Transient Compartmentalization dynamics in the presence of mutations and noise

Alex Blokhuis[a,b,*], Philippe Nghe[b], Luca Peliti[c], David Lacoste[a]

[a] *Gulliver Laboratory, UMR CNRS 7083, PSL Research University, ESPCI, 10 rue Vauquelin, F-75231 Paris, France*

[b] *Laboratory of Biochemistry, PSL Research University, ESPCI, 10 rue Vauquelin, F-75231 Paris, France*

[c] *SMRI, 00058 Santa Marinella (RM), Italy*

Abstract

We extend a recently introduced framework for transient compartmentalization of replicators with selection dynamics, by including the effect of mutations and noise in such systems. In the presence of mutations, functional replicators (ribozymes) are turned into non-functional ones (parasites). We evaluate the phase diagram of a system undergoing transient compartmentalization with selection. The system can exhibit either coexistence of ribozymes and parasites, or a pure parasite phase. If the mutation rate exceeds a certain level called the error threshold, the only stable phase is the pure parasite one. Transient compartmentalization with selection can relax this error treshold with respect to a bulk quasispecies case, and even allow ribozymes to coexist with faster growing parasites.

In order to analyze the role of noise, we also introduce a model for the replication of a template by an enzyme. This model admits two regimes: a diffusion limited regime which generates a high noise, and a replication limited regime, which generates a low noise at the population level. Based on this model, we find that, since the ribozyme dynamics belongs to the replication limited regime, the effects of noise on the phase diagram of the system are mostly negligible. Our results underlines the importance of transient compartmentalization for prebiotic scenarios, and may have implications for directed evolution experiments.

Keywords: Error Catastrophe, Mutation, Parasites, Growth Noise, RNA World

---

*Corresponding author

Email address: alex_blokhuis@hotmail.com (Alex Blokhuis)

1. Introduction

Compartments play a central role in many biological processes of cells, in particular in organelles such as the ER or in the Golgi apparatus [1]. Cells use compartments to organize chemical reactions in space: compartments eliminate the risk of losing costly catalysts which are essential for biochemical reactions, they also accelerate chemical reactions, while reducing the risk of cross-talks due to other side reactions.

In the early 20th century, Oparin suggested that membrane-less compartments, which he called coacervates, could have played a central role in the origin of life [2]. Many compartments in cells are bounded by membranes, but following the recent discovery of the so-called P-granules in C elegans embryo [3], biologists noticed that membrane-less compartments abound in living organisms. These membrane-less compartments are of interest for physicists who want to understand better the active non-equilibrium phase separation which creates them [4]; but also for chemists who are trying to synthetize and control them *in vitro* [5, 6].

After the discovery of the structure of DNA, the coacervates scenario for the origin of life got less popular, and was replaced by replication scenarios [7, 8]. In the sixties, Spiegelman showed that RNA could be replicated by an enzyme called $Q\beta$ RNA replicase, in the presence of free nucleotides and salt [9]. After a series of serial transfers, he observed the appearance of shorter RNA polymers, which he called parasites. Typically, these parasites are non-functional molecules which replicate faster than the RNA polymers introduced at the beginning of the experiment. In 1971, Eigen conceptualized this observation by proving theoretically that for a given accuracy of replication and a relative fitness of parasites, there is a maximal genome length that can be maintained without errors [10]. This result led to the following paradox: to be a functional replicator, a molecule must be long enough. However, if it is long, it cannot be maintained since it will quickly be overtaken by parasites. This puzzle eventually played a central role in origin of life studies [11, 12].

In the eighties, a theoretical solution emerged, the Stochastic corrector model [13] inspired by ideas of group selection [14]. The idea is to use compartmentalization and selection to maintain functional replicators (ribozymes). The compartments in that model undergo cell division, which is a sophisticated feature that strongly constrains the allowed prebiotic scenarios.

In order to address this point, and also to assess the role of transient compartmentalization

using a quantitative theoretical model, we introduced a general class of multilevel selection with transient compartmentalization [15]. This class includes several scenarios for the origin of life based on various types of compartments (lipid vesicles [16], pores [17, 18], inorganic compartments [19], ..) or various protocols of transient compartmentalization [20, 21] and a recent experiment, in which small droplets containing RNA in a microfluidic device [22] were used as compartments.

The related issue of cooperation between producers and non-producers has been discussed before [23]. When compartments are not well defined, their role can be played by spatial clustering, which can favor the survival of cooperating replicators [24, 25]. These ideas were combined in a recent study of a population of individuals growing in a large number of compartmentalized habitats, called demes [26]. Another recent related study on transient compartmentalization quantifies co-encapsulation effects in the context of directed evolution experiments [27].

In this paper, we go beyond the analysis carried out in our previous work [15] by including the effect of mutations and noise. The motivation of including mutations comes from experiments, since mutations play a role in the RNA droplet experiment [22] which inspired us. The motivation of discussing noise in more details comes from the realization that replication is inherently stochastic when a small number of replicators are present in compartments. Therefore, the deterministic approach used in our previous work [15] may seem like a severe approximation. In fact, it is important to appreciate that although our deterministic model neglects some sources of noise such as fluctuations in the growth rates, it includes fluctuations due to the smallness of the number of replicators present in the initial condition. This is a major source of noise because at the end of the exponential phase, the compartments no longer contain small numbers of molecules and therefore fluctuations become again negligible. A similar effect occurs when considering protein aggregation kinetics in small volumes: the fluctuations in the mass concentration of proteins at some later time are mainly due to the smallness of the number of proteins present in the initial condition [28].

In Sec. 2, we recapitulate the essence of the mutation-free model which we have introduced in [15], then in Sec. 3, we introduce an extension of that model to include deterministic mutations. The effect of noise is then covered in Sec. 4, which contains, in particular, in Sec. 4.1 a simple model for the replication of a single template by an enzyme, and in Sec. 4.4 an analysis of the growth noise in a population of such replicating molecules. The latter model is finally used to analyze the effect of noise on the transient compartmentalization dynamics introduced in the first

section.

2. The mutation-free model

2.1. Definition of the model

We start from a large pool of molecules, which contains functional molecules called ribozymes and non-functional ones called parasites. Let the fraction of ribozymes in this pool be x. These molecules then seed a large number of compartments, which is treated in the infinite limit. A given compartment will contain n molecules, out of which m will be ribozymes and the remaining ones parasites. It follows that n is a random variable drawn from a Poisson distribution of parameter λ, while the number m follows a binomial distribution $B_m(n, x)$. The resulting probability distribution for seeded compartments is then

$$P_\lambda(n, m, x) = \text{Poisson}(\lambda, n) B_m(n, x). \tag{1}$$

In addition to the replicating molecules, a large amount of replication enzymes $n_{Q\beta}$ and activated nucleotides n_u is supplied. We assume that all compartments contain the same amount of replication enzymes and activated nucleotides.

After seeding, the numbers of ribozymes m and parasites y grow exponentially, what in a deterministic model leads to

$$\bar{m} = m e^{\alpha T}, \tag{2}$$

$$\bar{y} = (n - m) e^{\gamma T}, \tag{3}$$

with T the time at the end of exponential growth phase, $\bar{m}$ the number of ribozymes and $\bar{y}$ the number of parasites at time T. At the end of this growth phase, we have $n_{Q\beta} \approx N = \bar{m} + \bar{y}$, at which point further growth will be limited by the number of replication enzymes. As a result, after time T, the growth will be linear instead of exponential, but in any case, the system composition defined here by the relative fraction of ribozymes, will not change. For this reason, we focus on the final composition at time T which is controlled by the ratio $\Lambda = e^{(\gamma - \alpha)T}$. Here, we do not describe precisely the crossover between the exponential and the linear regime, which could be done using the notion of carrying capacity [29]. In that case, the growth would be described by logistic equations and the carrying capacity would be comparable to $n_{Q\beta}$. Note also that the exact

4

time T may depend on m, n, but since in practice $n_{Q\beta} \gg m, n$ this dependence has a small effect on the results of the model as we have checked in the Suppl. Mat. of Ref. [15].

In any case, the ribozyme fraction at the end of growth phase can be well approximated as

$$\bar{x}(n, m) = \frac{\bar{m}}{N} = \frac{m}{n\Lambda - (\Lambda - 1)m}. \tag{4}$$

If parasites grow faster, we have $\gamma > \alpha$, and thus $\Lambda > 1$, which is the regime considered in Ref. [15]. In Sec 3, we also consider regimes in which $\gamma < \alpha$.

We now implement selection at the compartment level. In practice, selection could be autonomous or non-autonomous. For instance, in the experiment of Ref [22], the selection was non-autonomous : a measurement of the synthesis of a dye molecule by photodetection was used to promote or reject compartments according to the outcome of that measurement. Selection can in general be described by a selection function $f(\bar{x}) \geq 0$. In our work, we have assumed that the selection function only depends on the final composition $\bar{x}$ of the compartment.

For the ribozyme-parasite scenario, a natural choice for f is a monotonically increasing function of $\bar{x}$. As an example, we will use the sigmoidal function

$$f(\bar{x}) = 0.5 \left(1 + \tanh\left(\frac{\bar{x} - x_{th}}{x_w} \right) \right), \tag{5}$$

where x_{th} and x_w are dimensionless parameters, which describe respectively a threshold in the composition and the steepness of the function.

The compartments which have passed the selection step are then pooled together, forming a new pool of molecules from which future compartments can be seeded. The ribozyme fraction x' of this new ensemble is the average of $\bar{x}$ among the selected compartments

$$x' = \frac{\langle \bar{x} f(\bar{x}) \rangle}{\langle f(\bar{x}) \rangle}, \tag{6}$$

which is equivalent to

$$x'(\lambda, x) = \frac{\sum_{n,m} \bar{x}(n, m) f(\bar{x}(n, m)) P_\lambda(n, x, m)}{\sum_{n,m} f(\bar{x}(n, m)) P_\lambda(n, x, m)}. \tag{7}$$

The transient compartmentalization cycle is then repeated, starting with the seeding of new compartments from that pool of composition x'.

Upon repetition of this protocol, the pool composition typically converges to a fixed point x^*, which is a solution of

$$x = x'(\lambda, x). \tag{8}$$

The stability of the fixed point x^* changes when

$$\left.\frac{dx'}{dx}\right|_{x=x^*} = 1. \tag{9}$$

It is implicitly assumed that $x'(x)$ is a sufficiently smooth function of x for this derivative to be defined.

2.2. Main dynamical regimes

Although finding a fixed point x^* is generally difficult, our ribozyme-parasite model contains two simple fixed points: $x = 0$ and $x = 1$. By evaluating the stability of these two fixed points, four regimes can be distinguished, which are shown in the phase diagram in Fig 1. If $x = 1$ is stable and $x = 0$ unstable, ribozymes are stabilized and parasites are purged. If $x = 0$ is stable and $x = 1$ unstable, parasites deterministically invade the pool and purge ribozymes. If both $x = 0$ and $x = 1$ are unstable, trajectories from either side are attracted to a stable third fixed point $0 < x^* < 1$, leading to stable coexistence between parasites and ribozymes. Finally, if $x = 0$ and $x = 1$ are stable, their basins of attraction are separated by a third fixed point $0 < x^* < 1$, which is unstable, and in this case we have a bistable regime in which the initial composition determines the fate of the system.

These conclusions can only be drawn provided there are not other fixed points besides ($x = 0, x = 1, x = x^*$). Extra fixed points come in pairs (one stable, one unstable) and matter only if they are situated within $(0, 1)$, in which case a stable coexistence and a bistable phase would be added to the behavior inferred from the other fixed points. For simple monotonically increasing selection functions, we find that extra fixed points are a rare occurrence. Nevertheless a case where this occurs has been discussed in the Suppl. Mat. of Ref. [15].

2.3. Comparison to experiments

In addition to predicting the phase diagram associated with the long-time compositions reached by this transient compartmentalization dynamics, our theoretical model makes also predictions regarding the evolution of the ribozyme fraction as function of the round number, *i.e.*

6

Type	Length (nt) 2	$T_d(s)$	Relative r	Λ
Ribozyme	362	25.0	1.00	1
Parasite 1	245	20.7	1.21	13
Parasite 2	223	17.1	1.46	107
Parasite 3	129	14.6	1.71	473

Table 1: Lengths and doubling times for the parasites and ribozyme observed in Ref. [22], together with their relative aggressivity mesured by their relative growth rate r, and the corresponding values of Λ.

the number of completed cycles of compartmentalization. The model correctly reproduces that this fraction quickly goes to zero as function of the round number in bulk, less quickly with compartmentalization and no selection and even less quickly in the case of compartmentalization with selection. In the latter case, a finite fraction can be maintained for an infinite number of rounds provided λ is sufficiently small, corresponding to the coexistence region of the phase diagram.

In order to compare precisely the predictions of the model to the experiments of Ref. [22], it is important to know the value of key parameters such as Λ. Table 1 reports the experimental parameters measured in Ref. [22] for the ribozyme and three different parasites. The nucleotide length, its doubling time (T_d), its relative replication rate (r) from which we infer Λ in the final column. The doubling time T_d for the ribozyme is related to the growth rate α by $T_d = \ln(2)/\alpha$, and similarly the doubling times of the parasites is $T_d = \ln(2)/\gamma$.

In the experiment, a typical compartment contains λ RNA molecules that can be ribozymes or parasites, $2.6 \cdot 10^6$ molecules of Qβ replicase, and $1.0 \cdot 10^{10}$ molecules of each NTP. Replication takes place by complexation of RNA with Qβ replicase, which uses NTPs to make a complementary copy. This copy is then itself replicated to reproduce the original. There is a large amount of nucleotides, so that exponential growth of the target RNA proceeds until $N \approx n_{Q\beta}$. This large quantity of enzymes also means that in practice, the noise due to fluctuations in the number of enzymes should be very small. Starting from a single molecule, it takes $n_D = \log_2 n_{Q\beta} = 21.4$ doubling times to reach this regime. In a parasite-ribozyme mixture, we can estimate Λ using the relative r:

$$\Lambda = \frac{2^{n_D}}{2^{n_D/r}} = 2^{n_D(1-\frac{1}{r})}. \tag{10}$$

3. A modified model with deterministic mutations

In the deterministic model, we assume that a fraction μ of replicated ribozyme strands mutate into parasites. Thus, the equations describing the evolution of m and y in the growth phase assumes the form

$$
\begin{align}
\dot{m} &= \alpha m - \mu m = (\alpha - \mu)m \tag{11}\\
\dot{y} &= \gamma y + \mu m,
\end{align}
$$

which yields for the first equation

$$
\bar{m} = m e^{(\alpha - \mu)T}, \tag{12}
$$

where $\bar{m}$ is again the number of ribozymes at the end of the growth phase and m the value at the initial time. Now substituting Eq. (12) into the equation for y, one finds

$$
\bar{y} = \left(n - m + \mu m \frac{e^{(\alpha - \gamma - \mu)T} - 1}{\alpha - \mu - \gamma} \right) e^{\gamma T}. \tag{13}
$$

The ratio between the number of daughters of one parasite molecule and the number of daughters of a ribozyme molecule is now renormalized by the rate μ: $\bar{\Lambda} = e^{(\gamma + \mu - \alpha)T} = e^{\mu T} \Lambda$, where Λ is the relative growth of parasites introduced previously in the mutation-free model.

The fraction of ribozymes at the end of the exponential phase is now given by

$$
\bar{x}(n, m) = \frac{\bar{m}}{N} = \frac{m}{n\bar{\Lambda} - (\bar{\Lambda} - 1)(1 + \delta)m}, \tag{14}
$$

where $\delta = \mu/(\alpha - \mu - \gamma)$. We call δ the mutation ratio, which is a dimensionless measure of mutation versus relative growth (competition). When $\delta \to 0$, we recover the mutation-free model, if $|\delta| \gg 0$ mutations become dominant.

Selected compartments are then pooled together, and the new average fraction of ribozymes becomes $x'(x, \lambda, \delta, \bar{\Lambda})$. Note that for nonzero mutation rate ($\mu > 0$), $x' = 1$ ceases to be a fixed point in this deterministic approach, since parasites will always appear at sufficiently long times. Therefore, the pure ribozyme (R) phase is no longer present in the phase diagram of fig. 2.

The fixed point $x' = 0$ however is still present. If this fixed point is stable, we have a pure parasite phase. If it is unstable, there is stable coexistence at a fixed composition. If more fixed points appear, multiple stable compositions are in principle be possible.
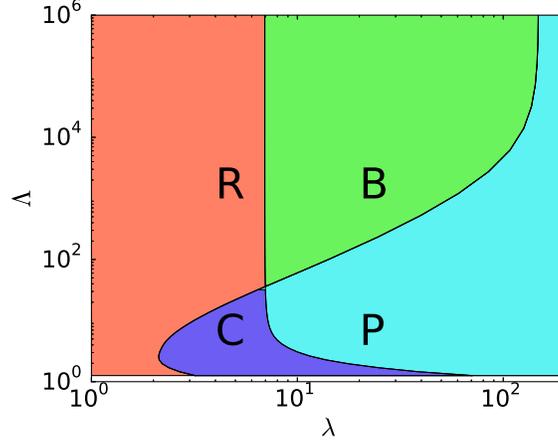
Figure 1: Original phase diagram of the mutation free model, taken from Ref[15]. The various phases are pure ribozyme (R), bistable (B), coexistence (C), pure parasite (P).

3.1. The prolific parasites regime $(\bar{\Lambda} \geq 1)$

Prolific parasites have a better bulk reproductive success than ribozymes, when $\bar{\Lambda} \geq 1$, which is equivalent to $\alpha \leq \mu + \gamma$ and $\delta < 0$. In a mutation-free model, this would imply necessarily a faster growth of parasites ($\alpha < \gamma$), but in the present case, we could also allow for slower parasites as compared to ribozymes (i.e. $\alpha > \gamma$), provided parasites are aided by a sufficiently high mutation rate μ.

The phase diagram is evaluated by testing the stability of the fixed point $x' = 0$. We find an asymptote behaving like $1/\lambda$ for large λ, and plateaus for small λ. The ends of these plateaus locate in the limit $\delta \to 0$ at the position of the vertical line separating the ribozyme and bistable phase in the original phase diagram.

Let us first derive the right asymptote in the $\lambda \gg 1$ limit. In this limit, we evaluate x' by considering compartments of size λ

$$x' = \frac{\lambda x \bar{x} f(\bar{x})}{(1-x)f(0) + \lambda x f(\bar{x})}. \tag{15}$$

The fixed point stability condition $dx'/dx|_{x=0} = 1$ leads to

$$\left. \frac{dx'}{dx} \right|_{x=0} = \frac{\lambda \bar{x} f(\bar{x})}{f(0)}. \tag{16}$$
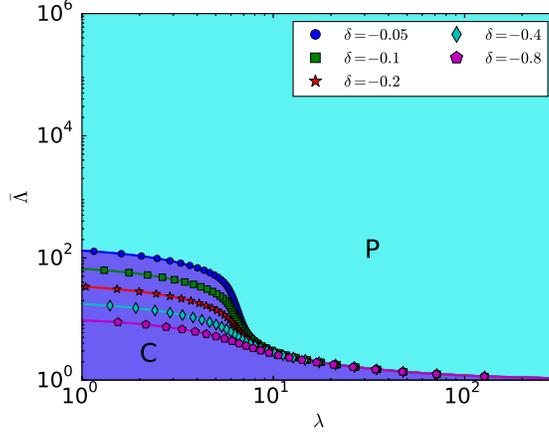
9

Figure 2: Phase diagram of the model with mutation in the case of prolific parasites. The selection function is given in Eq. (5). Phases are colored for $\delta = -0.05$, other separatrices are plotted for various mutation strengths δ. The possible phases are coexistence (C), pure parasite (P).

Upon substituting Eq. (14) evaluated at $m = 1, n = \lambda$ and approximating $f(\bar{x}) \approx f(0) + f'(0)\bar{x}$, (for $\lambda \gg 1, \bar{x} \ll 1$) we find a quadratic equation for $\bar{\Lambda}$, whose only physical solution ($\bar{\Lambda} \geq 1$) is

$$\bar{\Lambda} = \frac{\lambda - 2\delta - 2 + \sqrt{\lambda\left(4\frac{f'(0)}{f(0)} + \lambda\right)}}{2(\lambda - \delta - 1)}. \tag{17}$$

Since we consider monotonically increasing selection functions, $f'(0) > 0$. For $\lambda \gg -\delta$, we find

$$\bar{\Lambda} = 1 + \frac{f'(0)}{f(0)(\lambda - \delta - 1)} \approx 1 + \frac{f'(0)}{f(0)\lambda}, \tag{18}$$

which is the same expression as the one found in the mutation-free phase diagram [15]. This explains why there is a single asymptote as μ is varied in the $\lambda \gg 1$ limit.

The plateaus extend to very low values of λ. We can find their location by considering only compartments of size $n = 1$. In that case, the final compositions can be $\bar{x}(1,0) = 0$ or

$$\bar{x}(1,1) = \frac{1}{1 + \delta - \delta\bar{\Lambda}}. \tag{19}$$

We then have for the composition recursion

$$x' = \frac{x\bar{x}f(\bar{x})}{(1-x)f(0) + xf(\bar{x})}. \tag{20}$$

Evaluating the derivative of $x'(x)$, we find
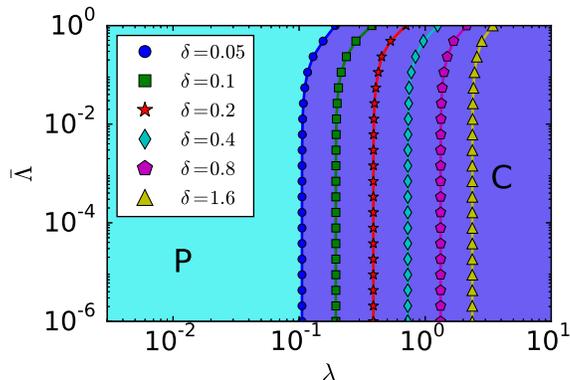
$$\frac{\bar{x}f(\bar{x})}{f(0)} = 1. \tag{21}$$

10

Figure 3: Phase diagram in absence of selection function for prolific ribozymes ($\bar{\Lambda} \leq 1$). Phases are colored for $\delta = 0.05$, separatrices are plotted for various mutation strengths δ. C: coexistence, P: pure parasite.

Substituting (19), we find that the location of plateaus obeys the implicit equation

$$\bar{\Lambda} = 1 + \frac{f(0) - f(\bar{x})}{f(0)\delta}. \tag{22}$$

3.2. The prolific ribozymes regime ($\bar{\Lambda} \leq 1$)

We now consider the opposite case where parasites are less prolific than ribozymes. This means $\alpha \geq \mu + \gamma$ and is equivalent to $\bar{\Lambda} \leq 1, \delta > 0$. This implies that $\alpha > \gamma$ (less aggressive parasites) and is reminiscent of a quasipecies scenario in which a fit ribozyme succesfully outcompetes its parasites in bulk [10]. Since this can already happen in the absence of selection, we consider here the case where there is no selection, *i.e.* $f(\bar{x}) = 1$.

To analyze this regime we again assess the fixed point stability of $x' = 0$. We locate numerically the separatrix as shown in Fig 3. We obtain separatrices that for $\bar{\Lambda} \to 0$ tend to a fixed value of λ.

Let us start by observing that when $\bar{\Lambda} \to 0$, there are only two final compartment compositions for nonempty compartments: $\bar{x}(n, 0) = 0$ or $\bar{x}(n, m) = 1/(1+\delta)$ for $m > 0$. We can now distinguish between three initial compartment compositions: (i) only parasites, (ii) no parasites, no ribozymes,

11

and (iii) containing at least one ribozyme. Their associated seeding probabilities are:

$$
\begin{aligned}
p_{\text{para}} &= \sum_{n=1}^{\infty} \frac{(1-x)^n \lambda^n}{n!} e^{-\lambda} = (e^{\lambda(1-x)} - 1)e^{-\lambda} \\
p_{\text{zero}} &= e^{-\lambda} \\
p_{\text{ribo}} &= 1 - p_{\text{para}} - p_{\text{zero}} = 1 - e^{-\lambda x}
\end{aligned}
\tag{23}
$$

In that case, we can write the composition recursion equation as

$$
x' = \frac{1}{1+\delta} \frac{p_{\text{ribo}}}{p_{\text{para}} + p_{\text{ribo}}},
\tag{24}
$$

The condition $dx'/dx|_{x=0} = 1$ yields the expression

$$
\lambda = (1+\delta)(1 - e^{-\lambda}),
\tag{25}
$$

for the asymptote. For $\lambda \ll 1$, we obtain using (25)

$$
\lambda = \frac{2\delta}{1+\delta},
\tag{26}
$$

which agrees very well with Fig 3.

Notice that here the coexistence phase is located to the right of the asymptotes, and the parasite phase to the left, whereas in Fig 2 it is the other way around. An intuitive way to understand this is to consider the limit $\lambda \to 0$. In this limit, nonempty compartments start with either a parasite or a ribozyme. The former will grow to a fully parasitic compartment, whereas the latter will contain ribozymes plus some parasites acquired by mutations. Therefore, at low λ, the ribozyme's capacity to outgrow parasites (competition) cannot be exploited, leading to ribozyme extinction. It is only when ribozymes and parasites are seeded together that the differential growth rate becomes important, which becomes increasingly likely for higher λ. The phase boundaries in Fig. 3 mark the point where enough compartments engage in competition to allow for ribozyme survival. The mutation strength δ compares mutation rate to competition. When $\delta \to 0$, there is enough competition to ensure coexistence for all λ.

3.3. Error catastrophe

An error catastrophe corresponds to a situation where the accumulation of replication errors eventually causes the disappearance of ribozymes. Since there are only a parasite (P) and a coexistence phase (C) in the model with mutations, the error catastrophe means that the coexistence
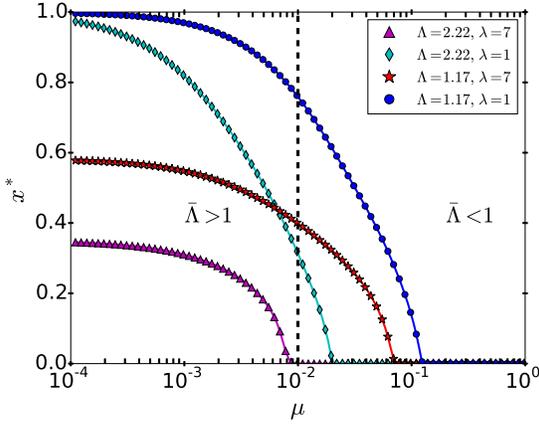
Figure 4: Steady state composition x^* as function of μ, $\alpha = 0.99$, $\gamma = 1.0$. Critical rates μ^* corresponds to separation between P and C phases in Fig. 5.
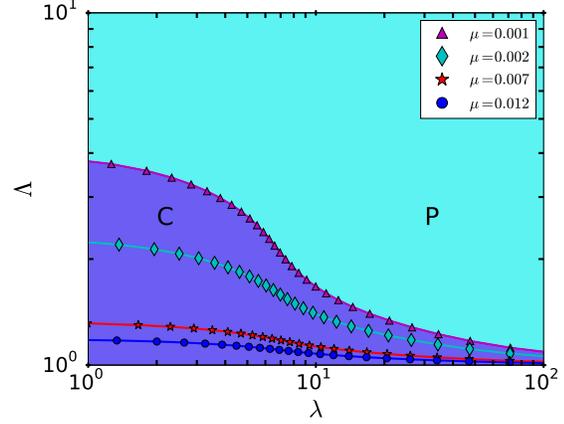
Figure 5: Phase diagram, drawn for f$\alpha = 0.99$, $\gamma = 1.0$. Separatrices are drawn for μ values close to μ^* in Fig. 4, corresponding to an error catastrophe.

region shrinks at the benefit of the parasite phase as the mutation rate increases. One sees this effect in Fig. 2, which corresponds to the prolific parasites regime ($\bar{\Lambda} \geq 1$) discussed above. In this figure, we see a larger coexistence region in the small λ region, because there the compartmentalization is efficient to purge parasites. As the mutation rate increases however, this region shrinks because the compartmentalization fails to purge the more numerous parasites.

In Fig. 4, a particular example is provided where α and γ are fixed, such that $\bar{\Lambda}$ is fixed, and μ is varied. Since competition is fixed, we have $\mu \propto \delta$. The resulting steady-state value $x = x^*$ then decreases monotonically with μ, and reaches $x = 0$ when crossing the phase boundary in Fig 5. For small values of λ, this boundary corresponds to the plateau region, for larger values, this corresponds to the $1/\lambda$ asymptote. As can be seen in Fig 4, coexistence is stable for much higher values of the mutation rate μ when the compartment size λ is small. This means that compartmentalization with selection leads to a relaxed error treshold with respect to the bulk.

The error catastrophe was also studied in the absence of selection, and was shown to be in the prolific ribozymes regime ($\bar{\Lambda} \leq 1$). In Fig. 6, an example of this case is shown, and there too, we see that the steady-state value of the ribozyme fraction x^* decreases as μ is increased, until it reaches the phase boundary in Fig 7. In contrast to Fig. 4, where the error threshold decreases as the size of compartments increases, the trend is just the opposite in Fig. 6, which is expected
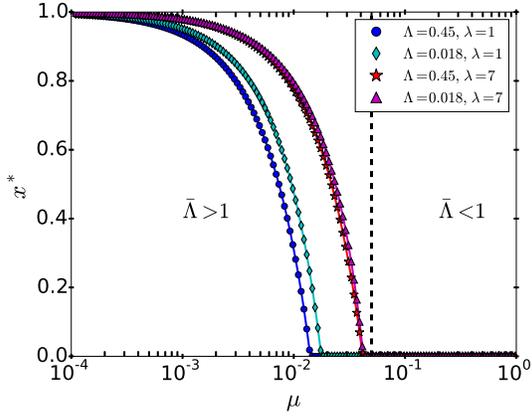
13

Figure 6: Steady state composition x^* as function of μ, $\alpha = 1.0, \gamma = 0.95$, in absence of selection ($f = 1.0$). Critical rates μ^* corresponds to separation between P and C phases in Fig. 7 .
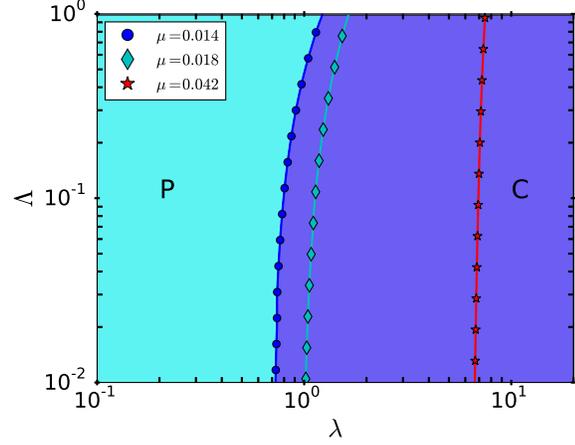
Figure 7: Phase diagram in absence of selection function ($f = 1.0$), drawn for $\alpha = 1.0, \gamma = 0.95$. Separatrices are drawn for μ values close to μ^* in Fig. 6.

since the role of ribozymes and parasites are exchanged here as compared to the prolific parasites regime.

In the prolific parasites regime, $\bar{\Lambda} \leq 1$ with selection, it is interesting to recast the error threshold as a constraint on the length of a polymer to be copied accurately, as done in the original formulation of the error threshold [10]. Let us introduce the error rate per nucleotide, ϵ. Then, for a sequence of length L, we have $\alpha - \mu = \alpha(1-\epsilon)^L$. Since $\epsilon \ll 1$, it follows from this that $\mu = \alpha\epsilon L$. When $\alpha \simeq \gamma$, we have $\ln \bar{\Lambda} = \alpha\epsilon LT$. Using Eq. (22), we find that the condition to copy the polymer accurately is

$$L \leq \frac{\ln(s)}{\epsilon \alpha T}, \tag{27}$$

where $s = f(\bar{x})/f(0)$ and $\alpha T/\ln 2$ is the number of generations. This criterium has a form similar to the original error threshold [10], namely

$$L \leq \frac{\ln(s')}{\epsilon}, \tag{28}$$

where $s' = \alpha/\gamma$ represents the selective superiority of the ribozyme. In our model, the equivalent of s' is s which characterizes the compartment selection.

14

4. Noise in growth

For deterministic growth, given by Eqs. (2)-(3), fluctuations in the growth rates, denoted α for the ribozymes and γ for the parasites, have been neglected. In order to estimate whether fluctuations in the growth rates of ribozymes and parasites could have a strong effect, we introduce in the next section a model for the replication process of such molecules by a replicating enzyme. This model includes noise due to the stochastic binding of the replicating enzyme to template molecules and the noise due to the stochasticity of monomer addition once the enzyme is bound to a template. Importantly, this model assumes that the replicase once bound stays always active until completion of the copy of the template, therefore the possibility that the replicase falls off the template before completion of the copy is neglected. Similarly, any effects associated with the interaction of multiple replicases on the same template are neglected. In fact, when the replicase falls off of its template, the copying process is aborted and the shorter chain which has been produced in this way becomes a parasite. We can therefore describe such a process as a mutation using the framework of the previous section. To separate the effects due to mutations and noise clearly, we disregard from now on the possibility of mutations, and we focus in the following on the description of the noise associated with replication. Such a noise can stabilize the ribozyme phase at the expense of coexistence, and the coexistence phase at the expense of the parasite phase. The noise of replication becomes very small when the rate-limiting step is nucleotide incorporation, in which case one can use a deterministic approach.

### 4.1. A minimal model for the replication process

The replication of an RNA strand A by a replicase can be considered to proceed through two stages. In the first stage, an RNA molecule A binds to a polymerase E, to form a complex X_0

$$A + E \xrightarrow{\kappa_C} X_0, \tag{29}$$

with the rate κ_C.

Subsequently, activated nucleotides X are incorporated in a stepwise fashion to the complementary strand. A complex of E and A with a complementary strand of length n will be denoted

by X_n, and the strand grows until the final length L is achieved, such that

$$X_n + \quad X \quad \xrightarrow{\kappa} X_{n+1}, \quad 0 \le n \le L - 2 \tag{30}$$

$$X_{L-1} + \quad X \quad \xrightarrow{\kappa} 2A, \tag{31}$$
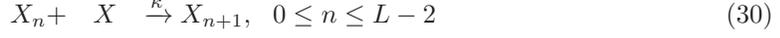
where for simplicity we have assumed the same rate k for both reactions. Let us denote by t the total time to yield $2A$ from A, which is the sum of the time associated with the step of complex formation, t_C and with the step of L nucleotide incorporations t_L. We thus have

$$t = t_C + t_L, \tag{32}$$

with $t_L = \sum_{i=0}^{L} t_i$ and t_i the time for adding one monomer, which we assumed is distributed according to

$$f(t_i) = \kappa e^{-\kappa t_i}. \tag{33}$$

For simplicity, we choose a single value κ for all monomer additions. The time for the formation of the complex, t_C is similarly distributed according to

$$f(t_C) = \kappa_C e^{-\kappa_C t_C}, \tag{34}$$

where $\kappa_C = 1/\langle t_C \rangle$.

Let us denote the moment generating function of t_C by $M_C(s)$ and similarly for t_L by $M_L(s)$ with :

$$
\begin{aligned}
M_C(s) &= \int_0^\infty dt_C \exp(-st_C) f(t_C) \\
&= \frac{\kappa_C}{s + \kappa_C}, \tag{35} \\
M_L(s) &= \int_0^\infty dt_L \exp(-st_L) f(t_L) \\
&= \prod_{i=1}^{L} \left[\int_0^\infty dt_i \exp(-st_i) f(t_i) \right] \\
&= \left(\frac{\kappa}{s + \kappa} \right)^L. \tag{36}
\end{aligned}
$$

From M_L one obtains the distribution of replication time $f(t_L)$ by performing an inverse Laplace transform:

$$f(t_L) = \mathcal{L}^{-1} \left[M_L(s) \right] = \frac{\kappa^L t_L^{L-1} e^{-\kappa t_L}}{\Gamma(L)}, \tag{37}$$

16

where $\mathcal{L}^{-1}$ represents the inverse Laplace transform. This equation shows that the replication time distribution of one strand of length L follows a Gamma distribution [30]. For $L = 1$, Eq. (37) becomes a simple exponential distribution, which is a memoryless distribution. For $L > 1$, this distribution has memory and the growth in the number of RNA strands can no longer be described as a simple Markov process. Note that the Gamma distribution is peaked around the mean value of t_L, namely L/κ for $L \gg 1$. In this limit, the replication time has very small fluctuations. This feature has recently been exploited to construct a single-molecule clock, in which the dissociation of a molecular complex occurs after a well-controlled replication time[31].

4.2. Coefficient of variation of the replication time

Let us now study the coefficient of variation of the full time t, which includes the diffusion of the replicase and the replication step. The generating function of t is clearly $M(s) = M_D(s)M_L(s)$. Thus, the cumulant-generating function defined as $K(s) = \ln M(s)$, yields the two moments of the distribution of t, namely the mean $\langle t \rangle$ and the variance σ_t^2. We have

$$\langle t \rangle = \langle t_C \rangle + \langle t_L \rangle = \frac{1}{\kappa_C} + \frac{L}{\kappa}, \tag{38}$$

$$\sigma_t^2 = \sigma_C^2 + \sigma_L^2 = \frac{1}{\kappa_C^2} + \frac{L}{\kappa^2}. \tag{39}$$

Thus the coefficient of variation of the replication time, namely $\sigma_t/\langle t \rangle$ is given by

$$\frac{\sigma_t}{\langle t \rangle} = \frac{\sqrt{\frac{1}{\kappa_C^2} + \frac{L}{\kappa^2}}}{\frac{1}{\kappa_C} + \frac{L}{\kappa}}. \tag{40}$$

Fig 8 shows this quantity as function of the length L and of the ratio of the rates (κ_C/κ).

There are two regimes: on one hand, when $L/\kappa \gg 1/\kappa_C$, the time taken by the replication step dominates over the time for the replicase to diffuse to its target. If in addition $\sigma_L^2 \gg \sigma_C^2$, the coefficient of variation of the time t scales as $1/\sqrt{L}$ and therefore becomes very small for long strands. This power-law regime in indeed visible as plateaus in Fig 8.

On the other hand, when $1/\kappa_C \gg L/\kappa$, the time to form a complex between the replicase and its template dominates over the replication time. This regime has a large coefficient of variation since $\sigma_t \simeq \langle t \rangle$ as also seen in Fig 8.
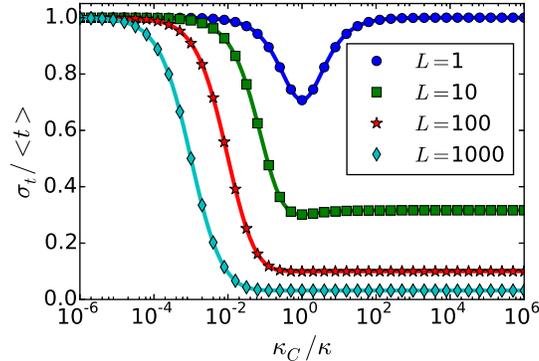
Figure 8: Waiting time variability $\sigma_t/\langle t \rangle$ for various polymer lengths L, as a function of the ratio of typical times for replication and complex formation

4.3. The generations representation

Let us look at these two growth regimes in a generations representations, where by generations we mean an event of copy of the template by the replicase. The diffusion-limited regime corresponds to Fig. 9, while the replication- limited regime corresponds to Fig. 10. In this representation, the differences in the two growth regimes become very clear. In the replication-limited regime, generations remain synchronized, until enough noise has accumulated over multiple generations. For two independent strains, generations become desynchronized after about $\sqrt{L}$ generations. In contrast, in the diffusion-limited regime, fluctuations are very large due to lack of synchronicity in the growth.

These figures have been obtained by simulating the growth of a replicating mixture starting from a single strand. The simulation follows k RNA-enzyme complexes, and for each the variable n_k measures the length of the growing complementary strand. For every nucleotide incorporation event, a strand i is chosen with probability $1/k$, after which its number of nucleotides is updated from n_i to $n_i + 1$. When $n_i + 1 = L$, we set $n_i = 0$, we update k to $k + 1$, and then we introduce an extra strand variable n_{k+1} for the new strand. Both the replication-limited regime and the diffusion-limited regime can be modeled using this simulation. In the latter case, we need to choose $L = 1$, which corresponds to exponentially distributed replication times.
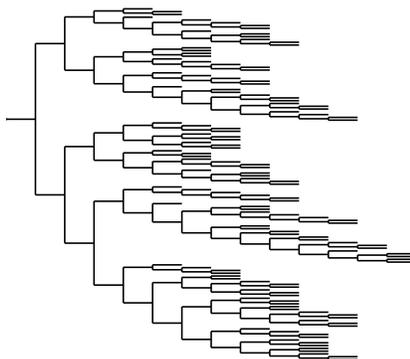
18

Figure 9: Generations representation of growth in the diffusion-limited regime. The simulation ends when the population size has reached 128. The horizontal axis corresponds to the generation number.
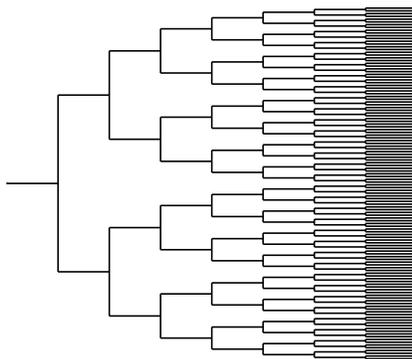

Figure 10: Generations representation of growth in the replication-limited regime. The simulation ends when the population size has reached 128. The horizontal axis corresponds to the generation number.

4.4. Population-level noise

In sec 4.2, we have analyzed the noise associated with the replication of a single strand. Ultimately, we wish to quantify the compositional variation of the final population. In order to do so, we turn to the theory of branching processes with variable lifetimes taken randomly from a fixed distribution [32]. As explained in Appendix A, this framework describes theoretically a population that grows exponentially starting from a single individual. In our molecular system, this single individual plays the role of the single molecule present in the initial condition before the replication starts; while the distribution of the lifetimes is the replication time distribution $f(t_L)$ obtained in Eq (37).

For $t_L \gg L/\kappa$, we find that the average population (starting from a single individual) $\mu^{(1)}$ scales as $\mu^{(1)}(t) = \mu^* e^{\alpha t}$, with a growth rate $\alpha \simeq \kappa \ln(2)/L$. The coefficient of variation of the population size $\sigma^{(1)}/\mu^{(1)}$ is

$$\frac{\sigma^{(1)}}{\mu^{(1)}} \approx \frac{\sqrt{2}\ln(2)}{\sqrt{L}}. \tag{41}$$

The renewal theory on which these results are based, can be generalized to the case that there are n individuals in the initial condition as shown in Appendix B. The full solution is rather complicated due to correlations between the subpopulations generated by the different molecules present in the initial condition. In the following, we neglect these correlations: therefore the n initial molecules generate n independent subpopulations, which all start at size 1 and follow the branching process described above and in Appendix A. In that case, each subpopulation now has a mean $\mu^{(1)} = \mu^{(n)}/n$ and a standard deviation $\sigma^{(1)} \approx \mu^{(1)}/\sqrt{L}$. This then allows to write

$$\sigma^{(n)} \approx \sqrt{n}\sigma^{(1)} = \frac{\mu^{(1)}}{\sqrt{nL}}. \tag{42}$$

We show in Fig. 11 that the corresponding coefficient of variation, $\sigma^{(n)}/\mu^{(n)}$, agrees well with simulations of the branching process. The 2000 simulation runs were stopped after a time t^* such that $\langle N(t^*)\rangle \simeq 5000$.

4.5. Fluctuations in logistic growth

The problem of two species competing for the same resources has been studied in the literature and offers a complementary perspective on the role of noise in a growing population, which has been studied in the previous section. Let us consider two such species, which typically start with a
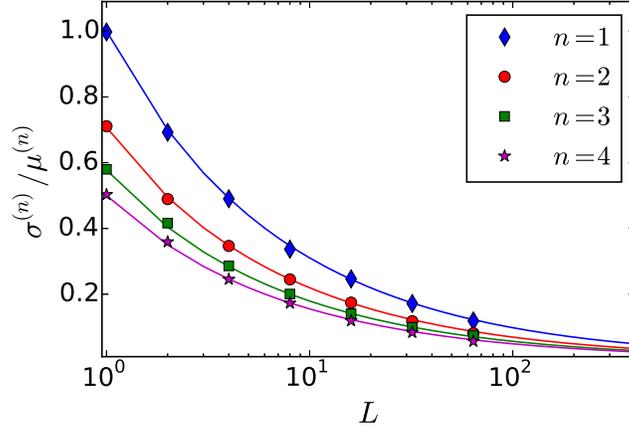
Figure 11: Coefficient of variation of the population size N as function of the initial population size n. The results have been averaged over 2000 runs. The solid lines represent the theoretical prediction: $1/\sqrt{nL}$.

few individuals and then grow according to logistic noise. As shown in Ref. [29], when the carrying capacity is reached, the number of each species is subject to giant fluctuations (the coefficient of variation is of the order of unity) when the two species have similar growth rates. In the terminology introduced in previous section, this model applies to the diffusion-limited regime ($L \to 1$), where a Markov description of the population dynamics is applicable.

Keeping the notations of the first section, we denote by n the initial number of molecules, which splits into m ribozymes and y parasites, and by N the final number of molecules in the compartment. In the neutral case ($\alpha = \gamma$), the moments of the number of ribozymes $\bar{m}$ are found to be [29] :

$$\langle \bar{m} \rangle = N \frac{m}{n}, \tag{43}$$

$$\sigma_{\bar{m}} = \sqrt{\frac{my}{n^2} \frac{N(N-n)}{n+1}}, \tag{44}$$

with again $y = n - m$. Since N remains fixed, $\sigma_{\bar{x}} = \sigma_{\bar{m}}/N$. This means that

$$\sigma_{\bar{x}} \approx \frac{1}{n} \sqrt{\frac{my}{n}} \tag{45}$$

for $N \gg n$, which means that the noise in the composition depends primarily on the number of individuals in the initial condition. Let us denote $s = \alpha/\gamma - 1 \ll 1$, with $s \ll 1$ and $\rho = \ln(N/n)$.

21

In Ref. [29], it was shown that

$$\frac{\sigma_{\bar{m}}}{\langle \bar{m} \rangle} = \sqrt{\frac{y}{m(n+1)}} \left(1 - \frac{\rho s n(m+1)}{(n+1)(n+2)} \right) \tag{46}$$

In general, the dynamics of the composition has a large variability for: (i) small compartments ($n \sim O(1)$), (ii) mixed compartments ($m, y > 0$), and for $m \approx y$, (iii) comparable growth rates ($s \to 0$).

Such a coefficient of variation is asymptotically constant on long times and the constant only depends on the initial number of molecules. A similar scaling for the coefficient of variation holds in a number of other physical situations, such as for the fluctuations in the number of protein filament formed in small volumes [28].

4.6. Noise in transient compartmentalization

Let us now apply the results of the section 4.4 to analyze the effect of the growth noise on our transient compartmentalization dynamics. Let us assume that the length of the ribozymes is L_α and that of the parasites L_γ. For experimental values of these parameters we refer the reader to Table 1. In section 2, we have defined m, y to be the initial number of ribozymes and parasites and $\bar{m}, \bar{y}$ to be the final mean number of ribozymes and parasites at the end of the growth phase in a given compartment. Using Eqs. (41)-(42), we obtain

$$\begin{aligned}
\frac{\sigma_{\bar{m}}}{\langle \bar{m} \rangle} &\simeq \frac{1}{\sqrt{L_\alpha m}}, \\
\frac{\sigma_{\bar{y}}}{\langle \bar{y} \rangle} &\simeq \frac{1}{\sqrt{L_\gamma y}}.
\end{aligned} \tag{47}$$

Since the ribozyme fraction $\bar{x}$ at the end of the exponential phase is given by $\bar{x}(n, m) = \bar{m}/N$ and $N \simeq n_{Q\beta}$, the noise on $\bar{x}(n, m)$ takes the following form :

$$\begin{aligned}
\sigma_{\bar{x}} &= \sqrt{\left(\frac{\partial \bar{x}}{\partial \bar{m}} \right)^2 \sigma_{\bar{m}}^2 + \left(\frac{\partial \bar{x}}{\partial \bar{y}} \right)^2 \sigma_{\bar{y}}^2)}, \\
&\simeq \sqrt{\left(\frac{\bar{y}}{N^2} \right)^2 \frac{\bar{m}^2}{mL_\alpha} + \left(\frac{-\bar{m}}{N^2} \right)^2 \frac{\bar{y}^2}{yL_\gamma}}, \\
&\simeq \bar{x}(1 - \bar{x}) \sqrt{\left(\frac{1}{mL_\alpha} + \frac{1}{yL_\gamma} \right)},
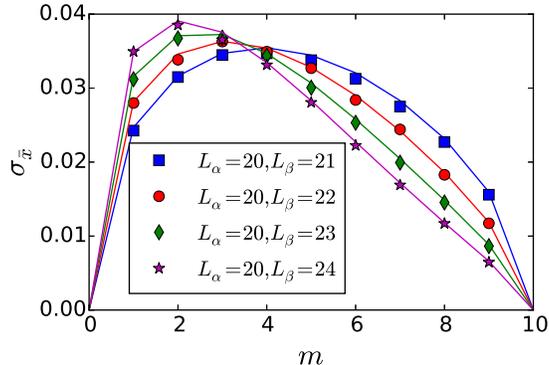\end{aligned} \tag{48}$$

22

Figure 12: Standard deviation of the ribozyme fraction, $\sigma_{\bar{x}}$, as predicted from simulations (symbols), and compared with predictions from Eq. (48) (solid lines). For each initial composition (m, n), 10000 simulations were performed until a time t^* such that $\langle N(t^*)\rangle \simeq 5000$ and by choosing $\alpha/\gamma = L_\gamma/L_\alpha$.

where we have used Eq. (42) with $\mu_{\bar{m}} = \bar{m}, \mu_{\bar{y}} = \bar{y}$. The factor $\bar{x}(1 - \bar{x})$ is largest for $\bar{x} = 1/2$ and vanishes for pure parasite and pure ribozyme compartments, which means that this noise can be neglected when $\Lambda \gg 1$ or $\Lambda \ll 1$. Note that if we choose $\alpha = \gamma$ (and thus $\bar{x} = m/n$), and $L_\alpha = L_\gamma = 1$, Eq. (48) becomes

$$\sigma_{\bar{x}} \simeq \frac{1}{n}\sqrt{\frac{my}{n}} \tag{49}$$

which is consistent with Eq. (45) which was found using a different formalism[29].

Using the parameters of Table 1 and (41), we can quantify the level of noise in the number of ribozymes or parasites. We find from this table that the ribozyme size was $L = 362$, and that the experiment should be in the replication-limited regime because the diffusion time scale should be approximately over $2 \cdot 10^4$ times smaller than replication times of the order of 10s. The noise in composition should be maximal when we start with one ribozyme and one parasite of equal length, and with $\alpha = \gamma$, which on average gives $\bar{x} = 1/2$. Consequently, the noise in composition is at most $\sigma_{\bar{x}} \approx 0.02$.

4.7. A weak noise approach

The growth equations given by Eqs. (2) and (3) are deterministic in nature, which means that a given initial condition (n, m) yields a unique final composition $\bar{x}(n, m)$. In contrast to that in a stochastic approach, a given n and m lead to many different trajectories, which means

that $\bar{x}(n, m)$ is a random variable with a probability distribution $p(\bar{x}(n, m))$. Consequently, the ribozyme fraction after one round is

$$x' = \frac{\sum\limits_{n,m} \int_0^1 d\bar{x} \bar{x}(n, m) p(\bar{x}(n, m)) f(\bar{x}) P_\lambda(n, x, m)}{\sum\limits_{n,m} \int_0^1 d\bar{x} p(\bar{x}(n, m)) f(\bar{x}) P_\lambda(n, x, m)}. \tag{50}$$

This expression is computationally demanding to evaluate for $\lambda \gg 1$, but it can be simplified significantly in the weak noise limit.

In order to construct a phase diagram in this limit, we simplify Eq. (50), by considering $p(\bar{x}(n, m)) \approx \mathcal{N}(\bar{x}, \sigma_{\bar{x}})$, where $\mathcal{N}$ denotes a normal distribution with mean $\bar{x}$ and standard deviation defined by Eq. (48). From Eq. (48) we expect the effect of noise to be largest when λ, L and Λ are close to 1 (if $\Lambda \gg 1, \bar{x} \to 0$). In Fig. 13, the original phase diagram from Ref. [15] is shown together with the modified phase boundaries (dotted lines) due to the presence of Gaussian noise using Eq. (50) for the case that $L_\alpha = L_\gamma = 3$.

Given that the amplitude of this type of noise should rapidly diminish for larger L, and that $L \sim O(100)$ in the experiment, we expect our ribozyme-parasite scenario to be well-described by a deterministic dynamics. We also see that the noise stabilizes the pure ribozyme phase (R) with respect to the coexistence phase (C) because in the presence of noise, the R region has grown at the expense of the C region. Similarly, the noise stabilizes the coexistence region (C) against the parasite region (P).

5. Conclusion

In this paper, we have carried out two important extensions of our previous work on transient compartmentalization [15], by including the effect of mutations and noise in such systems. This new study confirms one result of our previous work, namely that transient compartmentalization alone can stabilize functional replicators in the absence of a division of compartments of the kind considered in the Stochastic corrector model [13]. We can now add to that, that this property is robust with respect to mutations and noise, an important aspect for Origin of life studies.

In the presence of mutations, we have found that the phase diagram of long-time composition of this system only contains the parasite and the coexistence phases. The case where ribozymes
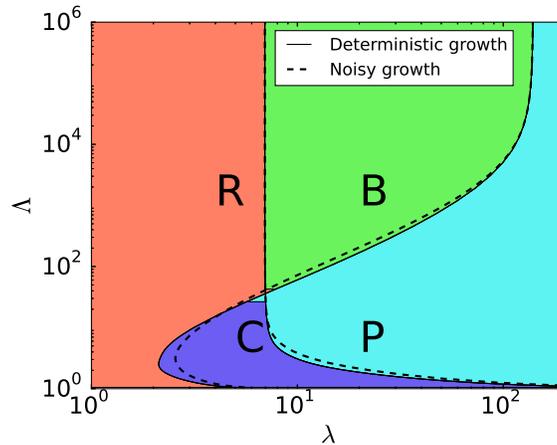
Figure 13: Phase diagram for ribozyme-parasite scenario in presence of noise given by Eq. (48), for $L_\alpha = L_\gamma = 3$.

grow faster than the parasites can be analyzed in terms of a modified error threshold, which interestingly now depends on the dynamics of compartmentalization and selection.

In order to analyze the role of noise in this system, we have introduced a simple model for the replication of a template by an enzyme. In the replication limited regime of that model, which should correspond to our experimental conditions, a low noise should be present at the population level, which we have quantified using tools borrowed from the theory of branching processes. In the end, we have studied the modified phase diagram of our model in this weak noise limit.

Of course, the two effects that we have studied here separately, namely mutations and noise, could be present simultaneously. We cannot also exclude that a more detailed modeling of the molecular replication or a different form of compartmentalization dynamics could lead to features not captured by the present treatment. Nevertheless, we think that the present framework represents a basis on which further studies could be built. We hope that our work may not only contribute to studies on the Origin of life but also to future developments on related important experimental techniques such as digital quantitative PCR [33] or Directed Evolution [34].

Acknowledgements

Appendix A. Population-level noise generated by a single individual in the initial condition

Let us consider an age-dependent renewal process, in which the probability density of branching at age t is given by $f(t)$, and upon branching, the probability of having k offspring is given by ϕ_k (assumed to be age-independent for simplicity). We would like to evaluate the behavior of the number $N(t)$ of individuals at time t. Let us define the function $h(s)$ by

$$h(s) = \sum_{k=0}^{\infty} \phi_k s^k. \tag{A.1}$$

We also define the generating function for the process $N(t)$ by

$$G(s,t) = \sum_{k=0}^{\infty} p_k(t) s^k, \tag{A.2}$$

where $p_k(t)$ is the probability that $N(t) = k$. We assume that $p_k(0) = \delta_{k1} = p_k^0$, i.e., that we start from a single object. We can then evaluate $p_k(t)$ by adding the probability that no branching has occurred between 0 and t, which is given by $1 - \int_0^t \mathrm{d}t\, f(t) = 1 - F(t)$, with the effect of the first branching at time u, such that $0 < u < t$. We obtain

$$p_k(t) = p_k^0 \left(1 - F(t)\right) + \sum_l \phi_l \int_0^t \mathrm{d}u\, f(u) \sum_{\{n_i\}} \delta_{\sum_{i=1}^{\ell} n_i, k} \prod_{i=1}^{\ell} p_{n_i}(t-u). \tag{A.3}$$

Multiplying by s^k and summing, we obtain

$$G(s,t) = s\left(1 - F(t)\right) + \int_0^t \mathrm{d}u\, f(u)\, h\left(G(s,t-u)\right). \tag{A.4}$$

Taking the derivative with respect to s at $s = 1$, we obtain the following equation for the average $\mu(t) = \sum_k k\, p_k(t)$:

$$\mu(t) = 1 - F(t) + m \int_0^t \mathrm{d}u\, \mu(t-u)\, f(u), \tag{A.5}$$

where $m = h'(1) = \sum_k k\, \phi_k$ is the average number of daughters upon branching.

To solve this equation in the limit $t \to \infty$, let us multiply both sides by $e^{-\alpha t}$ and take the limit. Since $\lim_{t \to \infty} F(t) = 1$, we obtain

$$
\begin{aligned}
\mu^* = \lim_{t \to \infty} \mu(t) e^{-\alpha t} &= \lim_{t \to \infty} \int_0^\infty du \, m\mu(t - u) \, e^{-\alpha(t-u)} \, e^{-\alpha u} \, f(u) \\
&= \mu^* m \int_0^\infty du \, e^{-\alpha u} f(u).
\end{aligned}
\tag{A.6}
$$

This equation allows for a solution different from 0 and ∞ if α is chosen to satisfy

$$
m \int_0^\infty du \, e^{-\alpha u} f(u) = 1.
\tag{A.7}
$$

Then, making use of a result by Smith [?], we obtain

$$
\frac{1}{\mu^*} = \frac{\alpha m^2}{m - 1} \int_0^\infty du \, u \, e^{-\alpha u} f(u).
\tag{A.8}
$$

As a consequence, we have

$$
\mu(t) \approx \mu^* \, e^{\alpha t}.
\tag{A.9}
$$

In the case we are considering we have

$$
f_L(t) = \frac{1}{\Gamma(L)} \kappa^L t^{L-1} e^{-\kappa t},
\tag{A.10}
$$

and $m = 2$, which yields

$$
\alpha = \kappa(2^{1/L} - 1) \approx \frac{\kappa \ln 2}{L},
\tag{A.11}
$$

giving, as long as $L \gg 1$,

$$
\mu(t) \approx \frac{2^{\kappa t/L}}{2 \ln 2}.
\tag{A.12}
$$

We can use this framework to also evaluate higher moments of the population size, and from that obtain the coefficient of variation of the population size which characterizes the amplitude of the noise. Let us denote the second derivative of the generating function with respect to s by ζ

$$
\zeta(t) = \left. \frac{d^2 G(s, t)}{d^2 s} \right|_{s=1} = \sum_{k=1}^\infty (k(k - 1)) p_k(t).
\tag{A.13}
$$

At large times, $\zeta(t) \approx \zeta^* e^{2\alpha t}$. The variance of the population size σ^2 follows from the standard relation:

$$
\sigma^2 = \zeta + \mu - \mu^2 \simeq \zeta - \mu^2.
\tag{A.14}
$$

27

For the specific case we are considering, we find

$$\zeta^* = \frac{2\mu^{*2}}{(2^{\frac{L+1}{L}} - 1)^L - 2}.$$ (A.15)

After extracting the leading contribution in the large L limit, we find:

$$\frac{\sigma}{\mu} \approx \frac{\sqrt{2}\ln(2)}{\sqrt{L}},$$ (A.16)

which is numerically close to $1/\sqrt{L}$ since $\sqrt{2}\ln 2 = 0.980.. \approx 1$.

Appendix B. Population-level noise generated from n individuals in the initial condition

If we start from n individuals rather than just one, we can write the probability to have k individuals at time t, $p_k^{(n)}(t)$, in terms of the subpopulations generated by n single individuals,

$$p_k^{(n)}(t) = \sum_{\{m_1,\ldots,m_n\}} \delta_{\sum_j m_j, k} \prod_{j=1}^n p_{m_j}^{(1)}(t).$$ (B.1)

Here, $p_k^{(1)}(t)$ denotes the probability of having a population size of k at time t, starting from one individual, which was considered in Appendix A. Note that we have added an addition superscript (1) to the notation used in Appendix A to emphasize the initial condition. From this equation, the new generating function follows :

$$G^{(n)}(s,t) = [G^{(1)}(s,t)]^n.$$ (B.2)

From this equation, we obtain the average,

$$\mu^{(n)}(t) = n\mu^{(1)}(t),$$ (B.3)

which expresses the average with n initial strands in terms of the average with one initial strand. For the second moment, we obtain

$$\zeta^{(n)} = n(n-1)[\mu^{(1)}]^2 + n\zeta^{(1)},$$ (B.4)

We can then extract $\sigma^{(n)}$ by using Eq. (A.14), which yields

$$\begin{aligned}[\sigma^{(n)}]^2 &= n(n-1)[\mu^{(1)}]^2 + n\mu^{(1)} - n^2[\mu^{(1)}]^2 + n\zeta^{(1)} \\ &= n\zeta^{(1)} + n\mu^{(1)}(1 - \mu^{(1)}).\end{aligned}$$ (B.5)

28

Together with Eq. (B.3), this leads to

$$\frac{\sigma^{(n)}}{\mu^{(n)}} \simeq \frac{\sqrt{\zeta^{(1)} - [\mu^{(1)}]^2}}{\sqrt{n}\mu^{(1)}} = \frac{\sigma^{(1)}}{\sqrt{n}\mu^{(1)}} \tag{B.6}$$

which is the coefficient of variation found previously for a single individual in the initial condition, divided by $\sqrt{n}$ as expected for the growth from independent individuals. This confirms the scaling found in Eq. (42).

References

[1] A. Jaiman, M. Thattai, Algorithmic biosynthesis of eukaryotic glycans, BioRxiv`doi:10.1101/440792`.

[2] A. I. Oparin, Origin of Life, Dover, 1952.

[3] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, A. A. Hyman, Germline p granules are liquid droplets that localize by controlled dissolution/condensation, Science 324 (5935) (2009) 1729–1732. `doi:10.1126/science.1172046`.

[4] D. Zwicker, R. Seyboldt, C. A. Weber, A. A. Hyman, F. Jülicher, Growth and division of active droplets provides a model for protocells, Nat. Phys. 13 (4) (2017) 408–413. `doi:10.1038/nphys3984`.

[5] E. Brinke, J. Groen, A. Herrmann, H. A. Heus, G. Rivas, E. Spruijt, W. T. S. Huck, Dissipative adaptation in driven self-assembly leading to self-dividing fibrils, Nature Nanotechnology 13 (9) (2018) 849–855. `doi:10.1038/s41565-018-0192-1`.

[6] K. K. Nakashima, J. F. Baaij, E. Spruijt, Reversible generation of coacervate droplets in an enzymatic network, Soft Matter 14 (2018) 361–367. `doi:10.1039/c7sm01897e`.

[7] P. G. Higgs, N. Lehman, The RNA world: molecular cooperation at the origins of life, Nat. Rev. Genet. 16 (1) (2015) 7–17. `doi:10.1038/nrg3841`.

[8] F. Dyson, Origins of life, Cambridge University Press, 1985.

[9] S. Spiegelman, I. Haruna, I. B. Holland, G. Beaudreau, D. Mills, The synthesis of a self-propagating infectious nucleic acid with a purified enzyme, Proc. Natl. Acad. Sci. USA 54 (3) (1965) 919–927. `doi:10.1073/pnas.54.3.919`.

[10] M. Eigen, Self-organization of matter and the evolution of biological macromolecules, Naturwissenschaften 58 (10) (1971) 465–523. `doi:10.1007/BF00623322`.

[11] J. Maynard Smith, E. Szathmáry, The Major Transitions in Evolution, Freeman, Oxford, 1995.

[12] N. Takeuchi, P. Hogeweg, Evolutionary dynamics of rna-like replicator systems: A bioinformatic approach to the origin of life, Physics of Life Reviews 9 (3) (2012) 219 – 263. `doi:10.1016/j.plrev.2012.06.001`.

[13] E. Szathmáry, L. Demeter, Group selection of early replicators and the origin of life, J. Theor. Biol 128 (4) (1987) 463–86. `doi:10.1016/S0022-5193(87)80191-1`.

[14] D. S. Wilson, A theory of group selection, Proc. Natl. Acad. Sci. USA 72 (1) (1975) 143–146. `doi:10.1073/pnas.72.1.143`.

[15] A. Blokhuis, D. Lacoste, P. Nghe, L. Peliti, Selection dynamics in transient compartmentalization, Phys. Rev. Lett. 120 (2018) 158101. `doi:10.1103/PhysRevLett.120.158101`.

[16] P. L. Luisi, P. Walde, T. Oberholzer, Lipid vesicles as possible intermediates in the origin of life, Curr. Op. Coll. Int. Sci. 4 (1) (1999) 33–39. `doi:10.1016/S1359-0294(99)00012-6`.

[17] M. Kreysing, L. Keil, S. Lanzmich, D. Braun, Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length, Nature chemistry 7 (2015) 203–208. `doi:10.1038/nchem.2155`.

[18] P. Baaske, F. M. Weinert, S. Duhr, K. H. Lemke, M. J. Russell, D. Braun, Extreme accumulation of nucleotides in simulated hydrothermal pore systems, Proc. Natl. Acad. Sci. U.S.A. 104 (22) (2007) 9346–9351. `doi:10.1073/pnas.0609592104`.

[19] E. V. Koonin, W. Martin, On the origin of genomes and cells within inorganic compartments, Trends Genet. 21 (12) (2005) 647–654. `doi:10.1016/j.tig.2005.09.006`.

[20] B. Damer, D. Deamer, Coupled phases and combinatorial selection in fluctuating hydrothermal pools: A scenario to guide experimental approaches to the origin of cellular life, Life 5 (1) (2015) 872–887. `doi:10.3390/life5010872`.

[21] T. Furubayashi, N. Ichihashi, Sustainability of a compartmentalized host-parasite replicator system under periodic washout-mixing cycles, Life 8 (3) (2018) 10. `doi:10.3390/life8010003`.

[22] S. Matsumura, A. Kun, M. Ryckelynck, F. Coldren, A. Szilágyi, F. Jossinet, C. Rick, P. Nghe, E. Szathmáry, A. D. Griffiths, Transient compartmentalization of RNA replicators prevents extinction due to parasites, Science 354 (6317) (2016) 1293–1296. `doi:10.1126/science.aag1582`.

[23] J. S. Chuang, O. Rivoire, S. Leibler, Simpson's paradox in a synthetic microbial system, Science 323 (5911) (2009) 272–275. `doi:10.1126/science.1166739`.

[24] A. S. Tupper, P. G. Higgs, Error tresholds for rna replication in the presence of both point mutations and premature termination errors, J. Theor. Biol. 428 (2017) 34–42. `doi:10.1016/j.jtbi.2017.05.037`.

[25] Y. E. Kim, P. G. Higgs, Co-operation between polymerases and nucleotide synthethases in the rna world, PLoS computational biology 12 (11) (2016) e1005161. `doi:10.1371/journal.pcbi.1005161`.

[26] L. Geyrhofer, N. Brenner, Coexistence and cooperation in structured habitats, bioRxiv`doi:10.1101/429605`.

[27] A. Zadorin, Y. Rondelez, Natural selection in compartmentalized environment with reshuffling, arXiv: 1707.07461.

[28] T. C. T. Michaels, A. J. Dear, T. P. J. Knowles, Stochastic calculus of protein filament formation under spatial confinement, New Journal of Physics 20 (5) (2018) 055007. `doi:10.1088/1367-2630/aac0bc`.

[29] B. Houchmandzadeh, Giant fluctuations in logistic growth of two species competing for limited resources, Phys. Rev. E 98 (2018) 042118. `doi:10.1103/PhysRevE.98.042118`.

[30] D. L. Floyd, S. C. Harrison, A. M. van Oijen, Analysis of kinetic intermediates in single-particle dwell-time distributions, Biophys. J. 99 (2) (2010) 360–366. `doi:10.1016/j.bpj.2010.04.049`.

[31] A. Johnson-Buck, W. M. Shih, Single-molecule clocks controlled by serial chemical reactions, Nano Letters 17 (12) (2017) 7940–7944. `doi:10.1021/acs.nanolett.7b04336`.

[32] S. Karlin, H. M. Taylor, A first course in stochastic processes, Academic Press, 1975. `doi:10.1016/C2009-1-28569-8`.

[33] B. J. Hindson, K. D. Ness, D. A. Masquelier, P. Belgrader, N. J. Heredia, A. J. Makarewicz, I. J. Bright, M. Y. Lucero, A. L. Hiddessen, T. C. Legler, et al., High-throughput droplet digital pcr system for absolute quantitation of dna copy number, Analytical chemistry 83 (22) (2011) 8604–8610. `doi:10.1021/ac202028g`.

[34] A. Dramé-Maigné, I. Golovkova, A. Zadorin, Y. Rondelez, Quantifying the performance of high-throughput directed evolution protocols, ArXiv: 1811.05288v1.