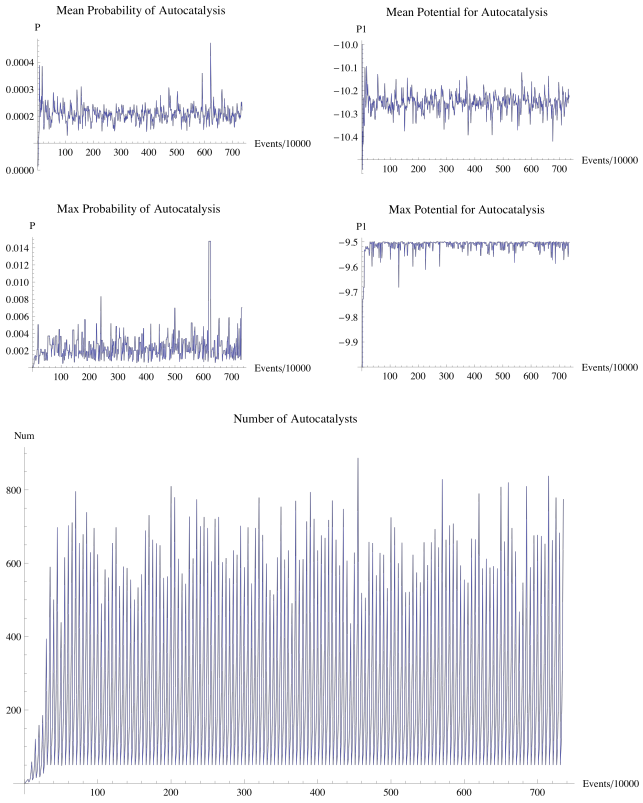

Figure 4. With uncorrelated $\bar{p}$**,** there is no improvement in autocatalysts over many serial dilutions. Selection is for the compartment with the largest number of autocatalysts.

Figure 4 shows the behaviour with an entirely uncorrelated potential for autocatalysis between successive autocatalysts $\bar{p}_c = e^{rand(-10,-9.5)}$. The probability p no longer tends to higher values because whilst a parental catalyst may occasionally produce an offspring with high p, this offspring has no tendency to itself produce offspring with high p. There exist autocatalysts in the population that do have high values of $\bar{p}$**,** but the mean value of $\bar{p}$ does not increase, as can be seen in the plot of mean $\bar{p}$ in Figure 4.
These results suggest that if the structural variability parameter $\bar{p}$ is not capable of being inherited, then there will be no tendency for the population of autocatalysts to tend towards becoming informational replicators.

Conclusions

The simple but fundamental principle demonstrated above is an example of the evolution of evolvability (Conrad 1990; Clune, Misevic et al. 2008; Parter, Kashtan et al. 2008), namely, that natural selection can act to select variants that are not of immediate benefit to the individual replicator, but confer improved variability properties, i.e. increase the chance that offspring will be fit. If there is variation (within generation differences) in variability (the capacity to produce variants during propogation) then there can be selection for variability properties that are beneficial to the lineage. This has been called lineage selection (Aboitiz 1991), and second order selection (Tenaillon, Taddei et al. 2001). Mark Toussaint has formalized the process of structuring phenotypic exploration distributions (Toussaint 2003) due to non-trivial neutrality, i.e. the capacity for the same phenotype p to be due to different genotypes $\bar{p}$**.** If some genotypes $\bar{p}$ tend to produce better variations in the phenotype p then those genotypes can be selected for. In this model it is shown that the capacity for non-trivial heritable neutral variation of $\bar{p}$ can allow increasing p.

The question remains, in chemistry, is there ever a circumstance in which $\bar{p}$ could be heritable within a lineage of autocatalysts? A conservative answer is sometimes yes, sometimes no. However, in this situation, the network dynamics would exhibit a tendency to select for that class of autocatalyst that did exhibit heredity of $\bar{p}$**.**

It is therefore proposed that experimentally it would be a matter of acute interest to take a rich generative chemistry such as that of Miller capable of producing a combinatorial explosion of polymers, and to take samples from the reactor once it had had a chance to generate this molecular diversity. These samples (selecting for the sample with the highest number of autocatalysts if possible) would be used to inoculate a new reactor. This cycle would be repeated for as many generations as possible. Each epoch should permit the generation of a new set of autocatalysts. This simple model predicts that such a protocol should be capable of generating informational replicators.

There are several simplifying assumptions of this model that must be examined. First we have ignored the fact that mass is finite. This means that exploration of the autocatalytic network may become limited if the mass of the reactor is used up producing non-autocatalytic molecules. Secondly we have completely ignored the existence of cross-catalytic interactions which may produce reflexive autocatalytic structures that can act as informational units. However, reflexive structures are only an intermediate step in what must eventually be selection for heretable $\bar{p}$ in the origin of

microevolution from macroevolution. An interesting addition to the model would be to allow species to be both autocatalytic and cross-catalytic with some probability. The interactions of the reactor would be described by a replication matrix. Adding a new species would involve producing a new row and column in this matrix. In addition to this matrix, each species would be described by structural parameters that determined the entries in the new row and column of the replication matrix for species that were produced in side reactions with it. Thirdly, the form of the structural parameter $\bar{p}$ (acting as a mean of a lognormal distribution to produce p') is somewhat arbitrary. A much more realistic method of describing the structural tendency for autocatalysis would be desirable.

Recent work by Ben Davis's group in Oxford has succeeded in enclosing a Formose cycle metabolism within lipid compartments. They are able to select for those compartments with certain chemical compositions (Gardner, Winzer et al. 2009). This paper is of some significance to them. If they were to simply choose small samples of each compartment and continue to test each sample for distinct autocatalytics, we predict that over many generations, one should find a greater diversity of independent autocatalysts.

Acknowledgements. Thanks to a Marie Curie Fellowship to work at Collegium Budapest, and the FP7 E-FLUX grant for evolutionary microfluidics.

Aboitiz, F. (1991). "Lineage selection and the capacity to evolve." Medical Hypotheses **36**: 155-156.

Clune, J., D. Misevic, et al. (2008). "Natural Selection Fails to Optimize Mutation Rates for Long-Term Adaptation on Rugged Fitness Landscapes." PLoS Computational Biology **4**(9): e1000187.

Conrad, M. (1990). "The geometry of evolution." BioSystems **24**(61-81).

Farmer, J. D., S. A. Kauffman, et al. (1986). "Autocatalytic Replication of Polymers." Physica D **22**: 50-67.

Fernando, C. and J. Rowe (2007). "Natural Selection in Chemical Evolution. ." Journal of Theoretical Biology **247**: 152-167

Fernando, C. and J. Rowe (2008). "The origin of autonomous agents by natural selection. ." Biosystems **91**: 355-373.

Fernando, C., M. Santos, et al. (2005). "Evolutionary potential and requirements for minimal protocells. ." Top. Curr. Chem. **259**: 167-211.

Gardner, P. M., K. Winzer, et al. (2009). "Sugar synthesis in a protocellular model leads to a cell signalling response in bacteria." Nature Chemistry **1**: 377-383.

Kauffman, S. A. (1986). "Autocatalytic Sets of Proteins." Journal of Theoretical Biology **119**: 1-24.

Miller, S. L. (1953). "Production of Amino Acids Under Possible Primitive Earth Conditions." Science **117**: 3046.

Parter, M., N. Kashtan, et al. (2008). "Facilitated Variation: How Evolution Learns from Past Environments to Generalize to New Environments." PLoS Computational Biology **4**(11): e1000206.

Szathmary, E. (2000). "The evolution of replicators. ." Phil. Trans. Roy. Soc. Lond. B **355**: 1669–1676.

Szathmáry, E. (2006). "The origin of replicators and reproducers." Philos. Trans. R. Soc. London. B. Biol. Sci. **361**(1474): 1761-1776.

Szathmáry, E. and J. Maynard Smith (1997). "From replicators to reproducers: the first major transitions leading to life." Journal of Theoretical Biology **187**(4): 555-571.

Tenaillon, O., F. Taddei, et al. (2001). "Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation." Res. Microbiol **152**: 11-16.

Toussaint, M. (2003). The evolution of genetic representations and modular adaptation. ND 04, 44780 Bochum---Germany.

Zachar, I. and E. Szathmáry (2010). "A New Replicator: A theoretical framework for analysing replication." BMC Biology **8**: 21.