



Prelife Catalysts and Replicators

Citation

Ohtsuki Hisashi, Martin A. Nowak. 2009. Prelife catalysts and replicators. Proceedings of the Royal Society B 276(1674): 3783-3790

Published Version

doi:10.1098/rspb.2009.1136

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:3804485>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Prelife catalysts and replicators

Hisashi Ohtsuki^{1,2,*} & Martin A. Nowak³

¹PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan

²Department of Value and Decision Science, Tokyo Institute of Technology, Tokyo 152-8552, Japan

³Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology,
Department of Mathematics, Harvard University, Cambridge MA 02138, USA

* Corresponding author (ohtsuki.h.aa@m.titech.ac.jp)

Life is based on replication and evolution. But replication cannot be taken for granted. We must ask what there was prior to replication and prior to evolution. How does evolution begin? We have proposed prelife as a generative system that produces information and diversity in the absence of replication. We model prelife as a binary soup of active monomers that form random polymers. Prevolutionary dynamics can have mutation and selection prior to replication. Some sequences might have catalytic activity and thereby enhance the rates of certain prelife reactions. We study the selection criteria for these prelife catalysts. Their catalytic efficiency must be above certain critical values. We find a maintenance threshold and an initiation threshold. The former is a linear function of sequence length, and the latter is an exponential function of sequence length. Therefore it is extremely hard to select for prelife catalysts that have long sequences. We compare prelife catalysis with a simple model for replication. Assuming fast template-based elongation reactions we can show that replicators have selection thresholds which are independent of their sequence length. Our calculation demonstrates the efficiency of replication and provides an explanation of why replication was selected over other forms of prelife catalysis.

Keywords: evolutionary dynamics, origin of life, evolution, replication, selection threshold, mathematical biology

Short title for page headings: Prelife catalysts

1 Introduction

The defining feature of biological systems is evolution. Biological organisms are products of evolutionary processes and are capable of undergoing further evolution. We think of the evolutionary process as modifying the traits of living systems. But how does evolution get started? How can we formulate a dynamical system that leads to the origin of evolution? What is there just before evolution begins? This paper is an extension of earlier work that tries to approach such questions (Nowak & Ohtsuki 2008, Manapat et al. 2009). In these papers, we have defined ‘prelife’ as a chemical system that can lead to information and diversity, that is capable of selection and mutation, but does not yet have replication. We have modeled prelife as a soup of active monomers, which can give rise to polymers. Here we assume that some polymers have catalytic activity: they increase the rate of certain reactions in prelife. We study the criteria for the selection of prelife catalysts. We compare prelife catalysts with replicators, that have the ability to make copies of themselves.

The origin of life is a transition from chemistry to biology. There have been many theoretical and empirical studies concerning the origin of life (Oparin 1953, Crick 1968, Orgel 1968, 1992, Eigen 1971, Eigen & Schuster 1982, Dyson 1982, 1999, Kuppers 1983, Stein & Anderson 1984, Farmer et al. 1986, Szathmary & Demeter 1987, Sievers & von Kiedrowski 1994, Fontana & Schuster 1998, Luther et al. 1998, Lifson & Lifson 1999, de Duve 2005, 2007). One line of research attempts to understand how chemical processes on early Earth can spontaneously synthesize the basic building blocks of life (Miller 1953, Allen & Ponnamperuma 1967, Miller & Orgel 1974, Hargreaves et al. 1977, Rao et al. 1982, Rushdi & Simoneit 2001, Benner et al. 2002, Ricardo et al. 2004, Benner & Ricardo 2005, Waechterhaeuser 2007). RNA has the ability to store genetic information and catalyze chemical reactions. Therefore, the proposal has been made that early life existed in an ‘RNA world’ (Orgel 1986, Joyce 1989, 2002, Ellington & Szostak 1990, Cech 1993, Johnston et al. 2001, Steitz & Moore 2003, Hughes et al. 2004). Bartel & Szostak (1993) discovered an RNA sequence that can catalyze RNA polymerization.

Some critics, however, argue that RNA is too complicated and fragile to arise spontaneously and

that the origin of life must have been based on simpler molecules, metabolic networks or compositional genomes (Shapiro 1984, 2006, 2007, Kauffman 1986, Morowitz et al. 1988, Segre et al. 1998, 2000). Sometimes this debate is called ‘RNA first’ versus ‘metabolism first’. Our own position is the following. All currently known biological organisms use RNA or DNA. At some time such a system must have evolved. Therefore, it is a valid program to investigate the principles that govern the emergence of a biological polymer which carries information. When this event took place, complicated chemical cycles must have been present, which generate the compounds needed for the biological polymers. In this sense, ‘metabolism first’ is certainly true, but an RNA-like system is needed for the emergence of genetic evolution.

A crucial step in the origin of life is the formation of the first cell (Szostak et al. 2001, Hanczyc et al. 2003, Chen & Szostak 2004a,b, Chen et al. 2004, 2005, Chen 2006). Fatty acids are simple molecules that can be synthesized under prebiotic conditions. They can self-assemble into bilayer vesicles, which can undergo growth and division. A decisive question is whether cells preceded information carrying polymers or vice versa. In the context of our theory, the ordering of these two events affects the population structure. If polymers came first, then their emergence can be studied in well-mixed populations. If cells came first, then the emergence of polymers should be studied in structured meta-populations containing ensembles of dividing sub-populations. From the perspective of mathematical analysis the logical first step is to study well-mixed populations (as we will do here) and later move to evolutionary dynamics in structured populations (Nowak & May 1992, Rousset 2004, Traulsen & Nowak 2006, Ohtsuki et al. 2006, Taylor et al. 2007, Tarnita et al. 2009).

Eigen & Schuster (1977, 1979) developed a hugely influential molecular theory of chemical evolution. Their quasispecies theory studies the competition of different replicators (McCaskill 1984, Eigen et al. 1989, Nowak & Schuster 1989, Nowak 1992). Hypercycles are cooperative interactions between two or more replicators. In contrast, our theory of prelife does not begin with the presence of replicators; instead we study mutation and selection prior to replication (Nowak & Ohtsuki 2008, Manapat et al. 2009). Therefore we study the origin of evolution and the competition between life (which is based on replication) and prelife (chemistry without replication). Fontana & Buss (1994ab) use λ calculus to study a generative chemistry with and without replication.

This paper is structured as follows. In Section 2 we present prelife and fully symmetric prelife. In Section 3 we discuss partial and perfect prelife catalysts. They give rise to hysteresis (bistability). In Section 4 we discuss a simple replicator. Section 5 is a brief summary of our findings.

2 Prelife

We consider two types of activated monomers, 0^* and 1^* . They are produced by prebiotic chemistry, and they decay at certain rates. They can also become deactivated to generate inactivated monomers, 0 and 1 . Activated monomers participate in co-polymerization reactions. Let i denote a binary string. We consider the following chemical reactions: $i + 0^* \rightarrow i0$ and $i + 1^* \rightarrow i1$. These chemical reactions can generate all binary strings. Inactivated monomers cannot be used for the elongation reactions, but they can react with active monomers; for example $0 + 1^* \rightarrow 01$.

The chemical kinetics of prelife are described by the following system of linear differential equations

$$\dot{x}_i = a_i x_{i'} - (d + a_{i0} + a_{i1}) x_i \quad i = 0, 1, 00, 01, \dots \quad (1)$$

The index i represents all binary strings (or sequences). The abundance of sequence i is denoted by x_i . Longer strings are produced from shorter ones by adding either a 0^* or a 1^* on the right side. Each string, i , has one precursor, denoted by i' , and two followers, denoted by $i0$ and $i1$ (see Fig. 1a). For example, 010 is the precursor of 0101 . The two followers of 0101 are 01010 and 01011 . For the precursors of strings 0 and 1 we set $x_{0'} = x_{1'} = 1$. The rate constants a_i denote the rate at which string i is formed from string i' by addition of an activated monomer (which is either 0^* or 1^*). Equation (1) assumes that the concentration of activated monomers are at constant steady state levels. This happens, for example, when the decay rate of activated monomers is larger than the rate at which they are used up in prelife reactions. In the following we think that the steady state density of activated monomers are already subsumed in rate constants. All strings are removed (decay) at rate d .

Prelife dynamics define a tree (more precisely a double tree) with the two roots, 0 and 1 . This ‘tree of prelife’ has infinitely many lineages (Fig. 1a). The half of all lineages starts from 0 , the other half starts from 1 . A lineage is a sequence of infinitely many strings that are followers of each other.

For example, one such lineage contains all all-0 strings: $0, 00, 000, \dots$. Another lineage contains alternating sequences (that start with 0): $0, 01, 010, 0101, \dots$. We could also consider prelife with more than two types of monomers, but this extension is not necessary for the purpose of this paper.

For fully symmetric prelife we assume $a_0 = a_1 = \lambda/2$ and $a_i = a$ for all other sequences, i . In this case, all sequences of length n have the same equilibrium abundance, $[\lambda/2a][a/(2a + d)]^n$. The total abundance of all strings is λ/d .

3 Prelife catalysis

Prelife catalysis means that some sequences have the ability to enhance the rates of certain prelife reactions. For example, sequence j might catalyze the reaction $i + 0^* \rightarrow i0$ at rate c (see Fig. 1b). In this case the rate of formation of sequence $i0$ can be written as $a_{i0}x_i + cx_ix_j$. The first term denotes the rate of the uncatalyzed reaction. The second term denotes the rate of the catalyzed reaction, which is proportional to the abundance of the catalyst, x_j . In a subsequent paper we plan to study sets of prelife catalysts, but here we focus on the dynamics of individual catalysts. We consider a prelife catalyst that enhances some (or all) of its upstream reactions (Fig. 1a). Our aim is to calculate the equilibrium abundance of such a catalyst. Therefore, we can study the conditions for selection of catalyzed over uncatalyzed prelife.

Let us consider fully symmetric prelife. Without loss of generality we assume that the catalyst is the all-0 sequence of length n , which we denote by 0^n . There are $n - 1$ upstream reactions in the lineage leading from 0 to 0^n . Each reaction, $0^k + 0^* \rightarrow 0^{k+1}$, is enhanced by c_k times the abundance of 0^n . The parameter c_k can be either zero or positive. In order to understand this system we study the abundances of sequences of the form 0^k , where $k = 1, 2, \dots$. We change our previous notation and let x_k denote the abundance of 0^k . We have the following system of ordinary differential equations:

$$\begin{cases} \dot{x}_1 = \lambda/2 - (2a + d)x_1 - c_1x_1x_n \\ \dot{x}_k = ax_{k-1} - (2a + d)x_k + c_{k-1}x_{k-1}x_n - c_kx_kx_n & (2 \leq k \leq n - 1) \\ \dot{x}_n = ax_{n-1} - (2a + d)x_n + c_{n-1}x_{n-1}x_n. \end{cases} \quad (2)$$

We are interested in the equilibrium abundance of the prelife catalyst, which we denote by \hat{x}_n . A

straightforward calculation shows that it is given as a root of the following polynomial equation:

$$x = \frac{\lambda}{2(2a + d)} \prod_{k=1}^{n-1} \frac{a + c_k x}{2a + d + c_k x}. \quad (3)$$

3.1 Partial catalysis

Imagine a prelife catalyst of length n that catalyzes m ($1 \leq m \leq n - 1$) of its $n - 1$ upstream reactions. For analytical simplicity we assume that c_k is either c or 0. That is, m entries of the vector (c_1, \dots, c_{n-1}) are c and the others are zero. In this case, the equilibrium abundance, \hat{x}_n , is given as a root of the equation

$$x = \frac{\lambda}{2(2a + d)} \left(\frac{a}{2a + d} \right)^{n-1-m} \left(\frac{a + cx}{2a + d + cx} \right)^m. \quad (4)$$

Note that equation (4) does not depend on which particular m reactions out of the $n - 1$ upstream reactions are enhanced. For a general c , equation (4) cannot be solved explicitly. Nevertheless, we obtain the following result. There exists a critical threshold of m , denoted by m_{cr} . If $m \leq m_{\text{cr}}$ then the equilibrium abundance, \hat{x}_n , is a monotone increasing function of the catalytic activity, c . If $m > m_{\text{cr}}$ then we observe a hysteresis effect: for an interval of intermediate c values, equation (4) has three positive roots; two of them correspond to stable equilibria and one to an unstable equilibrium. Which of the two stable equilibria is reached depends on the initial abundance of the catalyst. For a detailed analysis, see Appendix A.

3.2 Perfect catalysis

As a special case, let us study a sequence that enhances the rates of all of its upstream reactions. Therefore, we have $m = n - 1$. The equilibrium abundance, \hat{x}_n , is given as a root of the polynomial equation

$$x = \frac{\lambda}{2(2a + d)} \left(\frac{a + cx}{2a + d + cx} \right)^{n-1}. \quad (5)$$

For $c = \infty$, we obtain the maximum abundance, $\hat{x}_n = \lambda/2(2a + d)$ ($\equiv \hat{x}_n^{\text{max}}$). For a general c we obtain the following result. There exists a threshold for the length of the catalyst, n_{cr} . When $n \leq n_{\text{cr}}$ (Fig. 2), the equilibrium abundance \hat{x}_n is a monotone increasing function of c . When $n > n_{\text{cr}}$ (Fig. 3), we find the two branches of stable equilibria (the solid lines in Fig. 3a) and one unstable equilibrium

between them (the dotted line in Fig. 3a). The upper branch exists for $c \geq c_1$, while the lower branch exists for $c \leq c_2$. For $c_1 \leq c \leq c_2$, the equilibrium abundance, \hat{x}_n , depends on its initial abundance. If the catalyst is initially rare, then it will reach the lower equilibrium (Fig. 3b). If the catalyst is initially present at high abundance, then it will reach the higher equilibrium (Fig. 3c).

The first threshold, c_1 , is the critical value of c that is needed to maintain the catalyst at high abundance. The second threshold, c_2 , is the critical value that is needed to initiate high abundance of the catalyst when it is not common in the beginning. Therefore we call c_1 and c_2 ‘maintenance threshold’ and ‘initiation threshold’, respectively. For large n we obtain

$$\begin{aligned} c_1 &\approx \frac{2e(2a+d)(a+d)}{\lambda} \cdot n \\ c_2 &\approx \frac{2a^2(2a+d)}{e\lambda(a+d)} \cdot \left(\frac{2a+d}{a}\right)^n \frac{1}{n} \end{aligned} \quad (6)$$

Here $e = 2.718281 \dots$ (see Appendix B). The ‘maintenance threshold’, c_1 , grows as a linear function of the sequence length, n . The ‘initiation threshold’, c_2 , grows (approximately) as an exponential function of the sequence length, n . Therefore, it is extremely difficult to select for a catalyst that has a long sequence. At the same time it is unlikely that short sequences have good (or any) catalytic activity.

An intuitive biological summary is the following. The system has two equilibria, E_1 and E_2 . At E_1 the catalyst has low abundance; all sequences have almost the same abundances as in uncatalyzed prelife. At E_2 the catalyst has high abundance; it ‘dominates’ the population (see Figure 3). We say that at equilibrium E_2 the catalyst has been selected over uncatalyzed prelife. If the catalytic activity, c , is less than the threshold c_1 , then only E_1 is stable. If c is greater than c_2 , then only E_2 is stable. If c is between c_1 and c_2 then both equilibria are stable. Which one will be chosen depends on the initial condition. Therefore, if the prelife catalyst is already present at high abundance, then it will remain so as long as c is greater than c_1 . On the other hand, if the catalyst is initially not present at high abundance, then it will gain high abundance only if c is greater than c_2 . This ‘chemical hysteresis’ is caused by the bistability of our system.

4 Replication

4.1 The primer is a monomer

Imagine that a sequence i can make a copy of itself by using activated monomers. For fully symmetric prelife, we can once again assume without loss of generality that the replicator is the all-0 sequence of length n , denoted by 0^n . The replication starts from the primer, 0, and incorporates activated monomers 0^* for elongation.

The difference between the perfect prelife catalyst and the replicator is the following. The prelife catalyst can attach to a sequence and increase the rate at which the activated monomer is added. Afterwards the catalyst dissociates from the elongated sequence. In contrast, the replicator attaches to a primer and then holds on to the growing sequence. Therefore the catalytic activity of the replicator can ‘walk along’ the entire sequence. In both cases we assume that the catalyzed elongation step is not rate limiting. Consequently for the replicator a single rate limiting bimolecular reaction is sufficient (attaching between template and primer). For the perfect prelife catalyst we need $n - 1$ rate limiting bimolecular reactions. See Figure 1b.

As before, let x_k be the abundance of the sequence in the form of 0^k ($k = 1, \dots, n$). The consumption of primers is described by the term $-rx_1x_n$. If we assume perfect replication, two copies of replicators are produced from one primer and one replicator. Therefore the production of replicators is described by the term rx_1x_n . In a general case, we obtain the following system of differential equations:

$$\begin{cases} \dot{x}_1 = \lambda/2 - (2a + d)x_1 - rx_1x_n \\ \dot{x}_k = ax_{k-1} - (2a + d)x_k & (2 \leq k \leq n - 1) \\ \dot{x}_n = ax_{n-1} - (2a + d)x_n + \delta rx_1x_n. \end{cases} \quad (7)$$

Here the parameter δ represents the efficacy of replication. A perfect replication leads to $\delta = 1$. If replication is always unsuccessful we have $\delta = -1$, because replicators are consumed in vain. In general, δ takes a value between -1 and 1. In Appendix C, we provide a derivation of eq.(7) by examining the detailed mechanism of the replication process. A key assumption there is that the template-based elongation steps are not rate limiting. In the following we study $\delta > 0$, otherwise replicators are never selected.

From eq.(7) it is easy to see that the equilibrium abundance of the replicator, \hat{x}_n , is given as the positive root of the following quadratic equation:

$$2r(2a+d)x^2 + \{2(2a+d)^2 - \delta r\lambda\}x - \lambda a \left(\frac{a}{2a+d}\right)^{n-2} = 0. \quad (8)$$

For large r we obtain $\hat{x}_n^{\max} = \delta\lambda/2(2a+d)$, which agrees with \hat{x}_n^{\max} in the case of $c = \infty$ for prelife catalysts (see Section 3.2), but up to the factor δ . However, the dependence of the equilibrium abundance on r is qualitatively different from that on c in prelife catalysts. It is shown that if the efficacy of replication exceeds $\delta^* = (\frac{a}{2a+d})^{n-1}$ the equilibrium abundance \hat{x}_n monotonically increases with r . Bistability is never observed (Fig. 4). There exists a critical threshold of r given by

$$r^* = \frac{2(2a+d)^2}{(1-f)\delta^2\lambda} \left\{ \delta - \frac{1}{f} \left(\frac{a}{2a+d}\right)^{n-1} \right\}. \quad (9)$$

If $r > r^*$ holds, the equilibrium abundance of the replicator is more than a fraction f ($0 < f < 1$) of its theoretical maximum, *i.e.* $\hat{x}_n > f\hat{x}_n^{\max}$. Interestingly, the threshold eq.(9) converges to a fixed value, $\frac{2(2a+d)^2}{(1-f)\delta\lambda}$, for large n . In contrast to prelife catalysts, long replicators can be selected over prelife.

4.2 The primer is not a monomer

Now we consider a scenario where the primer of replication is not a monomer, but a sequence of length $\ell (> 1)$. As before, suppose that the replicator is 0^n . The primer of the replication is given by 0^ℓ ($1 < \ell < n$). Replication is described by the term $rx_\ell x_n$. Taking into account the efficacy of replication, we obtain the following system of differential equations:

$$\begin{cases} \dot{x}_1 = \lambda/2 - (2a+d)x_1 \\ \dot{x}_k = ax_{k-1} - (2a+d)x_k & (2 \leq k \leq \ell-1) \\ \dot{x}_\ell = ax_{\ell-1} - (2a+d)x_\ell - rx_\ell x_n \\ \dot{x}_k = ax_{k-1} - (2a+d)x_k & (\ell+1 \leq k \leq n-1) \\ \dot{x}_n = ax_{n-1} - (2a+d)x_n + \delta rx_\ell x_n. \end{cases} \quad (10)$$

A calculation shows that the equilibrium abundance of the replicator, denoted by \hat{x}_n , is given by the positive root of the quadratic equation

$$2r(2a + d)x^2 + \left\{ 2(2a + d)^2 - \delta r \lambda \left(\frac{a}{2a + d} \right)^{\ell-1} \right\} x - \lambda a \left(\frac{a}{2a + d} \right)^{n-2} = 0. \quad (11)$$

The equilibrium abundance of replicators monotonically increases with r if and only if the efficacy exceeds

$$\delta^* = \left(\frac{a}{2a + d} \right)^{n-\ell}. \quad (12)$$

Therefore for a fixed length of the replicator, n , the required efficacy grows exponentially with the length of the primer, ℓ . The replicator that requires a longer primer is less likely to be selected.

Suppose eq.(12) holds. We obtain $\hat{x}_n = [\delta\lambda/2a] \cdot [a/(2a + d)]^\ell (\equiv \hat{x}_n^{\max})$ at $r \rightarrow \infty$. The critical threshold of the replication constant, denoted by r^* , is given by

$$r^* = \frac{2(2a + d)^2}{(1 - f)\delta^2\lambda} \left\{ \delta - \frac{1}{f} \left(\frac{a}{2a + d} \right)^{n-\ell} \right\} \left(\frac{2a + d}{a} \right)^{\ell-1}. \quad (13)$$

This threshold means that if $r > r^*$ then the equilibrium abundance of the replicator exceeds a fraction f ($0 < f < 1$) of its theoretical maximum, *i.e.* $\hat{x}_n > f\hat{x}_n^{\max}$. For a fixed primer length, ℓ , the threshold (13) tends to a constant, $r^* = \frac{2(2a+d)^2}{(1-f)\delta\lambda} \left(\frac{2a+d}{a} \right)^{\ell-1}$ for large n . Thus the critical threshold (13) converges to a fixed value for increasing n , which is consistent with the result found in Section 4.1. The intuitive explanation for this finding is that the catalyzed elongation steps of the replication process are not rate limiting. Therefore the length of the replicator does not affect the rate of replication.

5 Discussion

We have studied the selection criteria for prelife catalysts and replicators. By prelife catalysts we mean sequences that can enhance certain reactions in prelife. The perfect prelife catalyst is a (hypothetical) sequence that enhances the rates of all reactions in its own production lineage. We show that even for a perfect prelife catalyst it is very difficult to achieve a high equilibrium abundance, because the catalytic activity has to exceed a threshold value that grows exponentially with the sequence length. In contrast sequences that can replicate can achieve high equilibrium

abundance even if they have considerable length. The critical replication rate is almost independent of the length of the replicator. But the required efficacy of replication grows with the length of the primer.

Our selection thresholds arise, because there is competition between prelife and catalytic prelife, on one hand, and between prelife and replication (life), on the other hand. The latter is especially interesting because prelife is needed to build the sequences for replication (the replicator and the primer), but then prelife and life compete for the same resources (activated monomers). This tension between prelife and life leads to the origin of evolution.

Acknowledgments

Support from the John Templeton Foundation, the NSF/NIH joint program in mathematical biology (NIH grant R01GM078986), the Bill and Melinda Gates Foundation (Grand Challenges grant 37874), and J. Epstein is gratefully acknowledged.

Appendices

Appendix A: Thresholds for bistability in prelife catalysts

First we will study eq.(5). Equation (5) is rewritten as

$$c = \frac{1}{x} \left[\frac{a + d}{1 - \left\{ \frac{2(2a+d)}{\lambda} x \right\}^{1/(n-1)}} - (2a + d) \right]. \quad (14)$$

Therefore we can regard c as a single-valued function of x . Let $c(x)$ be the right hand side of eq.(14).

Its derivative with respect to x is

$$c'(x) = \frac{1}{x^2} \left[\frac{a + d}{(n-1)\xi^2} - \frac{n(a+d)}{(n-1)\xi} + (2a + d) \right], \quad (15)$$

where $\xi \equiv 1 - [2(2a + d)x/\lambda]^{1/(n-1)}$. As c is non-negative, from eq.(14) we need $0 < \xi \leq (a + d)/(2a + d)$. Solving $c'(x) = 0$ leads to

$$(n-1)(2a+d)\xi^2 - n(a+d)\xi + (a+d) = 0. \quad (16)$$

Let D be the discriminant of the quadratic equation of ξ , eq.(16). D vanishes at

$$n_{\text{cr}} \equiv \frac{2 \left(2a + d + \sqrt{a(2a + d)} \right)}{a + d}. \quad (17)$$

Also, D is strictly negative at $n = 2$. Thus, if $2 \leq n \leq n_{\text{cr}}$ then $D \leq 0$, which means that $c'(x)$ is always non-negative. Therefore $c = c(x)$ is a monotone increasing function of x , so is its inverse function $x = x(c)$. If $n > n_{\text{cr}}$ then $D > 0$, which means that eq.(16) has two distinct roots. We can prove that these two roots satisfy $0 < \xi < (a + d)/(2a + d)$. Therefore, $c = c(x)$ has one local maximum and one local minimum, leading to the S-shaped curve in Figure 3a.

Next we study eq.(4). Equation (4) can be rewritten in the same form as eq.(5) by setting $n' \equiv m + 1$ and $\lambda' \equiv \lambda[a/(2a + d)]^{n-1-m}$. Therefore similar conclusions can be drawn. If $n' > n_{\text{cr}}$, or equivalently, if

$$m > m_{\text{cr}} \equiv \frac{2 \left(2a + d + \sqrt{a(2a + d)} \right)}{a + d} - 1 \quad (18)$$

holds, then the system shows bistability. There are the ‘maintenance threshold’, c_1 , and the ‘initiation threshold’, c_2 .

Appendix B: Asymptotic values of c_1 and c_2

First we study a perfect catalyst which catalyzes all of its upstream reactions. When $n > n_{\text{cr}}$, solving eq.(16) yields

$$\xi_{\pm} = \frac{n(a + d) \pm \sqrt{n^2(a + d)^2 - 4(n - 1)(2a + d)(a + d)}}{2(n - 1)(2a + d)} \quad (\xi_- < \xi_+). \quad (19)$$

Remember that we have defined ξ as $\xi = 1 - [2(2a + d)x/\lambda]^{1/(n-1)}$, so $x_{\pm} = [\lambda/2(2a + d)](1 - \xi_{\pm})^{n-1}$. Note that $x_+ < x_-$. The function $c = c(x)$ has its local maximum at $x = x_+$ and its local minimum at $x = x_-$. We obtain $c_1 = c(x_-)$ and $c_2 = c(x_+)$ (see Fig. 3a). A direct calculation shows the asymptotic estimates of these values shown in the main text. We use $(1 + \frac{1}{n})^n \approx e = 2.718281\dots$ for large n .

Similarly, for a catalyst that catalyzes a fraction $\theta(= m/(n - 1))$ of its upstream reactions with

the catalytic activity c , we obtain the following asymptotic estimates of c_1 and c_2 for large n :

$$\begin{aligned} c_1 &\approx \frac{2\theta e(2a+d)^\theta a^{1-\theta}(a+d)}{\lambda} \cdot \left[\left(\frac{2a+d}{a} \right)^{1-\theta} \right]^n \\ c_2 &\approx \frac{2a^2(2a+d)}{\theta e \lambda (a+d)} \cdot \left(\frac{2a+d}{a} \right)^n \frac{1}{n}, \end{aligned} \quad (20)$$

where $0 < \theta \leq 1$. Therefore, the two thresholds grow (approximately) exponentially with n when the catalyst enhances some of its upstream reactions ($0 < \theta < 1$). Only when the catalyst enhances *all* of its upstream reactions ($\theta = 1$) does the maintenance threshold, c_1 , grow linearly with n .

Appendix C: A detailed derivation of equation (7) in the main text

Here we explain the underlying mechanics of replication and provide a detailed derivation of equation (7). Let 0^n denote the replicator. As in the main text, we denote the abundance of sequence 0^k by x_k ($k = 1, \dots, n$). We assume direct as opposed to complementary replication. The replication process starts when a (inactivated) monomer 0 , which is a primer, attaches to a replicator, which is a template. This reaction is described by the term $\alpha x_1 x_n$. The resulting complex between the template and the primer grows in length by incorporating activated monomers 0^* one by one until it becomes the full double strand of 0^n . We call these steps elongation reactions. Let y_k denote the abundance of the complex between the template (of length n) and the growing sequence which has reached length k . The abundance of the full double strand is given by y_n . For simplicity, we assume that the reaction rate of each elongation step is constant and given by β . The full double strand separates at rate γ (for example, via temperature oscillations). All sequences and complexes decay at rate d . We obtain the following system of differential equations:

$$\begin{cases} \dot{x}_1 = \lambda/2 - (2a+d)x_1 - \alpha x_1 x_n \\ \dot{x}_k = ax_{k-1} - (2a+d)x_k & (2 \leq k \leq n-1) \\ \dot{x}_n = ax_{n-1} - (2a+d)x_n - \alpha x_1 x_n + 2\gamma y_n \\ \dot{y}_1 = \alpha x_1 x_n - \beta y_1 - dy_1 \\ \dot{y}_k = \beta y_{k-1} - \beta y_k - dy_k & (2 \leq k \leq n-1) \\ \dot{y}_n = \beta y_{n-1} - \gamma y_n - dy_n. \end{cases} \quad (21)$$

Remember that the stationary density of activated monomers is subsumed in the rate constants, λ , a and β . We assume that the rate of template-based elongation, β , is much faster than other rate constants such as a , d and γ . For the quasi-equilibrium abundance of full double strands we obtain

$$\hat{y}_n = \frac{\alpha}{\gamma + d} x_1 x_n, \quad (22)$$

and therefore

$$2\gamma\hat{y}_n = 2\alpha\frac{\gamma}{\gamma + d}x_1x_n. \quad (23)$$

Substituting eq.(23) into the first three lines of eq.(21) yields

$$\begin{cases} \dot{x}_1 = \lambda/2 - (2a + d)x_1 - \alpha x_1 x_n \\ \dot{x}_k = ax_{k-1} - (2a + d)x_k & (2 \leq k \leq n - 1) \\ \dot{x}_n = ax_{n-1} - (2a + d)x_n + \frac{\gamma - d}{\gamma + d}\alpha x_1 x_n \end{cases} \quad (24)$$

Rewriting parameters as $r = \alpha$ and $\delta = (\gamma - d)/(\gamma + d)$ reproduces equation (7) in the main text.

We note that the assumption of fast elongation (large β) is entirely consistent with our model for prelife catalysis, which also contains an implicit assumption of a fast ‘elongation’ step. The prelife catalyst, 0^n , binds its target sequence, 0^k , to form a complex $[0^n 0^k]$. This complex reacts very fast with an activated monomer, 0^* , to give rise to $[0^n 0^{k+1}]$. Subsequently the complex dissociates into 0^n and 0^{k+1} . Equation (2) assumes that the elongation reaction is not rate limiting. Therefore a replicator with a fast elongation reaction is the proper comparison for the prelife catalyst described by eq.(2). The difference between the replicator and the prelife catalyst is the following: the catalytic activity of the replicator ‘walks along’ the sequence, while the prelife catalyst can accelerate only a single elongation step and subsequently dissociates.

References

- Allen, W.V., Ponnampereuma, C. 1967. A possible prebiotic synthesis of monocarboxylic acids. *Curr. Mod. Biol.* **1**, 24-28.
- Bartel D.P., Szostak J.W. 1993. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411-1418.
- Benner, S.A., Caraco, M.D., Thomson, J.M., Gaucher, E.A. 2002. Planetary biology-paleontological,

- geological, and molecular histories of life. *Science* **296**, 864-868.
- Benner, S.A., Ricardo, A. 2005. Planetary systems biology. *Mol. Cell.* **17**, 471-472.
- Cech, T.R. 1993. The efficiency and versatility of catalytic RNA: implications for an RNA world. *Gene* **135**, 33-36.
- Chen, I.A. 2006. The emergence of cells during the origin of life. *Science* **314**, 1558-1559.
- Chen, I.A., Roberts, R.W., Szostak, J.W. 2004. The emergence of competition between model protocells. *Science* **305**, 1474-1476.
- Chen, I.A., Salehi-Ashtiani, K., Szostak, J.W. 2005. RNA catalysis in model protocell vesicles. *J. Am. Chem. Soc.* **127**, 13213-13219.
- Chen, I.A., Szostak, J.W. 2004a. A kinetic study of the growth of fatty acid vesicles. *Biophys. J.* **87**, 988-998.
- Chen, I.A., Szostak, J.W. 2004b. Membrane growth can generate a transmembrane pH gradient in fatty acid vesicles. *P. Natl. Acad. Sci. U.S.A.* **101**, 7965-7970.
- Crick, F.H. 1968. The origin of the genetic code. *J. Mol. Biol.* **38**, 367-379.
- de Duve, C. 2007. Chemistry and selection. *Chem. Biodivers.* **4**, 574-583.
- de Duve, C. 2005. The onset of selection. *Nature* **433**, 581-582.
- Dyson, F. 1999. *Origins of life*, Cambridge, U.K./New York: Cambridge University Press.
- Dyson, F.J. 1982. A model for the origin of life. *J. Mol. Evol.* **18**, 344-350.
- Eigen, M. 1971. Molecular self-organization and the early stages of evolution. *Q. Rev. Biophys.* **4**, 149-212.
- Eigen, M., McCaskill, J., Schuster, P. 1989. The molecular quasi-species. *Adv. Chem. Phys.* **75**, 149-263.
- Eigen, M., Schuster, P. 1982. Stages of emerging life—five principles of early organization. *J. Mol. Evol.* **19**, 47-61.
- Eigen, M., Schuster, P. 1979. *The hypercycle, a principle of natural self-organization*, Berlin/New York: Springer-Verlag.
- Eigen, M., Schuster, P. 1977. The hyper cycle. A principle of natural self-organization. Part A: Emergence of the hyper cycle. *Naturwissenschaften* **64**, 541-565.

- Ellington, A.D., Szostak, J.W. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822.
- Farmer, J.D., Kauffman, S.A., Packard, N.H. 1986. Autocatalytic replication of polymers. *Physica D* **22**, 50-67.
- Fontana, W., Buss, L.W. 1994a. What would be conserved if 'the tape were played twice'? *P. Natl. Acad. Sci. U.S.A.* **91**, 757-61.
- Fontana, W., Buss, L.W. 1994b. 'The arrival of the fittest': toward a theory of biological organization. *B. Math. Biol.* **56**, 1-64.
- Fontana, W., Schuster, P. 1998. Continuity in evolution: on the nature of transitions. *Science* **280**, 1451-1455.
- Hanczyc, M.N., Fujikawa, S.M., JW Szostak, J.W. 2003. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science* **302**, 618-622.
- Hargreaves, W.R., Mulvihill, S., Deamer, D.W. 1977. Synthesis of phospholipids and membranes in prebiotic conditions. *Nature* **266**, 78-80.
- Hughes, R.A., Robertson, M.P., Ellington, A.D., Levy, M. 2004. The importance of prebiotic chemistry in the RNA world. *Curr. Opin. Chem. Biol.* **8**, 629-633.
- Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., Bartel, D.P. 2001. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319-1325.
- Joyce, G.F. 1989. RNA evolution and the origins of life. *Nature* **338**, 217-224.
- Joyce, G.F. 2002. The antiquity of RNA-based evolution. *Nature* **418**, 214-221.
- Kauffman, S.A. 1986. Autocatalytic sets of proteins. *J. Theor. Biol.* **119**, 1-24.
- Koppers, B. 1983. *Molecular theory of evolution: outline of a physico-chemical theory of the origin of life*, Berlin/New York: Springer-Verlag.
- Lifson, S., Lifson, H. 1999. A model of prebiotic replication: survival of the fittest versus extinction of the unfittest. *J. Theor. Biol.* **199**, 425-433.
- Luther, A., Brandsch, R., von Kiedrowski, G. 1998 Surface-promoted replication and exponential amplification of DNA analogues. *Nature* **396**, 245-248.

- Manapat, M., Ohtsuki, H., Bürger, R., Nowak, M. A. 2009 Originator dynamics. *J. Theor. Biol.* **256**, 586-595.
- McCaskill, J. 1984. A stochastic theory of macromolecular evolution. *Biol. Cybern.* **50**, 63-73.
- Miller, S.L. 1953. A production of amino acids under possible primitive earth conditions. *Science* **117**, 528-529.
- Miller, S.L., Orgel, L.E. 1974. *The origins of life on the earth*, Englewood Cliffs, N.J.: Prentice-Hall.
- Morowitz, H.J., Heinz, B., Deamer, D.W. 1988. The chemical logic of a minimum protocell. *Origins Life Evol. B.* **18**, 281-287.
- Nowak, M. A., Schuster, P. 1989. Error thresholds of replication in finite populations. Mutation frequencies and the onset of Muller's ratchet. *J. Theor. Biol.* **137**, 375-395.
- Nowak, M. A. 1992. What is a quasispecies? *Trends Ecol. Evol.* **7**, 118-121.
- Nowak, M. A., May, R. M. 1992. Evolutionary games and spatial chaos. *Nature* **359**, 826-829.
- Nowak, M. A., Ohtsuki, H. 2008. Prevolutionary dynamics and the origin of evolution. *P. Natl. Acad. Sci. U.S.A.* **105**, 14924-14927.
- Ohtsuki, H., Hauert, C., Lieberman, E., Nowak, M. A. 2006. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502-505.
- Oparin, A.I. 1953. *The origin of life*, New York: Dover.
- Orgel, L.E. 1968. Evolution of the genetic apparatus. *J. Mol. Biol.* **38**, 381-393.
- Orgel, L.E. 1986. RNA catalysis and the origins of life. *J. Theor. Biol.* **123**, 127-149.
- Orgel, L.E. 1992. Molecular replication. *Nature* **358**, 203-209.
- Rao, M., Eichberg, J., Oro, J. 1982. Synthesis of phosphatidylcholine under possible primitive Earth conditions. *J. Mol. Evol.* **18**, 196-202.
- Ricardo, A., Carrigan, M.A., Olcott, A.N., Benner, S.A. 2004. Borate minerals stabilize ribose. *Science* **303**, 196.
- Rousset, F. 2004. Genetic structure and selection in subdivided populations. Princeton University Press.
- Rushdi, A.I., Simoneit, B.R. 2001. Lipid formation by aqueous Fischer-Tropsch-type synthesis over a temperature range of 100 to 400 degrees C. *Origins Life Evol. B.* **31**, 103-118.

- Segre, D., Ben-Eli, D., Lancet, D. 2000. Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. *P. Natl. Acad. Sci. U.S.A.* **97**, 4112-4117.
- Segre, D., Lancet, D., Kedem, O., Pilpel, Y. 1998. Graded autocatalysis replication domain (GARD): kinetic analysis of self-replication in mutually catalytic sets. *Origins Life Evol. B.* **28**, 501-514.
- Shapiro, R. 2007. A simpler origin for life. *Sci. Am.* **296**, 46-53.
- Shapiro, R. 2006. Small molecule interactions were central to the origin of life. *Q. Rev. Biol.* **81**, 105-125.
- Shapiro, R. 1984. The improbability of prebiotic nucleic acid synthesis. *Origins Life Evol. B.* **14**, 565-570.
- Sievers, D., von Kiedrowski, G. 1994. Self-replication of complementary nucleotide-based oligomers. *Nature* **369**, 221-224.
- Stein, D.L., Anderson, P.W. 1984. A model for the origin of biological catalysis. *P. Natl. Acad. Sci. U.S.A.* **81**, 1751-1753.
- Steitz, T.A., Moore, P.B. 2003. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.* **28**, 411-418.
- Szathmary, E., Demeter, L. 1987. Group selection of early replicators and the origin of life. *J. Theor. Biol.* **128**, 463-486.
- Szostak, J.W., Bartel, D.P., Luisi, P.L. 2001. Synthesizing life. *Nature* **409**, 387-390.
- Tarnita, C.E., Antal, T., Ohtsuki, H., Nowak, M.A. 2009. Evolutionary dynamics in set structured populations. *P. Natl. Acad. Sci. U.S.A.* **106**, 8601-8604.
- Traulsen, A., Nowak, M. A. 2006. Evolution of cooperation by multilevel selection. *P. Natl. Acad. Sci. U.S.A.* **103**, 10952-10955.
- Taylor, P.D., Day, T., Wild, G. 2007. Evolution of cooperation in a finite homogeneous graph. *Nature* **447**, 469-472.
- Wächtershäuser, G. 2007. On the chemistry and evolution of the pioneer organism. *Chem. Biodivers.* **4**, 584-602.

Figure Captions

Figure 1: (a) The tree of prelife. Activated monomers, 0^* and 1^* , form (random) polymers. Activated monomers can become deactivated, $0^* \rightarrow 0$ and $1^* \rightarrow 1$. Activated monomers can attach to the end of strings. For simplicity, we assume that all strings grow only on one side. Therefore, each string has one immediate precursor and two immediate followers. Each sequence has exactly one production lineage. The arrows indicate all the chemical reactions of prelife (up to binary strings of length 4). For catalyzed prelife we assume that some strings have the ability to catalyze certain reactions. There can be chemical hysteresis and multiple steady states. The perfect prelife catalyst is a string which enhances the rates of all chemical reactions in its own lineage (as shown in red for the string 0100). Partial catalysis occurs if a string catalyzes some reactions in its own lineage (as shown in blue for the string 1000). (b) Reaction mechanisms of prelife catalysis and replication. The prelife catalyst, sequence j , reacts with sequence i to form the complex ji . Then sequence i is extended by addition of an active monomer, 0^* . Subsequently the complex dissociates. For replication, the template n binds to the primer ℓ . Then the primer is extended by addition of active monomers. The catalytic activity of the template walks along the growing primer. Finally the completed double strand dissociates. The rate constants of replication are discussed in Appendix C.

Figure 2: The equilibrium abundances of the all-0 strings, $0^1, 0^2, 0^3, \dots$, are shown as a function of the catalytic activity, c . The catalyst, 0^4 , is shown in red. Shorter sequences are shown in blue, longer sequences in black. We use $a = 1, d = 1$ and $\lambda = 1$. For these parameters the critical length of the catalyst is given by $n_{\text{cr}} = 3 + \sqrt{3} = 4.732\dots$. The length of the catalyst in this example, $n = 4$, is below this threshold. Therefore, the equilibrium abundance of the catalyst (red curve) increases monotonically with c .

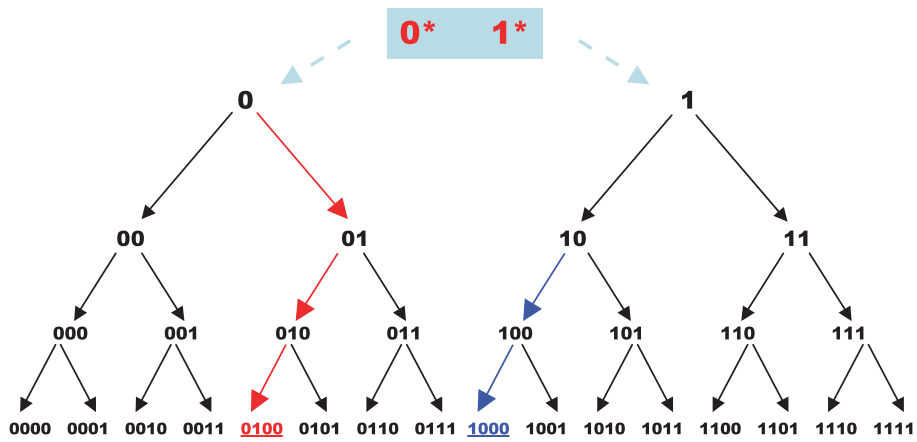
Figure 3: The equilibrium abundance of the catalyst, \hat{x}_n (red), is shown as a function of its catalytic activity, c . The catalyst enhances all of its upstream reactions. We use the same parameter as for Figure 2, but the length of the catalyst, $n = 7$, is above the critical value n_{cr} in this case. (a) The system shows bistability for $c_1 \leq c \leq c_2$. The solid lines in red represent stable equilibria. The

dotted line in red represents unstable equilibria. The red arrows represent the direction of the change from initial abundance to final abundance of the catalyst. **(b)** When the catalyst is initially rare, we observe a discontinuous jump in the abundance at $c = c_2$. The blue lines represent the abundances of sequences 0^1 to 0^6 (upstream). The red line represents the sequence, 0^7 . The black lines represent sequences 0^8 and longer (downstream). For $c < c_2$ shorter sequences have higher abundance. For $c > c_2$, the catalyst is most abundant. **(c)** When the catalyst is initially abundant, we observe a discontinuous jump in the abundance at $c = c_1$. For $c < c_1$ shorter sequences have higher abundance. For $c > c_1$ the catalyst is most abundant.

Figure 4: The equilibrium abundances of the replicator (red), shorter sequences (blue) and longer sequences (black) are shown as functions of the replication rate constant, r . We use $a = 1, d = 1, \lambda = 1, \delta = 1$ and $n = 7$. From eq.(9), the threshold value of r for the equilibrium abundance of replicators to exceed 10% of its theoretical maximum is predicted as $r^* = 19.7256 \dots$

Figure 1

(a)



(b)

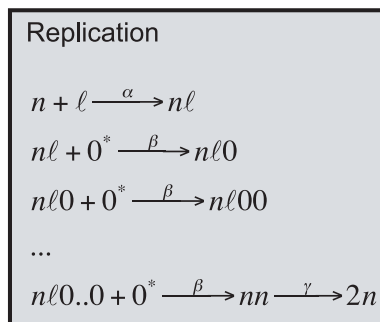
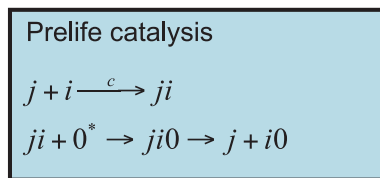


Figure 2

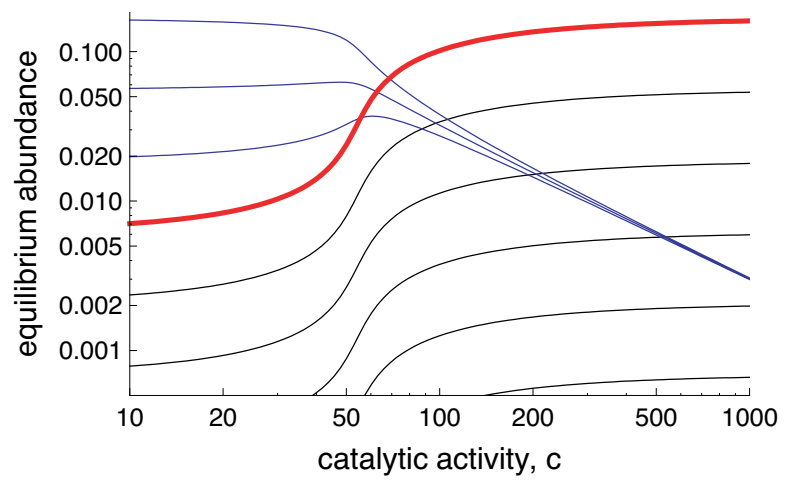
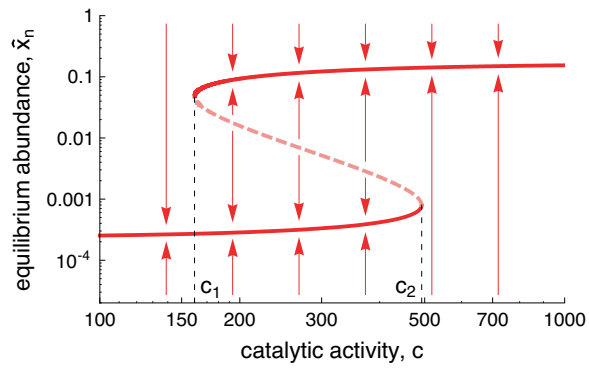
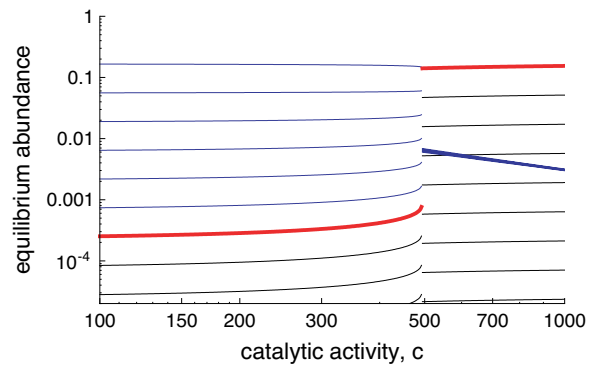


Figure 3

(a)



(b)



(c)

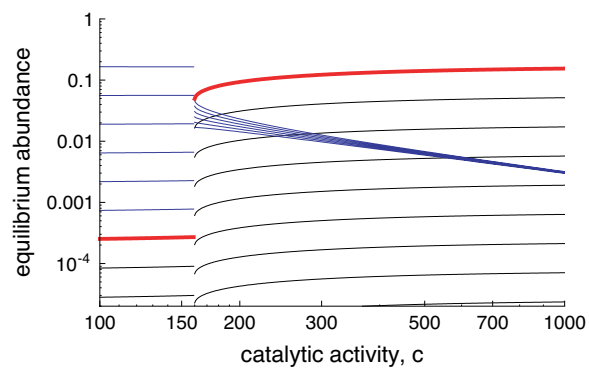


Figure 4

