Coevolution of reproducers and replicators at the origin of life and the conditions for the origin of genomes

Sanasar G. Babajanyan[1], Yuri I. Wolf[1], Andranik Khachatryan[2], Armen Allahverdyan[2], Eugene V. Koonin[1,*]

1 National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health of the USA, Bethesda, MD 20894, USA.

2 Alikahanyan National Laboratory (Yerevan Physics Institute), 2 Alikhanyan Brothers Street, Yerevan 0036, Armenia.

For correspondence: koonin@ncbi.nlm.nih.gov

Abstract

There are two fundamentally distinct but inextricably linked types of biological evolutionary units, reproducers and replicators. Reproducers are cells and organelles that reproduce via various forms of division and maintain the physical continuity of compartments and their content. Replicators are genetic elements (GE), including genomes of cellular organisms and various autonomous elements, that both cooperate with reproducers and rely on the latter for replication. All known cells and organisms comprise a union between replicators and reproducers. We explore a model in which cells emerged via symbiosis between primordial 'metabolic' reproducers (protocells) which evolved, on short time scales, via a primitive form of selection and random drift, and mutualist replicators. Mathematical modeling identifies the conditions, under which GE-carrying protocells can outcompete GE-less ones, taking into account that, from the earliest stages of evolution, replicators split into mutualists and parasites. Analysis of the model shows that, for the GE-containing protocells to win the competition and to be fixed in evolution, it is essential that the replication rate of the GE is coordinated with the rate of protocell division. At the early stages of evolution, random, high-variance cell division is advantageous compared to symmetrical division because the former provides for the emergence of protocells containing only mutualists, preventing takeover by parasites. These findings illuminate the likely order of key events on the evolutionary route from protocells to cells that involved the origin of genomes, symmetrical cell division and anti-parasite defense systems.

Significance

The origin of life, which is equivalent to the origin of cells, is arguably the greatest enigma in today's biology. The remarkable complexity characteristic of even the simplest extant cells could only evolve from simpler, pre-biological entities. Reconstructing that pre-cellular stage of evolution is a hard challenge. We present an evolutionary scenario in which cells evolved via symbiosis between protocells that harbored protometabolic reaction networks, could divide and were subject to selection, but lacked genomes, and primordial genetic elements. Mathematical modeling reveals conditions for the survival of such symbionts and the origin of modern-type genomes, in particular, coordination of the rates of protocell division and replication of genetic elements as well as random division of protocells.

Introduction

Replication of genetic information is naturally considered a fundamental – or, often, the central – property of evolving biological entities, both cellular organisms and genetic parasites. All these entities possess genomes that are often also called replicators (1-3). Evidently, however, life is not limited to information transmission. An adequate supply of energy and building blocks, which depends on spatial compartmentalization, is essential for the evolution of replicators. Hence the fundamental split of all propagating biological entities into reproducers and replicators (3). Cells are reproducers: their propagation is not limited to the replication of the genome but rather, involves reproduction of the entire cellular organization that provides the niche for the replicators. Although the genome carries the instructions for the production of all cell components, it is in itself insufficient for reproduction: *Omnis cellula e cellula*. The entire evolutionary history of life is an uninterrupted, physically continuous tree of cell divisions (in which the dead branches, evidently, overwhelmingly outnumber the growing ones). By contrast, all diverse genetic elements (GE) including both cellular and organellar genomes and genetic parasites (viruses, transposons) are replicators that recruit cellular molecular machinery for some of the key functions required for their replication, in particular, translation of the GE genes (4).

The extant reproducers (cells) that necessarily host a mutualistic replicator (the genome) are compartments bounded by phospholipid membranes permeated by diverse proteins containing hydrophobic transmembrane segments (5). Compartmentalization is essential not only for preventing diffusion of small molecules into the environment and thus keeping their concentrations inside the reproducer at levels sufficient to sustain metabolic reaction networks as well as replication, but also to maintain the integrity of selectable units that consist of reproducers together with the replicators inside them. The membranes perform essential transport functions, that is, selectively import molecules and ions that are required for cell reproduction (including replication), and expel toxic molecules and ions, often in an energy-dependent manner. In respiring cells, membranes also harness the energy released in oxidation reactions to produce ion gradients that are then transformed into the energy of the macroergic phosphodiester bond of ATP. Replicators lack such active, energizable membranes and thus depend on the reproducers for energy and building blocks required for replication. Furthermore, replicators and reproducers dramatically differ in terms of the chemistries involved in their propagation. The replication process requires only narrowly focused chemistry, namely, nucleotide polymerization. Evidently, this process is underpinned by the far more complex reactions of nucleotide biosynthesis, but these are supplied by the reproducer. In contrast, even the simplest reproducers exercise a rich repertoire of chemical reactions, with at least 1000 distinct small molecules metabolized by any cell type (6).

All replicators are hosted by reproducers and depend on the hosts for energy and building blocks. However, in terms of their relationships with the host reproducer, replicators span the entire range from (near) full cooperativity and a mutualistic relationship with the host reproducer in the case of cellular genomes through commensalism in the case of plasmids and transposons, to aggressive parasitism, in the case of lytic viruses (7, 8). Arguably, only a mutualistic, obligatory union of a host reproducer with a resident replicator(s), the genome that carries instructions for the reproduction of the host, can be considered a life form (organisms), in the crucial sense of temporally continuous, robust reproduction and the ensuing evolutionary autonomy. Thus, life took off when replicators evolved to encode components of the host reproducers, providing for the long-term persistence and evolution of the latter.

Origin of life often has been discussed in terms of competing ‘metabolism first’ vs ‘replication first’ scenarios (9), which can be reformulated as “reproducers first or replicators first?” The question seems formidable and resembles a chicken and egg problem. Indeed, the fundamental differences between

reproducers and replicators notwithstanding, these two types of evolving biological entities are inextricably linked. To the best of our knowledge, there are no 'pure' reproducers in the extant biosphere: all modern cells as well as some organelles, such as mitochondria and chloroplasts, comprise an obligatory mutualistic union of a reproducer and a replicator(s). We submit, however, that this obligatory relationship did not exist at the primordial, pre-biological stage of evolution, which started with primitive reproducers and eventually led to the emergence of the mutualistic reproducer-replicator systems. Indeed, emergence of replication is inconceivable without a steady supply of energy and building blocks, which only can be provided by a proto-metabolism with sufficient temporal stability, that is, by some form of primordial reproducers. Realistically, pre-biological evolution must have started with reproducers that initially did not carry any replicators within them, but rather comprised self-sustaining proto-metabolic circuits confined within membrane vesicles (10, 11). Reconstruction of these primordial reaction networks is a separate, challenging task that has been attempted in many studies (12-14) and is beyond the scope of this work. Several plausible scenarios for the abiotic emergence of membranes have been proposed (15, 16) (17). Regardless of the details of the primordial chemistry, the key feature of these proto-metabolic systems would have been simultaneous production and/or accumulation of both nucleotides and amino acids; nucleobases and simpler amino acids, at least, are readily synthesized abiogenically, under various conditions (18-22). Nucleotides, evolutionary antecedents of modern coenzymes, would function as catalysts of some of the reactions in the proto-metabolic networks, whereas other reactions could have been catalyzed by amino acids, peptides and metal clusters. Already at this stage, ATP would serve as the universal convertible energy currency. The source of energy for the primordial reproducers is a major conundrum without an unequivocal solution. The primordial membranes are unlikely to have been ion-tight as required for maintaining gradients that are converted into chemical energy in modern cells (23), so the primordial reproducers most likely were heterotrophs that made ATP by substrate-level phosphorylation.

If relatively high concentrations of nucleotides and amino acids were reached within the primordial metabolizing vesicles, synthesis of both oligonucleotides and oligopeptides at non-negligible rates could have become possible. Certain oligonucleotides can be efficient catalysts of different reactions, that is, the first, simple ribozymes. A notable case in point are self-aminoacylating mini-ribozymes which can be as small as pentanucleotides that, strikingly, catalyze self-aminoacylation almost as efficiently as modern, protein aminoacyl-tRNA synthetases (24, 25). The catalytic capacity of ribozymes is sequence-dependent, and therefore, fixation and amplification of the sequences of catalytically efficient ribozymes could become the principal driver of the evolution of replicators. Templated synthesis and ligation of oligonucleotides catalyzed by ribozymes have been demonstrated as well (26-29) although an efficient, processive ribozyme polymerase remains an outstanding goal. Such a process could give rise to the first proto-replicators, initially, most likely, oligoribonucleotides. Even such a primitive replication process would be sufficient to kick off natural selection of proto-replicators hosted by reproducers whereby efficient catalysts would be selected along with the reproducers containing them. These ribozyme proto-replicators would become symbionts of the host reproducers. Such symbionts can be either mutualists that benefit the reproducer or parasites that exploit the reproducer. The mutualists would provide catalytic capacities that become sustainable and evolvable thanks to replication, whereas the host reproducer provides compartmentalization, resources and energy. At this stage, however, the mutualistic relationship between reproducers and replicators likely would be facultative rather than essential. The rate of RNA replication, including ribozyme-catalyzed one, is sequence-dependent, just like the rates of other ribozyme-catalyzed reactions, and hence, a different type of selection would emerge, selfish selection for the replication rate alone. Thus, parasitic replicators would inevitably evolve concomitantly with the mutualists (30, 31). These parasites would not enhance the reproduction of the host reproducers, on the contrary, decreasing their fitness through competition for limited resources, but could

be problematic, if not outright impossible, to purge, in the long term.

Here we analyze an agent-based mathematical model of the co-evolution of reproducers and replicators, in an attempt to illuminate salient aspects of the evolution of the replicator-reproducer mutualism and the origin of genomes.

Results

An agent-based model of primordial coevolution of reproducers and replicators

The premises of the model

Our conceptual scenario for the origin of life as a symbiosis between reproducers and replicators is outlined in Figure 1. Reproducers are a central ingredient in this scenario. The long-term persistence of proto-metabolic networks requires some form of reproduction of the compartments encasing the reacting molecules. Division of growing lipid vesicles that strikingly resembles the division of wall-less bacteria, such as L-forms, has been demonstrated (32, 33). This is a simple, purely physico-chemical mechanism, stemming from basic physical principles, whereby vesicles become unstable after reaching a critical size and divide, not requiring complex molecular machineries that are involved in cell division in modern cells (16, 17). Although, in evolutionary biology, selection is habitually linked to replication of digital information carriers (nucleic acids), primordial reproducers, arguably, would have been subject to a primitive form of selection. Evidently, the reproduction of the proto-metabolizing vesicles would be a far cry from the high precision process of modern cell division that is coupled to genome replication. Rather, it would be a stochastic assortment of components among the daughter vesicles (Figure 1). With this type of reproduction, random drift would necessarily play a major role in evolution, and the entire evolutionary process would resemble the stochastic corrector model that was originally proposed to describe the reproduction of primitive cells that, supposedly, contained multiple, unlinked genes (34, 35). We extend the stochastic corrector idea back to the prebiotic evolution stage that was, we surmise, the era of (pure) reproducers. Even in the evolution of such reproducers, notwithstanding the major role of drift, natural selection could set in, through the survival of the fittest vesicles, that is, the most temporally persistent ones, thanks to higher stability and/or faster growth (36).

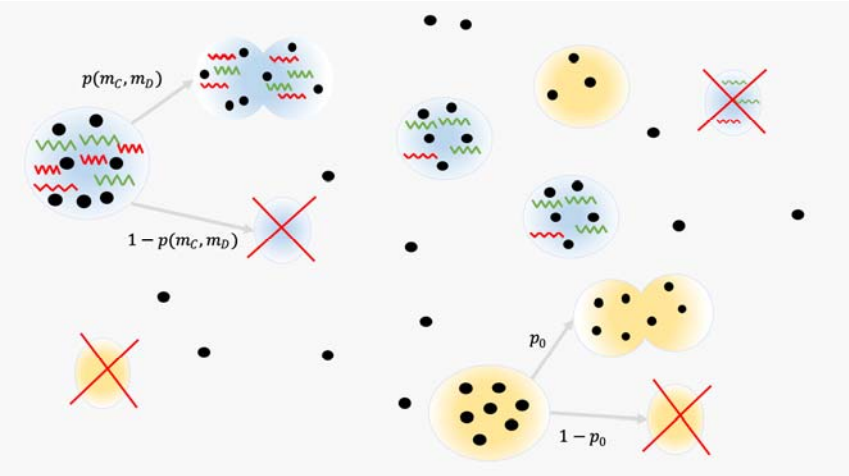


FIG. 1: **Pre-biological co-evolution of reproducers and replicators: a conceptual scenario and model framework.**
Protocells with (blue) and without (yellow) genetic elements (GE) compete for common resources (black circles). The GE in the protocells are either autonomous (red) or non-autonomous (green). Autonomous elements replicate themselves if the resources are present. Non-autonomous GEs replicate by interacting with autonomous elements. Both types of protocells can reproduce once their resources exceed some threshold value (in the depicted case, the arbitrarily set threshold is 7 units of resources for both protocell types). Successful reproduction results in two daughter cells. The probability of successful reproduction depends on the composition of GE in the reproducing protocell $p(m_{\mathcal{A}}, m_{\mathcal{N}})$, where $m_{\mathcal{A}}$ and $m_{\mathcal{N}}$ are the numbers of autonomous and non-autonomous GE, respectively (blue cell in upper left). The probability of successful reproduction is fixed for yellow cells (bottom yellow cell). If reproduction is successful, then daughter cells inherit the resources of the mother protocell, as well as GE, in the case of blue protocells. A reproducing cell dies in case of unsuccessful reproduction (upper left and bottom right), resulting in the dissipation of all resources and extinction of all GE. Protocells can also die due to the lack of resources (upper right blue protocell and bottom left yellow protocell).

In the symbiotic reproducer-replicator systems, competition and selection would occur at two levels: between replicators within a protocell, and between protocells carrying different complements of replicators. With respect to these two levels of selection, there would be 4 classes of replicators: 1) capable of autonomous replication and beneficial to the protocell (autonomous mutualists), 2) depending on other replicators for replication but beneficial to the protocell (non-autonomous mutualists), 3) capable of autonomous replication but useless to the protocell and incurring cost on the latter (autonomous parasites), 4) depending on other replicators for replication and useless to the protocell, thus incurring cost both on the mutualists and the protocell (non-autonomous parasites). The interactions between these distinct classes of replicators and between different replicators and protocells (reproducers) would shape the dynamics of pre-biological evolution.

A simple model of protocell population growth

We first consider evolutionary dynamics of cell-like reproducers (hereafter protocells) capable of resource metabolism and reproduction. There are no replicators at this stage. We assume that the protocell population is placed in a finite volume. A fixed amount of resources $R = \text{const}$ is constantly supplied to that entire volume. The model is discrete, such that the time units for the protocell-level dynamics correspond to the resource supply rounds. In the given round, each protocell can acquire one unit of resources at most, thereby decreasing the total amount of resources available in the environment in the given round. The remaining resources are removed from the volume at the end of each round. Therefore, each protocell in the population is described by a resource balance (stored resources) B^i, $i = 1,2,..N$, where N is the total number of cells in the population. We assume that proper functionality of the protocell demands a fixed housekeeping cost ΔE, which is subtracted from the resource balance of the protocell at the end of feeding phase of each round. A given protocell i dies if $B^i - \Delta E \leq 0$ and reproduces when its resource balance exceeds a fixed threshold value E_{tr}. Reproduction of the protocell is stochastic with a given probability p, that is, reproduction of the given cell ends up with two cells with the probability p and no cells (mother protocell dies) with the probability $1 - p$. The resource of the mother cell is divided between the daughter cells either randomly or symmetrically. In the case of random division, the resources of mother protocells allocate randomly, with

uniform distribution, between the daughter cells. The resources of mother cell are halved between the progenies for the case of symmetric division.

The total number of protocells can be approximated by considering the time variation of the total resource balance of the population of protocells (see SI Appendix A for derivation), and satisfies the following relation

$$N^*(p, \Delta E) \approx \frac{R(1+p)}{2\,\Delta E} \qquad (1)$$

The estimate of the total population size given by (1) fails for low reproduction probabilities $p \leq 1/2$. Indeed, in this case, each reproduction event yields less than one progeny on average, so that the total number of protocells declines over time, and thus, (1) fails. Also, (1) fails for $\Delta E \approx 1$, when the housekeeping cost is almost equal to the acquired resources in the given round. The estimate (1) well describes total number of the protocells in the equilibrium for the random division case. For the symmetric division of protocells, the average number of protocells in the equilibrium is slightly greater than (1), since the average (over the population) resource balance of the protocells is greater than $\frac{E_{tr}}{2}$ that has been used to obtain (1) (SI Appendix A).

According to (1), the threshold value of the resources necessary for reproduction E_{tr} does not affect the total population size of the protocells. However, the threshold value explicitly enters in the estimate in the presence of stochastic protocell death events that are not due to the lack of resources or reproduction failure (see SI Appendix A for derivation), that is, at each round a protocell might die with probability ν. In the presence of random death events, the estimate (1) takes the form $N^*(p, \Delta E) = \frac{R(1+p)}{2\,\Delta E + \nu\, E_{tr}}$. The threshold values play a crucial role in the competition between different populations of protocells (SI Appendix B for further discussion).

We assume that any increase of the probability of successful reproduction δp is accompanied by an increase in the housekeeping cost δE. The housekeeping cost is a proxy for (adaptive) protocell complexity, and we assume that more complex protocells have a higher probability of successful reproduction. Therefore, for a given population of protocells, the change $(\delta p, \delta E)$ will be evolutionarily neutral in terms of the total population size at equilibrium, that is $N^*(p_0 + \delta p, \Delta E_0 + \delta E) = N^*(p_0, \Delta E_0)$ defined by (1), if the following holds for $(\delta p, \delta E)$

$$\frac{\delta p}{\delta E} = \frac{1+p_0}{\Delta E_0} \qquad (2)$$

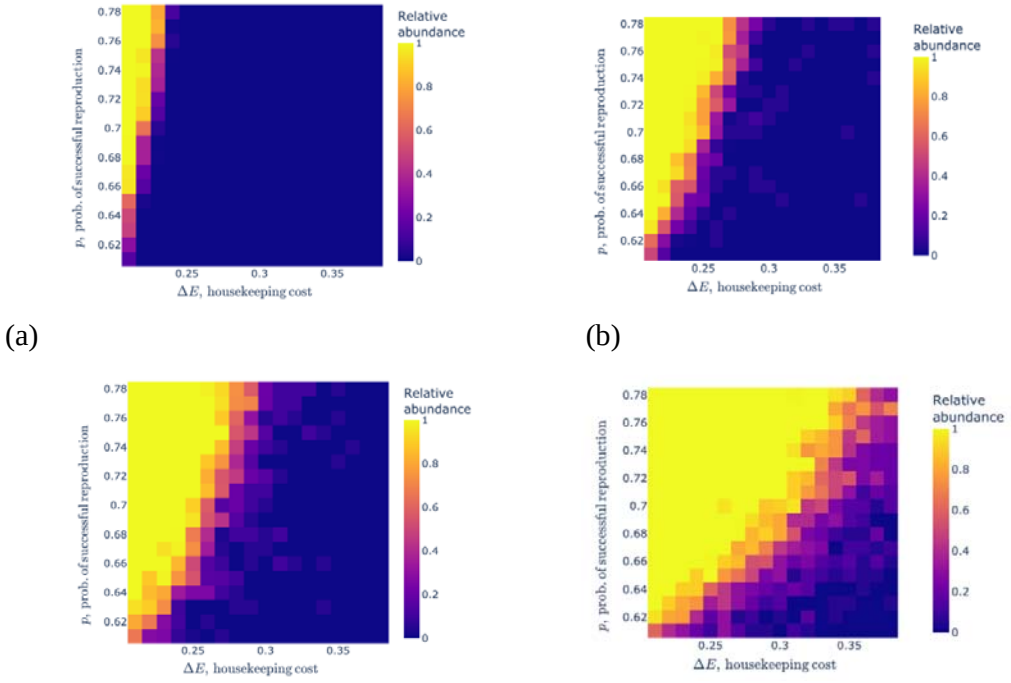
where $(p_0, \Delta E_0)$ are the base values of the probability of successful reproduction and housekeeping cost of the protocells, respectively. Assuming p is a function of ΔE, from (2) we obtain $p(\Delta E) = c \cdot \Delta E - 1$, where c is constant defined by the initial conditions.

In the absence of competition, the change of the successful reproduction probability and the housekeeping cost $(p_0 + \delta p, \Delta E_0 + \delta E)$ would be advantageous over the initial state $(p_0, \Delta E_0)$ if $\frac{\delta p}{\delta E} > \frac{1+p_0}{\Delta E_0}$, for a given population of protocells. The growth of a population is constrained only by the carrying capacity of the environment (the limited resource supply) and the housekeeping cost.

Suppose a protocell population with parameters $(p_0 + \delta p, \Delta E_0 + \delta E)$ invades an environment inhabited by a pre-existing population with parameters $(p_0, \Delta E_0)$. Eq. (1) was deduced for a single population only, and does not describe the competition between two competing populations. We simulated competition between two populations of protocells with different parameters and constructed the phase space plot showing the success or failure of the invaders (Fig. 2). The proxy for the invaders' success was their relative abundance $\frac{N_I(T)}{N_I(T)+N_b(T)}$ at round $T = 3000$, where $N_I(T)$ and $N_b(T)$ denote, respectively, the

number of invading protocells with parameters and the pre-existing protocells with parameters . We assume that the initial population sizes are equal . In the absence of stochastic death, in the case of symmetric protocell division, invaders with a higher probability of successful reproduction and greater cost win the competition only within a narrow range of parameters (Fig. 2a). There is a sharp asymmetry between the housekeeping cost and the probability of successful reproduction: the invaders need a sharp increase in the reproduction probability to win, for the given increase of housekeeping cost.

For random division of protocells, the required increase of the reproduction probability is smaller than in the case of symmetric division (Fig.2b). This threshold increase of the successful reproduction probability of protocells further decreases in the presence of stochastic death events for both symmetric and random division of the protocells (Fig.2c,d). Thus, the protocells with higher successful reproduction probability and higher metabolic cost are most competitive when both the protocell division and cell death occur stochastically. Stochastic death favors the increase in the reproduction probability of protocells because cells can die even if there is no lack of the resources. In this situation, the protocells to benefit from reproducing fast by increasing the reproduction probability at the expense of the concomitant increase of the housekeeping cost. Random division of protocells can result in asymmetric resource allocation, so that one of the daughter protocells is larger than the other. The smaller protocell will be vulnerable due to the deficit of resources. The large protocell will reproduce sooner than a cell produced by symmetric division. A failure of the larger protocell to reproduce will cause resource loss for the whole population greater than the resource loss would be for symmetric division. Thus, when both competing populations of protocells divide randomly, increase of the reproduction probability, even though accompanied by an increase of housekeeping cost, prevents the failure of larger protocells which may compensate the loss of resources associated with the appearance of smaller protocells that are vulnerable to the scarcity of resources.



(c) (d)

FIG. 2**: Competition of two populations of protocells with different growth parameters.** Relative abundance of protocells with parameters $(p', \Delta E')$ (invaders) vs the base population $(p_0, \Delta E_0) = (0.6, 0.2)$. The upper row panels show the outcome of the competition for symmetric division (a) and random division (b) of protocells for $\nu = 0$ (no stochastic death). The bottom panels show the outcomes in case of symmetric division (c) and random division (d) for non-zero death rate $\nu = 0.01$ The step sizes of the grid are $\delta p = \delta E = 0.01$. The color of each cell on the grid represents the relative abundance of the invaders averaged over $M = 20$ independent simulations. The number of steps in each axis is $L = 18$. Therefore, a cell in the heatmap with the relative abundance > 0.5 means that the invaders, parameterized by the coordinates of the heatmap cells, win the competition against the base population, characterized by the coordinates at the origin, more than half of the time. The remaining parameters are $R = 30, E_{tr} = 5$. It is seen that for the given increase of metabolic cost, the improvement of p necessary for invaders to outcompete the base population is higher for symmetric division of protocells than for random division of protocells. The necessary improvement of p of invaders visibly decreases even for a small death rate $\nu = 0.01$ for both symmetric and random division cases.

To summarize the model results for the evolution of pure reproducers, we show that accounting for the metabolic cost is essential for increasing the probability of successful reproduction of protocells. Protocells with relatively low reproduction probability (and accordingly low metabolic cost) are more likely to be outcompeted by those with a slightly higher reproduction probability (and higher metabolic cots) than protocells with an already high reproduction probability. We found that the increment of the reproduction probability required for an invading protocell population to win over the pre-existing population of protocells has to be larger for symmetric division than for random division of protocells, for the given increment of housekeeping cost. Thus, random division shifts the trade-off between the advantages of higher replication fidelity and the increasing metabolic cost towards favoring the former. Stochastic protocell death that are not caused by resource scarcity or failure of protocell reproduction further decreases the required increment of the successful reproduction probability. These conclusions are robust and stem from the fact that a population of protocells competing for a limited, common resource saturates at a regime where the metabolic cost is paid constantly to keep the protocells alive, whereas reproduction occurs slowly, in comparison, according to the reproduction probability. These observations on pure reproducers are directly relevant for the origin of the mutualistic symbiosis between protocells and GE because the invading population consist of GE-containing protocells as discussed in the next sections. Indeed, it is the trade-off between the reproduction fidelity and metabolic cost that enables advantageous feedback mechanisms between the protocells and GE, whereby GE improve the error-prone reproduction process of the primordial reproducers (protocells), while the protocells provide resources for the replication of GE, thus, increasing the housekeeping cost, as discussed in the subsequent sections.

Evolution of genetic elements in the absence of (proto)cells

Before addressing the coevolution of genetic elements with protocells, we consider the idealized (even if biologically unrealistic) case of evolution of GE in the absence of protocells (compartments). In the model, there are two types of GE, autonomous and non-autonomous replicators. The autonomous GE can replicate themselves as well as the non-autonomous GE, whereas the non-autonomous GE can replicate

only through interaction with the autonomous ones and do not contribute to the replication of the latter. The replication of both types of GE is assumed to be possible only in the presence of the necessary resources, and along with replication, GE can also die.

The GE are placed in a well-mixed volume and directly consume the resources which are supplied at a fixed rate. The following elementary processes describe resource supply and birth and death of GE:

$$\begin{aligned} &*\xrightarrow{r} S,\ S+\mathcal{A}\xrightarrow{r_{\mathcal{A}}} 2\mathcal{A},\ S+\mathcal{A}+\mathcal{N}\xrightarrow{r_{\mathcal{N}}}\mathcal{A}+2\mathcal{N} \\ &\mathcal{A}\xrightarrow{d_{\mathcal{A}}} 0,\qquad \mathcal{N}\xrightarrow{d_{\mathcal{N}}} 0 \end{aligned} \tag{3}$$

Here, S denotes the resources available in the environment, $\mathcal{A}$ and $\mathcal{N}$ stands for autonomous and non-autonomous replicators, respectively. From (3), we can obtain the deterministic description of the evolution of the number of GE and the amount of resources

$$\frac{d\,S}{d\,\tau} = 1 - r_{\mathcal{A}} m_{\mathcal{A}} S - r_{\mathcal{N}} m_{\mathcal{A}} m_{\mathcal{N}} S, \tag{4}$$

$$\frac{d\,m_{\mathcal{A}}}{d\,\tau} = r_{\mathcal{A}} m_{\mathcal{A}} S - d_{\mathcal{A}}\, m_{\mathcal{A}}, \tag{5}$$

$$\frac{d\,m_{\mathcal{N}}}{d\,\tau} = r_{\mathcal{N}} m_{\mathcal{A}}\, m_{\mathcal{N}} S - d_{\mathcal{N}} m_{\mathcal{N}}, \tag{6}$$

where $m_{\mathcal{A}}$ and $m_{\mathcal{N}}$ are the number of autonomous and non-autonomous GE, respectively. Note that the rates of elementary processes and time τ are scaled by r, that is $\tau \equiv r\,\tau, r_{\mathcal{A}} \equiv \frac{r_{\mathcal{A}}}{r}$ and so on.

The dynamical system (4-6) has two equilibria:

$$(S^*, m_{\mathcal{A}}^*, m_{\mathcal{N}}^*) = \left(\frac{d_{\mathcal{A}}}{r_{\mathcal{A}}}, \frac{1}{d_{\mathcal{A}}}, 0\right) \tag{7}$$

$$(S^*, m_{\mathcal{A}}^*, m_{\mathcal{N}}^*) = \left(\frac{d_{\mathcal{A}}}{r_{\mathcal{A}}}, \frac{r_{\mathcal{A}}\, d_{\mathcal{N}}}{r_{\mathcal{N}}\, d_{\mathcal{A}}}, \frac{1}{d_{\mathcal{N}}} - \frac{r_{\mathcal{A}}}{r_{\mathcal{N}}}\right) \tag{8}$$

The first of these (7) corresponds do the state free of non-autonomous replicator, whereas (8) describes a state where the two types of GE coexist, under the condition $\frac{1}{d_{\mathcal{N}}} - \frac{r_{\mathcal{A}}}{r_{\mathcal{N}}} > 0$ that is necessary for the final number of non-autonomous replicators to be positive. We assume that, in the absence of non-autonomous GE, the number of the autonomous GE in the equilibrium, given by (7), is $m_{\mathcal{A}}^* = \frac{1}{d_{\mathcal{A}}} > 1$.

The state defined by (7) is a stable equilibrium if $\frac{1}{d_{\mathcal{N}}} - \frac{r_{\mathcal{A}}}{r_{\mathcal{N}}} < 0$ (see SI Appendix C), that is, the opposite of the necessary condition for the coexistence of both types of GE (8). We further focus on the case (8) and $\frac{1}{d_{\mathcal{N}}} - \frac{r_{\mathcal{A}}}{r_{\mathcal{N}}} > 0$, that is, coexistence of autonomous and non-autonomous GE.

For the replication and death rates, we impose the conditions: $r_{\mathcal{A}} < r_{\mathcal{N}}$ and $d_{\mathcal{A}} < d_{\mathcal{N}}$, that is, the replication rate of the non-autonomous GE is greater than that of the autonomous GE, but the autonomous GE are more stable (lower death rate). Depending on the death rates, under these constraints, autonomous GE still can outcompete the non-autonomous GE, despite the faster replication of the latter, such that the equilibrium population size of the autonomous GE can be greater than that of the non-autonomous ones, $m_{\mathcal{A}}^* > m_{\mathcal{N}}^*$. However, below, we consider the region of the parameter space corresponding to the takeover by the non-autonomous GE $m_{\mathcal{A}}^* \ll m_{\mathcal{N}}^*$. That is, we require that

$$m_{\mathcal{A}}^{*} \ll m_{\mathcal{N}}^{*} \Rightarrow \frac{r_{\mathcal{N}}}{r_{\mathcal{A}}} \gg d_{\mathcal{N}}(1 + d_{\mathcal{N}}/d_{\mathcal{A}}) \quad (9)$$

Finally, we also require that the coexistence state (8) is a stable equilibrium of (4-6). The necessary and sufficient conditions for that are given by the following conditions (see SI Appendix C for derivation)

$$r_{\mathcal{N}} r_{\mathcal{A}}(d_{\mathcal{A}} + d_{\mathcal{N}}) - d_{\mathcal{N}}(d_{\mathcal{A}}^2(r_{\mathcal{N}} - r_{\mathcal{A}} d_{\mathcal{N}}) + r_{\mathcal{A}}^2 d_{\mathcal{N}}) > 0, \qquad \frac{1}{d_{\mathcal{N}}} - \frac{r_{\mathcal{A}}}{r_{\mathcal{N}}} > 0 \quad (10)$$

Equations (9,10) define the domain of the parameter space corresponding to the takeover by the non-autonomous GE.

Competition between protocells containing and lacking genetic elements

In the previous sections, we separately described the competition among protocells lacking GE (pure reproducers) and the competition among GE in the absence of protocells, where resources are supplied directly. We now explore the case most relevant for the origin of life, when the interactions among GE, described by (3), occur inside protocells and there is feedback between protocell reproduction and GE replication. In this case, the GE use the resources of the host protocell, and the replication of any GE is associated with an additional cost $E_c \ll 1$ (much smaller than the acquired resources per round) which is subtracted from the protocell's resource balance. Therefore, intracellular replication of GE increases the housekeeping cost for the host protocells. Obviously, under these conditions, protocells that harbor GE will lose the competition against those that lack them unless at least some of the GE are beneficial to the protocells. We therefore assume that the presence of mutualist GE in a protocell increases the probability of successful reproduction whereas parasitic GE only incur cost. The interplay between the opposing effects of GE on the reproduction of protocells defines the evolutionary outcome for the entire system. In this section, we analyze the interaction and evolution of only two of the four classes of GE defined above, Class 1 (autonomous mutualists) and Class 4 (non-autonomous parasites) GE. The case of non-autonomous mutualists and autonomous parasites is addressed in the next section.

The initial number of GE in protocells is given by the Poisson distribution with parameter μ

$$m_0^i = m_{\mathcal{A}0}^i + m_{\mathcal{N}0}^i = Poisson(\mu), i = 1, \ldots N_{g0} \quad (11)$$

where N_{g0} is the initial number of protocells containing GE (at the start of the first round of resource supply). The initial number of mutualists in each protocell is defined by a randomly chosen integer from $[0, m^i(0)]$. We assume that there is a time-scale difference between the protocell reproduction and competition, on the one hand, and the intracellular birth-death process (replication) of the GE, on the other hand. The protocell-level competition is governed by the rounds of resource supply to the environment. Let us denote by K the number of GE replication and death events (3) (excluding resource update) that occur in a protocell in each round of resource supply. In the considered region of the parameter space, given by (9) and (10), most of these events correspond to replication (when resources are sufficient in the protocells) due to the greater rate of replication compared to death. Also, K reflects the amount of resources available to the GE from the cell resource balance B^i. In a given round, the GE can use KE_c resources of the protocell at most. Let us consider the extreme case $K \to \infty$ for a given protocell, that is, the intracellular replication of the GE goes to steady state in each round. The resources

of the protocell will be exhausted in the steady state (this situation can be described by (4-6) except for the resource supply term in (4) because, for the given round t, the intracellular dynamics can be approximated by the deterministic description for $K \to \infty$). The change of the ratio of parasites to mutualists in a protocell is given by

$$\frac{d}{d\tau}\frac{m_{\mathcal{N}}^i}{m_{\mathcal{A}}^i} = \frac{m_{\mathcal{N}}^i}{m_{\mathcal{A}}^i}\left((r_{\mathcal{N}}m_{\mathcal{A}} - r_{\mathcal{A}})B^i - (d_{\mathcal{N}} - d_{\mathcal{A}})\right) \quad (12)$$

that is, in the presence of resources $B^i > 0$, the parasites grow faster than the mutualists. This process ends when $B^i = 0$ (resources are exhausted), after which the ratio of parasites to mutualists starts decreasing because $d_{\mathcal{N}} > d_{\mathcal{A}}$, followed by the eventual collapse of the entire system, both the protocells and the GE.

Both the GE and the resources are stochastically divided between the daughter protocells at reproduction, that is, the number of mutualists and parasites in one of the daughter protocells is randomly (with uniform distribution per type of element) pooled from the dividing mother cell. The resources are distributed proportionally to the total numbers of GE in a daughter protocell

$$m_{\mathcal{A}1} = Random\left[0, m_{\mathcal{A}}^i\right],\ m_{\mathcal{A}2} = m_{\mathcal{A}}^i - m_{\mathcal{A}1}, \quad (13)$$

$$m_{\mathcal{N}1} = Random\left[0, m_{\mathcal{N}}^i\right],\ m_{\mathcal{N}2} = m_{\mathcal{N}}^i - m_{\mathcal{N}1},$$

$$B_1 = \frac{m_{\mathcal{A}1}+m_{\mathcal{N}1}}{m_{\mathcal{A}}^i+m_{\mathcal{N}}^i}B^i, \qquad B_2 = B^i - B_1$$

The daughter protocells are denoted by 1 and 2, and m^i is the total number of GE in the mother protocell at the time of division. The division mechanism (13) is the most favorable among several other mechanisms of binary division, with respect to the appearance of mutualists-only protocells (37). Conversely, symmetric division (17) is the most unfavorable mechanism in this respect (37).

The time interval between two successive divisions of the protocells is not constant because of the constant threshold value E_{tr} and the increasing total number of protocells in the population. Therefore, E_{tr} has a critical impact on the survival and potential takeover of the mutualists.

The probability of successful reproduction of a GE-containing protocell linearly depends on the fraction of mutualists it contains

$$p\left(m_{\mathcal{A}}^i, m_{\mathcal{N}}^i\right) = \begin{cases} min\left[p_0\left(1 + \frac{m_{\mathcal{A}}^i}{m_{\mathcal{A}}^i+m_{\mathcal{N}}^i}\right), 1\right], & if\ m_{\mathcal{A}}^i + m_{\mathcal{N}}^i \neq 0 \\ p_0, & otherwise \end{cases} \quad (14)$$

Here p_0 is the probability of successful reproduction of the protocell in the absence of GE. The probability of successful reproduction is the same for the protocells containing only parasites and the GE-less protocells. In the absence of mutualists, parasites do not replicate on their own and, accordingly, consume no resources.

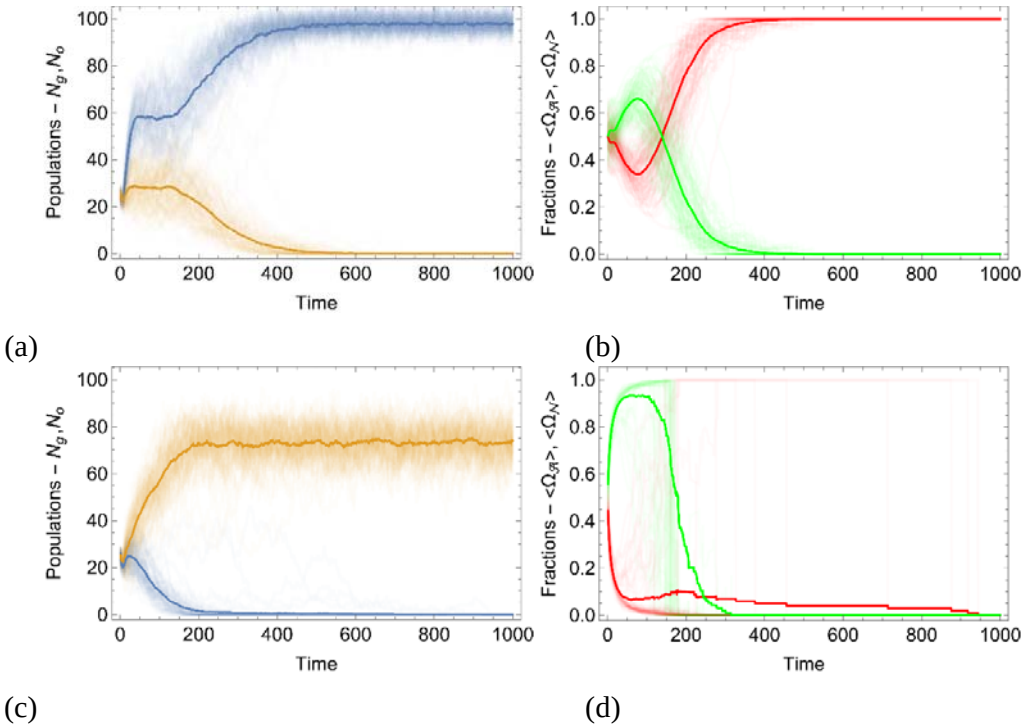


FIG. 3: **Competition of protocells containing genetic elements against protocells lacking genetic elements**.

Time dependency of the total number of the protocells with (blue lines) and without (yellow lines) GE, denoted by N_g and N_o respectively, for (a) $K = 1$ and (c) $K = 15$.
The thick lines show the size of the populations averaged over an ensemble of 100 realizations.
Time dependency of the fraction of autonomous mutualists (red lines) and non-autonomous parasites (green lines) for (b) $K = 1$ and (d) $K = 15$.
The thick lines show the behavior of fractions of genetic elements (15) averaged over the ensemble. The remaining parameters are as follows:
$r_{\mathcal{A}} = 0.04, r_{\mathcal{N}} = 0.05, d_{\mathcal{A}} = 0.005, d_{\mathcal{N}} = 0.01, R = 30, \Delta E = 0.3, E_c = 0.01, p_0 = 0.6, \mu = 100, E_{tr} = 5.$

Consider competition of GE-containing protocells against GE-less protocells (Figure 1). The protocells without GE divide randomly, that is, the resources are distributed randomly between the daughter protocells. If a GE-containing protocell loses those elements, it joins the population of GE-less protocells. The protocells that only contain parasites eventually lose them due to the deaths of GE and impossibility of the replication of parasites in the absence of mutualists. The basal housekeeping costs and threshold values are the same for both type of cells, $\Delta E_o = \Delta E_g$ and $E_{tro} = E_{trg}$.

Random allocation of GE among the progeny and death of GE eventually will result in the appearance of some protocells that carry only mutualists. The appearance of these mutualist-only protocells depends on the reproduction threshold value E_{tr} and the average number of intracellular elementary processes K in the given round of resource supply. Both quantities control the balance between mutualists and parasites. Larger values of each of these quantities lead to an increase of the fraction of parasites in the protocells, as follows from (12). Also, increase of K causes a concomitant increase of the housekeeping cost of the protocell by approximately KE_c.

The fractions of autonomous mutualists and non-autonomous parasites in the protocell population at the round t are

$$\Omega_{\mathcal{A}}(t) = \frac{\sum_i^{Ng(t)} m_{\mathcal{A}}^i}{\sum_i^{Ng(t)} m_{\mathcal{A}}^i + m_{\mathcal{N}}^i}, \qquad \Omega_{\mathcal{N}}(t) = \frac{\sum_i^{Ng(t)} m_{\mathcal{N}}^i}{\sum_i^{Ng(t)} m_{\mathcal{A}}^i + m_{\mathcal{N}}^i}. \tag{15}$$

All other parameters being equal, the outcome of the competition critically depends on K. For $K = 1$, that is, when GE replication is coupled with the protocell reproduction, GE-containing protocells win the competition (Fig. 3a), and following a brief initial surge of parasites, mutualist-only protocells take over (Fig. 3b). In contrast, for $K = 15$, that is, at a high replication rate that substantially exceeds the cell reproduction rate, GE-less protocells outcompete GE-containing ones (Fig. 3c), and after the GE pool becomes dominated by parasites, all GE die off (Fig. 3d).

The probability of successful reproduction (14) is greater in the initial phase of the dynamics due to the random distribution of the mutualists among the protocells. On average, the initial fractions of mutualists and parasites are almost equal in the population (Figs. 3c,d). The population of GE-containing protocells initially grows faster than the population of GE-less protocells, regardless of the eventual outcome (Figs. 3a,c) and similarly, the fraction of parasites initially grows in all cases (Figs. 3b,d). However, at low *K* values, the mutualist-only protocells emerge stochastically, due to the random protocell division, and eventually win the competition due to the selective advantage conferred by the mutualists (14). In contrast, at larger K values, parasites take over in the initial phase of the competition, effectively precluding the appearance of mutualist-only cells. As a result, GE-less protocells win the competition (Fig. 3c) whereas all GE die off (Fig. 3d). At a higher reaction rate, $K = 15$, the takeover by the parasites is much sharper than for $K = 1$ (Figs. 3c,d), decreasing the probability of the appearance of mutualist-only protocells. Even in this case, due to random division, a few mutualist-only protocells emerge, but they lose the competition due to the exhaustion of resources by the increasing number of GE-less cells.

Figure 4 shows the results of the competition between GE-containing and GE-less protocells depending on the values of E_{tr} and *K*. Here, we took snapshots of many independent simulations ($M = 10^3$) at time T (sufficiently large). For each simulation, the fraction of GE-containing protocells $\Gamma(T) = \frac{N_g(T)}{N_g(T)+N_o(T)}$ was calculated. The fraction of the protocells with genetic elements is then averaged over the ensemble of *M* independent simulations, that is, independent realizations of the population dynamics

$$< \Gamma(T) >= \frac{1}{M}\sum_k^M \frac{N_g^k(T)}{N_g^k(T)+N_o^k(T)} \tag{16}$$

The average fraction of GE-containing protocells, defined by (16), and the unbiased sample standard deviation $\sqrt{\frac{1}{M-1}\sum_k^M(\Gamma_k - < \Gamma >)^2}$ are shown in Fig.4 for different values of E_{tr} and K.

The results shown in Figure 4 demonstrate the crucial role of the time-scale difference between the protocell reproduction and GE replication for the outcome of the competition. Even for large reproduction threshold values, the GE-containing protocells can win the competition when reproduction and replication are on the same time-scale. The greater the disparity between the replication and reproduction rates, the less likely it is that GE-containing protocells take over. For the smaller values of the parameters, $E_{tr} = 2$ and $K = 1$, the behavior of independent simulations is nearly uniform, resulting in a small standard deviation. Increasing the values of these parameters ($E_{tr} = 5, K = 5$) results in an

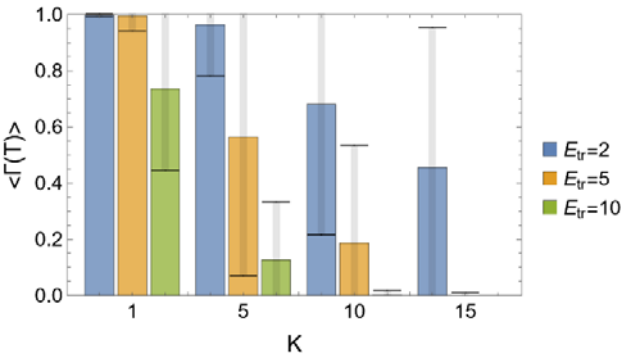


FIG. 4: **Competition between protocells containing and lacking genetic elements depending on K and E_{tr}.**

The bars show the average fraction of GE-containing protocells in the population at time T, for different values of E_{tr} and K. Averaging is carried out over the $M = 10^3$ independent simulations. The intervals at each bar show the standard deviation of the obtained values. The snapshots are taken at $T = 10^3$. $K = 1, 5, 10, 15$ for the given threshold value E_{tr}. All other parameters are the same as in Fig.3.

increased stochasticity of the competition outcome, and accordingly, a much greater standard deviation. Further increase of the parameter values again yields a uniform outcome, that is, elimination of the GE-containing cells (for $E_{tr} = 10$ and $K = 15$ there is no GE-containing cells at time T) and thus a small standard deviation.

Random division favors GE-containing protocells whereas symmetrical division (the resources and both types of genetic elements being allocated equally between the daughter protocells) is advantageous for GE-less protocells (see SI Appendix B). Indeed, with symmetrical division, GE-containing cells can fail to outcompete the GE-less cells even when replication and reproduction dynamics are on the same time-scale.

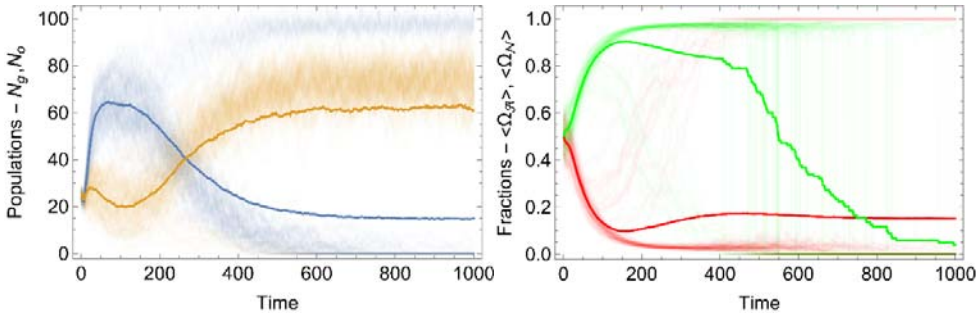
For symmetrical division, (13) is modified as follows

$$m_{\mathcal{A}1} = \left[\frac{m^i_{\mathcal{A}}}{2}\right], \qquad m_{\mathcal{A}2} = m^i_{\mathcal{A}} - m_{\mathcal{A}1}, \tag{17}$$

$$m_{\mathcal{N}1} = \left[\frac{m^i_{\mathcal{N}}}{2}\right], \qquad m_{\mathcal{N}2} = m^i_{\mathcal{N}} - m_{\mathcal{N}1},$$

$$B_1 = \frac{m_{\mathcal{A}1}+m_{\mathcal{N}1}}{m^i_{\mathcal{A}}+m^i_{\mathcal{N}}} B^i, \qquad B_2 = B^i - B_1.$$

where $\left[\frac{m^i_{\mathcal{A}}}{2}\right]$ is the integer part of the ratio $\frac{m^i_{\mathcal{A}}}{2}$ .With symmetrical division of GE-containing protocells, GE-less protocells almost always outcompete GE-containing ones (Fig. 5a) after parasites take over, in most realizations, and then, all GE die off in the vast majority of the cases (Fig. 5b).



(a) (b)

FIG.5: **Competition of protocells containing genetic elements against protocells that lack genetic elements for the case of symmetric division (17)**. Here $K = 1$ and $E_{tr} = 5$. All other parameters are the same as in Fig.3.

Nevertheless, in this case, GE-containing protocells do not always go extinct and mutualist-only protocells can persist when the replication and reproduction rates are coupled ($K = 1$) (Figs. 5a,b). Takeover by the mutualists in symmetrically dividing protocells is highly unlikely because the formation of mutualist-only protocells is far less likely than it is in the random division case where such protocells emerge much more often, and accordingly, the GE-containing protocells typically win the competition (compare Figs. 5a,b with Figs. 3a,b). Indeed, if parasites are present in the mother protocell at division time, both daughter protocells inherit half of those so that parasites proliferate again in both daughter protocells, precluding or at least delaying the appearance of mutualist-only protocells. For random division of protocells, the time at which mutualist-only cells appear is, on average, much shorter than it is in the case of symmetrical division. Let us denote the time (measured by resource update rounds) at which the first mutualist-only protocell appears in the population by Θ, the waiting time, which is a discrete random variable. A mutualist-only protocell can appear either due to the initial distribution of GE in the protocells (11), or as a result of random protocell division, or because of the death of all the parasites in the given protocell. If a mutualist-only protocell appears in the population due to the initialization (11), then $\Theta = 1$, that is, this protocell appears in the first round of resource supply. A mutualist-only protocell might not emerge at all during the simulation time T_s (all GE can even die out during this time, as exemplified in Fig.5a), in which case we assign $\Theta = T_s$. For convenience, we analyze the logarithm of the time of the first appearance of a mutualist-only protocell $Log\Theta$, with the ensemble average

$$< Log\Theta > = \frac{1}{M}\sum_{l=1}^{M} Log\Theta_l \qquad (18)$$

Here $M = 10^3$ is the total number of simulations, and Θ_l is the time at which the first mutualist-only cell appears in lth simulation. Under the worst-case outcome (for GE), when no mutualist-only protocells emerge in any of the M simulations, $< Log\Theta >= LogT_s$. Conversely, in the best-case outcome, a mutualist-only cell appears during the initialization of each simulation, $< Log\Theta >= 0$.

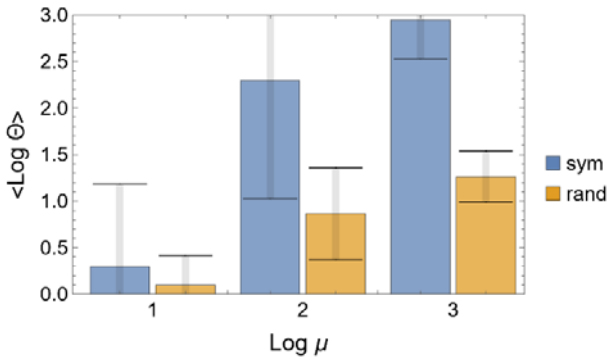


FIG. 6: **Kinetics of the appearance of mutualist-only protocells**.

The chart shows the dependency of the logarithm (base 10) of waiting time of the appearance of mutualist-only protocells on the value of the parameter μ which describes the initial number of GE in the protocells (11) for random (13) or symmetrical division (17). The simulation time is $T_s = 10^3$ and the number of independent simulations is $M = 10^3$. The intervals show the standard deviation of the samples of size M obtained for each pair of E_{tr} and K. All other parameters are the same as in Fig.3.

Figure 6 shows the dependency of the average waiting time (18) and the standard deviation of the samples on the Poisson parameter μ, the initial number of GE in each protocell (11), for the cases of random (14) and symmetrical (17) division of the protocells. Here $T_s = 10^3$, that, is the worst outcome is $< Log\Theta > = 3$. As expected, for the smaller initial number of GE (μ) in each protocell, the likelihood of the appearance of a mutualist-only protocell is higher than for the larger μ, that is, the time of the first appearance of mutualist-only cells is shorter than it is for larger μ. For all considered values of μ, the average waiting time of the appearance of mutualist-only protocells is substantially smaller in the case of random division than it is in the case of symmetrically dividing protocells. Indeed, for $\mu = 10^3$, in the case of symmetrical division, the time before the appearance of mutualist-only protocells is close to the simulation time T_s. This is the case because, for large μ, appearance of mutualist-only protocells due to the initialization (11) is combinatorially highly unlikely, in contrast to the smaller values of μ. The same holds for the case of random division. However, under random division, mutualist-only protocells appear in the population even in the case of large μ. Thus, random division of protocells substantially decreases the average waiting time $< Log\ \Theta >$ for all values of μ compared with the symmetrical division.

Non-autonomous mutualists and autonomous parasites

We further examined a distinct version of the model in which protocells contain non-autonomous mutualists (Class 2 replicators) and autonomous parasites (Class 4). In this case, the probability of successful reproduction of the protocells (14) increases with the fraction of (non-autonomous) mutualists (swapping $m^i_{\mathcal{A}} \to m^i_{\mathcal{N}}$ and $m^i_{\mathcal{N}} \to m^i_{\mathcal{A}}$ in (14)). Protocell-level selection, then, favors the appearance of mutualist-only protocells, but in this case, because the mutualists are incapable of replication, such protocells will lose GE because these mutualist cannot replicate in the absence of autonomous GE (4). The appearance of protocells containing only autonomous parasites is unfavorable as well because these GE provide no advantage to the protocell and, on the contrary, being able to replicate, incur a cost by consuming resources. In this case, random division is unfavorable because it increases the likelihood of

the appearance of mutualist-only and parasite-only protocells both of which are evolutionarily unstable. In contrast, symmetrical division reproduces the initial composition of GE in the protocells, and therefore, if both types of replicators were present initially, they will be present in the progeny protocells as well.

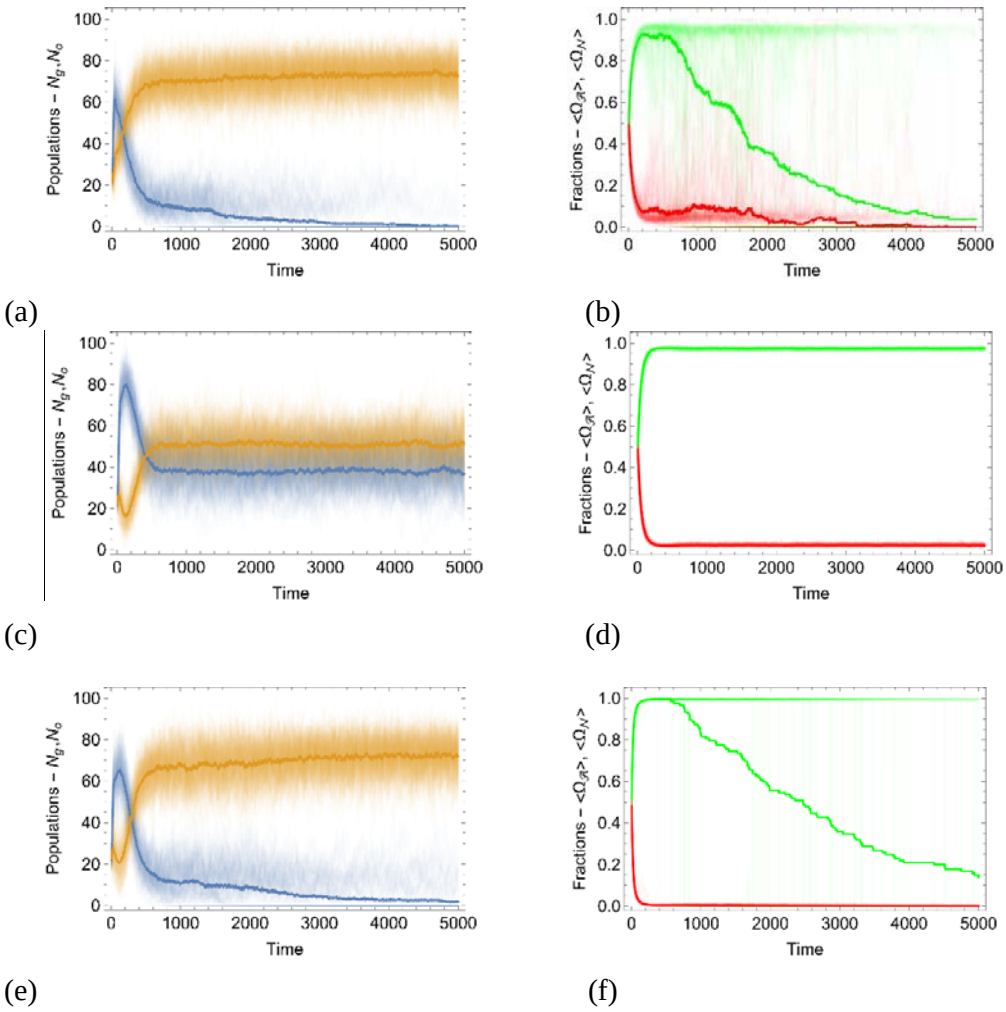


FIG. 7: **Competition between protocells with genetic elements and protocells without genetic elements for the case of non-autonomous mutualists and autonomous parasites**

Blue lines denote GE-containing protocells, and yellow lines denote GE-less protocells. Green lines show non-autonomous mutualists and red lines show autonomous parasites. The thick lines are the averages of independent simulations.

(a) Dynamics of the total number of protocells with and without GE for random protocell division for $K = 1$.

(b) Dynamics of the fractions of mutualists and parasites (15) averaged over the ensemble for random protocell division $K = 1$.
(c) Dynamics of the total number of protocells with and without GE for symmetrical protocell division for $K = 1$.
(d) Dynamics of the fractions of mutualists and parasites for symmetrical protocell division for $K = 1$.
(e) and (f) show the same as (c) and (d), respectively, for $K = 5$.

Here $E_{tr} = 5$. All other parameters are the same as in Fig.3.

Figure 7 illustrates the competition between GE-containing and GE-less protocells for the case of non-autonomous mutualists and autonomous parasites, under random and symmetrical division. In this case, to enjoy a sustainable competitive advantage, the GE-containing protocells have to carry both types of GE. The probability of successful reproduction (14) of such protocells will be lower than that of protocells containing only non-autonomous mutualists. However, the latter variety of protocells either lose the GE or die out due to the impossibility of replication in the absence of autonomous GE. Initially, the GE-containing protocells take over the competition (Figs. 7 a,c,e), however, only symmetric division and coupled replication of GE and reproduction of protocells ($K = 1$) allows for the long-term persistence of GE-containing protocells in the environment (Fig. 7c) because symmetrical division suppress the appearance of autonomous only or non-autonomous only protocells. Therefore, the survival of GE-containing protocells is strongly sensitive to the variation in K. Indeed, for $K = 5$. the GE-containing protocells lose the competition and die out (Fig. 7e.f). In the case of random division, GE-containing protocells either die or lose the GE due to the appearance of protocells containing only non-autonomous mutualists or only autonomous parasites (Figs. 7 b.).

The outcome of the competition will change if we assume that, in the absence of protocells, autonomous parasites and non-autonomous mutualists coexist, but the former are more abundant than the latter $m^*_{\mathcal{A}} \gg m^*_{\mathcal{N}}$, that is, the opposite of (9) holds, whereas (10) still applies. In this case, autonomous parasites will have a chance to survive within each protocell even in a random division case, providing the machinery for the replication of the mutualists. This scenario is illustrated in Fig.8 for symmetrical and random division. Both types of GE are now present in the population, ensuring the advantage of the GE-containing protocells and the availability of the replication machinery for both types of GE. In the case of symmetrical division, the GE-containing protocells outcompete the GE-lacking ones, whereas in the case of random division, GE-containing and GE-less protocells coexist. In the case of symmetrical division, the average fraction of mutualists is greater than it is in the case of random division. Thus, given the selective advantage provided by the non-autonomous mutualists, the probability of successful reproduction of symmetrically dividing protocells (14) is, on average, higher than that of randomly dividing protocells. Also, symmetrical division decreases the likelihood of the appearance of the parasite-only and mutualist-only protocells, which are in this case nonviable, compared to random division. Thus, for the same parameter region, symmetrical division is advantageous over random division of protocells for the case of autonomous parasites and non-autonomous mutualists.

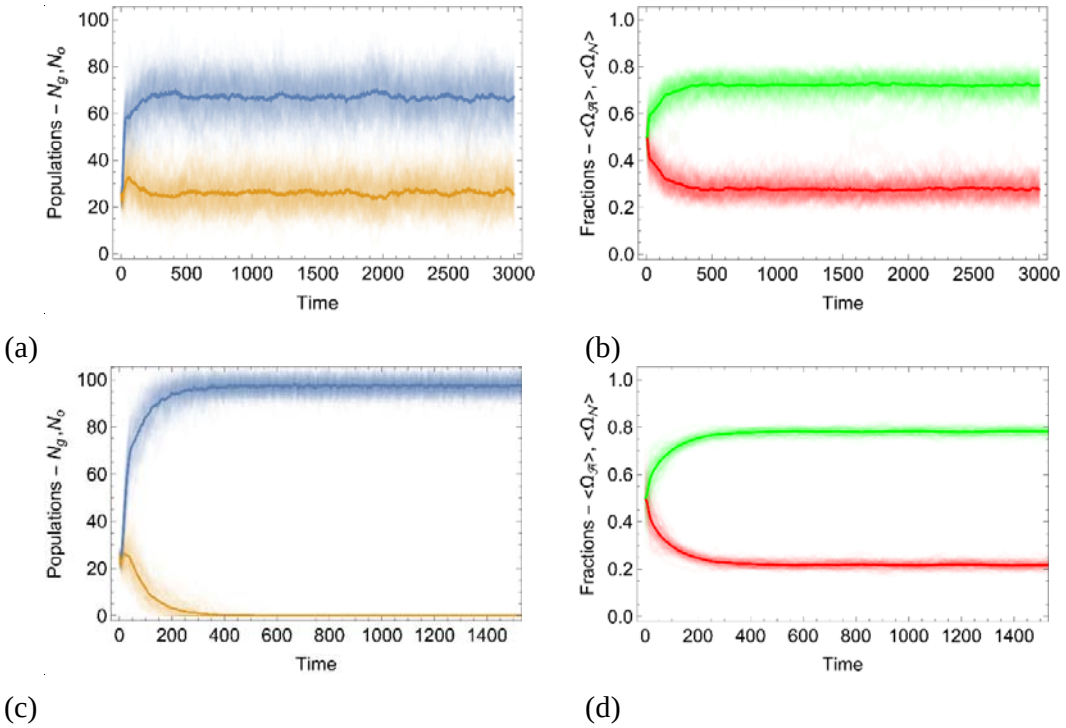


FIG. 8**: Competition between protocells with and without genetic elements for the cases of random division (a, b) and symmetrical division (c, d).** Here we $E_{tr} = 5$ and $K = 1$. The designations are as in Fig. 7. The model parameters are chosen such that the opposite of (9) is true. The rates of the intracellular GE dynamics are as follows $r_{\mathcal{A}} = 0.5, r_{\mathcal{N}} = 0.05, d_{\mathcal{A}} = 0.005, d_{\mathcal{N}} = 0.1$. All other parameters are as in Fig. 3.

Synopsis of the model results

Let us briefly summarize the results of our modeling of the coevolution of reproducers and replicators. We focus on the cost-benefit analysis of feedback mechanisms between protocells and GE. First, we developed a model that describes the population dynamics of protocells capable of resource metabolism and reproduction, in the absence of any GE. We considered the role of various parameters (threshold of the resources necessary for reproduction, cost of maintenance, probability of successful reproduction, symmetrical vs. asymmetrical reproduction of protocells) in the protocell-level dynamics. Then, we addressed the case of evolution of two types of GE, autonomous and non-autonomous, in the absence of protocells. In the model, the replication rate of non-autonomous genetic elements is greater than that of autonomous elements, but they are less stable than autonomous GE. We focus, in particular, on the part of the parameter space where non-autonomous GEs outcompete autonomous GE. Then, feedback is introduced between protocells and GE. We assume that GE interact among themselves within a protocell (including both competition and autonomous GE providing the replication machinery to non-autonomous ones) and also with the protocell. The GE use the resources of the host protocell for their replication, thus incurring an additional cost on the protocell. However, the presence of mutualists in the protocell increases the probability of successful reproduction. The centerpiece of this work is the competition

between GE-containing and GE-less protocells. We observed that selection favors GE-containing protocells when the protocell reproduction and GE replication occur on the same time scale, best of all, are precisely coupled. Random division is advantageous compared to symmetrical division for GE-containing protocells because randomness provides for the appearance of mutualists-only protocells that take over the population. We further considered the case of non-autonomous mutualists and autonomous parasites where replication-competent, autonomous GE carry no benefit for the protocell whereas the non-autonomous ones do. In this case, for the GE-containing protocells to enjoy a sustainable advantage, the autonomous parasites have to replicate faster than the non-autonomous mutualists. In this situation, symmetrical division is favored over random division because it ensures the perpetuity of the GE composition in protocells that have to contain both mutualists and parasites, that is, to maintain both the beneficial GE and the means for their replication.

Discussion

Origin of life, or more precisely, the origin of cells, the universal evolving units of life, remains a fundamental enigma (38). The information transmission pathways of modern cells that underpin evolution are exquisitely complex and themselves must have evolved under selection (39, 40). Motivated by these considerations and by the fundamental distinction between the two types of evolving biological entities, reproducers and replicators (4, 35), we modeled here the origin of cells as a mutualistic symbiosis between protocellular reproducers and primordial replicators. Protocells are commonly perceived as entities that already contained replicating RNA molecules within lipid membrane-bounded vesicles (41). We argue, however, that a stage preceding the origin of cells and involving selection for persistence (36) among pure reproducers devoid of any GE is inescapable in prebiological evolution. In our scenario, these protocellular reproducers become incubators for primordial replicators (GE), initially, likely, ribozymes that were selected for their catalytic activities (25, 42) that increased the fitness of the protocells carrying such replicators enhancing protocell reproduction, and only subsequently, assumed coding functions.

Replicators, however, present an inherent, major problem in that their evolution inevitably gives rise to parasitic elements that hijack the replications machinery of autonomous elements (30, 31). In homogeneous, well mixed systems, the parasites that replicate faster than autonomous elements tend to take over, leading to the collapse of the entire replicator ensemble (43). Compartmentalization can substantially change the evolutionary dynamics of autonomous elements and parasites, preventing the takeover by parasites and stabilizing the system (43-47). Furthermore, previous modeling studies strongly suggest that replicators are more likely to survive within protocells compared to surface-based spatial systems (13).

Here, we explored a mathematical model of evolution of replicators (GE) within protocellular reproducers seeking to define the conditions that favor the selection of protocells carrying GE and eventual emergence of genomes. Under the assumption that mutualist GE conferred selective benefits onto the reproducers, within which such GE replicated, we identified two key conditions for the fixation of the genetic system in evolution. First, the replication rate of the GE has to be coupled to the reproduction of the protocells. Counterintuitive as that might seem, GE replication at rates substantially higher than the reproduction rate of the protocells leads to the extinction of the GE. Informally, this requirement stems from the need to avoid exhaustion of the resources available for protocell reproduction

by uncontrolled replication of the GE. Second, the distribution of the GE between daughter cells has to be stochastic (random) rather than symmetrical, to enable the emergence of mutualist-only protocells. A similar conclusion has been previously reached for a model of group selection of replicators within cells (48).

We further investigated a version of the model, in which only parasites but not mutualists were endowed with the replication capacity. In this case, obviously, a reproducer-GE system could persist and be competitive only if it contained both parasites and mutualists. We found that to survive, such a system had to meet demanding criteria, namely, parasites replicating faster than mutualists and thus incurring substantial costs on the respective protocells, and furthermore, protocell division has to be symmetrical, to ensure the presence of both types of GE in the daughter protocells.

The simple, intuitive constraints on the early steps in the evolution of GE identified here have substantial implications for the origin of genomes and modern-type cells. The first of these pertains to the origin of large genomes, on the scale of the genomes of the extant bacteria and archaea. There is little doubt that the first GE were small, on the order of a kilobase, at most. Obviously, GE-containing protocells could stand a chance in the competition with GE-less ones only when all or at least a large fraction of the GE encoded their own replication machinery. The situation of autonomous parasites sustaining the replication of non-autonomous mutualists imposes a strict requirement of a high replication rate of the parasites, which contradicts the requirement for coupling the GE replication and protocell reproduction rates, and furthermore, incurs high cost on the protocells. Therefore, such composition of the GE pool in protocells appear highly unlikely. Thus, at the early stages of the genetic system evolution, the key role apparently belonged to GE resembling modern RNA viruses, especially, those that do not encode any structural proteins, but only the enzyme required for replication, such as narnaviruses or mitoviruses (49, 50). Protocells that harbored ensembles of mutualists would encompass multiple versions of the replication machinery. This excess of sequences dedicated to replication would engender selective pressure for joining genetic elements and eliminating the redundancy, saving resources and facilitation coordination of replication with protocell division. The second major corollary is the origin of dedicated defense systems against parasites. Defense systems are extremely abundant and diverse in modern prokaryotes where they account for a considerable, probably still underestimated fraction of the genome. (51, 52). At the primordial stages of evolution, when protocells divided stochastically, the mutualist-only protocells would win the competition against those lacking GE or those infested by parasites. However, upon the emergence of large genomes present in a single or few copies per cell, symmetrical division would evolve. However, symmetrical division makes (proto)cells vulnerable to parasite onslaught as parasites would persist after invading. Therefore, it appears that defense mechanisms, conceivably, those based on specific recognition of parasite sequences, would coevolve with symmetrical cell division mechanisms, being a pre-requisite for the long-term survival and evolution of such cells.

The present scenario for the origin of cells is consistent with the RNA world hypothesis (41, 42, 53). We stress, however, that the primordial RNA world must have evolved within pre-existing, metabolically active, membrane-bounded protocells (reproducers) as proposed previously by Copley, Smith and Morowitz (54).

Even if quite general, the model of the origin of life presented here suggests many avenues for experimental testing. In particular, experimental modeling of the origin of replicators within reproducers, that is, membrane vesicles encompassing proto-metabolic networks producing nucleotides and amino acids, and potentially, oligonucleotides and peptides, might not be far beyond the capability of modern

laboratories (55, 56).

Brief methods

For the protocell dynamics, the steps of the simulations are as follows. In the absence of GE, the protocell is described by the tuple of resource balance of the given cell and the type of the cell $\{B^i, type\}$ that defines the model parameters for that protocell ($\Delta E, p, E_{tr}$ and so on). GE-containing protocells are described by $\{B^i, type, m_C^i, m_D^i\}$. Thus, the whole population of protocells is described by the list of tuples. At the beginning of each round, the orderings of the tuples are randomly permuted to ensure swell mixed condition. A constant amount of resources R is supplied in the environment. The resource balance of the i-th cell of the population will be either $B_t^i = B_{t-1}^i + 1$ if there are still available resources in the environment before feeding, or will remain the same $B_t^i = B_{t-1}^i$ if the resources are already exhausted at the given round. After feeding, each protocell in the population pays the housekeeping cost defined by its type. If, after paying the housekeeping cost, the resource balance of the protocell is not positive $B_t^i - \Delta E \leq 0$, then the protocell is removed from the population.

The reproduction phase of the protocell starts when the resource balance is greater than the threshold value for reproduction defined by the type of the protocell, that is $B_t^i \geq E_{tr}$. The reproduction of the protocell ends up in two progeny protocells with probability p and no progeny with $1 - p$ (note, that for GE containing cells $p(m_{\mathcal{A}}^i, m_{\mathcal{N}}^i)$). That is, if the randomly generated number in the unit interval is less than p, then, new element is added in the list with the same type, and the resources of the reproducing protocell is either halved or randomly allotted. Then, the number of tuples of the given $type$ is selected from the main list representing the number of protocells of the given $type$ in the end of each round.

The intracellular dynamics of GE, that is, the birth and death process described by (4), is modeled using the Gillespie method, excluding the time of the occurrence of birth-death elementary processes because it is assumed that the number of elementary processes in each GE-containing protocell is equal to K at any given round of resource supply. The propensities of the elementary processes are constructed first. For the birth of autonomous GE, $k_1 = r_{\mathcal{A}}\, m_{\mathcal{A}}^i\, l^i$, where $m_{\mathcal{A}}^i$ is the number of autonomous GE in the protocell, and $l^i = Integer[\frac{B_t^i}{E_c}]$ describes the number of GE that can be made from the resources of the cell. A birth of a autonomous element results to the following changes $m_{\mathcal{A}}^i \to m_{\mathcal{A}}^i + 1$, and $l^i \to l^i - 1$. Similarly, for the birth of non-autonomous elements, the propensity is $k_2 = r_{\mathcal{N}}\, m_{\mathcal{N}}^i\, m_{\mathcal{A}}^i\, l^i$. The amount of resources and the numbers of GE and resources change according to $m_{\mathcal{N}}^i \to m_{\mathcal{N}}^i + 1$, $m_{\mathcal{A}}^i \to m_{\mathcal{A}}^i$ and $l^i \to l^i - 1$. The propensities of death processes are $k_3 = d_{\mathcal{A}}\, m_{\mathcal{A}}^i$ ($m_{\mathcal{A}}^i \to m_{\mathcal{A}}^i - 1$) and $k_4 = d_{\mathcal{N}}\, m_{\mathcal{N}}^i$ ($m_{\mathcal{N}}^i \to m_{\mathcal{N}}^i - 1$) for autonomous and non-autonomous elements, respectively. Then, a random number is generated from the $\epsilon \in [0,1]$ interval, and a reaction is chosen for which the following condition holds $\sum_{j=1}^{i-1} k_j < \epsilon \sum_{j=1}^{4} k_j \leq \sum_{j=1}^{i} k_j$, where $i = 1,..4$. Then, the number of GE in the protocell is updated according to the chosen process. The same steps are repeated K times in each round of the resource update in the environment. Note that the resource balance in the protocells is governed by the protocell level dynamics described above.

All other aspects of the simulations are presented in the Results section.

Acknowledgements

The authors thank Purificacion Lopez-Garcia for insightful discussions, crucial suggestions and critical reading of the manuscript. S.B., Y.I.W. and E.V.K. are supported by the Intramural Research Program of the National Institutes of Health of the USA (National Library of Medicine). A.A. and A.K. are supported by State Committee of Science of Armenia, grant No. 21AG-1C038.

Competing interests

The authors declare no competing interests.

References

1. Szathmary E (2000) The evolution of replicators. *Philos Trans R Soc Lond B Biol Sci* 355(1403):1669-1676
2. Szathmary E & Demeter L (1987) Group selection of early replicators and the origin of life. *J Theor Biol* 128(4):463-486
3. Szathmary E & Maynard Smith J (1997) From replicators to reproducers: the first major transitions leading to life. *J Theor Biol* 187(4):555-571
4. Koonin EV, Dolja VV, Krupovic M, & Kuhn JH (2021) Viruses Defined by the Position of the Virosphere within the Replicator Space. *Microbiol Mol Biol Rev* 85(4):e0019320
5. Alberts B, *et al.* (2014) *Molecular Biology of the Cell* (Garland Science, New York 6th Ed.
6. Stryer L (1995) *Biochemistry* (W.H.Freeman&Co, New York 4th Ed.
7. Jalasvuori M & Koonin EV (2015) Classification of prokaryotic genetic replicators: between selfishness and altruism. *Ann N Y Acad Sci* 1341:96-105
8. Koonin EV & Starokadomskyy P (2016) Are viruses alive? The replicator paradigm sheds a decisive light on an old but misguided question. *Stud Hist Philos Biol Biomed Sci* in press
9. Pross A (2004) Causation and the origin of life. Metabolism or replication first? *Orig Life Evol Biosph* 34(3):307-321
10. Zhu TF & Szostak JW (2009) Coupled growth and division of model protocell membranes. *J Am Chem Soc* 131(15):5705-5713
11. Higgs PG (2021) When Is a Reaction Network a Metabolism? Criteria for Simple Metabolisms That Support Growth and Division of Protocells. *Life (Basel)* 11(9)
12. Shirt-Ediss B, Sole RV, & Ruiz-Mirazo K (2015) Emergent chemical behavior in variable-volume protocells. *Life (Basel)* 5(1):181-211
13. Shah V, de Bouter J, Pauli Q, Tupper AS, & Higgs PG (2019) Survival of RNA Replicators is much Easier in Protocells than in Surface-Based, Spatial Systems. *Life (Basel)* 9(3)
14. Roy S & Sengupta S (2021) Evolution towards increasing complexity through functional diversification in a protocell model of the RNA world. *Proc Biol Sci* 288(1963):20212098
15. Lombard J, Lopez-Garcia P, & Moreira D (2012) The early evolution of lipid membranes and the three domains of life. *Nat Rev Microbiol* 10(7):507-515
16. Deamer D (2016) Membranes and the Origin of Life: A Century of Conjecture. *J Mol Evol* 83(5-6):159-168
17. Monnard PA & Deamer DW (2002) Membrane self-assembly processes: steps toward the first cellular life. *Anat Rec* 268(3):196-207
18. Miller SL (1953) Produciton of amion acids under possible primitive earth conditions. *Science* 117:528-529
19. Miller SL, Urey HC, & Oro J (1976) Origin of organic compounds on the primitive earth and in meteorites. *J Mol Evol* 9(1):59-72
20. Lazcano A & Miller SL (1996) The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell* 85(6):793-798
21. Becker S, *et al.* (2019) Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* 366(6461):76-82
22. Muchowska KB, Varma SJ, & Moran J (2020) Nonenzymatic Metabolic Reactions and Life's Origins. *Chem Rev* 120(15):7708-7744

23. Mulkidjanian AY, Galperin MY, & Koonin EV (2009) Co-evolution of primordial membranes and membrane proteins. *Trends Biochem Sci* 34(4):206-215
24. Turk RM, Chumachenko NV, & Yarus M (2010) Multiple translational products from a five-nucleotide ribozyme. *Proc Natl Acad Sci U S A* 107(10):4585-4589
25. Yarus M (2011) The meaning of a minuscule ribozyme. *Philos Trans R Soc Lond B Biol Sci* 366(1580):2902-2909
26. McGinness KE & Joyce GF (2003) In search of an RNA replicase ribozyme. *Chem Biol* 10(1):5-14
27. Vicens Q & Cech TR (2009) A natural ribozyme with 3',5' RNA ligase activity. *Nat Chem Biol* 5(2):97-99
28. Horning DP & Joyce GF (2016) Amplification of RNA by an RNA polymerase ribozyme. *Proc Natl Acad Sci U S A* 113(35):9786-9791
29. Tjhung KF, Shokhirev MN, Horning DP, & Joyce GF (2020) An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc Natl Acad Sci U S A* 117(6):2906-2913
30. Takeuchi N & Hogeweg P (2007) The role of complex formation and deleterious mutations for the stability of RNA-like replicator systems. *J Mol Evol* 65(6):668-686
31. Koonin EV, Wolf YI, & Katsnelson MI (2017) Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. *Biol Direct* 12(1):31
32. Mercier R, Kawai Y, & Errington J (2014) General principles for the formation and proliferation of a wall-free (L-form) state in bacteria. *Elife* 3
33. Wu LJ, *et al.* (2020) Geometric principles underlying the proliferation of a model cell system. *Nat Commun* 11(1):4149
34. Zintzaras E, Santos M, & Szathmary E (2002) "Living" under the challenge of information decay: the stochastic corrector model vs. hypercycles. *J Theor Biol* 217(2):167-181
35. Szathmary E (2006) The origin of replicators and reproducers. *Philos Trans R Soc Lond B Biol Sci* 361(1474):1761-1776
36. Doolittle WF & Inkpen SA (2018) Processes and patterns of interaction as units of selection: An introduction to ITSNTS thinking. *Proc Natl Acad Sci U S A* 115(16):4006-4014
37. Murray JD (2002) *Mathematical Biology: an Introduction* (Springer, New York.
38. Smith E & Morowitz HJ (2016) *The origin and Nature of Life on Earth: The Emergence of the Fourth Geosphere* (Cambridge University Press, Cambridge.
39. Penny D (2005) An interpretative review of the origin of life research. . *Biol. Philos.* in press
40. Wolf YI & Koonin EV (2007) On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol Direct* 2:14
41. Joyce GF & Szostak JW (2018) Protocells and RNA Self-Replication. *Cold Spring Harb Perspect Biol* 10(9)
42. Joyce GF (2002) The antiquity of RNA-based evolution. *Nature* 418(6894):214-221
43. Takeuchi N & Hogeweg P (2012) Evolutionary dynamics of RNA-like replicator systems: A bioinformatic approach to the origin of life. *Phys Life Rev* 9(3):219-263
44. Mizuuchi R & Ichihashi N (2021) Primitive Compartmentalization for the Sustainable Replication of Genetic Molecules. *Life (Basel)* 11(3)
45. Laurent G, Peliti L, & Lacoste D (2019) Survival of Self-Replicating Molecules under Transient Compartmentalization with Natural Selection. *Life (Basel)* 9(4)
46. Blokhuis A, Lacoste D, Nghe P, & Peliti L (2018) Selection Dynamics in Transient Compartmentalization. *Phys Rev Lett* 120(15):158101
47. Matsumura S, *et al.* (2016) Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science* 354(6317):1293-1296
48. Markvoort AJ, Sinai S, & Nowak MA (2014) Computer simulations of cellular group selection reveal mechanism for sustaining cooperation. *J Theor Biol* 357:123-133
49. Koonin EV, Senkevich TG, & Dolja VV (2006) The ancient Virus World and evolution of cells. *Biol Direct* 1:29
50. Koonin EV, *et al.* (2020) Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol Mol Biol Rev* 84(2)
51. Koonin EV, Makarova KS, & Wolf YI (2017) Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu Rev Microbiol* 71:233-261
52. Bernheim A & Sorek R (2020) The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol* 18(2):113-119
53. Gilbert W (1986) The RNA World. *Nature* 319:618
54. Copley SD, Smith E, & Morowitz HJ (2007) The origin of the RNA world: co-evolution of genes and metabolism. *Bioorg Chem* 35(6):430-443

55. Engelhart AE, Adamala KP, & Szostak JW (2016) A simple physical mechanism enables homeostasis in primitive cells. *Nat Chem* 8(5):448-453
56. O'Flaherty DK, *et al.* (2018) Copying of Mixed-Sequence RNA Templates inside Model Protocells. *J Am Chem Soc* 140(15):5171-5178