Biomathematical enzyme kinetics model of prebiotic autocatalytic RNA networks: degenerating parasite-specific hyperparasite catalysts confer parasite resistance and herald the birth of molecular immunity

Short title:

degenerating hyperparasite catalysts build anti-parasite immunity

Authors:

Magnus Pirovino[1], Christian Iseli[2], Joseph A. Curran[3], Bernard Conrad[4*]

Affiliations: [1]OPIRO Consulting Ltd., FL-9495 Triesen, Principality of Liechtenstein, [2]Bioinformatics Competence Center, EPFL and Unil, CH-1015 Lausanne, [3]Department of Microbiology and Molecular Medicine, and Institute of Genetics and Genomics of Geneva (iGE3), Geneva University, CH-1211 Genève 4, [4]Genesupport, Avenue de Sévelin 18, CH-1004 Lausanne

*Correspondence to: bernard.conrad@genesupport.ch

Catalysis and specifically autocatalysis are the quintessential building blocks of life. Yet, although autocatalytic networks are necessary, they are not sufficient for the emergence of life-like properties, such as self-propagation (replication) and adaptation. The ultimate and potentially fatal threat faced by molecular replicators is parasitism; if the polymerase error rate exceeds a critical threshold, even the fittest molecular species will disappear. Here we have developed an autocatalytic RNA early life mathematical network model based purely on enzyme kinetics, more specifically the steady-state approximation. We confirm previous models showing that these autocatalytic cycles are sustainable, provided there is a sufficient nucleotide pool. However, molecular parasites arise rapidly and become unsustainable unless they sequentially degenerate to hyperparasites (i.e. parasites of parasites). These hyperparasites acquire parasite binding specificity via two distinct temporal pathways. Our model is supported at three levels; firstly, ribozyme polymerases display Michaelis-Menten saturation kinetics and comply with the steady-state approximation. Secondly, ribozyme polymerases are capable of sustainable auto-amplification and of surmounting the fatal error threshold. Thirdly, with growing sequence divergence of host and parasite catalysts, the probability of self-binding increases and the trend towards cross-reactivity diminishes. Our model predicts that primordial host-RNA populations evolved via an arms race towards a host-parasite-hyperparasite catalyst trio that conferred parasite resistance within an RNA replicator niche. As such, it adds another mechanism – what's more, with biochemical precision – by which parasitism can be tamed and offers an attractive explanation for the universal coexistence of catalyst trios within prokaryotes and the virosphere, heralding the birth of a primitive molecular immunity.

Author Summary

The quintessential components of life comprise a potent mixture of naturally occurring, but improbable chemical reactions (catalysis), and the arrangement of such accelerated chemical reactions into closed loops (autocatalytic sets). This is required, but is not sufficient for such networks to self-propagate (amplification of the information carrier = host polymerization) and adapt (Darwinian evolution). As soon as self-propagation is attained, the next hurdle is parasitism. This typically involves shorter molecules (the products of replicative errors) that hitchhike the replicative potential of the host. They will invariably outcompete the regular amplification process, unless a solution is found. We have addressed this problem using a new model based on the mathematics of catalysis. This model confirms previous studies demonstrating that autocatalytic sets become self-sustaining, assuming that a sufficient pool of molecular building blocks is available. However, molecular parasitism is pervasive and potentially fatal for both host and parasite. In our model, we

allow these parasites to degenerate in a controlled fashion, giving rise to parasites of parasites (hyperparasites). As long as these hyperparasites acquire binding specificity for parasites, an attenuation of parasitism is observed. These parasite-hyperparasite cycles stabilize the host cycle, explaining why they are conserved, and why they are the likely reason behind the observation that all cellular hosts are associated with parasites (e.g. bacteria) and hyperparasites (e.g. viruses) across all kingdoms of life. Moreover, it provides a novel solution to the usually intractable problem of parasitism.

Introduction

The multistep transition process from inanimate to living systems has been dated to >3.9 Ga [1]. A non-exhaustive list of thresholds that needed to be progressively surmounted during this process has been drawn, comprising, among others, chirality symmetry breaking, spontaneous polymerization, and self-assembly of compartments. The catalytic closure threshold and Eigen's error threshold have received particular attention [2]. The latter posits that if the replication error exceeds a critical threshold value even the fittest replicators will not survive [3].

One central pillar in the theory of how early life could have materialized pertains to the "RNA world". This stipulates that the initial informational polymer was RNA, and that the same molecule was also the fundamental catalyst [4]. The underlying concept of autocatalytic sets was laid out more than half a century ago [3,5]. It was subsequently refined [6] and validated in a proof of principle experiment showing that RNA fragments could build self-replicating ribozymes from individual RNA fragments via cooperative catalytic cycles; networks with three members showed strong, highly cooperative growth [7]. In parallel, an increasingly efficient number of RNA polymerase ribozymes were generated by *in vitro* evolution. These exhibited high levels of processivity [8], and displayed Michaelis-Menten saturation kinetics [9]. *In fine*, one RNA polymerase ribozyme surpassed Eigen's critical error threshold, and successfully accomplished several rounds of cyclic self-replication [10,11]. Collectively, both theoretical considerations and experimental evidence support the view of a gradual evolution towards RNA self-catalysis. This started with self-assembly, followed by the formation of autocatalytic networks using self-assembled substrates. From this emerged different ribozymes (recombinases, ligases) and template based ligation, leading to a template based, self-replicase catalysed assembly of monomers [12].

The ribozyme catalysed RNA-polymerization process is even more error prone [4,8-11] than the proteinaceous RNA polymerases that notoriously lack proofreading activity [13]. This generates variant sequence space mutants called quasispecies [14,15]. This variant sequence-space intrinsically implies the obligate presence and evolutionary persistence of molecular parasites as observed in all

life forms, and can be argued essentially for thermodynamic reasons [16,17]. Two main mechanisms have been proposed to confer parasite resistance to a given habitat, namely spatial self-organization [18-24], and physical compartmentalisation most probable via lipid bilayers, and the construction of protocells [25-27]. A direct comparison indicates that the survival of RNA replicators most likely occurred in protocells [28].

Since catalysis and autocatalysis are central to life, from a mathematical perspective the Michaelis-Menten reaction scheme appears pertinent [29], and implicitly the quasi steady-state assumption [30]. These considerations lay the ground for modelling catalysis, autocatalysis and the interactions within primordial, template-based ecological RNA communities using a single biomathematical tool derived from enzyme kinetics. We model here, based on a previously established tripartite population scheme [31], how molecular parasite resistance and a primordial molecular antiparasite immunity gradually emerged from progressively degenerating hyperparasite catalysts acquiring parasite-specificity.

Results

Enzyme kinetics lies at the centre of the living world, and the intermediate complex lies at the centre of the centre. Consequently, in developing our current model we have used equations derived from enzyme kinetics, specifically the steady-state approximation by Briggs and Haldane [32]. Starting with a ribozyme polymerase exhibiting Michaelis-Menten saturation kinetics capable of beating the error-threshold [9-11], we demonstrate how template-based RNA autocatalytic cycles progressively degenerate to create successively parasite and hyperparasite cycles. Metabolic autocatalytic networks, arguably arising spontaneously and predating template-based polymerization [33], are assumed to be present and to provide sufficient energy and polymer precursors [23,31]. The previously established tripartite habitat framework, featuring a template-based RNA host and a molecular parasite that is itself parasitized (hyperparasite catalysts) [31] is instrumental for homeostasis and the emergence of molecular parasite resistance. It hinges upon hyperparasite catalysts acquiring parasite-specificity.

We provide mathematical models that describe the time evolution of autocatalytic RNA networks. The main goal of these models is to investigate the general conditions under which the stability of such networks could occur. The physical/chemical dimensions of the relevant input and output values (e.g. t = time, k_{cat}, k_{on}, k_{off}, K_M, RNA-concentration values, etc.) are assumed to be dimensionless. However, when available numerical values have been employed. Nevertheless, these values were not validated with respect to a specific experimental system.

Biomathematical Michaelis-Menten kinetics and the Briggs-Haldane steady-state approximation for autocatalytic RNA networks

It is assumed that the ribozyme polymerase represents the positive RNA strand R_1 with kinetic parameters capable of cyclic auto-amplification that surmount the error threshold (e.g. a k_{cat} of 0.1−1 min^{-1} and a K_M akin or superior to 0.1−1 μM) [9-11]. Its complementary negative RNA strand R_2 is considered to be catalytically inert (uniquely a template). A complete autocatalytic replication cycle comprises the initial production of R_2 by two R_1 ribozyme molecules, and the subsequent reproduction of R_1 by R_1 using template R_2 (Model 1, formalism described in Methods or Models, $R_1 \Rightarrow (R_1\_R_1) \rightarrow^c R_2 \Rightarrow (R_1\_R_2) \rightarrow^c R_1\ (1.0)$, see also S1 Supporting Information for scheme in Fig 2A). Figure 1 shows a graphical representation of this model assuming realistic starting kinetics and habitat restriction parameters (described in Methods or Models). As can be seen, the abundance of these two molecular species increases rapidly and becomes stable, supporting previous work showing the viability of ribozyme polymerases, assuming that sufficient energy in the form of molecular building blocks is available [23].

Molecular parasitism in autocatalytic cycles

Molecular parasites are an inherent feature of the replication process. For a number of reasons they are inseparably associated with all life forms [16,17], with the most important being the accuracy-rate trade-off [34,35]. In general, enzymes operate orders of magnitude (k_{cat}/K_M of $\sim 10^5$ $M^{-1}s^{-1}$) below the diffusion rate (10^8-10^9 $M^{-1}s^{-1}$). This implies promiscuous substrate binding, high catalytic activity, and as a secondary trade off, a high mutation rate generating a large variance in the quasispecies: i.e. speed being more important than enzyme accuracy [36]. Selection favours parasite molecules of subgenomic size [37-39], since the smaller they are the shorter their replication time, and the more overwhelmingly abundant they will become [38]. In addition, host ribozyme polymerases R_1 amplifying parasites are unavailable for the synchronous host amplification cycle $R_1\_R_2$. This is exactly what Model 2.1 shows (Fig 2A, see also S1 Supporting Information for scheme in Fig 2B). It is merely a matter of time before the molecular parasites $R_3 = P$ outcompete the host molecules R_1 and R_2. Furthermore, if molecular parasites are diffusing in from a neighbouring habitat, the process is even more rapid (Model 2.2, Fig 2B). Therefore, as expected, high parasite loads overpower a typical habitat driven by efficient, but promiscuous ribozyme polymerases with P populations. This conundrum begs a solution.

The emergence of hyperparasites: stable autocatalytic cycles in the presence of molecular parasites

The host ribozyme polymerase R_1 is by design promiscuous ensuring amplification of both the host cycle R_1-R_2, and the very efficient amplification of the typically smaller $R_3 = P$ population (see above).

Because of its high mutation rate, it will not only reproduce the original *P* species, but will inevitably also generate *P* mutants dubbed *F*, i.e. hyperparasites. Selection will preferentially amplify hyperparasite populations smaller than *P* [37-39] (Model 3, formally described in *Materials and Models*, $R_1 \Rightarrow (R_1\_P) \rightarrow^c F \ (1\&2 \rightarrow 3)$, see also S1 Supporting Information for scheme in Fig 2C). Short and loosely folded RNA sequences will be preferentially replicated (they are better templates), whereas longer and tightly folded ones may be better ribozymes and more resistant to hydrolytic decay [40]. As a corollary, *F* will be a highly efficient template, yet a poor or non-functional ribozyme capable of self-amplification. R_1 is already efficiently producing *P* and *F*. However, this process is not subject to evolutionary pressure and is bound to degenerate ($F \Rightarrow (P\_F) \rightarrow^c nil$, $F \Rightarrow (F\_F) \rightarrow^c nil \ (3.0)$). The hyperparasite will parasitize *P* for several reasons; firstly, while bound in intermediate *P_F* complexes it is not available to R_1, which can again engage in the amplification of the host cycle R_1-R_2. Secondly, in order for hyperparasitism to be highly efficient, it is required that *F* has a high specificity for *P*; i.e. in the presence of both R_1 and *P* it binds mainly or exclusively to the latter. Indeed, this is precisely what the accuracy-rate trade-off predicts, with selectivity implying the ability to discriminate between two templates when both are present [35], whereas rate denotes speed [35,36]. Ground state discrimination means that specificity is achieved mainly through substrate binding, which imposes strong accuracy/rate trade-offs [34,35]. Improvements in selectivity mediated by tighter cognate substrate binding invariably leads to lower catalytic efficiency (parallel decreases in the constants for the cognate substrates K_M^{cog} and k_{cat}^{cog} [34,35]). This is precisely what the model stipulates; *F* has parasite- and hyperparasite-selectivity, which comes at the cost of reduced catalytic activity as would be predicted for a smaller molecular species. As shown in Fig. 3A (Model 3.1), and assuming the standard parameters as applied in Model 1, with similar K_M values (in the μM range) for both the intermediate complexes (*P_F* and *F_F*) and R_1_R_2, hyperparasitism *per se* sustainably restores the abundance of the host cycle species R_1 and R_2. Still, the hyperparasite *F* is the most abundant species in the habitat, and the level of *P* remains important.

However, increasing the binding affinity of *F* for itself and *P* by one order of magnitude (K_M values for *P_F* and *F_F* now in the 0.1 μM range, Model 3.2, Fig. 3B, see also S1 Supporting Information for scheme in Fig 2D) fully restores the abundance of the host cycle RNA species R_1 and R_2, at levels indistinguishable from the starting condition in Model 1. Since binding affinity under these conditions impacts directly on K_D values and hence catalytic efficiency (k_{cat}/K_M) [35], incremental changes in the affinity of *F* for itself and for *P,* by varying either the on-rate (fast association) or off-rate (slow dissociation) can impact significantly on the systems equilibrium [41,42]. Given what is currently known about template selection, modulating the on-rates appears to be the favoured option [43,44].

Several lines of experimental evidence suggest that self-binding within both host and parasite catalysts, with little cross-reactivity, is likely to happen. Firstly, just like proteins, RNAs exhibit a wide range of association rates, although even the fastest are significantly slower than those observed with proteins [42]. Nonetheless, RNA dimer formation is among the most rapid [42]. Secondly, the host and parasite-hyperparasite catalyst populations will diverge considerably, in both size and sequence, making the formation of homodimers or -multimers much more likely than that of heterodimers- or -multimers. Indeed, self-similarity increases the likelihood of physical interactions and spatially similar molecular structures interact preferentially [45]. In conclusion, with growing sequence divergence of host and parasite catalysts the probability of self-binding increases, and the trend towards cross-reactivity diminishes, precisely as suggested by our model.

Two temporally distinct scenarios can be envisaged; the parasite and hyperparasite immediately loose polymerase function because of the inherently high mutation load associated with the elevated replication rates of parasitic templates, and acquire parasite binding specificity in a second step. Consistent with this notion, the same laws that govern substrate-enzyme interactions are also valid for protein-protein associations [41,47], and as an extension here for ribozyme polymerases [34,43,44]. This scenario is discussed in in the main model. Alternatively, as outlined in the model described in S1 Supporting Information, hyperparasite catalysts with higher binding affinity for parasites are initially selected, but rapidly loose catalytic activity because of the accuracy-rate trade-off [35]. Subsequently, *F*'s catalytic activity is bound to degenerate, since R_1 mediated replenishment of the parasite pool is far more efficient than the one resulting from the poor catalytic activity of *F*.

In sum, these simulations show that parasite resistance can be modelled with biochemical precision, and it relies on hyperparasites acquiring parasite binding specificity, via at least two temporally distinct pathways.

Discussion

We present a novel early-life model for autocatalytic ribozyme polymerase networks based on enzyme kinetics, more specifically the steady-state approximation. As a starting point, we defined a ribozyme polymerase with kinetic parameters drawn from in vitro models, which was capable of sustaining cyclic self-amplification and overcoming the error-threshold. We subsequently confirmed earlier models showing that these autocatalytic cycles are sustainable, as long as a sufficient nucleotide pool is available. Inevitably, an efficient but promiscuous polymerase ribozyme with inherently low fidelity will generate molecular parasites that will overwhelm the habitat. However, because molecular parasitism is a chain process, parasites of parasites (hyperparasites) will eventually

appear. As long as these hyperparasites acquire parasite binding-specificity, an increasingly solid and molecularly specific parasite resistance will ensue. We demonstrate that this parasite resistance can be modelled with biochemical precision. Collectively, these findings add one more mechanism by which parasites can be tamed. It offers an attractive explanation for the universal and tight presence of molecular trios within the cellular hosts from prokaryotes to the virosphere, and showcases the birth of parasite-resistance and a primitive molecular immunity (see below).

Once highly processive and efficient RNA catalysts are continuously provided to an experimental habitat, the host and parasite populations turn into an evolving ecosystem through Darwinian evolution, forming distinct host and parasite lineages that exhibit arms-race dynamics [37]. Resonating with this is the fact that self-replicating ribozyme networks consisting of three members display particularly high cooperative growth dynamics [7]. Intriguingly, there appears to be a trade-off between short and loosely folded sequences that are preferentially amplified (templates), and longer, tightly folded ones that are better ribozymes and are more resistant to hydrolytic decay [40]. This in-turn provides a rationale for splitting RNA communities into subgroups [7,37,40].

At the core of our current model is enzyme kinetics. The catalyst accuracy-rate trade-off analysis [35] applied to autocatalytic RNA networks [34] predicts the emergence of active ribozymes [10,11,46] with binding promiscuity and high catalytic activity, co-existing with subgenome-sized molecules characteristic of parasites with relatively lower catalytic activity [40]. This is precisely what our model predicts. Furthermore, this enzyme kinetics based model allows us to adapt these processes with high precision, since the input parameters can be varied at will. An incremental change in the affinity of the hyperparasite ribozyme for itself and for first generation parasite ribozymes can be achieved by varying association (fast association) and/or dissociation rates (slow dissociation) [41,42,47] although current models would favour the former option [43,44].

While primitive tripartite populations featuring rapidly degenerating catalysts likely represented a homeostatically stabilizing necessity, parasite communities within the population might have subsequently evolved more freely, with some becoming replicationally autonomous. Indeed, tripartite microbial populations composed of a host, a parasite and a hyperparasite are virtually ubiquitous [48], as evidenced for bacteria-bacteriophage-phage-satellites [49,50], host-RNA viruses-defective interfering RNA viruses [51], and eukaryotic host-NCLDVs-virophages [52]. Furthermore, it is striking that the appearance of microbial immunity is associated with hyperparasites. Several recent studies have shown that phages and their satellites encode a set of diverse antiphage systems [53,54], and defective interfering RNAs activate innate immunity [55]. Phage satellite-encoded immune systems protect bacteria from phage predation and were proposed to be an integral component of innate immunity [56]. This notion could easily be extended to

omnipresent, bona fide hyperparasites, such as endogenous retroviruses [57,58]. Therefore, there is likely a strong, ancestral link between hyperparasites and parasite immunity.

We previously developed a tripartite early-life host-parasite-hyperparasite framework based on Lotka Volterra (LV)-equations [31,59]. The present work arrives at similar results, showing that LV-based equations provide an adequate approximation for the more realistic, kinetics-based model developed here. Nonetheless, the present model has by design intrinsic limitations, notably the fact that we restrict the reaction to one intermediate polymerase-template complex [60]. Additionally, we assume that the catalytic efficiency of ribozyme polymerases will slowly approach that of proteinaceous holoenzymes and of mainstream, albeit complexed, multi-subunit enzymatic machineries. These aspects will be addressed in future studies.

Methods or Models

Definitions and Terms

1. Mathematical notation

If not alternatively specified, n-vectors are always column-vectors. If a real valued column-vector $x \epsilon \mathbb{R}^n$ is transposed to a row-vector we write x^T. Accordingly, also nxm matrices $A \epsilon \mathbb{R}^{nxm}$ can be transposed by the same symbol A^T. The i_{th} entry of a vector is denoted by x_i. The entry of the i_{th} row and the j_{th} column of a matrix A is denoted by A_{ij}.

2. Michaelis-Menten and Briggs-Haldane kinetics

The enzyme kinetics generally known as Michaelis-Menten kinetics [29,30] is introduced as follows:

$$E + S \underset{k_{off}}{\overset{k_{on}}{\leftrightharpoons}} E - S \overset{k_{cat}}{\rightharpoonup} E + P,$$

E being the enzyme, S the substrate, E_S the intermediate enzyme-substrate complex, and P the product. If all the symbols here are meant to be concentrations in a given environment (habitat), then the following equations apply:

$$\frac{dS}{dt} = -k_{on} E \cdot S + k_{off} E\_S$$

$$\frac{dE}{dt} = -k_{on} E \cdot S + (k_{off} + k_{cat}) E\_S$$

$$\frac{dE\_S}{dt} = k_{on} E \cdot S - (k_{off} + k_{cat}) E\_S$$

$$\frac{dP}{dt} = k_{cat} E\_S \qquad (1')$$

An important assumption is that the concentration of $E - S$ reaches the steady-state almost instantly (steady-state approximation). According to the Briggs-Haldane steady-state approximation [32], we therefore assume

$$\frac{dE\_S}{dt} = 0,$$

and thus:

$$k_{on} E \cdot S = (k_{off} + k_{cat}) E\_S$$

Here E is the free enzyme E, and S the free substrate. If we write

$$E_{tot} = E + E\_S \quad (2')$$

then the equation above yields, with the Michaelis constant $K_M = \frac{k_{off}+k_{cat}}{k_{on}}$,

$$(E_{tot} - E\_S) \cdot S = K_M E\_S \qquad (3')$$

and thus

$$E\_S = \frac{E_{tot} \cdot S}{K_M + S} \quad (4')$$

3. *Michaelis-Menten kinetics in the RNA-world*

Let us now consider a set of different RNA populations $\{R_1, R_2, \dots, R_n\}$ in a primordial molecular RNA landscape, with the supply of nucleotides being provided by metabolism [23,31]. This metabolism is not further specified, and we assume that there is a limitation on the number of RNA molecules $\{R_1, R_2, \dots, R_n\}$ that can be maintained (habitat restriction constant, see below). These RNA molecules may be ribozyme polymerases if they exert a catalytic activity or merely templates.

3.1 A simple three ribozyme system

In this RNA world, a catalytic activity is exerted if a free ribozyme polymerase exhibiting Michaelis-Menten saturation kinetics [9] from population R_1 encounters another from R_2. The ribozyme R_1 uses the free molecule R_2 as template, and by forming a polymerase-template complex, catalyses the production of a third ribozyme population R_3 (R_3 being the negative strand of R_2). We now apply Michaelis-Menten kinetics, which entails that not only R_1 is a ribozyme polymerase, but also the substrate R_2, and even the product R_3. The Michaelis-Menten formalism is valid for this special example. If R_1, R_2, R_3 denote the concentrations of the total ribozyme masses, and $R_{1_{free}}$, $R_{2_{free}}$, $R_{3_{free}}$ denote the concentrations of their respective free ribozyme masses, then $E = R_{1_{free}}$, $S = R_{2_{free}}$, and $P = R_3$. Then, equation $(2')$ reads as follows: $R_{1_{free}} = R_1 - R_1\_R_2$. But now, the free substrate $S = R_{2_{free}}$ is also a function of the total substrate R_2 and its enzyme-substrate complex $R_1\_R_2$, i.e., $R_{2_{free}} = R_2 - R_1\_R_2$. Equations $(1')$ and $(3')$ from above now read as follows:

$$\frac{dR_3}{dt} = k_{cat} R_1\_R_2 = \quad (1'')$$

$$(R_1 - R_1\_R_2) \cdot (R_2 - R_1\_R_2) = K_M R_1\_R_2 \quad (3'')$$

According to Briggs-Haldane it is assumed that the enzyme-substrate complex $R_1\_R_2$ reaches the steady-state very rapidly, $\frac{dR_1\_R_2}{dt} = 0$. Equation $(3'')$ is now a nonlinear quadratic equation, but still solvable for the unknown $R_1\_R_2$:

$$R_1\_R_2 = \frac{1}{2}(R_1 + R_2 + K_M) - \sqrt{\frac{1}{4}(R_1 + R_2 + K_M)^2 - R_1 R_2}$$

This solution of equation $(3'')$ is called the tight-binding or Morrison equation [61].

3.2 The general case of n ribozymes

In the above special case of three ribozymes, we have assumed that only ribozymes R_1 and R_2 are forming an enzyme-substrate complex $R_1\_R_2$. Ribozyme R_1 does not bind to ribozyme R_3 or to itself R_1, nor does ribozyme R_2 bind to R_1 or to R_3. In the general case, where we have n ribozymes $R_1, R_2, \dots, R_n$, every ribozyme polymerase R_i could potentially bind - through its role as enzyme - to a template (substrate) ribozyme R_j, forming an enzyme-template complex $R_i\_R_j$. In general, complex $R_i\_R_j$ is not equal to complex $R_j\_R_i$, since $R_i\_R_j$ is governed by the enzymatic activity of R_i, whereas $R_j\_R_i$ may be governed by the different enzymatic activity of R_j, which may not be the same.

Thus, in general, n ribozymes can bind to n^2 different enzyme-template complexes $R_i\_R_j, i, j = 1, \dots, n$.

$$R_{i_{free}} + R_{j_{free}} \underset{k_{off}^{i,j}}{\overset{k_{on}^{i,j}}{\leftrightharpoons}} R_i\_R_j \overset{k_{cat}^{i,j}}{\rightharpoonup} R_{j_{free}} + P$$

Equation $(1')$, describing the production rate of new products, now becomes

$$\frac{dP_{i,j}}{dt} = k_{cat}^{i,j} R_i\_R_j \qquad (1)$$

In an n ribozymes RNA world, the products $P_{i,j}$ can be found among the given ribozymes, i.e., $P_{i,j} \in \{R_1, R_2, \dots, R_n\}$. Equation $(2')$, describing the relation between the total (R_i) and the free $(R_{i_{free}})$ concentration of ribozymes, now turns into the following equation (2):

$$R_{i_{free}} = R_i - (\sum_{k=1}^{n} R_i\_R_k + R_k\_R_i) + R_i\_R_i \qquad (2)$$

Again, we assume according to Briggs-Haldane that all ribozyme-template complexes $R_i\_R_j$ reach the steady-state very rapidly, $\frac{dR_i\_R_j}{dt} = 0$, $i, j = 1, \dots, n$.

Given the Michaelis constants $K_M^{i,j} = \frac{k_{off}^{i,j} + k_{cat}^{i,j}}{k_{on}^{i,j}}$, equation $(3')$ turns now, for every $i, j = 1, \dots, n$, into

$$(R_i - (\sum_{k=1}^{n} R_i\_R_k + R_k\_R_i) + R_i\_R_i)(R_j - (\sum_{k=1}^{n} R_j\_R_k + R_k\_R_j) + R_j\_R_j) = K_M^{i,j} R_i\_R_j \quad (3)$$

Equation (3) is a nonlinear system of n^2 equations with n^2 unknown ribozyme-template complexes $R_i\_R_j$. Given the total concentrations, $R_1, R_2, \dots, R_n$, this equation is solvable through an appropriate nonlinear multidimensional multistep algorithm. For this purpose, we propose a Newton iteration outlined as follows.

3.3 Newton iteration for the solution of equation (3)

Let $R\_R \in \mathbb{R}^{n^2}$ be a real valued n^2 dimensional vector, with

$$R\_R = [R_1\_R_2, R_1\_R_2, \dots, R_1\_R_n, R_2\_R_1, R_2\_R_2, \dots, R_2\_R_n, R_3\_R_1 \ \dots, R_n\_R_n]^T$$

Let $F\colon \mathbb{R}^{n^2} \longrightarrow \mathbb{R}^{n^2}$, $R\_R \mapsto F(R\_R)$, be a n^2-multidimensional function

$F(R\_R) = [F_1(R\_R), F_2(R\_R), \dots, F_l(R\_R), \dots, F_{n^2}(R\_R)]^T$ with

$$F_l(R_R) =$$

$K_M^{i,j} R_i\_R_j$-$(R_i - (\sum_{k=1}^{n} R_i\_R_k + R_k\_R_i) + R_i\_R_i)(R_j - (\sum_{k=1}^{n} R_j\_R_k + R_k\_R_j) + R_j\_R_j)$,

l=1,…, n^2, where the indices i, j themselves are functions of index l. If we define these functions in the following way:

$$j(l) = \begin{cases} (l \bmod n) & , \ if \ (l \bmod n) > 0 \\ (l \bmod n) + n, & if \ (l \bmod n) = 0 \end{cases}$$

$$i(l) = \frac{(l - j(l))}{n} + 1,$$

then the lth component of vector R_R, is well-defined as $(R\_R)_l = R_{i(l)}\_R_{j(l)}$, and also the lth component of F, $F_l(R\_R)$ is well defined for every index $l = 1, \dots, n^2$. Now we state that a vector $R\_R = [R_1\_R_2, R_1\_R_2, \ \dots, R_n\_R_n]^T$ is a solution of (3), if an only if $F(R\_R) = 0$, i.e. $F_l(R\_R) = 0$ for every $l = 1, \dots, n^2$. The classic multidimensional Newton iteration algorithm needs an explicit expression for of the Jacobian J of $F(R\_R)$:

$$J_{l,m} = \frac{\partial F_l(R\_R)}{\partial (R\_R)_m}, \qquad l, m = 1 \dots n^2$$

The full derivation of the formula for the Jacobian is straight forward, therefore we just give here the final result. Let $\delta(i,j) = \begin{cases} 1, if \ i = j \\ 0, \quad else \end{cases}$ be the Kronecker symbol, then the explicit formula for the Jacobian is given as follows. For $l, m = 1, \dots, n^2$:

$$J_{l,m} = \frac{\partial F_l(R\_R)}{\partial (R\_R)_m} = K_M^{i(l),j(l)} \ \delta(l,m) +$$

$$+ R_{j_{free}} \big(\delta(i(m), i(l)) + \delta(j(m), i(l)) - \delta(i(m), i(l))\delta(j(m), i(l))\big) +$$

$$+ R_{i_{free}} \big(\delta(i(m), j(l)) + \delta(j(m), j(l)) - \delta(i(m), j(l))\delta(j(m), j(l))\big)$$

In this formula, the free ribozymes, $R_{i_{free}}$, are calculated according to (2).

Now we are able to write down the full Newton iteration algorithm. Given some initial set of total ribozyme concentrations $\{R_1, R_2, \dots, R_n\}$, we start assuming that no ribozyme-template complex has been built yet; i.e. $R_{i_{free}} = R_i, i = 1, .., n$, and $(R\_R)_l = 0$, l =1,..., n^2. The initial vector of ribozyme-template complexes therefore equals to zero: $(R\_R)^{(0)} = 0$. After k iterations, the resulting vector of ribozyme-template complexes is then calculated by:

$$(R\_R)^{(k+1)} = (R\_R)^{(k)} - (J)^{-1} F\left((R\_R)^{(k)}\right), \qquad k = 0,1,2, \dots$$

Here $(J)^{-1}$ is the is n^2x n^2 inverse matrix of the Jacobian J. Normally, this sequence of iterations convergences to a solution $(R\_R)^{(k)} \to (R\_R)^* = R\_R$, with $F((R\_R)^*) = 0$. The entries of the resulting vector R_R are then concentrations of all of the possible ribozyme-template complexes $(R\_R)_l = R_{i(l)}\_R_{j(l)}$, l =1,..., n^2, solving equation (3).

4. *Template based autocatalytic cycles*

4.1 Preamble

Reconsider the set of populations of different RNA molecules $\{R_1, R_2, \dots, R_n\}$ introduced above. The ribozymes form a catalytic cycle, if

$$R_1 \to^c R_2 \to^c R_3 \to^c \dots \to^c R_n \to^c R_1,$$

where the symbol "$A \to^c C$" means: ribozyme A catalyzes ribozyme C. The catalytic production of C is called *template based* if an encounter with a template ribozyme B is needed for this process, i.e. A needs to form, together with B, an enzyme-template intermediate complex A_B in order to produce C. We write "$A\_B \to^c C$". If the autocatalytic cycle reproduces not only all ribozymes carrying an enzymatic role, but also all ribozymes serving as templates along its way then we call it a *template based autocatalytic cycle*. The following example meets all the criteria of such a template based autocatalytic cycle:

$$R_1 \Rightarrow (R_1\_R_1) \to^c R_2 \Rightarrow (R_2\_R_1) \to^c R_3 \Rightarrow (R_2\_R_3) \to^c R_4 \Rightarrow (R_4\_R_1) \to^c R_1$$

This reads as follows, a ribozyme R_1 forms an enzyme-template complex with itself, $R_1\_R_1$, producing a new copy of ribozyme R_2. R_2 too forms an enzyme-template complex with R_1, $R_2\_R_1$, producing ribozyme R_3. R_2 now forms another enzyme-template complex with R_3, $R_2\_R_3$, producing R_4, which in turn produces again $R_1$by forming a catalytic enzyme-template complex $R_4\_R_1$ with R_1. Now we are able to model primitive template based autocatalytic cycles in an RNA landscape, with a given, but not further specified metabolism providing a limited supply of nucleotides.

4.2 Model 1: the host cycle

Let ribozyme R_1 be a population of positive strand ribozyme polymerases, and R_2 be the population of its negative strand, then we can model the following template based autocatalytic cycle:

Model 1

$$R_1 \Rightarrow (R_1\_R_1) \rightarrow^c R_2 \Rightarrow (R_1\_R_2) \rightarrow^c R_1 \quad (1.0)$$

The replication dynamics is governed by the following ODEs:

$$\frac{dR_1}{dt} = k_{cat}^{1,2}\, R_1\_R_2 \Pi - d_{R_1} R_1 \qquad (1.1)$$

$$\frac{dR_2}{dt} = k_{cat}^{1,1}\, R_1\_R_1 \Pi - d_{R_2} R_2 \qquad (1.2),$$

Here d_{R_i} are the decay rates of populations $R_i, i = 1, 2$, and $\Pi = \left(1 - \frac{R_1 + R_2}{K}\right)$ is the habitat restriction constant. K is the maximum number of R_1 or R_2 RNA molecules the habitat is able to maintain through the given metabolism. Given the total concentrations R_1 and R_2 we assume using Briggs-Haldane that the enzyme-template intermediate complexes $R_1\_R_1$ and $R_1\_R_2$ are reaching their steady-state at every point in time very rapidly. In other words, we assume that the process of attaining the steady-state for the enzyme-template intermediate complexes is by some orders of magnitudes faster than the production of new ribozymes. In order to calculate all of the possible enzyme-template complexes $R_i\_R_j$ through the Newton iteration algorithm, as outlined above, we need to attribute values to $k_{on}^{i,j}, k_{off}^{i,j}, k_{cat}^{i,j}$, along the Michaelis constants $K_M^{i,j} = \frac{k_{off}^{i,j} + k_{cat}^{i,j}}{k_{on}^{i,j}}, i, j = 1,2$. E.g.

$$k_{on} = \begin{bmatrix} .2 & .19 \\ \varepsilon & \varepsilon \end{bmatrix}, k_{off} = \begin{bmatrix} .1 & .1 \\ \eta & \eta \end{bmatrix}, k_{cat} = \begin{bmatrix} .1 & .08 \\ 0 & 0 \end{bmatrix}, K_M = \begin{bmatrix} 1.0 & .947 \\ \eta/\varepsilon & \eta/\varepsilon \end{bmatrix}$$

We further note that in this model ribozyme R_2 does not form an enzyme-template complex with R_1, nor with itself, assuming that R_2- the negative strand of R_1- does not have catalytic activity. Since the Newton algorithm needs Michaelis constants for all possible enzyme-template complexes $R_i\_R_j$, we assume here that $k_{on}^{2,1} = k_{on}^{2,2} = \varepsilon$ is a very low number greater than zero, and also $k_{off}^{2,1} = k_{off}^{2,2} = \eta > 0$. We choose ε and η so that formation of enzyme-template complexes $R_2\_R_1$, $R_2\_R_2$ is possible, but always at very low concentration and also dissociating very rapidly; e.g., $\varepsilon = 10^{-5}, \eta = 10^{-1}$. By choosing appropriate values for all other parameters (e.g. $d_{R_i} = d_{R_2} = 0.005$, $K = 20.0$, $R_{1_{initial}} = 1.0$, $R_{2_{initial}} = 1.0$). The ODEs $(1.1), (1.2)$ of Model 1 are solved numerically by simple timestep integration. Note that for every timestep $t_k = t_0 + k\Delta t$ a calculation of all the possible enzyme-template complexes $R_i\_R_j$ through the Newton iteration algorithm is needed.

4.3 Model 2: the emergence of parasitism

Autocatalytic host cycles (1.0) inevitably are subject to molecular parasitism [16,17]. In (1.0), the ribozyme polymerase R_1 is replicating R_2 and itself. Such a replicase cannot be so specific that it only replicates R_1 and R_2, since under such specific conditions R_1 would quickly lose its catalytic activity due to mutations it is subject to [16,17]. Therefore, ribozyme polymerase R_2 will inevitably be parasitized by RNA templates, typically of subgenomic size [37,38], say $R_3 = P$ (parasites):

$$P = R_3 \Rightarrow (R_2\_R_3) \rightarrow^c R_3 = P \quad (2.0)$$

This motivates following additional equation

$$\frac{dR_3}{dt} = k_{cat}^{1,3} R_2\_R_3 \Pi - d_{R_3} R_3 \qquad (2.1)$$

The habitat restriction term Π extends to

$$\Pi = \left(1 - \frac{R_1 + R_2 + {}^{R_3}/_{q_{R_3}}}{K}\right),$$

where $q_{R_3} > 1$, since parasites $R_3 = P$ are more efficient at using the habitat resources, because of their smaller size that confers a replicative advantage [38]. A ribozyme polymerase R_1 that engages in replicating $R_3 = P$, building an enzyme-template complex $R_1\_R_3$, cannot simultaneously form enzyme-template complexes with $R_1\_R_2$ or $R_1\_R_2$. Thus, R_3 parasitizes the whole autocatalytic cycle (1.0). Here, for simplicity, we do not distinguish between positive and negative strand parasites.

We model two parasite sources, the most probable being the imperfect fidelity of polymerase R_1. Rather than producing a viable RNA molecule R_2 or R_1, it may, at a specific mutation rate, $\mu > 0$, produce a non-viable mutant having more similarity to a parasite $P = R_3$, rather than to a viable RNA molecule of the host cycle (1.0). Parasite invasion from a neighbouring habitat can be viewed as an alternative.

The ODE equations for Model 2 are as follows:

$$\frac{dR_1}{dt} = (1-\mu) k_{cat}^{1,2} R_1\_R_2 \Pi - d_{R_1} R_1 \qquad (2.1)$$

$$\frac{dR_2}{dt} = (1-\mu) k_{cat}^{1,1} R_1\_R_1 \Pi - d_{R_2} R_2 \qquad (2.2),$$

$$\frac{dR_3}{dt} = (k_{cat}^{1,3} R_1\_R_3 + \mu(k_{cat}^{1,2} R_1\_R_2 + k_{cat}^{1,1} R_1\_R_1))\Pi + \alpha - d_{R_3} R_3 \qquad (2.3)$$

Here we assign values corresponding to the ones used in Model 1: $d_{R_1} = d_{R_2} = d_{R_3} = 0.005$, $K = 20.0, q_{R_3} = 5.0$, $R_{1_{initial}} = 1.0$, $R_{2_{initial}} = 1, R_{3_{initial}} = 0.0$, $\varepsilon = 10^{-5}, \eta = 10^{-1}$, and

$$k_{on} = \begin{bmatrix} .2 & .19 & .2 \\ \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon \end{bmatrix}, \qquad k_{off} = \begin{bmatrix} .1 & .1 & .1 \\ \eta & \eta & \eta \\ \eta & \eta & \eta \end{bmatrix}, \qquad k_{cat} = \begin{bmatrix} .1 & .08 & .1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$K_M = \begin{bmatrix} 1.0 & .947 & 1.0 \\ \eta/\varepsilon & \eta/\varepsilon & \eta/\varepsilon \\ \eta/\varepsilon & \eta/\varepsilon & \eta/\varepsilon \end{bmatrix}$$

4.4 Model 3: the emergence of hyperparasitism

Parasitism is a chain process, molecular parasites of even smaller size than the first generation parasites are bound to emerge over time [37,38]. The model stipulates that the ribozyme polymerase R_1 replicates host templates R_1 faithfully, the enzyme-template complexes $R_1\_R_1$ and $R_1\_R_2$ only catalyze R_2 and R_1, respectively. If replicase R_1 binds to a parasite template $R_1\_R_3$, it not only produces another copy of R_3, or a negative strand of R_3, respectively, which we again for simplicity do not distinguish here, and also another, smaller RNA molecule $R_4 = F$, which we call here the hyperparasite, F being smaller than R_3, but partially overlapping with it.

The hyperparasite cycle is triggered from the following process

$$R_1 \Rightarrow (R_1\_P) \rightarrow^c F \quad (1\&2 \rightarrow 3)$$

and has the following form:

$$F \Rightarrow (P\_F) \rightarrow^c F$$

$$F \Rightarrow (F\_F) \rightarrow^c F \quad (3.0)$$

The hyperparasite F binds to an enzyme-template complex P_F or to itself F_F, which implies that F is capable of replicating itself. Again for simplicity, we do not distinguish between positive and negative strand hyperparasites $R_4 = F$, assuming both are present. But since the production of F is already guaranteed through the trigger process $(1\&2 \rightarrow 3)$, such a process may degenerate to produce futile intermediate complexes [62] P_F or F_F:

$$F \Rightarrow (P\_F) \rightarrow^c nil$$

$$F \Rightarrow (F\_F) \rightarrow^c nil \quad (3.0)$$

$F = R_4$ is a hyperparasite, because it parasitizes the parasite P while bound to a complex P_F. A parasite molecule P bound to F is not able at the same time to bind to the ribozyme polymerase R_1, and therefore cannot be replicated anymore. In order to be effective for the host, the hyperparasite requires to have binding specificity for parasite molecules, $R_3 = P$: F should form intermediate enzyme-template complexes only with $R_3 = P$ molecules or with itself.

Now we are able to write down the ODE equations for Model 3, the host-parasite-hyperparasite model.

Model 3

$$\frac{dR_1}{dt} = (1-\mu)k_{cat}^{1,2}\, R_1\_R_2\, \Pi - d_{R_1} R_1 \qquad (3.1)$$

$$\frac{dR_2}{dt} = (1-\mu)k_{cat}^{1,1}\, R_1\_R_1\Pi - d_{R_2}R_2 \qquad (3.2),$$

$$\frac{dR_3}{dt} = (k_{cat}^{1,3}R_1\_R_3 + \mu(k_{cat}^{1,2}\, R_1\_R_2 + k_{cat}^{1,1}\, R_1\_R_1))\Pi + \alpha - d_{R_3}R_3 \qquad (3.3)$$

476 $$\frac{dR_4}{dt} = \beta\, k_{cat}^{1,3}R_1\_R_3\Pi - d_{R_4}R_4 \quad (3.4)$$

477

478 The habitat restriction parameter extends here to

$$\Pi = \left(1 - \frac{R_1 + R_2 + {R_3}/{q_{R_3}} + {R_4}/{q_{R_4}}}{K}\right)$$

479 Here we again assign values accordingly to the ones in Model 1 and 2: $d_{R_1} = d_{R_2} = d_{R_3} = d_{R_4} =$
480 0.005, with $q_{R_3} = 5.0, q_{R_4} = 10.0,$ $K = 20.0,\ R_{1_{initial}} = 1.0,\ R_{2_{initial}} = 1, R_{3_{initial}} = 0.0,$
481 $R_{4_{initial}} = 0.0$, mutation rate $\mu = .01$, parasite inflow per unit of time, $\alpha = 0.001$, $\beta = 1.5$ (β is
482 here a parameter quantizing the trigger process $(1\&2 \rightarrow 3)$), $\varepsilon = 10^{-5}, \eta = 10^{-1}$, and

483 $$k_{on} = \begin{bmatrix} .2 & .19 & .2 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & .1 \\ \varepsilon & \varepsilon & \varepsilon & .1 \end{bmatrix}, k_{off} = \begin{bmatrix} .1 & .1 & .1 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & .1 \\ \varepsilon & \varepsilon & \varepsilon & .1 \end{bmatrix}, k_{cat} = \begin{bmatrix} .1 & .08 & .1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

484 $$K_M = \begin{bmatrix} 1.0 & .947 & 1.0 & \eta/\varepsilon \\ \eta/\varepsilon & \eta/\varepsilon & \eta/\varepsilon & \eta/\varepsilon \\ \eta/\varepsilon & \eta/\varepsilon & \eta/\varepsilon & 1.0 \\ \eta/\varepsilon & \eta/\varepsilon & \eta/\varepsilon & 1.0 \end{bmatrix}$$

485 The values for Model 3 are exactly the same as in Model 2, except for the parameters affecting the
486 hyperparasite $R_4 = F$.

487

488 ***4.5 Model 3.1, µM-range Michaelis constants for P_F and F_F complexes***

489 Here we chose the following parameters for the binding of the hyperparasite $R_4 = F$ with $R_3 = P$ or
490 with itself: $k_{on}^{3,4} = k_{on}^{4,4} = .1,\ k_{off}^{3,4} = k_{off}^{4,4} = .1$, and hence the Michaelis constants $K_M^{3,4} = K_M^{4,4} =$
491 1.0.

492

493 ***4.6 Model 3.2, sub-µM-range Michaelis constants for P_F and F_F complexes***

494 Now if the binding of the complexes P_F, F_F is further strengthened, by e.g. increasing $k_{on}^{3,4} =$
495 $k_{on}^{4,4} = 1.0$, and so lowering the Michelis constants $K_M^{3,4} = K_M^{4,4} = 0.1$, the situation for the host cycle
496 can even be further stabilized, with a lower population of parasites P and less molecules bound to
497 the P_F and F_F complexes. Thus, if the hyperparasite F does not or only very weakly interact with
498 the larger host ribozyme polymerases R_1 and R_2, but does bind with higher affinity to parasite
499 molecules P and also to itself, then it is offering solid stability to the host cycle in the presence of
500 parasites P.

Acknowledgements

We thank Thomas Curran for discussions.

References

1. Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nat Ecol Evol. 2018;2: 1556-1562. Doi: 10.1038/s41559-018-0644-x. PMID:30127539
2. Jeancolas C, Malaterre C, Nghe P. Thresholds in Origin of Life Scenarios. iScience. 2020;23: 101756. Doi: 10.1016/j.isci.2020.101756. PMID:33241201
3. Eigen M. Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften. 1971;58: 465-523. PMID:4942363
4. Joyce GF, Szostak JW. Protocells and RNA Self-Replication. Cold Spring Harb Perspect Biol. 2018;10: a034801. Doi: 10.1101/cshperspect.a034801. PMID:30181195
5. Kauffman SA. Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. J Cybernet. 1971;1: 71-96.
6. Kauffman SA. Autocatalytic sets of proteins. J Theor Biol. 1986;119: 1-24. Doi: 10.1016/s0022-5193(86)80047-9. PMID:3713221
7. Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N. Spontaneous network formation among cooperative RNA replicators. Nature. 2012;491: 72-7. Doi: 10.1038/nature11549. PMID:23075853
8. Tjhung KF, Shokhirev MN, Horning DP, Joyce GF. An RNA polymerase ribozyme that synthesizes its own ancestor. Proc Natl Acad Sci U S A. 2020;117: 2906-2913. Doi: 10.1073/pnas.1914282117. PMID:31988127
9. Kakoti A, Joyce GF. RNA Polymerase Ribozyme That Recognizes the Template-Primer Complex through Tertiary Interactions. Biochemistry. 2023;62: 1916-1928. doi: 10.1021/acs.biochem.3c00091. PMID:37256719
10. Papastavrou N, Horning DP, Joyce GF. RNA-catalyzed evolution of catalytic RNA. Proc Natl Acad Sci U S A. 2024;121: e2321592121. doi: 10.1073/pnas.2321592121. PMID:38437533
11. Chen IA. RNA life on the edge of catastrophe. Proc Natl Acad Sci U S A. 2024;121: e2402649121. doi: 10.1073/pnas.2402649121. PMID:38478681

12. Pavlinova P, Lambert CN, Malaterre C, Nghe P. Abiogenesis through gradual evolution of autocatalysis into template-based replication. FEBS Lett. 2023;597: 344-379. Doi: 10.1002/1873-3468.14507. PMID:36203246
13. Steinhauer DA, Domingo E, Holland JJ. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. Gene. 1992;122: 281-8. doi: 10.1016/0378-1119(92)90216-c. PMID:1336756
14. Domingo E, García-Crespo C, Perales C. Historical Perspective on the Discovery of the Quasispecies Concept. Annu Rev Virol. 2021; 8: 51-72. Doi: 10.1146/annurev-virology-091919-105900. PMID:34586874
15. Bull JJ, Meyers LA, Lachmann M. Quasispecies made simple. PloS Comput Biol. 2005;1: e61. Doi: 10.1371/journal.pcbi.0010061. PMID:16322763.
16. Koonin EV, Wolf YI, Katsnelson MI. Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. Biol Direct. 2017;12: 31. Doi: 10.1186/s13062-017-0202-5. PMID:29202832
17. Iranzo J, Puigbò P, Lobkovsky AE, Wolf YI, Koonin EV. Inevitability of Genetic Parasites. Genome Biol Evol. 2016;8: 2856-2869. doi: 10.1093/gbe/evw193. PMID:27503291
18. Szabó P, Scheuring I, Czárán T, Szathmáry E. In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. Nature. 2002;420: 340-3. doi: 10.1038/nature01187. PMID:12447445
19. Könnyu B, Czárán T, Szathmáry E. Prebiotic replicase evolution in a surface-bound metabolic system: parasites as a source of adaptive evolution. BMC Evol Biol. 2008;8: 267. doi: 10.1186/1471-2148-8-267. PMID:18826645
20. Branciamore S, Gallori E, Szathmáry E, Czárán T. The origin of life: chemical evolution of a metabolic system in a mineral honeycomb? J Mol Evol. 2009;69: 458-69. doi: 10.1007/s00239-009-9278-6. PMID:19806387
21. Takeuchi N, Hogeweg P. Multilevel selection in models of prebiotic evolution II: a direct comparison of compartmentalization and spatial self-organization. PLoS Comput Biol. 2009;5: e1000542. doi: 10.1371/journal.pcbi.1000542. PMID:19834556
22. Shay JA, Huynh C, Higgs PG. The origin and spread of a cooperative replicase in a prebiotic chemical system. J Theor Biol. 2015 Jan 7;364:249-59. doi: 10.1016/j.jtbi.2014.09.019. PMID: 25245369.
23. Kim YE, Higgs PG. Co-operation between Polymerases and Nucleotide Synthetases in the RNA World. PLoS Comput Biol. 2016;12: e1005161. doi: 10.1371/journal.pcbi.1005161. PMID:27820829

568 24. Tupper AS, Higgs PG. Error thresholds for RNA replication in the presence of both point
569 mutations and premature termination errors. J Theor Biol. 2017;428: 34-42. doi:
570 10.1016/j.jtbi.2017.05.037. PMID:28606750
571 25. Zintzaras E, Santos M, Szathmáry E. Selfishness versus functional cooperation in a stochastic
572 protocell model. J Theor Biol. 2010;267: 605-13. doi: 10.1016/j.jtbi.2010.09.011.
573 PMID:20837027
574 26. Ma W, Yu C, Zhang W, Zhou P, Hu J. The emergence of ribozymes synthesizing membrane
575 components in RNA-based protocells. Biosystems. 2010;99: 201-9. doi:
576 10.1016/j.biosystems.2009.11.003. PMID:19961895
577 27. Bianconi G, Zhao K, Chen IA, Nowak MA. Selection for replicases in protocells. PLoS Comput Biol.
578 2013;9: e1003051. doi: 10.1371/journal.pcbi.1003051. PMID:23671413
579 28. Shah V, de Bouter J, Pauli Q, Tupper AS, Higgs PG. Survival of RNA Replicators is much Easier in
580 Protocells than in Surface-Based, Spatial Systems. Life (Basel). 2019;9: 65. doi:
581 10.3390/life9030065. PMID:31394866
582 29. Johnson KA. A century of enzyme kinetic analysis, 1913 to 2013. FEBS Lett. 2013;587: 2753-66.
583 doi: 10.1016/j.febslet.2013.07.012. PMID:23850893
584 30. Roussel MR. Heineken, Tsuchiya and Aris on the mathematical status of the pseudo-steady state
585 hypothesis: A classic from volume 1 of Mathematical Biosciences. Math Biosci. 2019;318:
586 108274. doi: 10.1016/j.mbs.2019.108274. PMID:31697965
587 31. Conrad B, Iseli C, Pirovino M. Energy-harnessing problem solving of primordial life: Modeling the
588 emergence of catalytic host-nested parasite life cycles. PloS One. 2023;18: e0281661. Doi:
589 10.1371/journal.pone.0281661. PMID:36972235
590 32. Briggs GE, Haldane JB. A Note on the Kinetics of Enzyme Action. Biochem J. 1925;19: 338-9. doi:
591 10.1042/bj0190338. PMID:16743508
592 33. Stubbs RT, Yadav M, Krishnamurthy R, Springsteen G. A plausible metal-free ancestral analogue
593 of the Krebs cycle composed entirely of α-ketoacids. Nat Chem. 2020;12: 1016-1022. Doi:
594 10.1038/s41557-020-00560-7. PMID:33046840
595 34. Janzen E, Blanco C, Peng H, Kenchel J, Chen IA. Promiscuous Ribozymes and Their Proposed Role
596 in Prebiotic Evolution. Chem Rev. 2020;120: 4879-4897. doi: 10.1021/acs.chemrev.9b00620.
597 PMID:32011135
598 35. Tawfik DS. Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic
599 efficiency? Curr Opin Chem Biol. 2014;21: 73-80. doi: 10.1016/j.cbpa.2014.05.008.
600 PMID:24954689

36. Fitzsimmons WJ, Woods RJ, McCrone JT, Woodman A, Arnold JJ, Yennawar M, Evans R, Cameron CE, Lauring AS. A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. PLoS Biol. 2018;16: e2006459. doi: 10.1371/journal.pbio.2006459. PMID:29953453

37. Furubayashi T, Ueda K, Bansho Y, Motooka D, Nakamura S, Mizuuchi R, Ichihashi N. Emergence and diversification of a host-parasite RNA ecosystem through Darwinian evolution. Elife. 2020;9: e56038. doi: 10.7554/eLife.56038. PMID:32690137

38. Mills DR, Peterson RL, Spiegelman S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. Proc Natl Acad Sci U S A. 1967;58: 217-24. doi: 10.1073/pnas.58.1.217. PMID:5231602

39. Marshall KA, Ellington AD. Molecular parasites that evolve longer genomes. J Mol Evol. 1999;49: 656-63. doi: 10.1007/pl00006586. PMID:10552046

40. Szilágyi A, Könnyű B, Czárán T. Dynamics and stability in prebiotic information integration: an RNA World model from first principles. Sci Rep. 2020;10: 51. doi: 10.1038/s41598-019-56986-8. PMID:31919467

41. Schreiber G, Haran G, Zhou HX. Fundamental aspects of protein-protein association kinetics. Chem Rev. 2009;109: 839-60. doi: 10.1021/cr800373w. PMID:19196002

42. Gleitsman KR, Sengupta RN, Herschlag D. Slow molecular recognition by RNA. RNA. 2017;23: 1745-1753. doi: 10.1261/rna.062026.117. PMID:28971853

43. Lawrence MS, Bartel DP. Processivity of ribozyme-catalyzed RNA polymerization. Biochemistry. 2003;42: 8748-55. doi: 10.1021/bi034228l. PMID:12873135

44. Martin LL, Unrau PJ, Müller UF. RNA synthesis by in vitro selected ribozymes for recreating an RNA world. Life (Basel). 2015;5: 247-68. doi: 10.3390/life5010247. PMID:25610978

45. Lukatsky DB, Shakhnovich BE, Mintseris J, Shakhnovich EI. Structural similarity enhances interaction propensity of proteins. J Mol Biol. 2007;365: 1596-606. doi: 10.1016/j.jmb.2006.11.020. PMID:17141268

46. McRae EKS, Wan CJK, Kristoffersen EL, Hansen K, Gianni E, Gallego I, Curran JF, Attwater J, Holliger P, Andersen ES. Cryo-EM structure and functional landscape of an RNA polymerase ribozyme. Proc Natl Acad Sci U S A. 2024;121: e2313332121. doi: 10.1073/pnas.2313332121. PMID:38207080

47. Qin S, Pang X, Zhou HX. Automated prediction of protein association rate constants. Structure. 2011;19: 1744-51. doi: 10.1016/j.str.2011.10.015. PMID:22153497

48. Koonin EV, Dolja VV, Krupovic M, Kuhn JH. Viruses Defined by the Position of the Virosphere within the Replicator Space. Microbiol Mol Biol Rev. 2021;85: e0019320. doi: 10.1128/MMBR.00193-20. PMID:34468181

635 49. de Sousa JAM, Fillol-Salom A, Penadés JR, Rocha EPC. Identification and characterization of
636 thousands of bacteriophage satellites across bacteria. Nucleic Acids Res. 2023;51: 2759-2777.
637 doi: 10.1093/nar/gkad123. PMID:36869669
638 50. Eppley JM, Biller SJ, Luo E, Burger A, DeLong EF. Marine viral particles reveal an expansive
639 repertoire of phage-parasitizing mobile elements. Proc Natl Acad Sci U S A. 2022;119:
640 e2212722119. doi: 10.1073/pnas.2212722119. PMID:36256808
641 51. Vignuzzi M, López CB. Defective viral genomes are key drivers of the virus-host interaction. Nat
642 Microbiol. 2019;4: 1075-1087. doi: 10.1038/s41564-019-0465-y. PMID:31160826
643 52. Katzourakis A, Aswad A. The origins of giant viruses, virophages and their relatives in host
644 genomes. BMC Biol. 2014;12: 51. doi: 10.1186/s12915-014-0051-y. PMID:25184667
645 53. Ibarra-Chávez R, Brady A, Chen J, Penadés JR, Haag AF. Phage-inducible chromosomal islands
646 promote genetic variability by blocking phage reproduction and protecting transductants from
647 phage lysis. PLoS Genet. 2022;18: e1010146. doi: 10.1371/journal.pgen.1010146.
648 PMID:35344558
649 54. Rousset F, Depardieu F, Miele S, Dowding J, Laval AL, Lieberman E, Garry D, Rocha EPC, Bernheim
650 A, Bikard D. Phages and their satellites encode hotspots of antiviral systems. Cell Host Microbe.
651 2022;30: 740-753.e5. doi: 10.1016/j.chom.2022.02.018. PMID:35316646
652 55. Genoyer E, López CB. The Impact of Defective Viruses on Infection and Immunity. Annu Rev Virol.
653 2019;6: 547-566. doi: 10.1146/annurev-virology-092818-015652. PMID:31082310
654 56. Fillol-Salom A, Miguel-Romero L, Marina A, Chen J, Penadés JR. Beyond the CRISPR-Cas
655 safeguard: PICI-encoded innate immune systems protect bacteria from bacteriophage predation.
656 Curr Opin Microbiol. 2020;56: 52-58. doi: 10.1016/j.mib.2020.06.002. PMID:32653777
657 57. Stauffer Y, Marguerat S, Meylan F, Ucla C, Sutkowski N, Huber B, Pelet T, Conrad B. Interferon-
658 alpha-induced endogenous superantigen. a model linking environment and autoimmunity.
659 Immunity. 2001;15: 591-601. doi: 10.1016/s1074-7613(01)00212-6. PMID:11672541
660 58. Malfavon-Borja R, Feschotte C. Fighting fire with fire: endogenous retrovirus envelopes as
661 restriction factors. J Virol. 2015;89: 4047-50. doi: 10.1128/JVI.03653-14. PMID:25653437
662 59. Pirovino M, Iseli C, Curran J A, Conrad B. Biomathematical model of prebiotic autocatalytic RNA
663 networks shows birth of adaptive immunity from degenerating molecular parasite catalysts.
664 bioRxiv 2023.05.25.542273; doi: https://doi.org/10.1101/2023.05.25.542273
665 60. Toti P, Sbordone L, Sbordone C, Bauer C. Ribozymes: analytical solution of the one-substrate,
666 two-intermediate reversible scheme for enzyme reactions. J Biol Phys. 2006;32: 473-88. doi:
667 10.1007/s10867-006-9030-z. PMID:19669435

61. Morrison JF. Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. Biochim Biophys Acta. 1969;185: 269-86. doi: 10.1016/0005-2744(69)90420-3. PMID:4980133

62. Bar-Even A, Milo R, Noor E, Tawfik DS. The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness. Biochemistry. 2015;54: 4969-77. doi: 10.1021/acs.biochem.5b00621. PMID:26219075

Author contributions

Conception and design of the work: MP and BC. Data acquisition MP; data analysis MP, BC, JAC, CI; interpretation of data MP, BC, JAC, CI. Creation of new software used in the work: MP, CI. Drafted the work and revised it: BC, JAC and MP.

Data availability

The data, including scripts, is fully made available at the following sites (https://github.com/BICC-UNIL-EPFL/biomathematical-enzyme-kinetics-model [github.com], and the DOI: https://doi.org/10.5281/zenodo.11060828 [doi.org].

Competing interests

The authors declare no competing interests.

Figures and Figure Legends

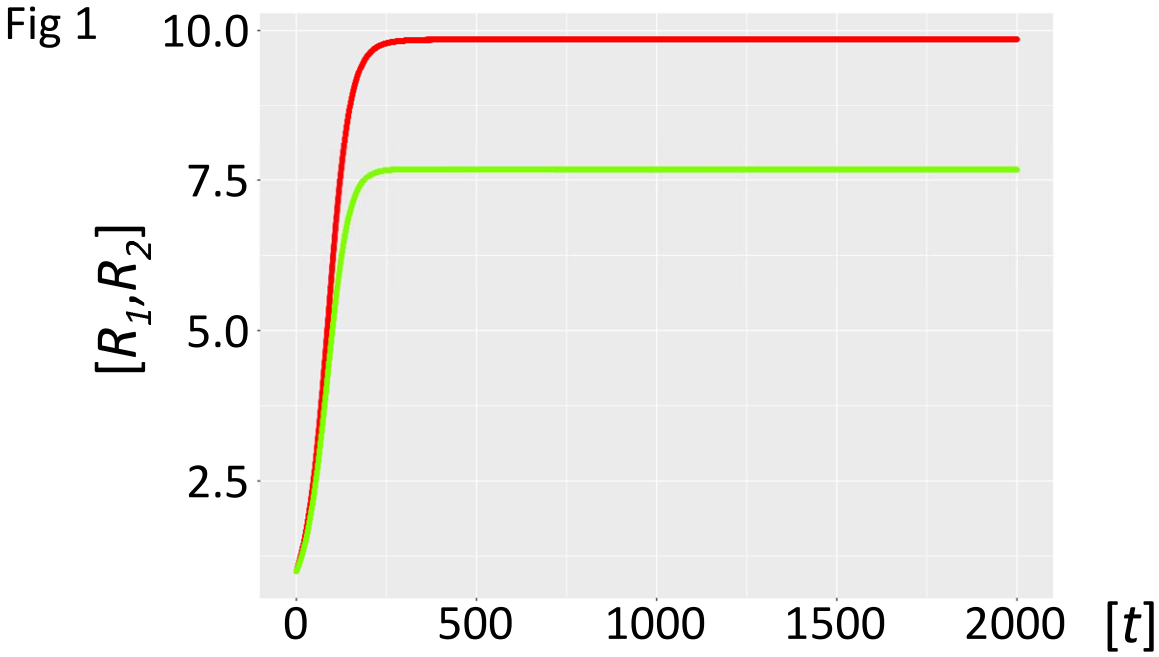


Figure 1. Habitat with autocatalytic host cycle (Model 1)

Fig 1. Ribozyme concentrations of defined populations R_1 and R_2 [units on unspecified scale] are plotted against time t [units on unspecified scale]. The host catalysts (ribozyme polymerase R_1, red, negative strand ribozyme R_2, light green) build a self-amplifying and sustained autocatalytic habitat, as long as sufficient nucleotides are provided.

Fig 2A

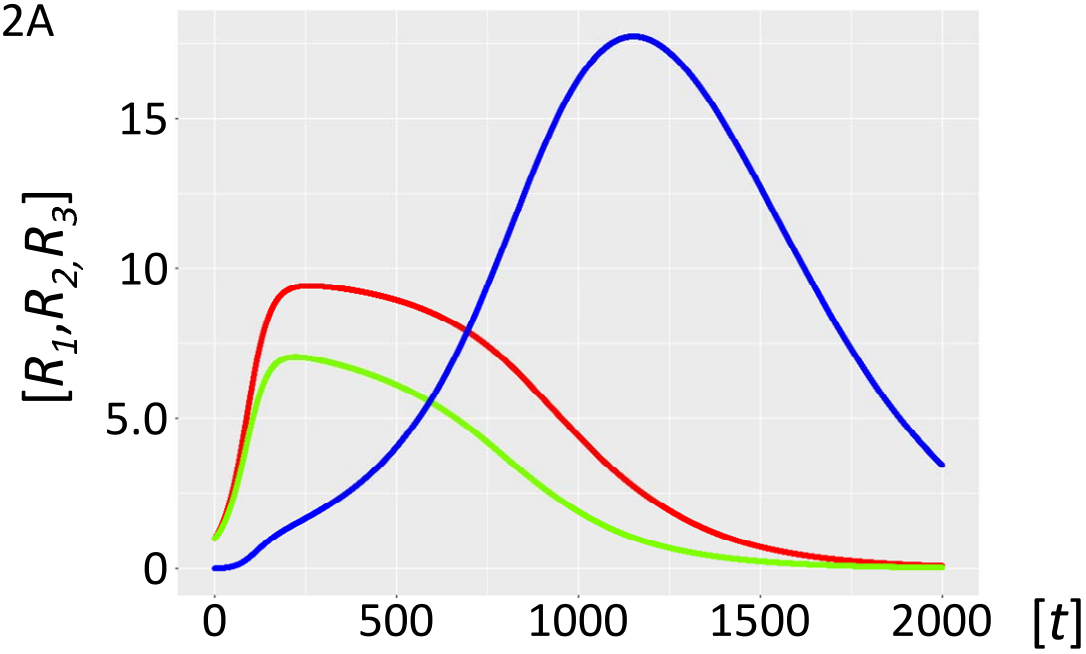
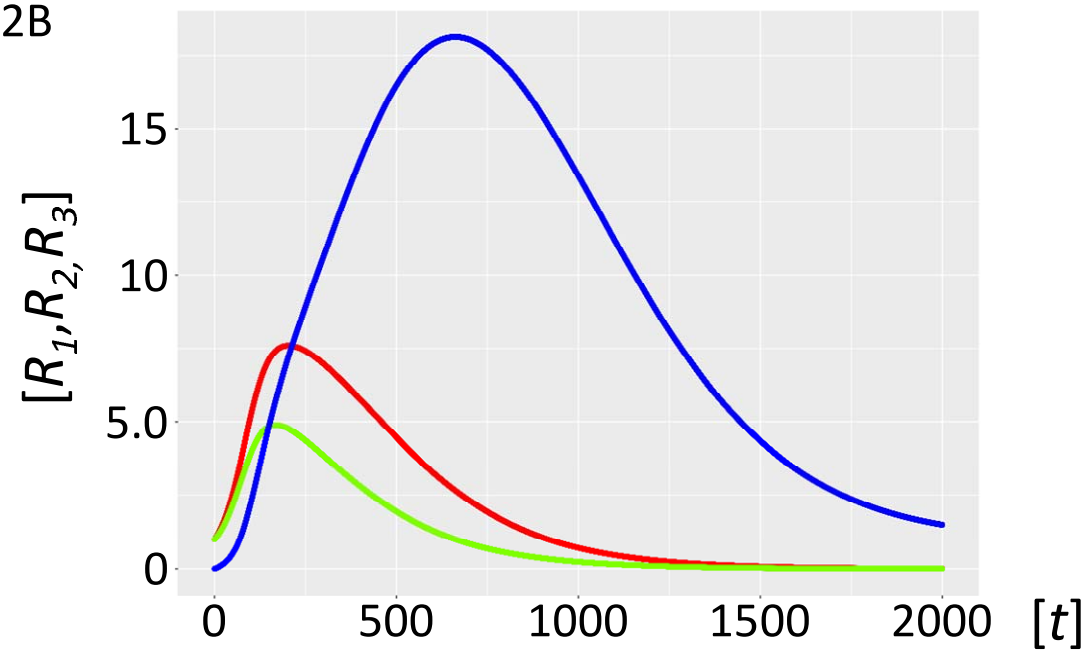


Figure 2A. Autocatalytic habitat with molecular parasite exposure (Model 2.1)

Fig 2A. Ribozyme concentrations of defined populations R_1, R_2 and R_3 [units on unspecified scale] are plotted against time t [units on unspecified scale]. The host catalysts (ribozyme polymerase R_1, red, negative strand ribozyme R_2, light green) are rapidly invaded by the parasite ribozyme species R_3 (blue), which is more efficiently amplified by the host ribozyme polymerase R_1.

765

Fig 2B



766
767

768 **Figure 2B. Autocatalytic habitat with molecular parasite inflow (Model 2.2)**

769 Fig 2B. Ribozyme concentrations of defined populations R_1, R_2 and R_3 [units on unspecified scale] are
770 plotted against time t [units on unspecified scale]. If parasite ribozymes R_3 (blue) diffuse in from a
771 neighbouring habitat, the host catalysts (ribozyme polymerase R_1, red), and the negative strand
772 ribozyme R_2 (light green) are even more rapidly outcompeted and disappear.

773
774
775
776
777
778
779
780
781
782
783
784
785
786

Fig 3A

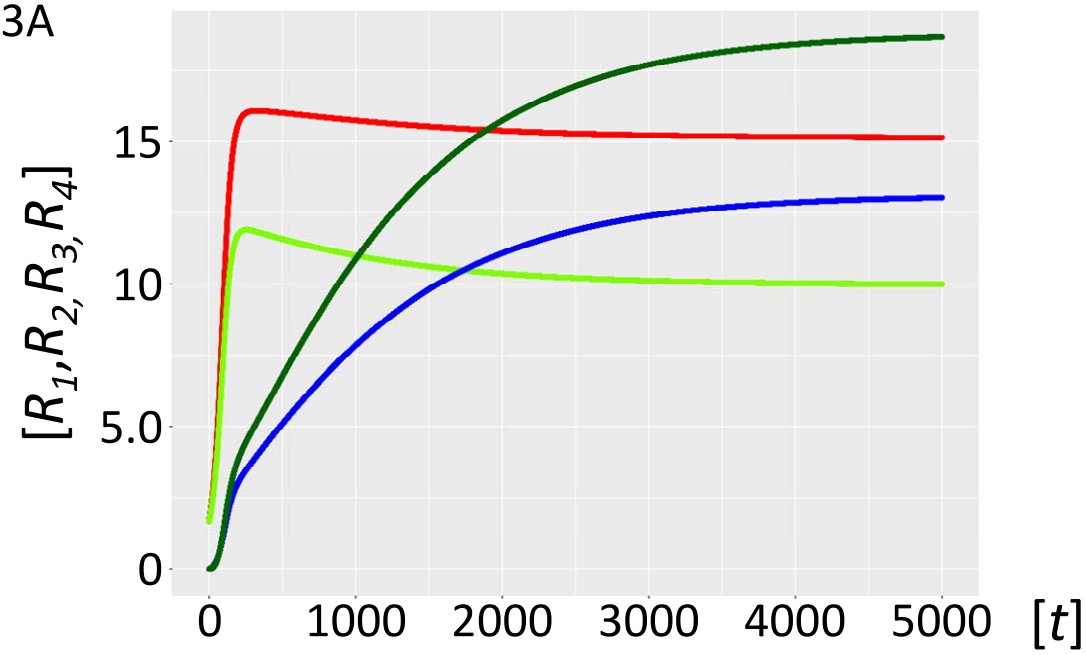


Figure 3A. Stabilized host-parasite-hyperparasite cycle with K_M in µM range

Fig 3A. Ribozyme concentrations of defined populations R_1, R_2, R_3 and R_4 [units on unspecified scale] are plotted against time t [units on unspecified scale]. A habitat with highly efficient catalysts (ribozyme polymerase R_1, red, negative strand ribozyme R_2, light green) becomes stabilized by the generation and persistence of hyperparasites (parasite of parasite) F (dark-green) that parasitize the parasites R_3 (blue). This already happens if intermediate complexes P_F and F_F have a K_M in the µM range (Model 3.1). However, the hyperparasite (dark-green) and parasite R_3 (blue) ribozymes remain prevalent under these conditions.

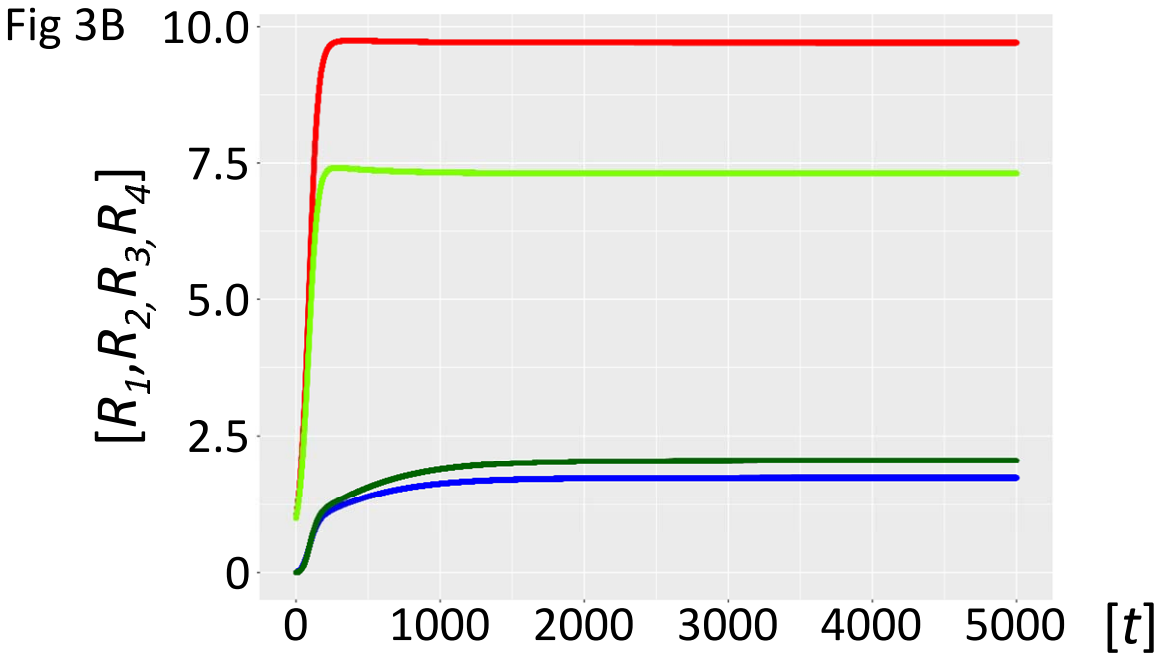


Figure 3B. Stabilized host-parasite-hyperparasite cycle with K_M in 0.1 µM range

Fig 3B. Ribozyme concentrations of defined populations R_1, R_2, R_3 and R_4 [units on unspecified scale] are plotted against time t [units on unspecified scale]. A habitat with highly efficient catalysts (ribozyme polymerase R_1, red, negative strand ribozyme R_2, light green) becomes fully stabilized in the presence of parasite-hyperparasite cycles, if the K_M of the intermediate complexes *P_F* and *F_F* is decreased one order of magnitude in the 0.1 µM range (Model 3.2). Parasitism now is completely tamed (*P* blue, *F* dark green).