Emergence of catalytic function in prebiotic information-coding polymers

Alexei V. Tkachenko
Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY 11973, USA

Sergei Maslov
Department of Bioengineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA
Department of Physics, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA and
Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

Life as we know it relies on the interplay between catalytic activity and information processing carried out by biological polymers. Here we present a plausible pathway by which a pool of prebiotic information-coding oligomers could acquire an early catalytic function, namely sequence-specific cleavage activity. Starting with a system capable of non-enzymatic templated replication, we demonstrate that even non-catalyzed spontaneous cleavage would promote proliferation by generating short fragments that act as primers. Furthermore, we show that catalytic cleavage function can naturally emerge and proliferate in this system. Specifically, a cooperative catalytic network with four subpopulations of oligomers is selected by the evolution in competition with chains lacking catalytic activity. The cooperative system emerges through the functional differentiation of oligomers into catalysts and their substrates. The model is inspired by the structure of the hammerhead RNA enzyme as well as other DNA- and RNA-based enzymes with cleavage activity that readily emerge through natural or artificial selection. We identify the conditions necessary for the emergence of the cooperative catalytic network. In particular, we show that it requires the catalytic rate enhancement over the spontaneous cleavage rate to be at least $10^2 - 10^3$, a factor consistent with the existing experiments. The evolutionary pressure leads to a further increase in catalytic efficiency. The presented mechanism provides an escape route from a relatively simple pairwise replication of oligomers towards a more complex behavior involving catalytic function. This provides a bridge between the information-first origin of life scenarios and the paradigm of autocatalytic sets and hypercycles, albeit based on cleavage rather than synthesis of reactants.

I. INTRODUCTION

One of the most intriguing mysteries in science is the origin of life. Despite extensive research in this area, we are still far from understanding how life has emerged on Earth. One promising hypothesis is the RNA world theory [1–6], inspired by the discovery of ribozymes [7], i.e. RNA molecules capable of enzymatic activity. According to this hypothesis, all processes in early life were carried out by the RNA, which was used both to store information and to catalyze biochemical reactions. In particular, specific ribozymes could have catalyzed the self-replication of arbitrary RNA sequences, a function currently performed by specialized protein-based enzymes. However, based on the results of the existing experiments [8, 9], such a catalytic function requires rather long and carefully designed RNA sequences, which are highly unlikely to arise spontaneously. In contrast, one of the simplest catalytic activities of ribozymes is their ability to cleave an RNA sequence at a specific site. Indeed, such ribozymes independently evolved in multiple branches of life [10] and have been shown to emerge rapidly and repeatedly from artificial selection [11, 12]. DNA molecules have also been shown to be capable of site-specific cleavage targeting either RNA [13] or DNA sequences [14, 15].

In this paper, we consider a population of information-carrying polymers capable of templated non-enzymatic replication [16, 17]. This may have been the state of the proto-RNA world before the emergence of ribozymes. This could involve heteropolymers chemically distinct from present-day RNA [17] and/or inorganic catalysts such as, e.g. mineral surfaces [18, 19]. We demonstrate that (i) even spontaneous cleavage promotes replication by generating short fragments used as primers for templated growth; (ii) catalytic cleavage activity naturally emerges in these populations and gets selected by the evolution; (iii) this mechanism is robust with respect to replication errors.

In a series of previous studies, we have shown that non-enzymatic templated replication can lead to the formation of longer chains [20] as well as to a reduction in sequence entropy [21]. Such a reduction in entropy has subsequently been observed experimentally for templated ligation of DNA oligomers [22]. This selection in sequence space can be seen as a first step towards Darwinian evolution. However, this does not necessarily imply the emergence of a catalytic function. In this paper, we build on these findings and further investigate the potential for the evolution of catalytic activity in the proto-RNA world.

II. MODEL AND RESULTS

In our model, we consider the population dynamics of a pool of heteropolymers analogous to the familiar nucleic acids (RNA or DNA) but capable of enzyme-free templated polymerization. The basic processes in this

scenario are similar to those in Polymerase Chain Reaction (PCR), where the system is driven out of equilibrium by cyclic changes in the environment (e.g. temperature), which we refer to as "night" and "day" phases. During the night phase, heteropolymers hybridize with each other following Watson-Crick-like complementarity rules. If the terminus of one chain hybridizes with the middle section of another chain the former can be gradually elongated by the virtue of non-enzymatic templated polymerization [23–26]. We will refer to this type of hybridization as "end-to-middle", the former chain as the "primer" and the latter chain as the "template". During the day phase, the hybridized structures melt and all the heteropolymers separate from each other. During the next night, they hybridize with new partners, providing them with the opportunity to elongate further. Unlike the classical PCR process, we assume that the polymerization in our system occurs without any assistance from enzymes and may proceed in either direction along the chain. Equivalently, instead or polymerization, the elongation could rely on ligation with very short chain segments which is naturally bi-directional. The elongation of primers naturally leads to the copying of information from the template's sequence. The obvious limiting factor for this process is the availability of primers and the likelihood of end-to-middle hybridization resulting in elongation. The key observation behind our model is that the breakup (cleavage) of a chain creates a new pair of potential primers. Each of them could be elongated during subsequent nights. Thus, somewhat counter-intuitively, breakup of chains results in their proliferation.

Our previous theoretical [20, 21] and experimental [22] results demonstrated that templated-assisted replication of heteropolymers has a generic tendency to substantially reduce their sequence entropy. Such reduction has important consequences in the context of the current work: it significantly increases the likelihood of end-to-middle hybridization of chains during the night phase. That in turn creates an evolutionary pressure to further decrease sequence entropy. A detailed study of this fascinating mechanism falls beyond the scope of the current study. However, below we will assume the logical end of this dynamics where the pool of sequences is composed of fragments of one or several nearly non-overlapping master sequences and their complementaries. Note that for any two overlapping chains, such that the sequence of the first one is a fragment of the master sequence while the other is a fragment of the complementary master sequence, the end-to-middle binding is essentially guaranteed. The exception to this rule is when both chains terminate at the same points so that they are exact complements of each other.

A. Random cleavage model

Based on the argument presented above, we focus on the case of a system populated with chains that are fragments of a single master sequence or its complement. We denote the total concentration of fragments of the master sequence as $c(t)$, while the concentration of all fragments of the complementary sequence as $\bar{c}(t)$. Our system operates in a chemostat, i.e. a reservoir constantly supplied with fresh monomers at the concentration m_0 and diluted at the rate δ. Let $M(t)$ (respectively $\bar{M}(t)$) be the concentration of all monomers incorporated into chains of the subpopulation $c(t)$ (respectively $\bar{c}(t)$). The concentration of free monomers not incorporated into any chains is given by:

$$m = m_0 - M - \bar{M} \tag{1}$$

It is convenient to introduce a minimal length l_0 of a chain that would hybridize with its complementary partner during the night phase, and use it as the unit of chain length, instead of a single monomer. In effect, this leads to the renormalization of all monomer concentrations as $m = [m]/l_0$, where $[m]$ is the conventional molarity. M, $\bar{M}$ and m_0 are similarly renormalized, while the polymer concentration remains unmodified: $c = [c]$. In what follows, we renormalize all lengths and concentrations so that $l_0 = 1$.

We assume that nights are sufficiently long and that the hybridization rate is sufficiently fast so that early on during each night phase, most chains find partners with a complementary overlap. That assumes that the total concentration of all master sequence fragments, c is lower than the total concentration of all fragments in the complementary subpopulation, $\bar{c}$. If this is not the case, i.e. if $c > \bar{c}$, only a fraction $\bar{c}/c$ of all fragments of the master sequence find a partner, while the rest of the c-subpopulation remains unpaired and thus does not elongate. For simplicity, our model neglects the possibility of the formation of hybridized complexes involving more than two chains.

It is well recognized that self-replication based on templated polymerization or ligation is vulnerable to product inhibition, i.e. to the re-hybridization of the products that are meant to act as templates [16, 27]. Indeed, since full-length templates and their complements would bind each other stronger than any shorter fragment, the primers would typically be displaced from the templates by longer chains. In Appendix A we analyze the binding kinetics during the night phase and come to a rather surprising conclusion that despite the strand displacement, a finite and substantial fraction of all the primers will stay hybridized with their respective templates throughout the long enough night phase.

As discussed earlier, two hybridized chains whose sequences are fragments of the master sequence and its complementary would typically have two ends that undergo templated growth at a certain rate proportional to monomer concentration m (see Figure 1). The exception to this rule is when these two hybridized chains terminate at exactly the same point. In our model, the average primer elongation rate is $r \cdot m(t)$. The rate parameter r accounts for the finite probability of primer

binding to a template during the night phase as well as for a finite night-to-day ratio. Note that the value of r does not change after our renormalization $l_0 \to 1$). The $M(t)$ dynamics is thus given by

$$\frac{dM}{dt} = rm \cdot \min(c, \bar{c}) - \delta \cdot M \tag{2}$$

The equation for the complementary subpopulation is obtained by replacing M with $\bar{M}$. Note that m includes only monomers, and therefore is not subdivided into two complementary subpopulations.

We assume that the breakup of chains is completely random and happens at a constant rate β_0 at any internal bond along the chain. The concentration of these breakable bonds is given by M. Because of our choice of the unit length l_0, $\beta_0 = l_0\beta_0'$, where β_0' is the cleavage rate of a single bond. Since each fragmentation of a chain creates one new primer, the equation governing the overall concentration of chain fragments in the subpopulation c is given by:

$$\frac{dc}{dt} = \beta_0 M - \delta c \tag{3}$$

Once again, the equation for the complementary subpopulation is obtained by replacing c with $\bar{c}$ and M with $\bar{M}$.

Combining equations (2) and (3) we observe that the steady state is a symmetric mixture $c = \bar{c}$. The average length (in units of l_0) of all chain fragments in the subpopulation c is given by:

$$\langle L \rangle = \frac{M}{c} = \frac{\delta}{\beta_0} = \frac{rm^*}{\delta}. \tag{4}$$

This in turn determines the steady state concentrations:

$$m^* = \frac{\delta^2}{\beta_0 r} \tag{5}$$

$$c = \bar{c} = \frac{1}{2}\left(\frac{\beta_0}{\delta} m_0 - \frac{\delta}{r}\right) \tag{6}$$

To obtain Eq. (6) we combined Eqs. (4) and (1) Note that the mutually templating chains survive only when the concentration of free monomers supplied to the system m_0 exceeds m^*.

B. Model with catalyzed cleavage

In the model considered above, the random breakage of chains led to their proliferation. It is therefore reasonable to expect that the ability of a heteropolymer to catalyze cleavage would be selected by the evolution. Incidentally, some of the simplest known RNA-based enzymes (ribozymes) have exactly this function [7, 10–12, 28–31].

Here we consider a simple model in which a heteropolymer capable of catalyzing cleavage spontaneously emerges from a pool of mutually-templating chains. Our model is inspired by the real-world examples of naturally occurring hammerhead ribozyme [28, 29] as well as artificially selected DNA-cleaving DNA enzymes [15]. The minimal structure of the hammerhead ribozyme [30–32] consists of a core region of 15 (mostly) invariant nucleotides flanked by three helical stems formed by mutually complementary RNA sequences. The cleavage happens at a specific site of this structure, located immediately adjacent to one of these stems. While a classical hammerhead ribozyme consists of a single RNA chain capable of self-cleavage, the same structure could be assembled from two chains, one (labeled b in Figure 1) containing a hairpin and capable of hybridizing with and subsequently cleaving the other chain (labeled a in Figure 1). Furthermore, such two-chain structures are realized in certain DNA-cleaving DNA enzymes [15].

We consider a scenario in which a master sequence spontaneously emerges from a random pool and subsequently diverges into two closely related subpopulations a and b and their respective complementarities $\bar{a}$ and $\bar{b}$. The sequences of a and b are mostly identical except for a short insert in the chains b and $\bar{b}$, rendering them catalytically active. That is to say, when $\bar{b}$ is bound to a chain from the subpopulation a, it induces a cleavage at a specific site of that chain. Inspired by the hammerhead ribozyme we assume that the cleavage site in a is immediately adjacent to the start of the catalytic insert in b (see Figure 1).

We further assume that b-chains are capable of cleaving $\bar{a}$-chains at the same site. This symmetry most likely does not apply to highly-optimized hammerhead ribozymes, but it is reasonable in enzymes with only a modest level of catalytic efficiency. Let the cleavage within $\bar{b}$-a or b-$\bar{a}$ duplexes occur at rates β and $\bar{\beta}$ respectively.

The cleavage of a chain of type a by $\bar{b}$ produces two pieces, right and left. The concentrations of these fragments are referred to as a_R and a_L respectively. Inspired by the example of the hammerhead ribozyme (illustrated in Fig. 1B) we assume that one of them (a_L) can serve as a primer only for a. Conversely, the other fragment (a_R) can serve as a primer for either a or b depending on the first templating chain with which it will hybridize (respectively $\bar{a}$ or $\bar{b}$). Assuming a random first encounter, the probability for a_R to serve as a primer to b is thus given by $\phi = \bar{b}/(\bar{a} + \bar{b}) < 1$, while the probability of it to grow into a is $1 - \phi = \bar{a}/(\bar{a} + \bar{b})$. Similarly, conversion probabilities in the complementary subpopulation $\bar{a}_R$ are determined by $\bar{\phi} = b/(a + b)$

As in the random cleavage model, the concentration of hybridized duplexes is given by the smaller of two concentrations $a+b$ and $\bar{a}+\bar{b}$. This is captured by the factor $\chi = \min(a + b, \bar{a} + \bar{b})/(a + b) \leq 1$ in the elongation rate of a primer: $rm\chi$. Here, as before, this rate is measured in l_0 bases per unit time, and m is the free monomer concentration.

We observe that in order for a cleavage fragment to work as a primer, it needs to exceed the minimal primer

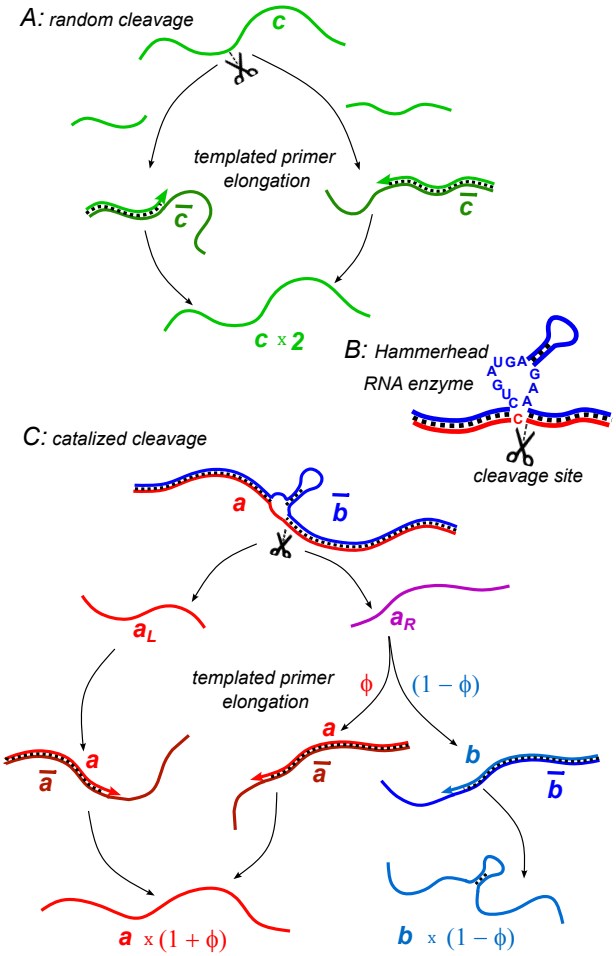

FIG. 1. (A) Random cleavage model. A random break in a chain of type c generates two primer fragments, which are elongated to give rise to two chains of type c. Elongation requires a complementary template of type $\bar{c}$. (B) An example of catalyzed cleavage given by hammerhead ribozyme [32]. Note that the right cleavage fragment is perfectly complementary to the blue sequence, while the left one contains an extra non-complementary base C. (C) Catalyzed cleavage model. A cleavage of the red chain a catalyzed by the blue chain $\bar{b}$ gives rise to two primers a_L (red) and a_R (purple). Because of an extra non-complementary base (see Panel B) the a_L primer can only elongate to a, while the a_R primer - to either a or b depending on its first hybridization partner. Similar processes involving complementary chains $\bar{a}$ and $\bar{b}$ (not shown) result in the replication of templates.

length l_0. Therefore, a newly formed cleavage product a_L needs to grow by at least l_0 bases before it can be considered a part of the a subpopulation. Indeed, if it has not grown by that length, another catalyzed cleavage at the same site would not increase the number of primers in the system. Therefore, the rate at which chains in the subpopulation a_L are converted to a is given by $rm\chi$. The rate of conversion of a_R to a is similar, up to the factor $1-\phi$ discussed above: $rm\chi(1-\phi)$. If a segment a_R was first hybridized to $\bar{b}$, it will eventually grow to be a part of the subpopulation b. However, in order to become functional, this chain has to grow at least by the length of the catalytic insert, which is distinct from l_0. Furthermore, the rate of elongation of b is slowed down by the presence of a hairpin in the catalytic domain of the $\bar{b}$ structure. Both effects can be captured by a factor λ in the conversion rate from a_R to b given by $rm\chi\phi/\lambda$, relative to that of a_R to a given by $rm\chi(1-\phi)$.

Altogether, the dynamics of our model is described by the following equations

$$\dot{c} = \beta_0 M + \beta\phi a - \delta c \tag{7}$$

$$\dot{b} = \beta_0 M\bar{\phi} + \frac{rm\chi\phi}{\lambda} a_R - \delta b \tag{8}$$

$$\dot{a}_L = \beta\phi a - rm\chi a_L - \delta a_L \tag{9}$$

$$\dot{a}_R = \beta\phi a - rm\chi(1-\phi)a_R - \frac{rm\chi\phi}{\lambda} a_R - \delta a_R \tag{10}$$

Here $c \equiv a+b+a_l+a_R$. Note that Eq. (7) is obtained by first writing the kinetic equation for a and then adding it up to the sum of Eqs. (8-10). The first terms in the r.h.s of Eqs. (7-8) represent random non-catalyzed cleavage that occurs at rate β_0 at any location of any chain (compare to Eq. 3). Similarly to Eq. (2) the dynamics of the number density M of monomers incorporated into chains a, a_L, a_R, and b is described by

$$\dot{M} = rm\chi c - \delta M. \tag{11}$$

This in turn determines the concentration of free monomers remaining in the solution as given by Eq. (1).

Additional equations are obtained by replacing variables a, a_L, a_R, b, and M with their complementary counterparts: $\bar{a}$, $\bar{a_L}$, $\bar{a_R}$, $\bar{b}$, and $\bar{M}$, respectively. In addition, β, ϕ, and χ should be replaced with $\bar{\beta}$, $\bar{\phi}$, and $\bar{\chi} = \min(a+b, \bar{a}+\bar{b})/(\bar{a}+\bar{b})$, respectively.

C. System dynamics

A crucial feature of this multi-component system is that catalytic cleavage depends on the cooperativity between all four subpopulations a, $\bar{a}$, b, and $\bar{b}$. To understand when such a cooperative steady state exists, we have numerically solved Eqs. (7-11) for $\beta_0 = 0.015$, $\lambda = 2$, $\delta = 1$ and different values of the catalytic cleavage rate β. Figures 2A-C show the dynamic trajectories of our system in a-b space for a wide range of initial conditions. For β too small (e.g. $\beta = 6$ in Fig. 2A) or too large (e.g. $\beta = 18$ in Fig. 2C), the only steady state solutions correspond to the survival of either the $a/\bar{a}$ or $b/\bar{b}$ subpopulation. In these non-cooperative fixed points, marked with red and blue stars in Fig. 2, one set of chains drives the other to extinction. Thus, they propagate by random rather than catalytic cleavage. The concentrations of monomers and a complementary pair of surviving chains at a non-cooperative fixed point are given by (5) and (6), respectively.

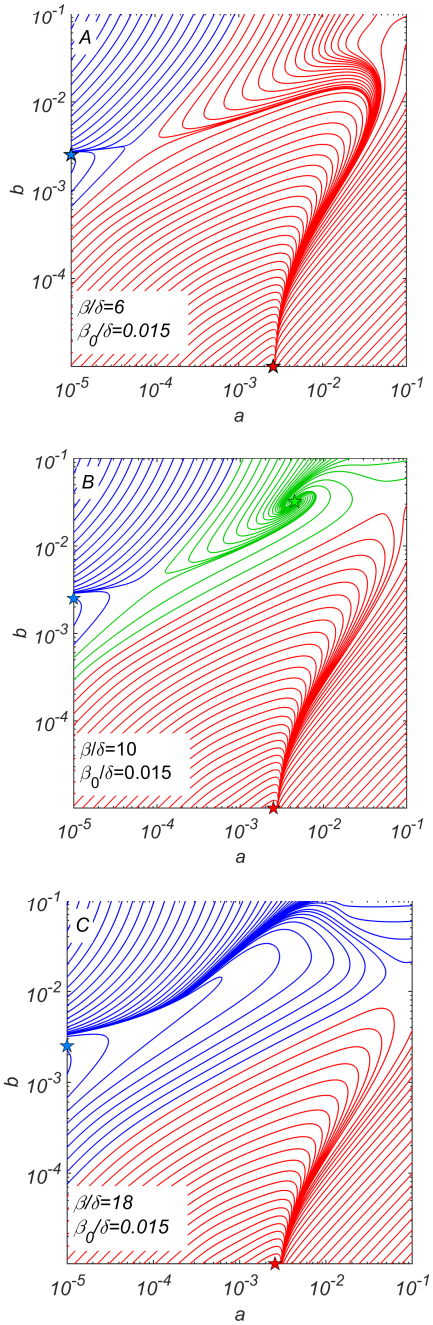

FIG. 2. Dynamical phase portraits for different catalytic cleavage rates β. (A) The phase portrait for a small catalytic cleavage rate $\beta = 6$ has two non-cooperative steady state solutions marked with red and blue stars corresponding to pure $a/\bar{a}$ and pure $b/\bar{b}$ subpopulations respectively. These solutions are maintained by random rather than catalytic cleavage. (B) The phase portrait for intermediate catalytic cleavage rate $\beta = 10$ in addition to two non-cooperative steady states marked with red and blue stars has a cooperative steady state marked with the green star in which all four subpopulations coexist. One can reach this state e.g. starting from the non-cooperative $b/\bar{b}$ steady state (the blue star) and adding a relatively small subpopulation of $a/\bar{a} > 2e-5$ crossing the saddle point separating blue and green trajectories. (C) The phase portrait for a large catalytic cleavage rate $\beta = 18$ again has only two non-cooperative cleavage steady states marked with red and blue stars. All three panels were obtained by numerically solving dynamical equations (7-11) with random cleavage rate $\beta_0 = 0.015$, elongation asymmetry factor $\lambda = 2$, and dilution factor $\delta = 1$.

For an intermediate value of β (e.g. for $\beta = 10$ shown in Fig. 2B) we observe the emergence of a new cooperative fixed point (marked by the green star), where all four subpopulations survive at concentrations $a = \bar{a} > 0$ and $b = \bar{b} > 0$. This fixed point is mainly maintained by catalytic cleavage.

The phase portrait of our system shown in Fig. 2B suggests a plausible scenario for the emergence and subsequent evolution of this cooperative system. This evolutionary scenario starts with $b/\bar{b}$ subpopulations existing alone. Eventually, copying errors might result in the emergence of a small subpopulation of a or $\bar{a}$ at a concentration $a \ll b$. If this concentration is greater than a certain threshold separating blue and green trajectories in Fig. 2B, the dynamics of the system would drive it to the cooperative fixed point (green star in Fig. 2B). For specific parameters used to generate Fig. 2B, the minimal ratio between concentrations a and b is roughly 0.01.

D. Properties of the cooperative steady state

To better understand the conditions for the existence of the cooperative regime, we analytically derived the steady state solutions of Eqs. (7-11). The key result is the relationship between the steady state monomer concentration, $m^*_{\rm coop}$, and the catalytic cleavage rate β for given values of β_0, λ, δ and r derived in the SI Appendix. Fig. 3A shows this dependence for $\lambda = 2$, and different values of β_0/δ alongside with data points (open circles) obtained by direct numerical solution of dynamical equations (7-11).

The stable fixed point corresponds to the monotonically increasing branch of the graph $m^*_{\rm coop}$ vs β (solid lines in Fig. 3A), while two decreasing branches (dashed lines in Fig. 3A) correspond to two dynamically unstable saddle points separating different steady state solutions (see Fig. 2B).

Increasing the parameter β_0/δ, e.g., by making the dilution rate δ smaller, makes the range of β for which the cooperative solution exists progressively smaller until it altogether disappears above $\beta_0/\delta \approx 0.057$ (for $\lambda = 2$).

Fig. 3B shows the ranges of β/δ and β_0/δ for which the cooperative solution exists (as before, for $\lambda = 2$). Solid lines of different colors correspond to three values of β_0 used in Fig. 3A.

While the full set of our analytical results described in the SI Appendix is rather convoluted, here we present a simplified expression for the range of values of β/δ, where the cooperative solution exists, and the corresponding range of $m^*_{\rm coop}$:

$$4\lambda < \frac{\beta}{\delta} < \frac{(\lambda-1)\delta}{4\beta_0} \tag{12}$$

$$\frac{2\lambda\delta}{(\lambda-1)\beta_0} < \frac{m^*_{\rm coop}}{m^*} < \frac{1}{2} \tag{13}$$

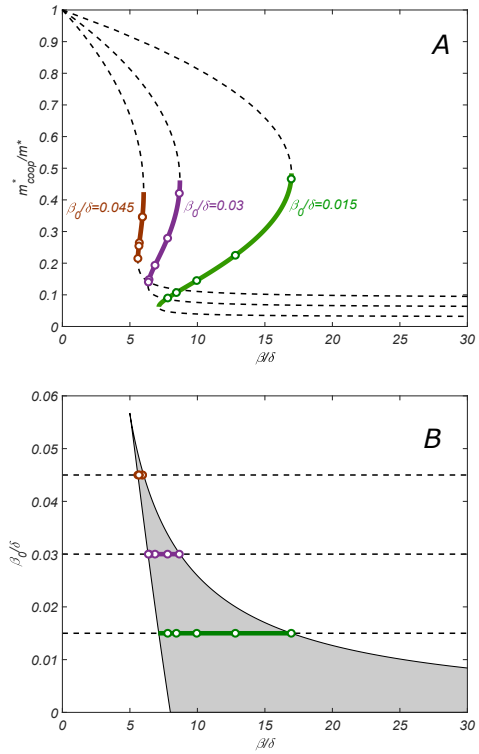

FIG. 3. (A) The relationship between parameters of the cooperative state. m^*_{coop}/m^* plotted vs β for $\lambda = 2$, $\delta = 1$ and increasing values of β_0: 0.015 (green), 0.03 (purple) and 0.045 (red). Lines are given by the parametric equation describing the state and derived in the SI Appendix (Eq. S14), while open circles are obtained by direct numerical solution of dynamical equations (7-11). Monotonically increasing branches (solid lines) correspond to the stable cooperative fixed point, while the decreasing branches (dashed lines) - to the dynamically unstable saddle points separating different steady state solutions in Fig. 2B. (B) Phase diagram of the cooperative state. The shaded region marks the values of β/δ and β_0/δ for which the cooperative solution exists. Green, purple and red lines show the ranges of β for which the cooperative solution exists for the corresponding value of β_0 shown in Panel A. Increasing the parameter β_0 makes the range of β for which the cooperative solution exists progressively smaller until it altogether disappears above $\beta_0/\delta \approx 0.057$.

These conditions were derived in the limit $\beta_0/\delta \ll 1/\lambda$ (see SI Appendix for details).

E. Evolutionary dynamics

The competition for monomers is the main mechanism of natural selection operating in our system. The steady state with the lowest level of monomer concentration is favored by evolution since the competing states would not be able to proliferate at that level of monomers. Eq. (13) implies that the monomer concentration m^*_{coop} in the cooperative solution is always less than a half of its value m^* in the absence of the catalytic cleavage. Thus, once the cooperative state emerges, it drives out all chain sequences that still rely on non-catalytic cleavage for replication. The continuously increasing "fitness" of the system can be quantified, e.g., by the ratio m^*/m^*_{coop}. Note that the non-cooperative solution has a fitness of 1, while the cooperative solution has a fitness higher than 2.

The fitness landscape of our system shown in Fig. 4 depends on two parameters of the catalytic cleavage: β and λ. Its preeminent feature is a relatively narrow fitness ridge (orange and red color in Fig. 4). For a fixed value of λ this ridge corresponds to a sharp fitness maximum located at the lowest possible catalytic cleavage rate β for which the cooperative state still exists (see the red region in Fig. 4B). In other words, for a given λ, the selective pressure would drive β down to the lower boundary separating cooperative and non-cooperative regions.

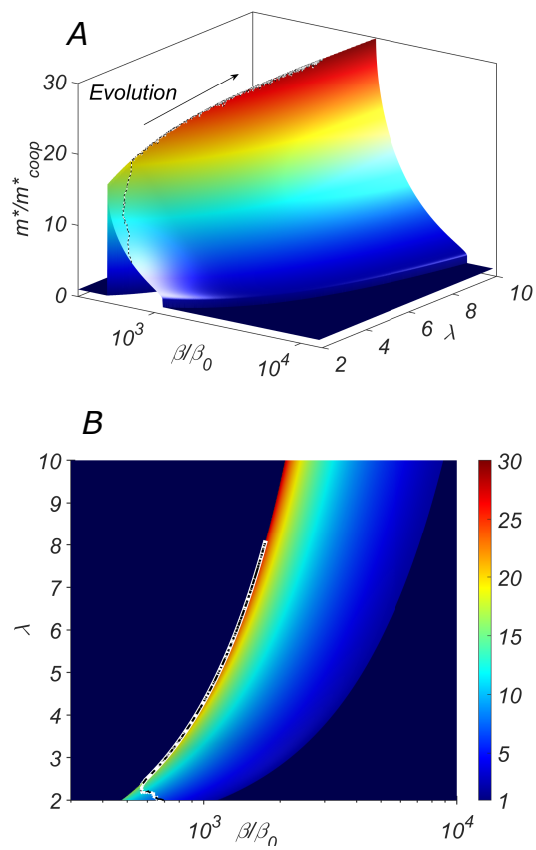

FIG. 4. The fitness landscape of our system. (A) the three-dimensional plot and (B) the heatmap of the fitness m^*/m^*_{coop} of the cooperative state in our system as a function of the catalytic cleavage rate enhancement β/β_0 and elongation asymmetry λ. The co-evolution of λ and β would increase together. A typical evolutionary trajectory in which small changes in these parameters are independent of each other is shown as a dashed line.

However, λ and β are expected to co-evolve together.

Indeed, $\lambda > 1$ quantifies the ratio by which the structural properties of the type-b chain (its excess length, the hairpin unzipping free energy, etc.) slow down its replication compared to that of the type-a chain. Thus, it is reasonable to assume that λ could be easily modified in the course of the evolution. As a consequence of the selective pressure, both λ and β are expected to increase in the course of the evolution driving the system up the ridge in Fig. 4. This intuition is confirmed by a direct numerical simulation in which we model the evolution as a Monte Carlo process with fitness playing the role of negative energy Parameters λ and β were allowed to vary randomly and independently of each other. A sample evolutionary trajectory is shown as a dashed line in Fig. 4.

One can imagine several possible pathways leading to this self-sustaining cooperative system. (i) A pair of mutually complementary non-catalytic chains (a and $\bar{a}$ in our notation) gains function due to a copying error, giving rise to a small subpopulation of "sister" chains (b and $\bar{b}$ in our notation) with nascent cleavage activity directed toward $\bar{a}$ and a respectively. (ii) A pair of catalytically active chains b and $\bar{b}$ emerges first, subsequently losing the catalytic inserts due to a copying error, thus giving rise to a small subpopulation of substrate chains a and $\bar{a}$. According to the dynamic phase portrait of our system illustrated in Fig. 2B the second pathway is more plausible than the first one. Indeed, at least for the parameters used in Fig. 2B, the conversion of only a few percent of b chains to a brings the systems into the basin of attraction of the cooperative fixed point marked green in the figure. The other scenario is less likely since it requires a much higher ratio of emergent over ancestral population sizes.

Our results indicate that the cooperative steady state emerges when the catalytic rate enhancement over the spontaneous cleavage rate is at least $10^2 - 10^3$. This is a relatively modest gain compared to the 10^9 enhancement reported for a highly-optimized hammerhead ribozyme [31]. However, it is comparable to the rate enhancement observed after only five rounds of in vitro selection from an unbiased sample of random RNA sequences [12]. In the course of subsequent evolution, two main parameters of our model β and λ are expected to increase in tandem. The first parameter, β, is the catalytic cleavage rate, which, as we know, can increase over multiple orders of magnitude [12]. The second parameter, λ, quantifies the relative delay in the elongation of b-types chains compared to a-type chains. It may be caused at least in part by the difficulty of unzipping the hairpin, thus λ can be dramatically increased by making the hairpin stem longer. One should note that our model requires the catalytic activity of both b chains and their complementary partner, $\bar{b}$ chains. Thus, sequences are expected to evolve to simultaneously optimize these two catalytic rates. The need for this compromise would likely prevent the catalysts from reaching the maximum efficiency observed, e.g. in fully optimized ribozymes [31]. The drive to further optimize cleavage activity might trigger a transition to more complex catalytic networks, e.g. to an increase in the number of chains involved.

III. DISCUSSION

The proposed scenario is certainly not the only plausible pathway that could lead to the emergence of functional heteropolymers in the prebiotic world. To illustrate some of the unique features of the presented model, it is useful to place it in the context of other recent proposals. One of the most intriguing possibilities is the virtual circular genome (VCG) model recently proposed in ref. [33]. It is based on the observation that a relatively long ancestral genome can be stored as a collection of short overlapping RNA fragments of a circular master sequence or its complement. The model assumes unidirectional non-enzymatic replication of these fragments. On the one hand, it explains how a collection of relatively short RNA fragments (10-12 nucleotides each) could store a large amount of genetic information. On the other hand, computer simulations indicate that the VCG is susceptible to so-called sequence scrambling, where the appearance of repeats in the master sequence results in the loss of integrity of an entire circular genome [34]. In fact, this can be seen as a special case of the classical error catastrophe [35]: a single self-replicating master sequence of a virtual circular genome (and its complement), is replaced by a pool of scrambled sequences. The proliferation of the VCG is quite sensitive to these errors, since the model assumes unidirectional polymerization, so that in order to return to copying a particular segment, one must copy the entire virtual circle.

Our model, while sharing some elements with the VCG model, differs in two important aspects: (i) it assumes that the non-enzymatic templated polymerization is bidirectional, and (ii) the functional activity of the heteropolymer is localized in a relatively short sequence region that is catalytically cleaved and thus replicated first. This implies that the errors accumulated outside the narrow functional region do not affect the viability of our autocatalytic system. In fact, our approach can be combined with the VCG model to generate a linear rather than a circular virtual genome. The sequence scrambling would still be present in this model in the form of branching of a sufficiently long master sequence, but it would not affect the ability of the functional region of the sequence to self-replicate. An alternative proposal that avoids sequence scrambling is the rolling cycle model [27, 36], which relies on strand displacement rather than the cyclic melting of complementary strands. It generally requires a more sophisticated setup, including a preselected sequence pool and the availability of cleavage enzymes. Our model provides a plausible pathway for how this catalytic function might emerge.

Our model aims to describe the early stages of the evolution of life on Earth based on non-enzymatic polymerization. While it may seem challenging to test it for conditions relevant to the origin of life, our main con-

clusions can still be verified experimentally. RNA or DNA can be used as model polymers in such experiments, as both have demonstrated catalytic abilities in cleavage reactions [10–15]. To simulate primordial polymerization driven by day/night cycling, the experiment would have to rely on enzymatic polymerization or ligation as used e.g. in Polymerase Chain Reaction (PCR) [37, 38] or Ligase Chain Reaction (LCR) [39]. However, it's important to note that our model assumes polymerization in both the 5'-to-3' and 3'-to-5' directions, unlike traditional PCR, which only adds new nucleotides in the 5'-to-3' direction. This problem of bidirectional polymerization was solved by evolution using Okazaki fragments[40]. Inspired by this discontinuous synthesis of the lagging strand of DNA, we propose a possible experimental implementation of our system based on ligation rather than polymerization enzymes. In this scenario, the system would be supplied with ultrashort random DNA segments. These segments, which are much shorter than the minimal primer length (l_0), would play the role of "monomers" and bidirectional primer extension would occur through a sequence of ligation steps connecting adjacent ultrashort segments to each other. Another important consideration for experimental implementation is the need to activate the nucleotides to provide free energy for polymerization. Thus, both the short fragments supplied to the system and the new primers formed by cleavage must be chemically activated.

IV. CONCLUSIONS

Our results suggest a plausible pathway by which a pool of information-carrying polymers could acquire the catalytic function, thereby bringing this system closer to the onset of the RNA world. We start with a pool of polymers capable of non-enzymatic templated polymerization and subjected to cyclic change of conditions (day/night cycles). The replication in our system is carried out during the night phase of the cycle via the elongation of primers hybridized with their complementary templates. We first observe that any cleavage of chains generates new primers and thus promotes their replication. The mutual replication of complementary chains is sustainable, as long as cleavage and elongation rates are large compared to the dilution rate: $\beta_0 m_0 r > \delta^2$. This suggests that a faster, catalyzed cleavage would be selected by the evolution. Furthermore, DNA or RNA sequences capable of catalyzing site-specific cleavage are known to be relatively simple and readily arise via either natural or artificial selection [10–15]. The oligomer replication based on catalyzed cleavage is not trivial, as it requires cooperativity between multiple chain types. Our study shows that a stable cooperative solution can be achieved with as few as four subpopulations of chains. Furthermore, we demonstrate that there is a wide range of conditions under which this catalytic network proliferates and significantly outcompetes non-catalytic ancestors of the constituent chains.

ACKNOWLEDGEMENTS

This research used resources of the Center for Functional Nanomaterials, which is a U.S. DOE Office of Science User Facility, at Brookhaven National Laboratory under Contract No. DE-SC0012704.

[1] W. Gilbert, nature **319**, 618 (1986).
[2] J. A. Doudna and J. W. Szostak, Nature **339**, 519 (1989).
[3] L. E. Orgel, Critical reviews in biochemistry and molecular biology **39**, 99 (2004).
[4] T. A. Lincoln and G. F. Joyce, Science **323**, 1229 (2009).
[5] M. P. Robertson and G. F. Joyce, Cold Spring Harbor perspectives in biology **4**, a003608 (2012).
[6] P. G. Higgs and N. Lehman, Nature Reviews Genetics **16**, 7 (2015).
[7] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech, cell **31**, 147 (1982).
[8] D. P. Bartel and J. W. Szostak, Science **261**, 1411 (1993).
[9] D. P. Horning and G. F. Joyce, Proceedings of the National Academy of Sciences **113**, 9786 (2016).
[10] M. de la Peña and I. García-Robles, Rna **16**, 1943 (2010).
[11] K. Williams, S. Ciafre, and G. Tocchini-Valentini, The EMBO journal **14**, 4551 (1995).
[12] K. Salehi-Ashtiani and J. W. Szostak, Nature **414**, 82 (2001).
[13] R. R. Breaker and G. F. Joyce, Chemistry & biology **1**, 223 (1994).
[14] S. K. Silverman, Nucleic acids research **33**, 6151 (2005).
[15] M. Chandra, A. Sachdeva, and S. K. Silverman, Nature chemical biology **5**, 718 (2009).
[16] J. W. Szostak, Journal of Systems Chemistry **3**, 1 (2012).
[17] S. C. Kim, D. K. O'Flaherty, C. Giurgiu, L. Zhou, and J. W. Szostak, Journal of the American Chemical Society **143**, 3267 (2021).
[18] J. P. Ferris, Elements **1**, 145 (2005).
[19] C. A. Jerome, H.-J. Kim, S. J. Mojzsis, S. A. Benner, and E. Biondi, Astrobiology **22**, 629 (2022).
[20] A. V. Tkachenko and S. Maslov, The Journal of Chemical Physics **143**, 07B612_1 (2015).
[21] A. V. Tkachenko and S. Maslov, The Journal of chemical physics **149**, 134901 (2018).

[22] P. W. Kudella, A. V. Tkachenko, A. Salditt, S. Maslov, and D. Braun, Proceedings of the National Academy of Sciences **118**, e2018830118 (2021).
[23] J. Sulston, R. Lohrmann, L. Orgel, and H. T. Miles, Proceedings of the National Academy of Sciences **59**, 726 (1968).
[24] B. Weimann, R. Lohrmann, L. Orgel, H. Schneider-Bernloehr, and J. Sulston, Science **161**, 387 (1968).
[25] R. Lohrmann, P. Bridson, and L. Orgel, Science **208**, 1464 (1980).
[26] D. Duzdevich, C. E. Carr, and J. W. Szostak, Nucleic acids research **48**, e70 (2020).
[27] A. S. Tupper and P. G. Higgs, Journal of Theoretical Biology **527**, 110822 (2021).
[28] G. A. Prody, J. T. Bakos, J. M. Buzayan, I. R. Schneider, and G. Bruening, Science **231**, 1577 (1986).
[29] C. J. Hutchins, P. D. Rathjen, A. C. Forster, and R. H. Symons, Nucleic acids research **14**, 3627 (1986).
[30] W. G. Scott, J. T. Finch, and A. Klug, Cell **81**, 991 (1995).
[31] W. G. Scott, L. H. Horan, and M. Martick, Progress in Molecular Biology and Translational Science **120**, 1 (2013).
[32] H. W. Pley, K. M. Flaherty, and D. B. McKay, Nature **372**, 68 (1994).
[33] L. Zhou, D. Ding, and J. W. Szostak, RNA **27**, 1 (2020).
[34] P. Chamanian and P. G. Higgs, PLOS Computational Biology **18**, e1010458 (2022).
[35] M. Eigen, Naturwissenschaften **58**, 465 (1971).
[36] F. Rivera-Madrinan, K. Di Iorio, and P. G. Higgs, Life **12** (2022), 10.3390/life12091373.
[37] K. B. Mullis and F. A. Faloona, in *Methods in enzymology*, Vol. 155 (Elsevier, 1987) pp. 335–350.
[38] R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich, Science **239**, 487 (1988).
[39] F. Barany, Proceedings of the National Academy of Sciences **88**, 189 (1991).
[40] R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto, R. Kainuma, A. Sugino, and N. Iwatsuki, in *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 33 (Cold Spring Harbor Laboratory Press, 1968) pp. 129–143.

APPENDIX A. EFFECT OF TEMPLATE REHYBRIDIZATION

Consider a simple model in which two mutually complementary sequences act as templates for each other's polymerization. Let C and $\bar{C}$ be the concentrations of these unhybridized template chains. Let c and $\bar{c}$ be the concentrations of unhybridized fragments of these two sequences, respectively. They may act as primers once bound to their complementary template. At the beginning of the "night" phase, we assume all the chains to be in an unhybridized state, i.e. $C = C_0$, $\bar{C} = \bar{C}_0$, $c = c_0$, and $\bar{c} = c_0$. We will describe the system's kinetics during the "night" phase by assuming that hybridization is essentially irreversible, except for the possibility of a primer strain displacement by a full chain. We only focus on the regime when the total concentration of primers, c_0 and $\bar{c}_0$ is much smaller than that of free templates, C and $\bar{C}$. Note that the concentrations of bound primers are $c_0 - c$, and $\bar{c}_0 - \bar{c}$, respectively.

The kinetics of this simple model is described by the following rate equations:

$$\dot{C} = -\kappa_0 C\bar{C} \tag{S1}$$

$$\dot{c} = -\kappa_1 \bar{C}c + \kappa_2 C(c_0 - c) \tag{S2}$$

The first equation accounts for the hybridization of mutually complementary template chains (assuming $c \ll C$, and $\bar{c} \ll \bar{C}$, κ_0 is the corresponding association rate). The second equation accounts for the hybridization of primers with their complementary substrates and for the primer strand displacement due to the hybridization of the template with its full-length complementary. The respective association rates for these two processes, κ_1 and κ_2, are dependent on the length and sequences of the chains involved, but generally comparable. The other two equations, for $\bar{C}$ and $\bar{c}$, are obtained by replacing all the concentrations with their complementaries.

We first consider the asymmetric case, when, $C_0 > \bar{C}_0$ (without loss of generality). This implies that for a long enough duration of the night, the number of hybridized chains is limited by $\bar{C}_0$. Thus, only one type of the template chain will remain in the solution, $C = C_0 - \bar{C}_0$, while the other will be completely hybridized, i.e. $\bar{C} = 0$. In turn, this implies that the situation will be even more drastic for the primers: $\bar{c} = 0$ (completely hybridized), while $c = c0$ (completely free). As a result, the minority fraction, $\bar{C}$ would replicate faster and the balance would be restored, so that $C_0 = \bar{C}_0$.

Now, we will analyze the above set of differential equations for a symmetric case, $C_0 = \bar{C}_0$. The fraction of free template chains vanishes with time, as a power law (rather than exponentially):

$$C(t) = \bar{C}(t) = \frac{C_0}{1 + C_0\kappa_0 t} \tag{S3}$$

As to the concentration of free fragments, it reaches a

steady state value between 0 and c_0:

$$c = \frac{c_0}{1 + \kappa_1/\kappa_2} \tag{S4}$$

Since the association rates κ_1 and κ_2 are comparable, we come to a surprising conclusion that a finite fraction of fragments, of the order of 1, will stay hybridized in the course of a long enough night, despite the effect of strand displacement by longer chains:

$$\frac{c_0 - c}{c_0} = \frac{1}{1 + \kappa_2/\kappa_1} \sim 1 \tag{S5}$$

APPENDIX B. COOPERATIVE STEADY STATE SOLUTION

The steady state of Eqs. (7-11) from the main text must satisfy the following set of equations:

$$\beta\phi a = \delta c - \beta_0 M \tag{S6}$$

$$\delta b = \beta_0 M\phi + \frac{rm\phi}{\lambda} a_R \tag{S7}$$

$$\beta\phi a = (rm + \delta)\, a_L \tag{S8}$$

$$\beta\phi a = \left(rm\left(1 - \phi + \frac{\phi}{\lambda}\right) + \delta\right) a_R \tag{S9}$$

$$rmc = \delta M \tag{S10}$$

Let us define

$$\mu = \frac{\delta}{rm}. \tag{S11}$$

By using Eq. (S10) to exclude variable M, and replacing b in Eq. (S7) with $\phi(a + b)$, the fixed point conditions can be rewritten as:

$$\beta\phi a = (\delta - \beta_0/\mu)c \tag{S12}$$

$$a + b = \frac{1}{\mu}\left(\frac{a_R}{\lambda} + \frac{\beta_0}{\delta}c\right) \tag{S13}$$

$$a_L = \frac{(\mu - \beta_0/\delta)c}{1 + \mu} \tag{S14}$$

$$a_R = \frac{(\mu - \beta_0/\delta)c}{1 + \mu - (1 - \lambda^{-1})\phi} \tag{S15}$$

As $a + b$, a_L, and a_R are now expressed in terms of c, we use its definition, $c = a + b + a_L + a_R$, to get

$$\left(1 - \frac{\beta_0}{\mu\delta}\right) c = \left(1 + \frac{1}{\lambda\mu}\right) a_R + a_L = \left(1 - \frac{\beta_0}{\mu\delta}\right) \cdot \cdot \left(\frac{1 + \lambda\mu}{1 + \mu - (\lambda - 1)\phi} + \frac{1}{1 + \mu}\right) c \tag{S16}$$

This yields a compact expression for ϕ in terms of μ and λ only, thus invariant with respect to both cleavage rates, β_0 and β :

$$\phi = \left(1 - \frac{\lambda\mu}{\lambda - 1}\right)(1 + \mu) \tag{S17}$$

Another relationship is obtained by using Eqs. (S13) and (S15) to express $a = (1-\phi)(a+b)$ in terms of c, and substituting it into Eq. (S12):

$$\frac{\delta}{\beta} = \phi(1 - \phi)\left(\frac{1}{\lambda\,(1 - \phi + \mu) + \phi} + \frac{1}{\mu\delta/\beta_0 - 1}\right) \tag{S18}$$

By substituting Eq. (S17) into Eq. (S18) one gets the analytic relationship that allows to compute β for arbitrary values of μ, β_0, δ and λ:

$$\frac{\delta}{\beta} = \frac{\mu}{\lambda - 1}\left(1 - \frac{\lambda\mu}{\lambda - 1}\right)\left(1 + (1 + \mu)(1 + \lambda\mu)\frac{\beta_0}{\mu\delta - \beta_0}\right) \tag{S19}$$

In practice, it is the monomer concentration that gets adjusted to its steady state value m^*, thus μ is the variable that has to satisfy the above equation, for given values of other parameters.

The stable fixed points only appear on the decreasing segment of the function $\beta(\mu)$. Thus, the limits of stability correspond to zero derivatives of the r.h.s. of Eq. (S19). This leads to the following condition:

$$\left(1 - \frac{2\lambda\mu}{\lambda - 1}\right) z^2 + \mu A'(\mu) z - A(\mu) = 0 \tag{S20}$$

Here we defined $z = \mu\delta/\beta_0 - 1$,

$$A(\mu) = \left(1 - \frac{\lambda\mu}{\lambda - 1}\right)(1 + \mu)(1 + \lambda\mu) = = 1 + \frac{(\lambda^2 - \lambda - 1)\mu - 2\lambda\mu^2 - \lambda^2\mu^3}{\lambda - 1} \tag{S21}$$

$$A'(\mu) = \frac{(\lambda^2 - \lambda - 1) - 4\lambda\mu - 3\lambda^2\mu^2}{\lambda - 1} \tag{S22}$$

The positive solution to the quadratic Eq. (S20) is given by

$$z(\mu) = \frac{(\lambda - 1)\mu A'(\mu)}{2(\lambda - 1 - 2\lambda\mu)} \cdot \cdot \left(\sqrt{1 + 4\left(1 - \frac{2\lambda\mu}{\lambda - 1}\right)\frac{A(\mu)}{\mu^2 A'(\mu)^2}} - 1\right) \tag{S23}$$

Now, by using Eq.(S19) and our definition of x, both β and β_0 can be parameterized by μ:

$$\beta_0(\mu) = \frac{\mu\delta}{z(\mu) + 1} \tag{S24}$$

$$\beta(\mu) = \frac{\delta(\lambda - 1)}{\mu\left(1 - \frac{\lambda\mu}{\lambda - 1} + \frac{A(\mu)}{z(\mu)}\right)} \tag{S25}$$

This parametric curve in (β, β_0) space defines the boundary of the region in which the non-trivial fixed point solution exists.

A simplified asymptotic relationship between β, β_0 and μ can be obtained in the limit of $\mu \ll 1$. In this case,

Eq.(S23) implies $z(\mu) \approx 1$. By substituting this result into Eqs. (S24)-(S25), one obtains

$$\mu \approx 2\beta_0/\delta \tag{S26}$$

$$\beta \approx \frac{(\lambda-1)\delta^2}{4\beta_0} \tag{S27}$$

This result gives the lower bounds for μ and the upper bound for β consistent with the cooperative solution, for a given β_0. The approximation is valid in the limit of $\mu \ll 1$, i.e. $\beta_0/\delta \ll 1$.

Another asymptotic result can be obtained for vanishingly small non-catalytic cleavage rate β_0. In that limit, Eq. (S19) turns into quadratic equation since $(1+\mu)(1+\lambda\mu)/\mu \ll \delta/\beta_0$. Its two solutions correspond to unstable and stable fixed points. Specifically, the stable branch is given by

$$\mu \approx \frac{\lambda-1}{2\lambda}\left(1-\sqrt{1-\frac{4\lambda\delta}{\beta}}\right) \tag{S28}$$

The solution only exists for $\beta/\delta \geq 4\lambda$, which sets the lower bound for β. This critical point corresponds to $\mu = \frac{\lambda-1}{2\lambda}$, which is the upper bound of μ. Note that this approximation is valid as long as

$$\frac{\beta_0}{\delta} \ll \frac{\mu}{(1+\mu)(1+\lambda\mu)} \approx \frac{2(\lambda-1)}{(3\lambda-1)(\lambda+1)} \tag{S29}$$

To summarize, the range of β for which the cooperative solution exists, and the respective range of the steady-state monomer concentration m^*_{coop}, can be approximated as:

$$4\lambda < \frac{\beta}{\delta} < \frac{(\lambda-1)\delta}{4\beta_0} \tag{S30}$$

$$\frac{2\lambda\delta}{(\lambda-1)\beta_0} < \frac{m^*_{\text{coop}}}{m^*} < 1/2 \tag{S31}$$

Here $m^* = \delta^2/(r\beta_0)$ is the steady state monomer concentration in the non-cooperative regime given by Eq. (5). The fitness parameter for the cooperative regime

$$\frac{m^*}{m^*_{\text{coop}}} = \frac{\mu\delta}{\beta_0} \tag{S32}$$