

Selection for Replicases in Protocells

Ginestra Bianconi¹, Kun Zhao², Irene A. Chen³, Martin A. Nowak^{4*}

1 School of Mathematical Sciences, Queen Mary University of London, London, United Kingdom, **2** Department of Physics, Northeastern University, Boston, Massachusetts, United States of America, **3** Department of Chemistry and Biochemistry, University of California, Santa Barbara, Santa Barbara, California, United States of America, **4** Program for Evolutionary Dynamics, Department of Mathematics, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America

Abstract

We consider a world of nucleotide sequences and protocells. The sequences have the property of spontaneous self-replication. Some sequences - so-called replicases - have enzymatic activity in the sense of enhancing the replication rate of all (or almost all) sequences. In a well-mixed medium, natural selection would not favor such replicases because their presence equally benefits sequences with or without replicase activity. Here we show that protocells can select for replicases. We assume that sequences replicate within protocells and that protocells undergo spontaneous division. This leads to particular population structures which can augment the abundance of replicases. We explore various assumptions regarding replicase activity and protocell division. We calculate the error threshold that is compatible with selecting for replicases.

Citation: Bianconi G, Zhao K, Chen IA, Nowak MA (2013) Selection for Replicases in Protocells. *PLoS Comput Biol* 9(5): e1003051. doi:10.1371/journal.pcbi.1003051

Editor: Michael Doebeli, University of British Columbia, Canada

Received: November 24, 2012; **Accepted:** March 20, 2013; **Published:** May 9, 2013

Copyright: © 2013 Bianconi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Support from the John Templeton Foundation (<http://www.templeton.org/>) and from the Foundational Questions in Evolutionary Biology Fund (<http://www.fas.harvard.edu/~fqeb/grants/>) - grant RFP-12-05 - is acknowledged. Publication fees were supported by the UCSB Open Access Fund Pilot Program (<http://www.library.ucsb.edu/scholarly-communication/ucsb-open-access-fund-pilot-program>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: martin_nowak@harvard.edu

Introduction

The origin of life must have required a series of transitions building new levels of molecular interaction. However, a tension often exists between the fitness of an individual sequence and the fitness of the collective [1,2]. This tension would be important for the earliest replicase enzymes (i.e., replicases), which would help other individuals replicate without helping themselves directly [3]. Indeed, replicase activity cannot be selected in a thoroughly mixed solution, as natural selection favors the evolution of sequences that parasitize the replicases. The proposed solution to this problem is to essentially create small groups of interactors, either by compartmentation or a lattice-like structure [4–12]. Selection among individuals in the group favors parasites, but selection at the level of the group favors groups with more replicases, thus allowing ‘altruistic’ replicases to survive [13,14].

Compartments, in the form of membrane vesicles, have become an important experimental model for protocells [15–18]. Amphiphilic molecules, such as fatty acids, that can form membrane boundaries can be produced abiotically [19–23] and are found in samples from carbonaceous chondrite meteorites [24–27]. Indeed, vesicles can be formed from meteoritic organic extracts dissolved in water [28]. Recent work on model protocell membranes has demonstrated that vesicles can grow through filamentous structures and divide spontaneously by mild shear forces or photochemical stimulation, a robust ‘pearling’ mechanism that produces many small daughter vesicles [29,30]. Interestingly, experimental studies of cell division mutants in bacteria also suggest that cells divide by pearling when the cell division machinery is eliminated [31]. Pathways for vesicle fission into two

daughter vesicles have also been observed, again stimulated by growth [32–34]. Ribozyme reactions and non-enzymatic polymerization reactions can be encapsulated inside experimental protocells [35,36]. Supramolecular assemblies might have a role in promoting polymerization, as demonstrated by the observation that ribozyme-catalyzed RNA polymerization is more efficient if confined to micelles [37]. Inspired by these promising protocell experiments, we focus on vesicles in the theoretical study that follows.

Previous models of compartmentation by vesicles have studied the ability of vesicles to enhance information storage and affect replicase selection. Prior models have usually assumed that the encapsulated genotypes influence vesicle replication (i.e., growth, survival, or division), causing selection among vesicles. Encapsulation has been shown to increase information capacity if vesicle survival depends on the simultaneous presence of multiple self-replicating (i.e., not necessarily replicase) ‘genes’ [4,10]. In a model by Hogeweg and Takeuchi [7], encapsulation could increase information capacity if the best self-replicators also enhanced vesicle replication, but not if vesicle replication was neutral (i.e., division occurred when the molecular population size reached a certain number). With respect to the evolution of enzymatic activity, replicase dynamics were studied in spatial grids [6], where replicases were found to evolve greater fidelity and information capacity, essentially due to reciprocal altruism in local clusters. However, the generality of this model is unclear, as certain tradeoffs were assumed to exist between replication fidelity, efficiency, and templating ability, and vesicular protocells were not studied. Takeuchi and Hogeweg studied the survival of replicase enzymes (and their parasites) in vesicles, in which

Author Summary

The origin of life, proceeding from chemical reactions to cells, must have included a critical transitional period in which catalytically active sequences arose. A fundamental problem exists for the first catalytic sequences: their activity would not enhance their own fitness directly, and might even decrease their own fitness relative to that of other molecules. Catalytic sequences are constantly encumbered by mutation and drift, limiting the amount of information that can be maintained. Population structures, such as cells, are known to be able to counter this problem. Here we introduce a simple model of the earliest cells to understand limits on information for catalysts with different properties. We find some parallels to information limits on replicators in free solution. Conditions that keep replicases together, or enhance their effect as their abundance increases, permit the evolution of catalytically active sequences.

parasites were also assumed to contribute to vesicle growth [11]. The dynamics were complicated, but in general selection at the compartment level could counter selection among replicases. Furthermore, stochastic fluctuations have been shown to be important for switching from a ribozyme-poor to a ribozyme-rich regime, a situation that could be enabled by compartmentation into protocells [38,39].

While recent progress has been made in evolving an RNA enzyme that can copy another RNA sequence [40–42], the difficulty of this task has prompted several suggestions for simpler enzymatic activities that might have preceded the polymerase. Any activity that could promote replication would be considered a replicase. For example, RNA sequences that catalyze ligation could stitch together short oligos in a template-directed manner [43–46]; an exonuclease could enhance speed and fidelity by removing dangling mismatched ends [47]; a permease could increase the rate of heterotrophic uptake [36]. Cooperative phenotypes may also characterize early autocatalytic replicator cycles, such as systems of ligases or recombinases [43,48–51]. Broadly speaking, in even simpler terms, a replicase might act through colligative properties rather than sequence-specific interactions. For example, an osmolyte might reduce evaporative loss, or a charged polymer might trap useful oppositely charged species. Osmotic pressure has already been shown to drive membrane growth [52]. Such simple chemical activities, while not enzymatic, are weakly altruistic in the sense that they help themselves and other molecules equally.

In light of recent experimental progress, we re-examine the conditions under which enzymatic activity can be selected, using a simple but plausible model of encapsulated replicases and inactive molecules. We first consider a scenario, in which the replicase helps all molecules within a protocell to replicate and is not itself impaired as a template. Then we consider a more altruistic enzyme, which can help other molecules but not itself directly. Vesicle division in our model occurs when the encapsulated population size reaches a certain threshold, but the replicases and vesicles are otherwise unlinked. We calculate the conditions under which altruistic enzymatic activity can be selected.

Results

We consider two types of sequences. Type *A* can act as replicase, potentially enhancing the replication of sequences of any type, but type *B* cannot. All sequences undergo spontaneous self-

replication, and moreover all sequences are targets of replicase activity. Thus all sequences benefit equally from the presence of type *A* sequences. Type *A* does not have an intrinsic preference to catalyze the replication of other type *A* sequences; it treats all targets equally. It is evident that natural selection would not augment the abundance of type *A* sequences in a well-mixed population. At best, the type *A* sequences would have the same fitness as all other molecules, so they have no selective advantage. However, *A* sequences can be erroneously copied to produce *B* sequences, which causes the population to drift toward an all-*B* state. Back mutation from *B* to *A* can be neglected, because a specific sequence is needed for replicase activity. Type *A* represents a small fraction of possible sequences, while all other sequences are of type *B*.

The probability that a type *A* sequence replicates without mutation is given by q . If a type *A* sequence replicates with mutation, the offspring will be a type *B* sequence; this happens with probability $1 - q$. We can think of a point mutation rate, u , and a number of positions, L , which must remain unchanged in order to retain replicase activity. For example, L has been estimated to be roughly 75% of the physical length of a functional RNA molecule [53]. Then we have $q = (1 - u)^L$. Replication of type *B* sequences always results in type *B*; thus we neglect back-mutation. In a well-mixed population type *A* sequences would become extinct for any positive mutation rate, $u > 0$.

Let us now study the evolutionary dynamics of *A* and *B* sequences within protocells. Denote by $A_i B_j$ a protocell, which contains i sequences of type *A* and j sequences of type *B*. If an *A* sequence replicates within this protocell without mutation we obtain $A_{i+1} B_j$. If an *A* sequence replicates with mutation, or if a *B* sequence replicates, we obtain $A_i B_{j+1}$.

We explore four different replicases that enhance the replication rate of the molecules within the protocell in different ways. In each case the sequence *A* represent a different type of replicase, which we label $R1$, $R2$, $R1\alpha$ and $R2\alpha$ (see Figure 1).

- Replicase $R1$ has the following property: the presence of at least one *A* sequence inside a protocell enhances the replication rate of all sequences within that protocell to a value a , which is greater than 1. In a protocell that contains only type *B* sequences the replication rate is 1.
- Replicase $R2$ has the following property: the presence of an *A* sequence within a protocell augments the replication rates of all other sequences in this protocell, but not its own. Thus, if there is only a single *A* sequence present in a protocell, then all other sequences have an increased replication rate, $a > 1$, while the *A* sequence has replication rate 1. If at least two *A* sequences are in a protocell, then all sequences in that protocell have an increased replication rate, a .
- Replicase $R1\alpha$ has the following property: the replication rate increases with the number of *A* sequences inside a protocell. In particular, we assume that if there are i *A* sequences inside a protocell, the replication rate of all sequences within that protocell is $1 + \alpha i$, where $\alpha > 0$. In a protocell that contains only type *B* sequences the replication rate is 1.
- Replicase $R2\alpha$ has the following property: the replication rate increases with the number of *A* sequences inside a protocell, but a single *A* sequence does not enhance its own replication rate. In particular, we assume that if there are i many *A* sequences inside a protocell, the replication rate of all *A* sequences within that protocell is $1 + \alpha(i - 1)$, and the replication rate of all *B* sequences within that protocell is $1 + \alpha i$, where $\alpha > 0$. For $R2\alpha$, *A* sequences receive less

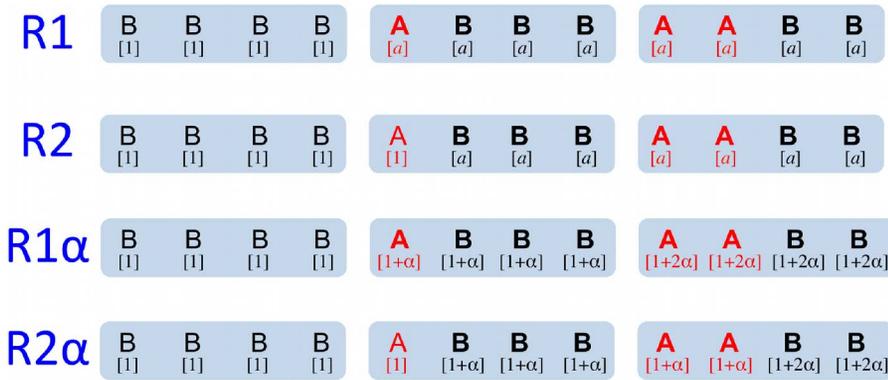


Figure 1. Effect of the replicase in different models. Three possible protocells are shown for each model, each containing four sequences, including 0, 1, or 2 sequences of type A (red) and the remainder being type B (black). The replication rate of each sequence is shown in brackets under the sequence. Sequences that receive benefit from the replicase(s) are shown in bold type (i.e., replication rate > 1). As in the main text, $a > 1$ and $\alpha > 0$. In *R1* and *R1α*, type A sequences enhance replication of all A and all B, such that all sequences in the cell have the same replication rate. In *R2* and *R2α*, type A molecules enhance replication of other molecules but not themselves. In *R1* and *R2*, the effect of a single A is maximal. In *R1α* and *R2α*, the presence of multiple As increases the rate enhancement. doi:10.1371/journal.pcbi.1003051.g001

advantage than B sequences, as might be expected if the replicase acts directly on other sequences.

In Figure 2 we show the reaction kinetics for all four types of replicases.

Replication within a protocell increases the number of sequences inside the protocell. We assume that the cell divides once a certain maximum number, m , of sequences has been reached. We consider two types of cell division. (i) Division into two: each sequence of the parent cell is given at random to one of the two daughter cells. (ii) Division into many: each daughter cell contains exactly one sequence. In both cases we do not need to keep track of empty cells. In Figure 3 we show how the different mechanisms for cell division work for a protocell of maximal size $m = 5$.

For replicase *R1* we obtain the following, surprisingly simple result irrespective of the number m and irrespective of whether cells divide into two or into many. A sequences can be maintained in the population if $q > q_c$ where

$$q_c = \frac{1}{a}.$$

If the probability of error-free replication, q , is greater than $1/a$, then replicases can be selected within protocells. The result is reminiscent of the error-threshold of quasispecies theory, which describes the selection of a master sequence (not a replicase) in a well-mixed medium [54–56].

For replicase *R2* it is harder to select for A sequences. The reason is that an A sequence can only help other sequences to reproduce but not itself. Again we find an error threshold, but this time we do not obtain a simple closed form expression. We derive a numerical solution, which is shown in Figure 4. We observe that division into two daughter cells leads to less restrictive conditions (for given m) than division into many. In this case if protocells divide into many daughter cells, then each sequence starts off alone within a cell; here single A sequences have no advantage over single B sequences. On the other hand, if protocells divide into two, then for larger m it is typically the case that each A sequences is together with other A sequences

after cell division and immediately benefits from the enzymatic activity.

For replicase *R1α*, we can prove (see Materials and Methods and Text S1) that the error threshold satisfies the following inequality

$$q_c < \frac{1}{1 + \alpha}.$$

Our numerical solution of the error threshold shown in Figure 4 demonstrates that this upper limit is tight for large values of α . These results are valid when we consider division into two as well as division into many.

Replicase *R2α* can only help other sequences to reproduce but not themselves. Therefore we find that *R2α* is more difficult to select than *R1α* (see Figure 4). We observe that division into two leads to less restrictive conditions (for given m) than division into many. We can explain this phenomenon with the same argument used to explain why it is more difficult to select for *R2* than for *R1* (see above).

We have conducted numerical calculations of the error threshold for values of m between 3 and 10. We observe that for *R1α* the error threshold is a decreasing function of the maximum number of encapsulated sequences m . We find a similar behavior of the error threshold for *R2* and *R2α* as long as the protocells split into many daughter cells. In this case, for larger m it is easier to select for these replicases. But if the protocells divide into two, we observe for *R2* and *R2α* that the error threshold for a given a (or α) does not always decrease with m . Instead there is an optimum cell size which favors selection of the replicase.

Discussion

We have studied the constraints on the information content of a replicase arising in protocells. Mutations that produce inactive variants of the replicase are an inevitable consequence of molecular replication, creating parasitic or commensal sequences that derive benefit from the presence of the replicase. The population structure imposed by the protocell membranes can prevent loss of the replicase. We investigated four types of

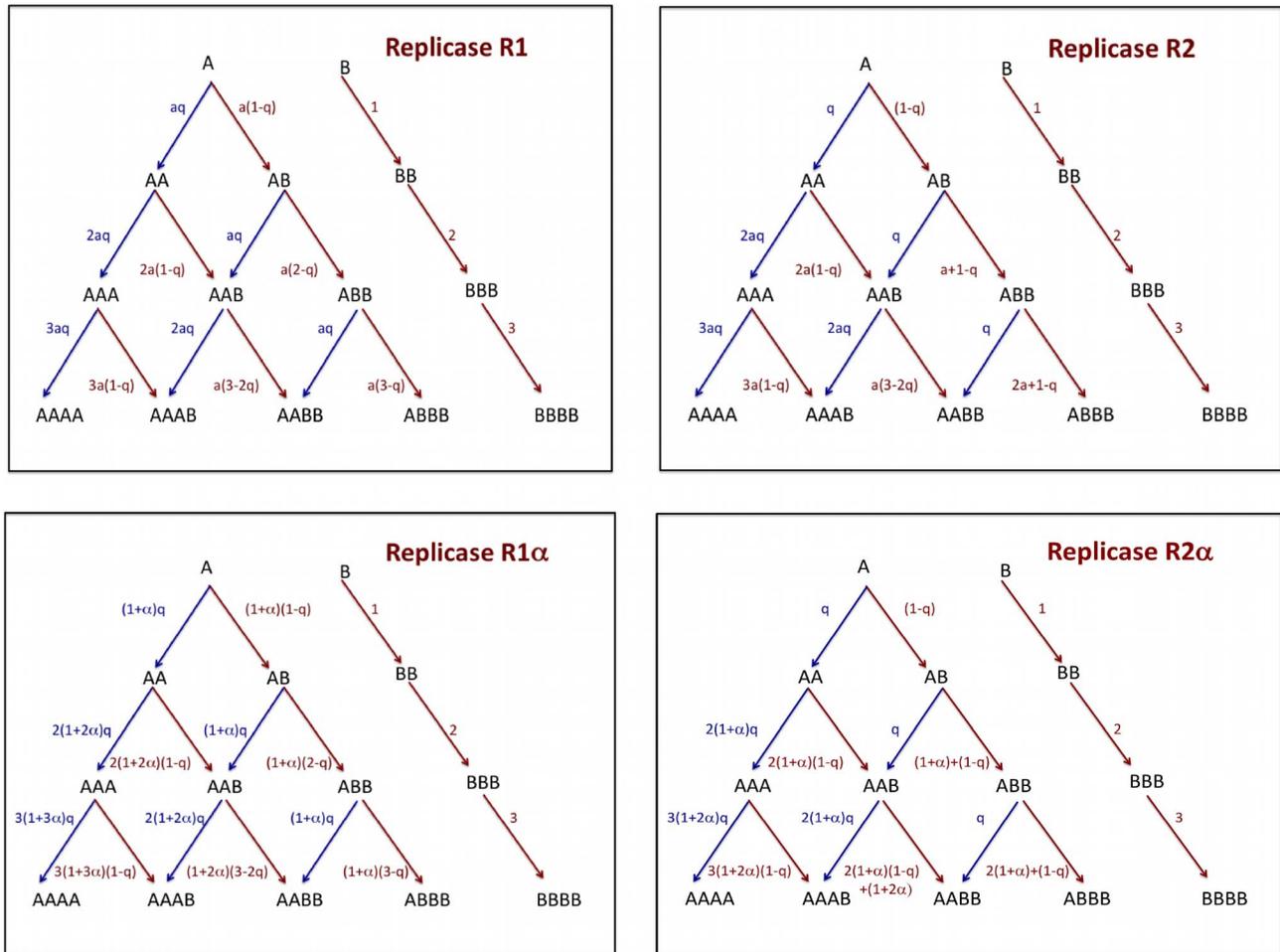


Figure 2. Reaction kinetics for protocells of different composition for the four replicases considered in this paper. The sequences of type A indicate the replicases. All sequences undergo spontaneous self-replication and are targets of replicase activity. The sequences of type B, in absence of sequences of type A in the protocell, replicate at rate 1. A type A sequence replicates with mutation with probability $1 - q$, and the offspring will be a type B sequence. If the sequences A encode the replicase R1, it is sufficient that at least one sequence of type A is present in the protocell for enhancing the replication of every sequence in the same protocell to $a > 1$. If the sequences A encode for the replicase R2, the sequences enhance the rate of replication of all the other sequences to $a > 1$. If the sequences A encode the replicase R1 α , and there are i A sequences in the protocell, the rate of all the sequences is given by $1 + \alpha i$ with $\alpha > 0$. If the sequences A encode the replicase R2 α , and there are i A sequences in the protocell, the rate of the A sequences is given by $1 + \alpha(i - 1)$ while the rate of replication of the B sequences is $1 + \alpha i$ with $\alpha > 0$. doi:10.1371/journal.pcbi.1003051.g002

replicases and two types of division. Table 1 summarizes our results, showing the maximum genome length L_{max} (i.e., number of invariant, informative sites), given experimental parameters for error rates and a prebiotically plausible replication enhancement from the presence of the replicase. The experimental error rates cover a reasonable range, from the misincorporation rate of non-enzymatic RNA replication (representing the lowest fidelity reaction one might consider) to the error rate of an RNA polymerase replicase producing full-length copies (a recently reported replicase with relatively high fidelity) [40,57]. The results also depend on the value of a or α . In Table 1, a low value was chosen ($a=10$, $\alpha=10$) to represent early, relatively poor replicases, but the absolute numbers would increase with greater replicase activity (See Table S1 in Text S1 for analogous calculations for $a=\alpha=200$). It is also important to note that L_{max} is the number of invariant sites, so the physical length of the molecule could be greater [53]. In addition, very small replicases

have been reported (as small as 5 nt; [58]), so it is conceivable that low-information sequences could potentially act as weak replicases.

There are many possible chemical functions that could enhance molecular replication within the cell. Two major categories of replicases are those whose presence helps all molecules in the cell, including itself (a commensal situation; replicases R1 and R1 α), and those whose presence helps other molecules in the cell but not itself (an altruistic situation; replicases R2 and R2 α). Commensal ‘replicases’ might have beneficial colligative properties. These represent a very early stage of evolution, in which sequences did not necessarily perform specific functions and could be poorly folded. For example, this situation might apply to the selection of the chemical backbone (e. g., RNA) itself. On the other hand, altruistic replicases might perform any number of specific functions, and indeed any RNA that folded into a stable structure would have a compromised fitness for template replication compared to poorly folded RNAs. Intuitively, it is therefore more

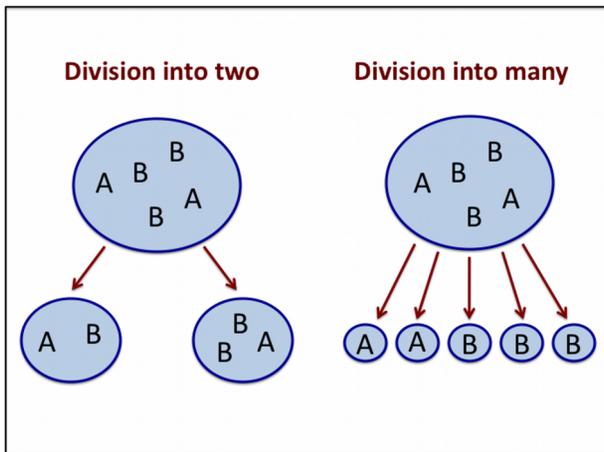


Figure 3. Division mechanism. When a protocell reaches the maximum size m , it splits. Here we consider two splitting mechanisms. In the first case the protocell splits into two daughter protocells, of random composition, with each protocell containing at least one sequence. In the second case, the protocell splits into m daughter protocells, and each daughter protocell contains a sequence. doi:10.1371/journal.pcbi.1003051.g003

difficult for an altruistic replicase to survive, so less information can be maintained, as we observe in our results (Table 1).

One analytical result of particular interest is the form of the error threshold for $R1$ in protocells ($q_c = 1/a$). This form is identical to the form of the classical error threshold considering a ‘master’ replicator sequence with fitness f competing with its mutants ($q_c = 1/f$) [59,60]. If enzymes are encapsulated in protocells, it seems that selection has effectively moved up to the next level, from competition among individual replicator sequences to competition among protocells based on the encapsulated genotypes. The collective advantage a takes the place of the individual advantage f , and the survival of the enzyme depends on the mutation rate just like a ‘master’ sequence would in free solution.

The dependence of replication enhancement on the number of replicases is likely to increase linearly initially, and then to saturate at some point. We examined these two regimes separately. To examine the saturated regime, we assumed that a single copy of $R1$ (commensal) or $R2$ (altruistic) produced the maximal effect on replication rates. $R1\alpha$ (commensal) and $R2\alpha$ (altruistic) represent the analogous initial regimes, respectively. These two regimes give similar limits on information, particularly at large values of α , but slightly more information could be maintained in the initial, non-saturated regime. Intuitively, if protocells containing multiple replicases have greater advantage, the overall benefit from the presence of the replicase is greater, allowing more information to be maintained at the same mutation rate.

Cell division typically proceeds via binary fission, or division into two daughter cells. In addition, some model protocells divide by fissioning into many daughter vesicles as described earlier. Bacteria lacking cell division machinery also appear to divide by fissioning into many small cells [31]. We therefore modeled two limiting scenarios for division mechanisms: division into two daughter vesicles, and division into many daughter vesicles (i.e., more daughter vesicles than encapsulated replicators, Figure 3). In general, binary fission is better in terms of maintaining genetic information. Intuitively, binary fission can keep replicases together so they benefit from one another, while division into many vesicles immediately isolates the replicases from each other.

Like previous theoretical models [7], we assume that division occurs upon reaching a particular size m (a number of encapsulated sequences). In general, larger m is more permissive to the replicases, allowing more information to be stored because replicases can group together more of the time, enhancing the mutual benefit. In our case, for $R1$, there is no disadvantage for isolated single replicase molecules, and we obtain that group size is irrelevant to the error threshold. For $R2$, the replicase is disadvantaged compared to the mutant sequences, but the addition of mutant sequences does not further decrease the fitness of the replicase; in addition, if two replicases are present, then the replicases do not suffer the disadvantage in the protocell.

The effect of cell size, m , is in contrast with group selection models based on cooperator-defector games, in which larger group size makes selection of the cooperator trait more difficult [9]. In those models, larger groups are more likely to generate defectors (by mutation), which then take over the entire group because of their intrinsic selective advantage. In our current model, larger cells are also more likely to produce non-catalytic (type B) sequences by mutation, but they do not have a selective advantage; they rely on mutation pressure and drift to take over a cell.

To summarize, an RNA replicase arising during the origin of life would be most able to resist mutational pressure under the following conditions: the ability to enhance its own fitness, compartmentalization (which permits selection of the enzymatic behavior), additive enhancement from multiple replicases, larger cell size, and binary fission of compartments. The replicases might correspond to a number of different possible chemical activities. For example, $R1$ and $R1\alpha$ could correspond to a bulk chemical activity (e.g., charged polymer) that enhances replication for all encapsulated sequences (e.g., by attracting oppositely charged ‘food’ molecules) without impacting its own replication. $R2$ and $R2\alpha$ could correspond to a ribozyme with a specific folded structure, which benefits other sequences but not itself directly, such as an RNA polymerase or a membrane transporter.

In conclusion, we have attempted to present the simplest possible models for the selection of enzymatic activity that are inspired by experimental protocells. We estimate the conditions that enable survival of the replicase trait. We focused on simple models in order to understand the underlying dynamics. However, this work could potentially be extended to include more realistic chemical detail, as found in other recent modeling [61]. Other processes could also be included, such as exchange of genetic material among protocells [62]. Another important consideration is that our modeling is deterministic, as a first step in understanding the system. Although the number of RNAs per protocell is small, the number of protocells may be large, justifying a deterministic approach. However, a stochastic approach would be more realistic and could highlight interesting phenomena [38,39]. Also, in our model, we assume that some A is present in the initial pool, and therefore survival of A depends on the error criterion. Because we neglect back-mutation of B into A , A cannot be generated de novo in our model; a more realistic model would include the possibility of back-mutation. In addition, our model includes the decay or degradation of protocells (and thus the replicases contained within them), but not of individual sequences within the protocells. This corresponds to the assumption that the removal of protocells (e.g., by dilution), rather than destruction of individual sequences, is the dominant process of decay. Thus, a system containing A 's could transition to an all- B system through loss of protocells containing A 's. A more realistic model would include differential decay of molecules within the protocells as well. Further studies would be needed to test the effects of such realistic modifications to the models. Notwithstanding additional

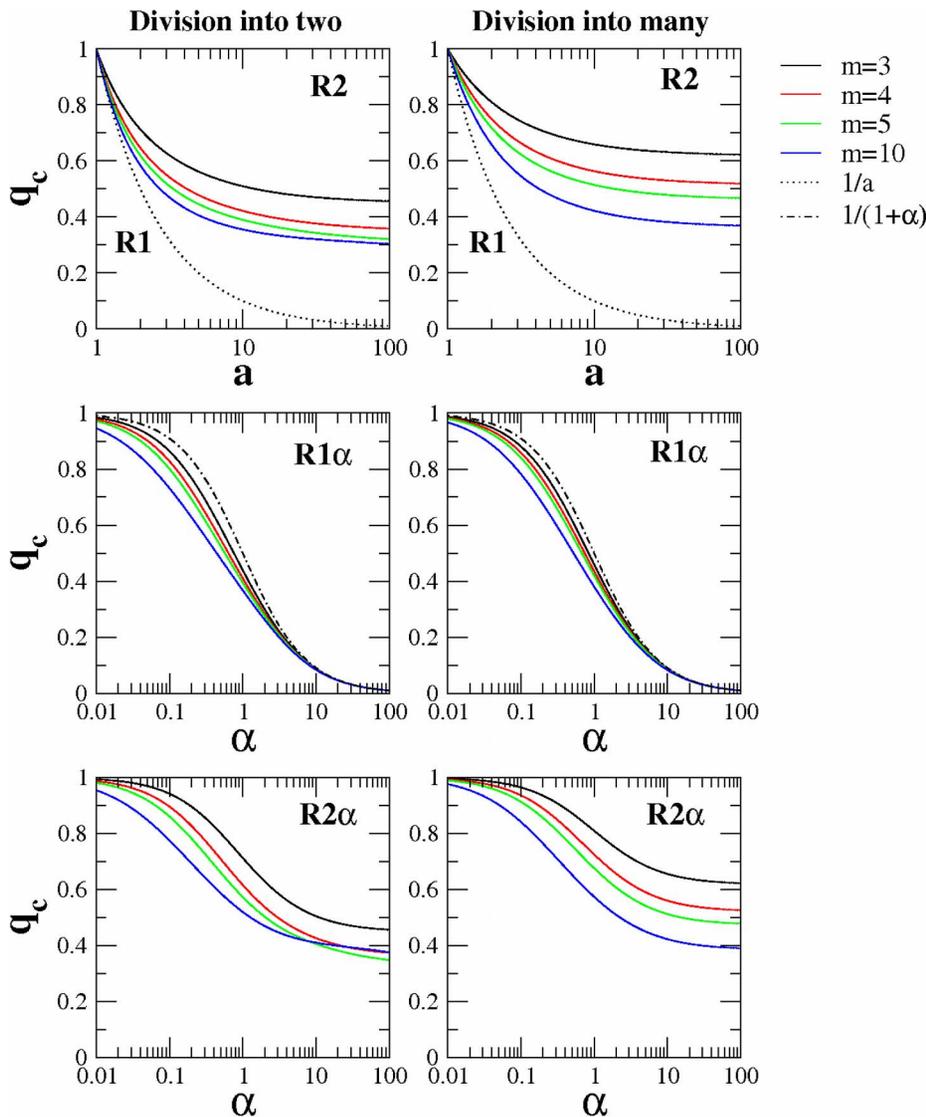


Figure 4. Error threshold for the four replicases under study. In the left panels the protocells of maximal size m divide into two daughter cells. In the right panels the protocells of maximal size m divide into many (m) daughter cells. Curves were generated by numerical simulation as described in the text.
doi:10.1371/journal.pcbi.1003051.g004

complexities, we find that replicases can be selected under a variety of assumptions. In the simplest case we observe an error threshold arising from protocell competition, in striking analogy to replicator competition. That is, the condition for replicase selection in protocells mirrors the classical condition for replicator selection [54,59], suggesting the emergence of a new level of selection in which protocells are a mathematical analog to replicators. In addition, we find that conditions that tend to keep replicases together, or enhance their effect as their abundance increases, permit evolution of more information.

Materials and Methods

The mutation-selection-cell division (MSCD) equations

We indicate with $x_{i,j}$ the frequency of protocells of composition $A_i B_j$. In the Text S1 we describe the mutation-selection-cell division (MSCD) equations for the general case. Here we show how the model reads for replicase *R1*. The reaction kinetics are

described in Figure 2. The MSCD equations for replicase *R1* read

$$\begin{aligned} \dot{x}_{1,0} &= -ax_{1,0} + d_{1,0} - \phi x_{1,0} \\ \dot{x}_{i,0} &= -aix_{i,0} + a(i-1)qx_{i-1,0} + d_{i,0} - \phi x_{i,0} \\ \dot{x}_{i,j} &= -a(i+j)x_{i,j} + a[i(1-q) + j-1]x_{i,j-1} + a(i-1)qx_{i-1,j} \\ &\quad + d_{i,j} - \phi x_{i,j} \quad i > 1 \ \& \ j > 0 \\ \dot{x}_{1,j} &= -a(i+1)x_{1,j} + a[(1-q) + j-1]x_{1,j-1} + d_{1,j} - \phi x_{1,j} \quad j > 0 \\ \dot{x}_{0,1} &= -x_{0,1} + d_{0,1} - \phi x_{0,1} \\ \dot{x}_{0,j} &= -jx_{0,j} + (j-1)x_{0,j-1} + d_{0,j} - \phi x_{0,j} \quad j > 1. \end{aligned}$$

In these equations $d_{i,j}$ denotes the rate at which protocells of composition $A_i B_j$ are formed as a consequence of the splitting of protocells of size m . For splitting into two daughter cells, $d_{i,j}$ can be written as

Table 1. Maximal length of the selected replicase L_{max} .

Division into two	m	$L_{max}(u=0.17)$	$L_{max}(u=0.0088)$
R1	any m	12	260
R2	3	3	76
R2	4	4	97
R2	5	5	106
R2	10	5	117
R1 α	3	13	275
R1 α	4	13	277
R1 α	5	13	278
R1 α	10	13	280
R2 α	3	3	77
R2 α	4	4	96
R2 α	5	4	102
R2 α	10	4	100
Division into many	m	$L_{max}(u=0.17)$	$L_{max}(u=0.0088)$
R1	any m	12	260
R2	3	2	47
R2	4	3	65
R2	5	3	75
R2	10	4	97
R1 α	3	13	274
R1 α	4	13	276
R1 α	5	13	277
R1 α	10	13	281
R2 α	3	2	47
R2 α	4	3	65
R2 α	5	3	75
R2 α	10	4	97

Maximal length of the selected replicase L_{max} calculated by imposing $(1-u)^{L_{max}} = q_c$ for the different models under consideration with $a = 10$ or $\alpha = 10$. The parameter m is the maximum number of sequences in the protocell. The parameter u is the mutation rate per base. The parameters a and α reflect the rate enhancement from type A sequences. L_{max} may be roughly 75% of the physical length of the molecule for functional RNA [53].

doi:10.1371/journal.pcbi.1003051.t001

$$d_{i,j} = \sum_{i' \geq i, j' \geq j, i'+j'=m} \frac{\binom{i'}{i} \binom{j'}{j}}{2^{m-1} - 1} r_{i',j'}$$

For splitting into many (m) daughter cells, $d_{i,j}$ can be written as

$$\begin{aligned} d_{1,0} &= \sum_{i \geq 1, i+j=m} i r_{i,j} \\ d_{0,1} &= \sum_{j \geq 1, i+j=m} j r_{i,j} \\ d_{i,j} &= 0 \quad (i > 1 \text{ or } j > 1). \end{aligned}$$

where the dissociation rates $r_{i,j}$ of protocells with $i+j=m$ are given by

$$\begin{aligned} r_{i,j} &= aiqx_{i,j-1}(1-\delta_{j,0}) + a[(i-1)(1-q) + j]x_{i-1,j}(1-\delta_{i,1}) \\ i+j &= m > 3 \text{ \& } i > 1 \\ r_{0,m} &= (m-1)x_{0,m-1} \end{aligned}$$

The error threshold for replicase R1

The frequency of sequences A ($\sum_{i,j} ix_{i,j}$) evolves according to the MSCD equations and can be written as

$$\sum_{i,j} i\dot{x}_{i,j} = qa \sum_{ij} ix_{i,j} - \phi \sum_{ij} ix_{i,j} = (aq - \phi) \sum_{ij} ix_{i,j}$$

Therefore if

$$aq > \phi,$$

the number of protocells with sequences A increases. On the other side the total number of sequences $\sum_{ij} (i+j)x_{ij}$ evolves according to the MSCD equation and it can be proved that independently of the splitting mechanism, it satisfies the following equation:

$$\sum_{ij} (i+j)\dot{x}_{ij} = a \sum_{ij|i \geq 1} (i+j)x_{i,j} + \sum_j jx_{0,j} - \phi \sum_{ij} (i+j)x_{i,j}$$

By setting $\sum_{ij}(i+j)\dot{x}_{ij}=0$ we obtain the value for ϕ that is needed to preserve the total number of sequences (i.e. $\sum_{ij}(i+j)x_{ij}=1$). We find therefore

$$\phi = a - (a-1) \sum_j j x_{0,j}.$$

Therefore $\phi \rightarrow 1$ if $\sum_j j x_{0,j} \rightarrow 1$. Substituting $\phi=1$ in the relation $aq > \phi$, we find that the configuration with $\sum_j j x_{0,j}=1$ is not stable and protocells with sequences A will be selected if

$$q > q_c = \frac{1}{a}.$$

This result is proved here for the case in which we assume that

References

- Lewontin RC (1970) The units of selection. *Annu Rev Ecol Syst* 1: 1–18.
- Nowak MA (2006) Five rules for the evolution of cooperation. *Science* 314: 1560–1563.
- Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. *Nature* 409: 387–390.
- Szathmari E, Demeter L (1987) Group Selection of Early Replicators and the Origin of Life. *Journal of theoretical biology* 128: 463–486.
- Nowak MA, May R (1992) *Nature* 359: 826–829.
- Szabo P, Scheuring I, Czaran T, Szathmari E (2002) In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* 420: 340–343.
- Hogeweg P, Takeuchi N (2003) Multilevel selection in models of prebiotic evolution: Compartments and spatial self-organization. *Origins of Life and Evolution of the Biosphere* 33: 375–403.
- Scheuring I, Czaran T, Szabo P, Karolyi G, Toroczkai Z (2003) Spatial models of prebiotic evolution: soup before pizza? *Origins of life and evolution of the biosphere* 33: 319–355.
- Traulsen A, Nowak MA (2006) Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences* 103: 10952–10955.
- Fontanari JF, Santos M, Szathmari E (2006) Coexistence and error propagation in pre-biotic vesicle models: A group selection approach. *Journal of theoretical biology* 239: 247–256.
- Takeuchi N, Hogeweg P (2009) Multilevel Selection in Models of Prebiotic Evolution II: A Direct Comparison of Compartmentalization and Spatial Self-Organization. *Plos Computational Biology* 5(10):e1000542.
- Hogeweg P, Takeuchi N (2012) Evolutionary dynamics of RNA-like replicator systems: A bioinformatic approach to the origin of life. *Physics of Life Reviews* 9: 219–263.
- Maynard Smith J and Szathmari E (1995) *The Major Transitions in Evolution*. New York: Oxford University Press.
- Maynard Smith J and Szathmari E (1999) *The Origin of Life*. New York: Oxford University Press, New York.
- Dzicziel AJ, Mann S (2012) Designs for life: protocell models in the laboratory. *Chemical Society reviews* 41: 79–85.
- Walde P (2010) Building artificial cells and protocell models: experimental approaches with lipid vesicles. *Bio Essays* 32: 296–303.
- Rasmussen S, Bedau M, Chen L, Deamer D, Krakauer D, et al. (2009) *Protocells: Bridging nonliving and living matter*. Cambridge: MIT Press.
- Deamer D, Szostak JW, editors (2010) *The Origins of Life*. Cold Spring Harbor: Cold Spring Harbor Laboratories Press.
- Allen WV, Ponnampertuma C (1967) A possible prebiotic synthesis of monocarboxylic acids. *Currents in modern biology* 1: 24–28.
- Dworkin J, Deamer D, Sandford S, Allamandola L (2001) Self-assembling amphiphilic molecules: Synthesis in simulated interstellar/precometary ices. *Proceedings of the National Academy of Sciences of the United States of America* 98: 815–819.
- Rushdi AI, Simoneit BR (2001) Lipid formation by aqueous Fischer-Tropsch-type synthesis over a temperature range of 100 to 400 degrees C. *Origins of life and evolution of the biosphere* 31: 103–118.
- Yuen GU, Lawless JG, Edelson EH (1981) Quantification of monocarboxylic acids from a spark discharge synthesis. *J Mol Evol* 17: 43–47.
- McCollom TM, Ritter G, Simoneit BR (1999) Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. *Origins of life and evolution of the biosphere* 29: 153–166.
- Yuen GU, Kvenvolden KA (1973) Monocarboxylic acids in Murray and Murchison carbonaceous meteorites. *Nature* 246: 301–303.
- Yuen G, Blair N, Des Marais DJ, Chang S (1984) Carbon isotope composition of low molecular weight hydrocarbons and monocarboxylic acids from Murchison meteorite. *Nature* 307: 252–254.
- Lawless JG, Yuen GU (1979) Quantification of monocarboxylic acids in the Murchison carbonaceous meteorite. *Nature* 282: 396–398.

the number of sequences in the system remains constant. Nevertheless the error threshold of the model remains the same if we impose that the number of protocells in the system is fixed. In the Text S1 we give full details of this derivation and we show how to solve the MSCD equations for the other replicases considered in this paper.

Supporting Information

Text S1 Detailed description of model and calculations. (PDF)

Author Contributions

Conceived and designed the experiments: GB MAN. Performed the experiments: GB KZ MAN. Analyzed the data: GB KZ IAC MAN. Wrote the paper: GB IAC MAN.

- Naraoka H, Shimoyama A, Harada K (1999) Molecular distribution of monocarboxylic acids in Asuka carbonaceous chondrites from Antarctica. *Origins of life and evolution of the biosphere* 29: 187–201.
- Deamer DW (1985) Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. *Nature* 317: 792–794.
- Zhu TF, Adamala K, Zhang N, Szostak JW (2012) Photochemically driven redox chemistry induces protocell membrane pearling and division. *Proceedings of the National Academy of Sciences of the United States of America* 109: 9828–9832.
- Zhu TF, Szostak JW (2009) Coupled growth and division of model protocell membranes. *J Am Chem Soc* 131: 5705–5713.
- Leaver M, Dominguez-Cuevas P, Coxhead JM, Daniel RA, Errington J (2009) Life without a wall or division machine in *Bacillus subtilis*. *Nature* 457: 849–853.
- Kurihara K, Tamura M, Shohda K I, Toyota T, Suzuki K, Sugawara T (2011) Self-reproduction of supramolecular giant vesicles combined with the amplification of encapsulated DNA. *Nature chemistry* 3:775–781.
- Markvoort AJ, Pflieger N, Staffhorst R, Hilbers PA, Van Santen RA, et al (2010) Self-reproduction of fatty acid vesicles: A combined experimental and simulation study. *Biophysical journal* 99: 1520–1528.
- Stano P, Wehrli E, Luisi PL (2006) Insights into the self-reproduction of oleate vesicles. *Journal of Physics: Condensed Matter* 18: S2231.
- Chen IA, Salehi-Ashtiani K, Szostak JW (2005) RNA catalysis in model protocell vesicles. *J Am Chem Soc* 127:13213–13219.
- Mansy SS, Schrum JP, Krishnamurthy M, Tobé S, Treco DA, et al (2008) Template-directed synthesis of a genetic polymer in a model protocell. *Nature* 454: 122–125.
- Müller UF, Bartel DP (2008) Improved polymerase ribozyme efficiency on hydrophobic assemblies. *RNA* 3:552–62.
- Wu M, Higgs PG (2009) Origin of self-replicating biopolymers: autocatalytic feedback can jump-start the RNA world. *J Mol Evol* 69: 541–54.
- Higgs PG, Wu M (2012) The Importance of Stochastic Transitions for the Origin of Life. *Orig Life Evol Biosph* 42(5):453–7.
- Wochner A, Attwater J, Coulson A, Holliger P (2011) Replicase-catalyzed transcription of an active replicase. *Science* 332: 209–212.
- Johnston W K, Unrau P J, Lawrence M S, Glasner M E, Bartel DP (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* 292: 1319–1325.
- Zaher HS, Unrau PJ (2007) Selection of an improved RNA polymerase replicase with superior extension and fidelity. *Rna* 13: 1017–1026.
- Lincoln TA, Joyce GF (2009) Self-sustained replication of an RNA enzyme. *Science* 323: 1229–1232.
- Doudna J A, Usman N, Szostak JW (1993) Replicase-catalyzed primer extension by trinucleotides: A model for the RNA-catalyzed replication of RNA. *Biochemistry* 32(8), 2111–2115.
- James KD, Ellington AD (1999) The fidelity of template-directed oligonucleotide ligation and the inevitability of polymerase function. *Origins of Life and Evolution of Biospheres* 29: 375–390.
- Jain SS, Anet FA, Stahle CJ, Hud NV (2004) Enzymatic Behavior by Intercalating Molecules in a TemplateDirected Ligation Reaction. *Angewandte Chemie International Edition* 43(15): 2004–8.
- Szostak JW (2011) An optimal degree of physical and chemical heterogeneity for the origin of life? *Philosophical Transactions of the Royal Society B: Biological Sciences* 366: 2894–2901.
- Kim DE, Joyce GF (2004) Cross-catalytic replication of an RNA ligase ribozyme. *Chem Biol* 11(11):1505–12.
- Paul N, Joyce GF (2002) A self-replicating ligase ribozyme. *Proc Natl Acad Sci U S A* 99(20):12733–40.
- Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, et al (2012) Spontaneous network formation among cooperative RNA replicators. *Nature* 491(7422):72–7.

51. Vasas V, Fernando C, Santos M, Kauffman S, Szathmary E. (2012) Evolution before genes. *Biol Direct* 7:1.
52. Chen IA, Roberts RW, Szostak JW (2004) The emergence of competition between model protocells. *Science* 305: 1474–1476.
53. Kun A, Santos M, Szathmary E (2005) Real ribozymes suggest a relaxed error threshold. *Nat Genet* 37(9):1008–11.
54. Eigen M, Schuster P (1977) the Hypercycle. A principle of Natural Self-Organization. Part A: Emergence of the Hypercycle. *Naturwissenschaften* 64:541–565.
55. Nowak M, P. Schuster (1989). Error thresholds of replication in finite populations. Mutation frequencies and the onset of Muller’s ratchet. *J theor Biol* 137: 375–395.
56. Nowak MA (1992). What is a quasispecies? *Trends Ecol Evol* 7: 118–121.
57. Leu K, Obermayer B, Rajamani S, Gerland U, Chen IA (2011) The prebiotic evolutionary advantage of transferring genetic information from RNA to DNA. *Nucl Acids Res* 39: 1–13.
58. Turk RM, Chumachenko NV, Yarus M (2010) Multiple translational products from a five-nucleotide replicase. *Proceedings of the National Academy of Sciences* 107: 4585–4589.
59. Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften* 58: 465–523.
60. Nowak MA (2006) *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press.
61. Mavelli F (2012) Stochastic simulations of minimal cells: the Ribocell model. *BMC Bioinformatics* 33:S10.
62. Zenisek SF, Hayden EJ, Lehman N (2007) Genetic exchange leading to self-assembling RNA species upon encapsulation in artificial protocells. *Artif Life* 13 : 279–89.

Selection for replicases in protocells: Text S1

Ginestra Bianconi, Kun Zhao, Irene A. Chen & Martin A. Nowak

I. THE GENERAL FRAMEWORK

We consider protocells containing self-replicating sequences of type A and type B. Self-replicating sequences of type A act as replicases and are able to speed up the replication of other sequences A and B in the same protocell. The rate of replication of sequences B in absence of sequences A is assumed to be a constant that we set to 1. The probability that a sequence A replicates without mutation is given by $q = (1 - u)^L$ where u is the probability of point mutations and L is the length of the sequence.

A. Transition rates

We distinguish between five types of replicases:

- **Replicase R1.**

If there is at least one sequence A (replicase $R1$) in the protocell, the replication of all the sequences occurs at rate $a > 1$.

- **Replicase R2.**

If there are two or more sequences A (replicases $R2$) in the protocell, the replication of all the sequences occurs at rate $a > 1$. If only one sequence A is present in the protocell, this sequence replicates at rate 1, while the B sequences in the protocell replicate at rate a .

- **Replicase $R1\alpha$.**

If there are i sequences A (replicases $R1\alpha$) in the protocell, the replication of all the sequences occurs at rate $1 + i\alpha$ with $\alpha > 0$.

- **Replicase $R2\alpha$.**

If there are i sequences A (replicases $R2\alpha$) in the protocell, the replication of all the A sequences occurs at rate $1 + (i - 1)\alpha$ and the replication of the B sequences occurs at rate $1 + \alpha i$ with $\alpha > 0$.

B. Division mechanisms

The division of the protocells can occur according to different dynamical rules. We have considered the following two division mechanisms:

- **i.** The protocell, when it reaches a certain cell size m , splits into two smaller protocells each one containing at least one sequence.
- **ii.** The protocell, when it reaches a certain size m , splits into m protocells each one containing a single sequence.

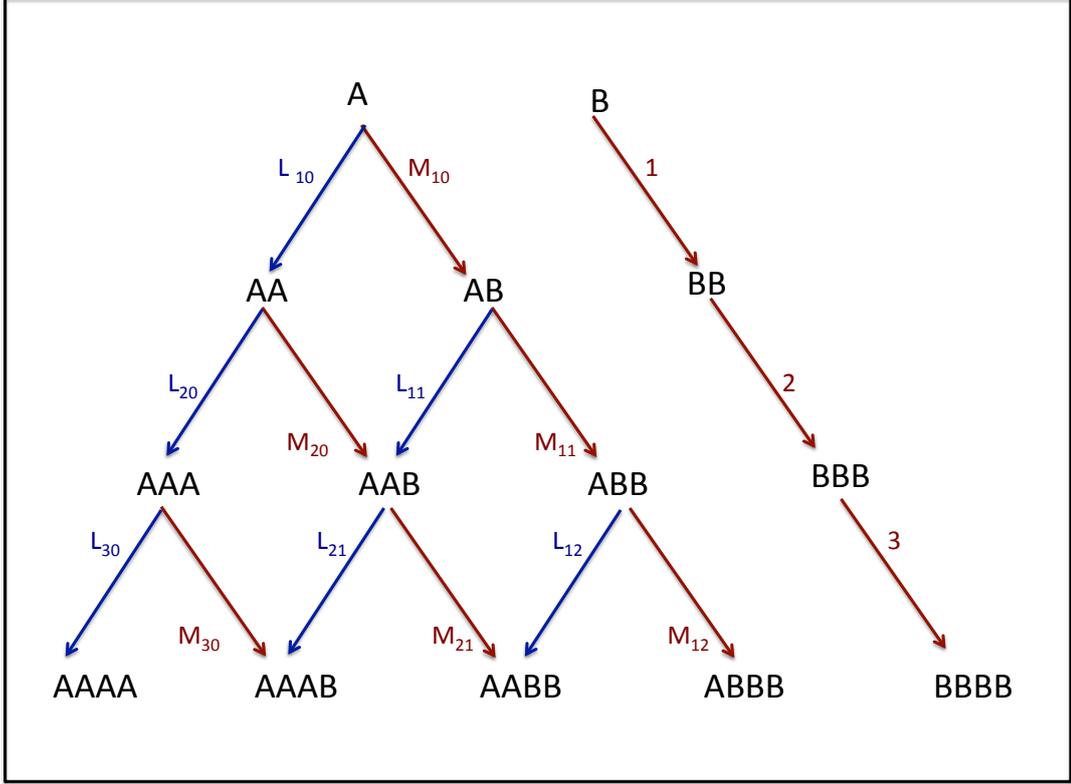


FIGURE S1 : General reaction kinetics for the evolution of protocells. The protocell composition includes both sequences of type A (replicases) and sequences of type B. Both sequences can self-replicate but sequences of type A (replicases) are able to speed up the replication rate of the other molecules inside the same protocell. When a sequence of type A replicates, a mutation occur with probability $1 - q$, giving rise to a sequence of type B. When a protocell reaches a maximum size, it splits.

II. GENERAL RESULTS

A. The mutation-selection-cell division (MSCD) equations

We indicate with $x_{i,j}$ the frequency of protocells of composition $A_i B_j$. The reaction kinetics for a general model of evolution of protocells is described in Figure S1. The MSCD equations read in the general case

$$\begin{aligned}
 \dot{x}_{1,0} &= -[M_{1,0} + L_{1,0}]x_{1,0} + d_{1,0} - \phi x_{1,0} \\
 \dot{x}_{i,0} &= -[M_{i,0} + L_{i,0}]x_{i,0} + L_{i-1,0}x_{i-1,0} + d_{i,0} - \phi x_{i,0} \\
 \dot{x}_{i,j} &= -[M_{ij} + L_{ij}]x_{ij} + M_{i,j-1}x_{i,j-1} + L_{i-1,j}x_{i-1,j} + d_{i,j} - \phi x_{i,j} \quad i > 1 \& j > 0 \\
 \dot{x}_{1,j} &= -[M_{1j} + L_{1j}]x_{1j} + M_{1,j-1}x_{1,j-1} + d_{1,j} - \phi x_{1,j} \quad j > 0 \\
 \dot{x}_{0,1} &= -x_{0,1} + d_{0,1} - \phi x_{0,1} \\
 \dot{x}_{0,j} &= -jx_{0,j} + (j-1)x_{0,j-1} + d_{0,j} - \phi x_{0,j} \quad j > 1.
 \end{aligned} \tag{1}$$

In these equations the rates $M_{i,j}, L_{i,j}$ depend on the particular model under consideration and they will be specified for any particular model taken in consideration in the following sections. Here we focus on the general aspects of the MSCD equations which are independent of the model specification. In Eqs. (1) $d_{i,j}$ denotes the rate at which protocells of composition $A_i B_j$ are formed as a consequence of the division of protocells of size m . For division into two daughter cells, $d_{i,j}$ can be written as

$$d_{i,j} = \sum_{i' \geq i, j' \geq j, i'+j'=m} \frac{\binom{i'}{i} \binom{j'}{j}}{2^{m-1} - 1} r_{i',j'}. \quad (2)$$

For division into many (m) daughter cells, $d_{i,j}$ can be written as

$$\begin{aligned} d_{1,0} &= \sum_{i \geq 1, i+j=m} i r_{i,j} \\ d_{0,1} &= \sum_{j \geq 1, i+j=m} j r_{i,j} \\ d_{i,j} &= 0 \quad (i > 1 \text{ or } j > 1). \end{aligned} \quad (3)$$

In Eq. (2) and Eqs. (3) the dissociation rates $r_{i,j}$ of protocells with $i + j = m$ are given by

$$\begin{aligned} r_{i,j} &= L_{i,j-1} x_{i,j-1} (1 - \delta_{j,0}) + M_{i-1,j} x_{i-1,j} (1 - \delta_{i,1}) \quad i + j = m > 3 \& i > 1 \\ r_{0,m} &= (m-1) x_{0,m-1} \end{aligned} \quad (4)$$

The parameter ϕ is the fitness of the population of protocell, and we can use that to normalize either the total number of protocell in the system $\sum_{ij} x_{ij} = 1$ or the total number of molecules $\sum_{ij} (i+j) x_{ij}$. The results on the error threshold are independent on the type of normalization adopted.

B. The error threshold: implicit equation

In order to find an equation for the error threshold we investigate the structure of the MSCD Eqs. (1). We define a vector \vec{y}^A including the frequency of protocells containing at least one sequence A,

$$\vec{y}^A = \{x_{i,j}\}_{i \geq 1}. \quad (5)$$

Similarly we define a vector \vec{y}^B including frequency of protocells containing only sequences B,

$$\vec{y}^B = \{x_{0,j}\}_{j \geq 1}. \quad (6)$$

The quasi-species equations given by Eqs.(1) can be written as

$$\begin{aligned} \dot{\vec{y}}^A &= C_{AA} \vec{y}^A - \phi \vec{y}^A \\ \dot{\vec{y}}^B &= C_{BA} \vec{y}^A + C_{BB} \vec{y}^B - \phi \vec{y}^B \end{aligned} \quad (7)$$

where C_{AA}, C_{BA}, C_{BB} are matrices whose explicit expression can be found by considering Eqs.(1). The stationary solution is reached when $\phi = \lambda_{max}$ where λ_{max} is the maximal eigenvalue of the eigenvalue problem

$$C_{AA} \vec{y}^A = \lambda \vec{y}^A$$

$$C_{BA}\vec{y}^A + C_{BB}\vec{y}^B = \lambda\vec{y}^B. \quad (8)$$

To calculate the error threshold $q_c(a)$, in the appendix A we show that one solution exist with $\vec{y}^A = \vec{0}$, $\vec{y}_i^B > 0 \forall i$ exists and the maximal eigenvalue associated with such solution is $\lambda_{max} = 1$. In the appendix we will also show that this result is independent on the type of splitting mechanism adopted.

Therefore the error threshold of the models are provided by the following implicit equation

$$|C_{AA} - \phi I| = |C_{AA} - I| = 0. \quad (9)$$

We have used this equation to find the error threshold for all the models under consideration. In the case in which this equation has multiple solutions $q = q_c \in (0, 1]$, according to the MSCD equation, we have to find the solution $q = q_c$ for which $\phi = 1$ is the maximal eigenvalue of the matrix C_{AA} . In order to give a concrete example on how this calculation is performed in appendix B we give the analytic solution of the error threshold for the replicase R2 with $m = 3$.

C. The error threshold: General relation

Let consider the problem of finding a non zero solution $\vec{y}^A \neq \vec{0}$ to the problem

$$C_{AA}\vec{y}^A = \vec{y}^A \quad (10)$$

as long as $q = q_c(a)$. The system of equations that we need to solve is then

$$x_{i,j} = -[M_{i,j} + L_{i,j}]x_{i,j} + M_{i,j-1}x_{i,j-1}(1 - \delta_{j,0}) + M_{i-1,j}x_{i-1,j}(1 - \delta_{i,1}) + d_{i,j} \quad i \geq 1 \quad (11)$$

Multiplying each of these equation by i and summing over i and j we can prove that at $q = q_c(a)$ the frequencies of protocells $x_{i,j}$ should satisfy

$$\sum_{i \geq 1, j} ix_{i,j} = \sum_{i \geq 1, j} L_{i,j}x_{i,j}. \quad (12)$$

In fact it can be shown that

$$\sum_{i,j} id_{i,j} = \sum_{i,j,i+j=m} ir_{i,j} \quad (13)$$

independently from the splitting method taken under consideration. Eq. (12) is a very useful relation to determine some bounds for the error-threshold of the model. Nevertheless this equation doesn't solve the entire MSCD model because it depends on the values of the frequency of the protocells $x_{i,j}$.

III. SPECIFIC MODELS

A. Replicase R1

In this model if there is at least one sequence A in the protocell, the replication of all the sequences occur at rate $a > 1$. We consider at the same time the two proposed mechanisms for protocell division.

The rates $L_{i,j}, M_{i,j}$ are given by

$$\begin{aligned} L_{i,j} &= iaq \\ M_{i,j} &= ia(1-q) + ja. \end{aligned} \quad (14)$$

Considering the general relation Eq. (12) and substituting the rates given by Eqs. (B12) we get that, at the error threshold, the frequencies $x_{i,j}$ satisfy

$$\sum_{i \geq 1, j} ix_{i,j} = aq_c \sum_{i \geq 1, j} ix_{i,j} \quad (15)$$

Therefore

$$q_c = \frac{1}{a} \quad (16)$$

independently of the maximal size of the protocell m and the division mechanism.

B. Replicase R2

In this model the enzyme A is able to speed up the reaction of other molecules (either sequences A or B) but not itself. Therefore the rates of duplications are given by

$$\begin{aligned} L_{i,j} &= q && \text{for } i = 1 \\ L_{i,j} &= iaq && \text{for } i \geq 2 \\ M_{i,j} &= aj + (1-q) && \text{for } i = 1 \\ M_{i,j} &= aj + ai(1-q) && \text{for } i \geq 2 \end{aligned}$$

By using the relation Eq. (12) valid at the error threshold, we get that

$$\sum_{i \geq 1, j} ix_{i,j} = aq_c \sum_{i \geq 1} ix_{i,j} - q_c(a-1) \sum_j x_{1,j} \quad (17)$$

Since we assume $a > 1$ we can show that with these rates the selection of an enzyme is more difficult than in the previous model and we have

$$q_c \geq \frac{1}{a}. \quad (18)$$

Moreover by solving the Eqs. (1) in the large a limit we can show that in this limit

$$\begin{aligned} x_{1,0} &= d_{1,0}/2 = \mathcal{O}(1) \\ x_{i,j} &= \mathcal{O}\left(\frac{1}{a}\right) && \text{for } i \geq 1 \&\& (i,j) \neq (1,0) \end{aligned}$$

Therefore, taking into account this scaling and Eq. (17) we can show that

$$q_c(a) \rightarrow q_c(\infty) > 0 \text{ as } a \rightarrow \infty \quad (19)$$

We have numerically calculated the error threshold for this model up to $m = 10$ considering both types of protocell division, i.e. the case in which each protocell of maximal size m splits into two protocells or the case in which it splits into many protocells. The results show, as the theory and Eq. (19) predict, that in this case the selection for

sequences of type A (replicases $R2$) is more difficult than the selection of replicases $R1$ considered in the previous section.

The simulation results conducted for the case in which the protocells of maximal size divide into two, show a crossing of the error threshold curves $q_c = q(a)$ calculated for different values of the maximal size of the protocell m . For example, we observe that, for $a = 200$, the error threshold for $m = 8$ is smaller than the error threshold for $m = 10$ (See Table S1), while for $a = 10$, we observe that the error threshold for $m = 8$ is greater than the error threshold for $m = 10$. This suggests that, for a value of $a = 200$, there is an optimum size of the protocell. We have performed the numerical calculation up to value of $a = 10,000$ for protocells of maximal size $m \in [3, 10]$. These results suggest that in the large a limit ($a \gg 1$) the error threshold q_c is a decreasing function of m .

The simulation results conducted for the case in which the protocells of maximal size divide into many, show that, for every value of a the error threshold is a decreasing function of m . Therefore if the protocells can become larger the selection for the replicase is easier.

C. Replicase $R1\alpha$

In this model the rate of enzymatic replication depends on the amount of A in the protocell and is given by $1 + i\alpha$ where i is the number of enzymes A in the protocell. The rate of duplications $M_{i,j}, L_{i,j}$ are given by

$$\begin{aligned} L_{i,j} &= i(1 + \alpha i)q \\ M_{i,j} &= i(1 + \alpha i)(1 - q) + j(1 + i\alpha) \end{aligned} \quad (20)$$

By using the relation Eq. (12) valid at the error threshold, we get that

$$\sum_{i \geq 1, j} ix_{i,j} = q_c \sum_{i \geq 1, j} i(1 + \alpha i)x_{i,j} \quad (21)$$

Therefore we have

$$q_c = \frac{\sum_{i,j} ix_{i,j}}{\sum_{i,j} ix_{i,j} + \alpha \sum_{i,j} i^2 x_{i,j}} \quad (22)$$

satisfying the following conditions

$$\frac{1}{1 + (m-1)\alpha} \leq q_c \leq \frac{1}{1 + \alpha}. \quad (23)$$

Therefore for large values of α the error threshold goes to zero.

We have numerically calculated the error threshold for this model up to $m = 10$ considering both types of protocell division. The results show, as the theory and Eq. (23) predict, that $q_c \leq 1/(1 + \alpha)$. Moreover, this bound is tight for large values of α . Finally, the simulation results conducted for the case in which the protocells of maximal size divide into two or into many, show that, for every value of α the error threshold is a decreasing function of m . Therefore if the protocells can become larger the selection for the replicase is easier.

D. Replicase R2 α

In this model the rate of enzymatic replication depends on the amount of A in the protocell. If i sequences of type A are present in the protocell, the A sequences replicate at rate $1 + \alpha(i - 1)$ and the B sequences replicate at rate $1 + \alpha i$. In this case the rates of duplication M_{ij}, L_{ij} are given by

$$\begin{aligned} L_{i,j} &= i[1 + \alpha(i - 1)]q && \text{for } i \geq 2 \\ M_{i,j} &= (1 + \alpha i)j + i[1 + \alpha(i - 1)](1 - q) \end{aligned}$$

By using the relation Eq. (12) valid at the error threshold, we get that

$$\sum_{i \geq 1, j} ix_{i,j} = q_c \sum_{i \geq 1, j} i[1 + \alpha(i - 1)]x_{i,j} \quad (24)$$

Therefore we have

$$q_c = \frac{\sum_{i,j} ix_{i,j}}{\sum_{i,j} i[1 + \alpha(i - 1)]x_{i,j}} \quad (25)$$

satisfying the condition

$$q_c > \frac{1}{1 + (m - 2)\alpha}. \quad (26)$$

Moreover by solving the Eqs. (1) in the large α limit we can show that in this limit

$$\begin{aligned} x_{1,0} &= d_{1,0}/2 = \mathcal{O}(1) \\ x_{i,j} &= \mathcal{O}\left(\frac{1}{\alpha}\right) \quad \text{for } i \geq 1 \& (i, j) \neq (1, 0) \end{aligned}$$

Therefore, taking into account this scaling and Eq. (17) we can show that

$$q_c(\alpha) \rightarrow q_c(\infty) > 0 \quad \text{as } \alpha \rightarrow \infty \quad (27)$$

We have numerically calculated the error threshold for this model up to $m = 10$ considering both types of protocell division (see figure S2). The simulation results, conducted for the case in which the protocells of maximal size divide into two, show a crossing of the error threshold curves $q_c = q(\alpha)$ calculated for different values of the maximal size of the protocell m , suggesting that, for a value of $\alpha = 100$, there is a optimum size of the protocell. We have conducted the numerical calculations up to values $\alpha = 10,000$. These results suggest that in the large α limit, $\alpha \gg 1$, the error threshold is a decreasing function of the maximal size of the protocell m .

The simulation results conducted for the case in which the protocells of maximal size divide into many, show that, for every value of α the error threshold is a decreasing function of m . Therefore if the protocells can become larger the selection for the replicase is easier.

TABLE S1 : Maximal length of the selected replicase L_{\max}

Division into two	m	$L_{\max}(u = 0.17)$	$L_{\max}(u = 0.0088)$
$R1$	any m	24	521
$R2$	3	4	89
$R2$	4	5	117
$R2$	5	6	130
$R2$	8	6	141
$R2$	10	6	138
$R1\alpha$	3	28	600
$R1\alpha$	4	28	600
$R1\alpha$	5	28	600
$R1\alpha$	8	28	600
$R1\alpha$	10	28	600
$R2\alpha$	3	4	89
$R2\alpha$	4	5	112
$R2\alpha$	5	5	121
$R2\alpha$	8	5	123
$R2\alpha$	10	5	115
Division into many	m	$L_{\max}(u = 0.17)$	$L_{\max}(u = 0.0088)$
$R1$	any m	24	521
$R2$	3	2	54
$R2$	4	3	74
$R2$	5	4	86
$R2$	8	5	106
$R2$	10	5	113
$R1\alpha$	3	28	599
$R1\alpha$	4	28	600
$R1\alpha$	5	28	600
$R1\alpha$	8	28	600
$R1\alpha$	10	28	600
$R2\alpha$	3	2	54
$R2\alpha$	4	3	73
$R2\alpha$	5	3	83
$R2\alpha$	8	4	100
$R2\alpha$	10	5	107

Maximal length of the selected replicase L_{\max} calculated by imposing $(1 - u)^{L_{\max}} = q_c$ for the different models under consideration with $a = 200$ or $\alpha = 200$. The parameter m is the size of the largest protocell.

Appendix A: The fitness of the solution with no A molecules , i.e. $\vec{y}^A = \vec{0}$

When there are no molecules of type A in the protocells, i.e. $\vec{y}^A = \vec{0}$, the eigenvector problem Eqs.(8) becomes

$$C_{BB}\vec{y}^B = \lambda\vec{y}^B \quad (\text{A1})$$

and λ can be obtained from equation

$$|C_{BB} - \lambda I| = 0. \quad (\text{A2})$$

The fitness ϕ is equal to the maximal eigenvalue of the eigenvalue problem in Eq. (A1), i.e. $\phi = \lambda_{max}$. In the case in which we consider the splitting into two daughter protocells the matrix C_{BB} can be explicitly written as

$$C_{BB} = \begin{pmatrix} -1 & & & & & \frac{(m-1)m}{2^{m-1}-1} \\ 1 & -2 & & & & \frac{(m-1)\binom{m}{2}}{2^{m-1}-1} \\ & \ddots & \ddots & & & \vdots \\ & & i-1 & -i & & \frac{(m-1)\binom{m}{i}}{2^{m-1}-1} \\ & & & \ddots & \ddots & \vdots \\ & & & & m-3 & -(m-2) & \frac{(m-1)\binom{m}{m-2}}{2^{m-1}-1} \\ & & & & & m-2 & -(m-1) + \frac{(m-1)\binom{m}{m-1}}{2^{m-1}-1} \end{pmatrix}_{(m-1) \times (m-1)} \quad (A3)$$

and therefore $|C_{BB} - \lambda I|$ can be explicitly expressed as

$$|C_{BB} - \lambda I| = (-1)^{m-2} \frac{\Gamma(\lambda + m)}{\Gamma(\lambda + 1)} \left[\frac{(m-1)!}{2^{m-1}-1} \sum_{i=1}^{m-1} \frac{\Gamma(\lambda + i)}{\Gamma(\lambda + m)} \cdot \frac{\binom{m}{i}}{(i-1)!} - 1 \right] \equiv (-1)^{m-2} \frac{\Gamma(\lambda + m)}{\Gamma(\lambda + 1)} \Phi(\lambda). \quad (A4)$$

One should note that

$$\begin{aligned} \Phi(1) &= \frac{(m-1)!}{2^{m-1}-1} \sum_{i=1}^{m-1} \frac{\Gamma(i+1)}{\Gamma(m+1)} \cdot \frac{\binom{m}{i}}{(i-1)!} - 1 \\ &= \frac{1}{m} \sum_{i=1}^{m-1} \frac{i \binom{m}{i}}{2^{m-1}-1} - 1 \\ &= \frac{1}{m} \frac{m2^{m-1} - m}{2^{m-1}-1} - 1 = 0. \end{aligned} \quad (A5)$$

Therefore $\lambda = 1$ is one eigenvalue. Since $\Phi(\lambda)$ is monotonically decreasing function, $\lambda = 1$ is the only real root of Eq.(A2). Similarly, if we consider the splitting of the protocells into many daughter protocells, C_{BB} can be written as

$$C_{BB} = \begin{pmatrix} -1 & & & & & & (m-1)m \\ 1 & -2 & & & & & \\ & \ddots & \ddots & & & & \\ & & i-1 & -i & & & \\ & & & \ddots & \ddots & & \\ & & & & m-3 & -(m-2) & \\ & & & & & m-2 & -(m-1) \end{pmatrix}_{(m-1) \times (m-1)}. \quad (A6)$$

Therefore we can explicitly calculate the following determinant, obtaining

$$|C_{BB} - \lambda I| = (-1)^{m-2} \frac{\Gamma(\lambda + m)}{\Gamma(\lambda + 1)} \left[\frac{\Gamma(\lambda + 1)}{\Gamma(\lambda + m)} \cdot m! - 1 \right] = (-1)^{m-2} \frac{\Gamma(\lambda + m)}{\Gamma(\lambda + 1)} \Psi(\lambda). \quad (A7)$$

The function $\Psi(\lambda)$ is a decreasing function of λ , with $\Psi(1) = 0$. Therefore $\lambda = 1$ is the maximal real solution of the system of equations. Finally we conclude that the fitness of a system of protocells in which the molecules A are not present is $\phi = \lambda_{max} = 1$ independently on the splitting mechanism under consideration.

Appendix B: Complete analytic solution of the error threshold for replicase R2 with $m = 3$

1. Splitting into two daughter protocells

The quasi-species equations for the frequency of protocells x_{ij} of protocells containing different type of molecules is given by

$$\begin{aligned}
 \dot{x}_{1,0} &= -x_{1,0} + \left(r_{3,0} + \frac{2}{3}r_{2,1} + \frac{1}{3}r_{1,2}\right) - \phi x_{1,0} \\
 \dot{x}_{0,1} &= -x_{0,1} + \left(\frac{1}{3}r_{2,1} + \frac{2}{3}r_{1,2} + r_{0,3}\right) - \phi x_{0,1} \\
 \dot{x}_{2,0} &= r_{3,0} + \frac{1}{3}r_{2,1} + qx_{1,0} - 2ax_{2,0} - \phi x_{2,0} \\
 \dot{x}_{1,1} &= \frac{2}{3}r_{2,1} + \frac{2}{3}r_{1,2} + (1-q)x_{1,0} - (1+a)x_{1,1} - \phi x_{1,1} \\
 \dot{x}_{0,2} &= r_{0,3} + \frac{1}{3}r_{1,2} + x_{0,1} - 2x_{0,2} - \phi x_{0,2}
 \end{aligned} \tag{B1}$$

where the rates $r_{3,0}, r_{2,1}, r_{1,2}$ and $r_{0,3}$ are the rate of dissociation of the protocells with three molecules and are given by

$$\begin{aligned}
 r_{3,0} &= 2aqx_{2,0} \\
 r_{2,1} &= 2a(1-q)x_{2,0} + qx_{1,1} \\
 r_{1,2} &= (a+1-q)x_{1,1} \\
 r_{0,3} &= 2x_{0,2}.
 \end{aligned} \tag{B2}$$

The fitness ϕ depends on the conservation law that we impose to the system. In fact,

- if the total number of protocells $\sum_{ij} x_{ij} = 1$ is conserved then

$$\phi = 2ax_{2,0} + (1+a)x_{1,1} + 2x_{0,2}; \tag{B3}$$

- if the total number of molecules $\sum_{ij} (i+j)x_{ij} = 1$ is conserved then

$$\phi = x_{1,0} + x_{0,1} + 2ax_{2,0} + (1+a)x_{1,1} + 2x_{0,2}. \tag{B4}$$

The error threshold of the model is independent on the conservation law that we imposed. To calculate the error threshold $q_c(a)$ we need to take advantage of the structure of the transition matrix shown in Figure .

Therefore let us define the vector \vec{y}^A of the frequency of protocells containing at least one molecule A

$$\vec{y}^A = \begin{pmatrix} x_{1,0} \\ x_{2,0} \\ x_{1,1} \end{pmatrix}$$

and the vector \vec{y}^B of concentration of protocells containing only molecules of type B,

$$\vec{y}^B = \begin{pmatrix} x_{0,1} \\ x_{0,2} \end{pmatrix}.$$

The quasi-species equations given by Eqs (B1) – (B2) read

$$\begin{aligned}\dot{\vec{y}}^A &= C_{AA}\vec{y}^A - \phi\vec{y}^A \\ \dot{\vec{y}}^B &= C_{BA}\vec{y}^A + C_{BB}\vec{y}^B - \phi\vec{y}^B.\end{aligned}\tag{B5}$$

The matrices C_{AA} , C_{BA} and C_{BB} are given by

$$\begin{aligned}C_{AA} &= \begin{pmatrix} -1 & \frac{2}{3}a(2+q) & \frac{1}{3}(1+a+q) \\ q & -\frac{4}{3}(1-q) & \frac{1}{3}q \\ 1-q & \frac{4}{3}a(1-q) & -\frac{1}{3}(1+a) \end{pmatrix} \\ C_{BB} &= \begin{pmatrix} -1 & 2 \\ 1 & 0 \end{pmatrix} \\ C_{BA} &= \begin{pmatrix} 0 & \frac{2}{3}a(1-q) & \frac{1}{3}[2(a+1)-q] \\ 0 & 0 & \frac{1}{3}(a+1-q) \end{pmatrix}.\end{aligned}$$

The stationary state is reached when $\phi = \lambda_{max}$ where $\lambda_{max} = 1$ is the maximal eigenvalue of the eigenvalue problem

$$\begin{aligned}C_{AA}\vec{y}^A &= \lambda\vec{y}^A \\ C_{BA}\vec{y}^A + C_{BB}\vec{y}^B &= \lambda\vec{y}^B.\end{aligned}\tag{B6}$$

One solution of this system of equation is $\vec{y}^A = \vec{0}$, $\vec{y}^B = (1/2, 1/2)$ associated with the maximal eigenvalue $\lambda = 1$. This solution becomes unstable when the maximal eigenvalue of the eigenvalue problem Eqs. (B6) become greater than one. This occurs at the error threshold $q = q_c(a)$. In order to study when this happens is sufficient to study the spectrum of the eigenvalue problem

$$C_{AA}\vec{y}^A = \lambda\vec{y}^A\tag{B7}$$

and to impose that the maximal eigenvalue of the matrix C_{AA} is equal to one, i.e. $\lambda_{max} = 1$. This happens when

$$\det(C_{AA} - 1) = 0.\tag{B8}$$

This equation is cubic in q and therefore will have three solutions. The only solution in the range $q \in [0, 1]$ provides the required solution $q_c = q_c(a)$ describing the error threshold. The solution to the equation $\det(C_{AA} - 1) = 0$ is given by

$$q_c(a) = \frac{2a^2 - 6a - 3}{6a} + \frac{1}{6a} [(P_1(a))^2 + 81P_2(a)]^{1/6} \left\{ \cos(\theta/3) - \sqrt{3} \sin(\theta/3) \right\}\tag{B9}$$

with

$$\begin{aligned}P_1(a) &= 27 + 162a + 2943a^2 + 4968a^3 - 558a^4 - 72a^5 - 8a^6 \\ P_2(a) &= -756a^2 - 5652a^3 - 63621a^4 - 173016a^5 + 366204a^6 + \\ &197520a^7 + 33484a^8 + 2784a^9 + 128a^{10}\end{aligned}$$

and

$$\tan \theta = \frac{9\sqrt{P_2(a)}}{P_1(a)} \text{ with } \theta \in [0, \pi).\tag{B10}$$

The limits of $q_c(a)$ are $q_c(a) \rightarrow 1$ as $a \rightarrow 1$ and $q_c(a) \rightarrow -2 + \sqrt{6} = 0.449\dots$ as $a \rightarrow \infty$. This error threshold is independent on the constraints that we are imposing (i.e. conservation of the total number of molecules or conservation of the total number of cells).

2. Splitting into many daughter protocells

The quasi-species equations for the frequencies x_{ij} of protocells containing different type of molecules is given by

$$\begin{aligned}
\dot{x}_{1,0} &= -x_{1,0} + (3r_{3,0} + 2r_{2,1} + 1r_{1,2}) - \phi x_{1,0} \\
\dot{x}_{0,1} &= -x_{0,1} + (r_{2,1} + 2r_{1,2} + 3r_{0,3}) - \phi x_{0,1} \\
\dot{x}_{2,0} &= qx_{1,0} - 2ax_{2,0} - \phi x_{2,0} \\
\dot{x}_{1,1} &= (1 - q)x_{1,0} - (1 + a)x_{1,1} - \phi x_{1,1} \\
\dot{x}_{0,2} &= x_{0,1} - 2x_{0,2} - \phi x_{0,2}
\end{aligned} \tag{B11}$$

where the rates $r_{3,0}, r_{2,1}, r_{1,2}$ and $r_{0,3}$ are the rate of dissociation of the protocell with three molecules and are given by

$$\begin{aligned}
r_{3,0} &= 2aqx_{2,0} \\
r_{2,1} &= 2a(1 - q)x_{2,0} + qx_{1,1} \\
r_{1,2} &= (a + 1 - q)x_{1,1} \\
r_{0,3} &= 2x_{0,2}.
\end{aligned} \tag{B12}$$

The parameter ϕ depends on the conservation law that we impose to the system.

In fact,

- if the total number of cells $\sum_{ij} x_{ij} = 1$ is conserved then

$$\phi = 4ax_{2,0} + 2(1 + a)x_{1,1} + 4x_{0,2}; \tag{B13}$$

- if the total number of molecules $\sum_{ij} (i + j)x_{ij} = 1$ is conserved then

$$\phi = x_{1,0} + x_{0,1} + 2ax_{2,0} + (1 + a)x_{1,1} + 2x_{0,2}. \tag{B14}$$

Proceeding as in previous section we get that the error threshold is given by

$$q_c(a) = \frac{-7a - 2a^2 + \sqrt{-12 - 4a + 121a^2 + 100a^3 + 20a^4}}{2(-1 + 2a + 2a^2)}. \tag{B15}$$

The limit for large values of a , i.e. $a \rightarrow \infty$ is given by $q_c(a) \rightarrow \frac{1}{2}(-1 + \sqrt{5})$.



[plos.org](#) [create account](#) [sign in](#)

Browse | For Authors | About Us

Search

advanced search

<p> OPEN ACCESS PEER-REVIEWED</p> <p>RESEARCH ARTICLE</p> <h2 style="margin: 0;">Selection for Replicases in Protocells</h2> <p style="margin: 0;">Ginestra Bianconi, Kun Zhao, Irene A. Chen, Martin A. Nowak </p>	<p>1,427</p> <p style="font-size: x-small;">VIEWS</p>	<p>5</p> <p style="font-size: x-small;">ACADEMIC BOOKMARKS</p>	<p>3</p> <p style="font-size: x-small;">SOCIAL SHARES</p>
--	---	--	---

Article	About the Authors	Metrics	Comments	Related Content
---------	-------------------	---------	----------	-----------------

Download

Print

Share

Reader Comments (3)

[Post a new comment](#) on this article

Correction to Materials and Methods MSCD equations

Posted by [irenechen](#) on 14 May 2013 at 19:59 GMT

In the rate equation for x_{1j} , j should replace i in the first term.

Also note that the rate equation for x_{i0} applies when $i > 1$.

No competing interests declared.

report a concern
respond to this posting

Comments

Bacillus subtilis still had a cell membrane with numerous transmembrane proteins, glycoproteins, etc.
Posted by GenotypePhenotype

Thank you for an excellent article.
Posted by GenotypePhenotype

Correction to Materials and Methods MSCD equations
Posted by irenechen

ADVERTISEMENT

VISIT US AT BOOTH 13

21st Annual International Conference on Intelligent Systems for Molecular Biology

July 21-23
Berlin

<p style="font-size: x-small; margin-top: 10px;">Ambra 2.7.0 Managed Colocation provided by Internet Systems Consortium.</p> <p style="font-size: x-small; margin-top: 10px;"> Privacy Policy Terms of Use Advertise Media Inquiries </p>	<p>Publications</p> <p style="font-size: x-small;"> PLOS Biology PLOS Medicine PLOS Computational Biology PLOS Currents PLOS Genetics PLOS Pathogens PLOS ONE PLOS Neglected Tropical Diseases </p>	<p>plos.org</p> <p>Blogs</p> <p>Collections</p> <p>Send us feedback</p>
---	---	---