

The Emergence of Life as a First-Order Phase Transition

Cole Mathis,¹ Tanmoy Bhattacharya,^{2,3} and Sara Imari Walker^{4,5}

Abstract

It is well known that life on Earth alters its environment over evolutionary and geological timescales. An important open question is whether this is a result of evolutionary optimization or a universal feature of life. In the latter case, the origin of life would be coincident with a shift in environmental conditions. Here we present a model for the emergence of life in which replicators are explicitly coupled to their environment through the recycling of a finite supply of resources. The model exhibits a dynamic, first-order phase transition from *nonlife* to *life*, where the life phase is distinguished by *selection* on replicators. We show that environmental coupling plays an important role in the dynamics of the transition. The transition corresponds to a redistribution of matter in replicators *and their environment*, driven by selection on replicators, exhibiting an explosive growth in diversity as replicators are selected. The transition is accurately tracked by the mutual information shared between replicators and their environment. In the absence of successfully repartitioning system resources, the transition fails to complete, leading to the possibility of many frustrated trials before life first emerges. Often, the replicators that initiate the transition are not those that are ultimately selected. The results are consistent with the view that life's propensity to shape its environment is indeed a universal feature of replicators, characteristic of the transition from nonlife to life. We discuss the implications of these results for understanding life's emergence and evolutionary transitions more broadly. Key Words: Origin of life—Prebiotic evolution—Astrobiology—Biopolymers—Life. Astrobiology 17, 266–276.

1. Introduction

LIFE AND THE EARTH SYSTEM are tightly coupled (Smith and Morowitz, 2016). A prominent example is the dramatic change in Earth's atmosphere due to the evolution of oxygenic photosynthesis (Schirmer *et al.*, 2015). An important open question is whether life's propensity to shape its environment is a *universal* feature of life, characteristic of the origin of life itself, or if the coupling between life and its environment observed in our biosphere is a product of evolutionary optimization that may perhaps evolve in some planetary contexts but not others. One of the most distinctive features of life is replication—the ability to make copies—which engenders living matter with the capacity to sustain stable patterns of non-equilibrium behavior. Accordingly, numerous theoretical studies for the emergence of life have focused on the appearance of the first replicators, including identifying the conditions under which replicators can be selected from a prebiotic milieu (Szathmari and Maynard Smith, 1997; Nowak and Ohtsuki, 2008; Manapat *et al.*, 2009; Ohtsuki and Nowak, 2009; Wu and Higgs, 2009, 2012;

Walker *et al.*, 2012). Here we study a computational model for the emergence of replicating polymers, which includes coupling to an environment through recycling of a finite supply of resources, to address the role of the *environment* in driving the transition from nonliving to living matter. We demonstrate that a key property of selection for replication in prebiotic systems under resource-limited conditions is the feedback between replicators and their environment.

To study the role of environmental coupling, we consider a model prebiotic “replicative chemistry” with a finite supply of monomers, which must be recycled through polymer degradation to replenish resources available for synthesis and replication of polymers. By contrast, the majority of theoretical models for the emergence of replicators thus far have implemented reactor flows with a constant flux of monomers into the system and removal of chemical species via dilution, and therefore do not explicitly include feedback from the environment. Examples include the transition from *prelife* to *life* as studied by Nowak and Ohtsuki, (2008), where *prelife* is defined as a generative chemistry with no replication (polymerization only), to be contrasted

¹Department of Physics, Arizona State University, Tempe, Arizona.

²Santa Fe Institute, Santa Fe, New Mexico.

³Los Alamos National Laboratory, Los Alamos, New Mexico.

⁴Beyond Center for Fundamental Concepts in Science and School of Earth and Space Exploration, Arizona State University, Tempe, Arizona.

⁵Blue Marble Space Institute of Science, Seattle, Washington.

with *life*, where replicators are selected (Nowak and Ohtsuki, 2008; Manapat *et al.*, 2009; Ohtsuki and Nowak, 2009). In their model, a transition from prelife to life is observed by externally modulating the replication rate of polymers: above a critical rate constant for replication, replicating polymers can be selected. Similar features have been noted by Wu and Higgs (2009, 2012) and Szathmary and Maynard Smith (1997). By explicitly coupling replicators to their environment, we show this transition can occur *spontaneously*, without tuning any relevant rate parameters externally, and is dynamically driven by the environment. We also demonstrate that the abrupt nature of the observed transition shares many features in common with first-order phase transitions as characterized in the physical sciences. Here the two phases, which we nominally call *nonlife* and *life*¹, are distinguished, not by the discovery of replication but by the absence and presence of *selection* for replication. The dynamics observed demonstrate many of the hallmarks of *dynamic kinetic stability* (DKS) (Pross, 2005), where the life phase is characterized by the kinetically driven stability of self-replication. We discuss the implications of these results for furthering our understanding of the emergence of life and evolutionary transitions more broadly.

2. Methods

2.1. Model description

We model the emergence and dynamics of replicators in an artificial prebiotic “chemistry” that consists of two monomer types denoted by “0” and “1.” The properties of our model chemistry are fully specified by the rate constants k_p , k_d , and k_r for polymerization, degradation, and replication, respectively, the finite constant abundances of the two monomeric components 0 and 1, a constant r specifying the minimal length of replicating sequences, and fitness landscapes associated with sequence-specific replication and stability introduced below.

Polymerization occurs via addition of monomers to the end of growing sequences. Polymers can degrade into shorter sequences, which can occur at any bond within a given sequence with equal probability. To simplify the computational model, we adopted a common approximation in models for prebiotic polymerization that the inverse process of two short, but non-monomeric, sequences ligating to produce a longer polymer is sufficiently rare to be neglected (which would be the case, for example, if monomers are much more common than dimers) (Nowak and Ohtsuki, 2008; Manapat *et al.*, 2009; Walker *et al.*, 2012). All sequences of length $L \geq r$ can self-replicate such that polymers must be sufficiently “complex” to copy themselves (von Kiedrowski, 1986, 1993; Paul and Joyce, 2004). In this study, we set $r=7$, such that the appearance of the first replicators is rare but not so rare that we never observe it (Wu and Higgs, 2009). Changing r changes the relative timescale of the transition but does not qualitatively effect the results presented herein.

¹We realize that defining life is a much more complex issue than simply defining life as the ability to replicate and use the terminology “life” heuristically to indicate a distinction between the two phases relevant to a key property of life, that is, its ability to make copies.

Since we are interested in the dynamics of replication in this work, and specifically the origins of life, we do not include the effects of mutation, which is well known to play an important role in evolution once life has already emerged (Eigen, 2000) but is not expected to alter the qualitative features of the transition in the simplified model reported here. Therefore, in our model, replication only functions to copy extant sequences and does not produce novelty. Polymerization, however, does produce new sequences, and in our model, novelty is solely introduced through the prebiotic recycling of monomers via degradation and polymerization.

Simulations were implemented using a kinetic Monte Carlo algorithm (Gillespie, 1976, 1977). For more detailed discussion of the implementation of that algorithm in prebiotic recycling chemistries, we refer the reader to the works of Walker *et al.* (2012) or Vaidya *et al.* (2013). In what follows, the polymerization, degradation, and replication rate constants were set to $k_p=0.0005$, $k_d=0.5000$, and $k_r=0.0050$, respectively, and the system was initialized with 500 monomers each of 0 and 1 with no polymers present, unless otherwise noted. Since the reaction network is a closed mass system, the initial conditions specify the bulk composition of the system for all time.

2.2. Two fitness landscapes: static and dynamic

To explicitly couple the properties of replicators to those of their environment, we model the fitness of replicators as determined by two factors:

- A *static* fitness associated with a trade-off between stability and replicative efficiency that is an intrinsic “chemical” property of individual polymer sequences.
- A *dynamic* fitness associated with resource availability in the environment.

The former concept of *static* fitness encompasses the components of the fitness landscape associated with the properties of specific polymer sequences, without taking any account of the availability of resources. We choose to model selection for replication and stability as a trade-off since in many real-world chemical systems molecules that fold well are typically not good self-replicators, and conversely good self-replicators often do not fold well and are thus less resistant to degradation (Szabó *et al.*, 2002). The latter concept of *dynamic*, environmentally dictated fitness accounts for the availability of resources (free monomers) in the system and is a unique feature of the resource-dependent replication model presented here (see also Walker *et al.*, 2012, or Vaidya *et al.*, 2013).

Static Fitness. An important question in any model for the emergence of life is how sensitive the observed dynamics are to model parameters or to the fitness landscape imposed in the absence of empirical data. In what follows, we therefore consider several static fitness landscapes, which vary in how well the composition of the “fittest” sequence (or sequences) matches the resource availability in the bulk environment. This allows us to determine whether it is selection on replicators generally or features specific to a particular fitness landscape (and its relationship to abiotic resource distributions) that drive the dynamics observed. We consider four cases:

- *Landscape I (LI): Replicators and stable sequences are rare.* For the first example, replicative efficiency increases with the number of 0 monomers in a sequence, and stability with the number of 1 monomers, such that

homogeneous all-0 sequences with $L \geq 7$ are the best replicators, and all-1 sequences are the most stable. Since the bulk composition of the prebiotic environment consists of an equal number of 0 and 1 monomers in our simulations, and polymerization does not favor any specific bond type, good replicators and stable sequences are very rarely produced abiotically, and their composition does not reflect the abiotic distribution of resources.

- **Landscape II (\mathcal{LII}): Replicators are rare; stable sequences are common.** For the second example, replicative efficiency increases with the number of 00 bonds in a given sequence, while 01 bonds increase stability. This allows replicators to be rare as determined by the rate of spontaneous polymerization, whereas stable sequences are more readily produced. Even in this case, the composition of the best replicators does not match the bulk composition of the environment, but the composition of the most stable sequences now does.
- **Landscape III (\mathcal{LIII}): Replicators and stable sequences are common.** For the third example, replicative efficiency increases with the number of 10 bonds in a given sequence, while 01 bonds increase stability. For this fitness landscape, both efficient replicators and stable sequences are readily produced via spontaneous polymerization processes, and both the best replicators and most stable sequences reflect the composition of the environment.
- **Landscape IV (\mathcal{LIV}): All sequences with $L \geq 7$ replicate with equal efficiency.** For the final example, all sequences of length $L \geq 7$ replicate with equal efficiency [as would occur if replication were environmentally driven (Walker *et al.*, 2012)]. We use this as a control to determine whether the features observed are intrinsic to a selection for *replicative fitness* or a more general property of selection for *replication* and the transition from prebiotic polymerization processes to more “life-like” template-based replication.

For \mathcal{LI} – \mathcal{LIII} , the mathematical form for the trade-off between replication and stability is quantified as

$$f(n) = 0.5 + \frac{n^2}{2(10+n^2)} \quad (1)$$

following the work of Szabó *et al.* (2002), who implemented a similar trade-off among attributes of replicating polymers. In our implementation, the replicative fitness of a sequence x_i with length L is quantified by scaling its replication rate k_r by a parameter $\alpha_r(x_i) = 1 + f(n)$ (which is a number between 1.5 and 2), where n is the quantity that confers replicative efficiency (*i.e.*, 0 monomers, 00 bonds, and 10 bonds in \mathcal{LI} – \mathcal{LIII} , respectively). Similarly, the stability of sequence x_i is determined by scaling the degradation rate k_d by the parameter $\alpha_s(x_i) = 1 - f(m)$ (which is a number between 0 and 0.5), where m is the quantity that confers stability (*i.e.*, 1 monomers for \mathcal{LI} , 01 bonds for \mathcal{LII} and \mathcal{LIII}). Sequences with $L < 7$ do not replicate, so only the stability landscape is relevant for short sequences. In each example, this establishes a fitness landscape *intrinsic* to a polymer’s specific sequence that is fixed within a given environmental context.

Dynamic Fitness. Since we are interested in the coupling between replicators and their environment, we also intro-

duce an *extrinsic*, dynamic term to replicative efficiency, which is determined by the availability of free monomers in the environment. To this end, the replication rate for sequence x_i is weighted by a factor $\beta(x_i) = \sum_{n_i}^{L-1} y_{n_i} y_{n_i+1}$ where y_n is the abundance of the monomer species at position n in sequence x_i . This term yields a computationally tractable resource-dependent replication rate that is also sequence-dependent. This term may be motivated as a sum over all possible nucleation events on a template [see, *e.g.*, Vaidya *et al.* (2013) for an explicit example as applies to ribozyme recycling]. As such, the replication rate of a given sequence x_i depends in part on how well its sequence composition matches the relative abundances of 0 and 1 monomers in the environment. Since the abundances of 0 and 1 monomers change over time as monomers are consumed via polymerization and replication and generated via degradation, this creates a dynamic, environmentally dictated fitness landscape that is a central feature of any resource-constrained dynamics. We expect qualitative features of the dynamics observed here to be a general feature of species-specific and resource-dependent replication, independent of the particular functional form of $\beta(x_i)$.

2.3. Tracking the selection of replicators with mutual information

To characterize the dynamics of the observed phase transition, we employ mutual information, a common tool in information theory, which measures the mutual dependence of two variables within a dynamic time series by quantifying how much information the two variables share in common. We use mutual information, \mathcal{I} , to measure the extent to which the composition of replicators is determined by their environment, and vice versa. We define the sets R and E , which contain ordered pairs that track the number of 0 and 1 monomers in replicators (R) and in free monomers in the environment (E), allowing us to measure $\mathcal{I}(R : E)$, defined as the mutual information shared between replicators and their environment. See the appendix for additional details.

3. Results

3.1. A first-order phase transition from nonlife to life

Two long-lived states are observed, which we nominally call *nonlife* and *life*. These two phases are dominated by polymer formation via polymerization or via replication, respectively. While the nonlife phase here shares features in common with prelife as previously characterized (Nowak and Ohtsuki, 2008), it also has some striking differences—we therefore use nonlife rather than prelife (in our system, many transitions fail to complete, so the emergence of replicators and life is not inevitable; thus nonlife is more appropriate).

For fixed values of k_p , k_d , and k_r , the system exhibits a spontaneous and abrupt transition from nonlife to life for all four fitness landscapes investigated. The transition is accurately tracked by measuring the mutual information $\mathcal{I}(R;E)$ between the composition of extant replicators and free monomer resources, as shown in Fig. 1. The details of the transition are dependent on the static fitness landscape chosen, but the transition exists *independent* of the nature of the landscape chosen. For \mathcal{LI} (Fig. 1, top left), mutual

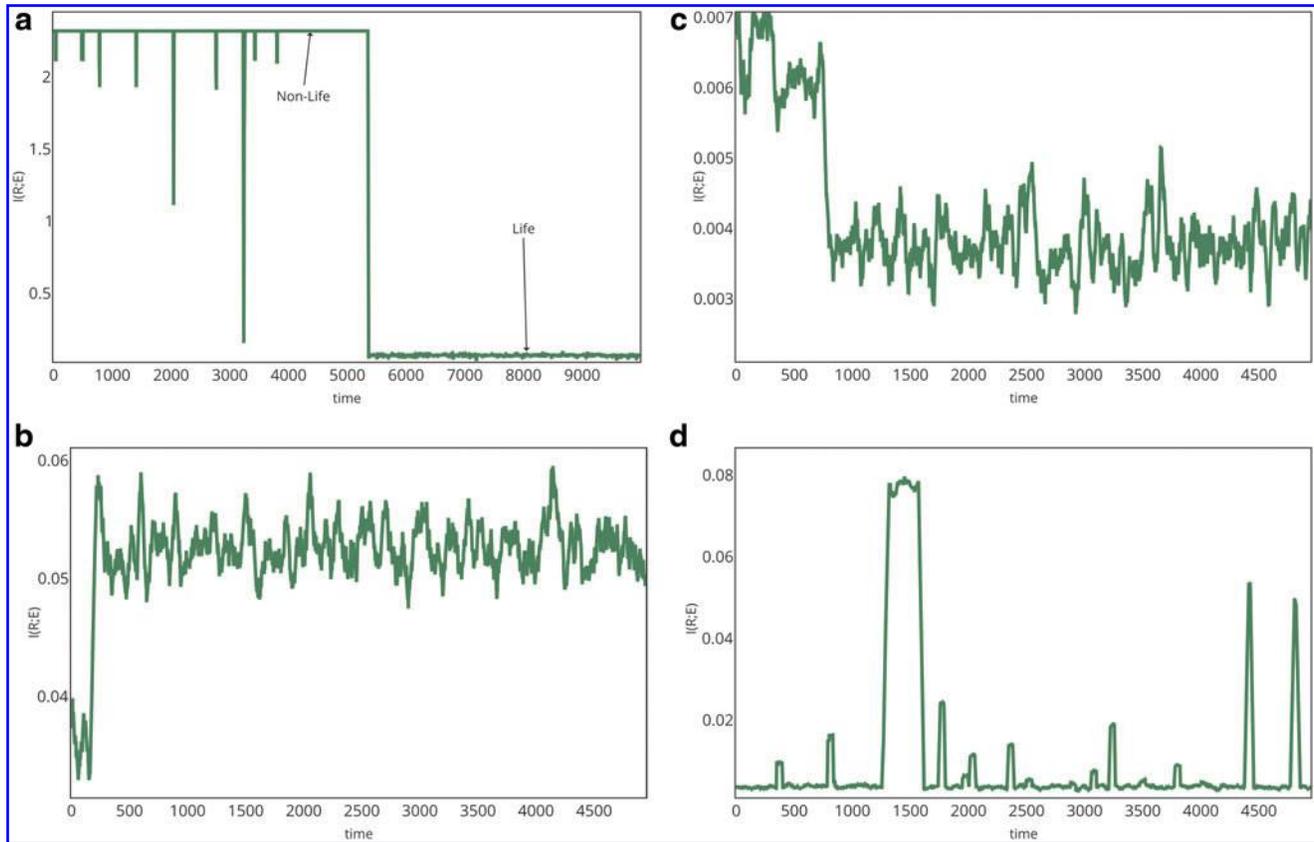


FIG. 1. Typical time series of the mutual information shared between replicators and their environment (free monomers) for each fitness landscape. Top left: $\mathcal{L}I$. Top right: $\mathcal{L}II$. Bottom left: $\mathcal{L}III$. Bottom right: $\mathcal{L}IV$. The phase transition is clearly evident in the abrupt shift (or shifts) in $\mathcal{I}(R;E)$ observable in each case.

information *decreases* through the transition, since the composition of selected replicators differs from that of their environment. This is true also of $\mathcal{L}II$ (Fig. 1, top right); however, the magnitude of the difference in $\mathcal{I}(R;E)$ between the two phases is less dramatic due to selection of stable, less efficiently replicating sequences whose composition does match bulk resources, in addition to efficient replicators whose composition does not. For $\mathcal{L}III$ (Fig. 1, bottom left), mutual information *increases* since the composition of selected replicators reflects the bulk composition of their environment. For $\mathcal{L}IV$ (Fig. 1, bottom right), where there is no selection on specific sequences, the transition is visible in the rapid changes in the mutual information between a state of high and a state of low mutual information, as stochastically determined replicators are selected and accumulate through exponential growth, until a large fluctuation leads to a different replicator that dominates. For $\mathcal{L}IV$, the transition is frustrated for the particular choice of parameters: sequences are transiently selected, but the system settles back before the sequences dominate. The absolute value of $\mathcal{I}(R;E)$ and the size of the fluctuations are sensitive to the fitness landscape imposed; however, the presence of an abrupt, spontaneous transition is apparent in each case and accurately tracks the onset of selection on the properties of replicators.

Since the transition is apparent for each landscape explored, we focus on $\mathcal{L}I$, as we expect this landscape to exhibit the most interesting dynamics, given that replicators and stable sequences are rare and not readily produced

prebiotically. We also expect this case to be the most realistic, as functional biopolymers are sparse in sequence space and often do not share their composition with the ambient environment.

The dynamics of the transition from nonlife to life displays many hallmarks of physical first-order phase transitions. The results of a Kolmogorov–Smirnov (KS) test suggest that the distribution of wait times until the transition successfully completes is consistent with the expected exponential distribution (Fig. 2, panel a) (Smith and Morowitz, 2016). The frequency that the transition occurs is dependent on both the composition of replicators and the environment (Fig. 2, panel b). This is distinct from other models that do not account for environmental feedback (Nowak and Ohtsuki, 2008; Wu and Higgs, 2009)—here, the transition is not coincident with the first “discovery” of a sequence capable of replication, since replicators can exist in nonlife. Instead, the transition occurs when replicators and the environment share similar resource distributions (high extrinsic fitness). Since both monomer species are equally abundant in the initial distribution of resources for the examples reported here, the nucleation event is typically mediated by a heterogeneous replicator (or replicators) composed of a roughly equal number of 0s and 1s (this is true for all four landscapes). These are not the sequences that are ultimately selected in the life phase for $\mathcal{L}I$, which include only the homogeneous, fit sequences. Given that the transition is spontaneous, abrupt, and exponentially distributed, we

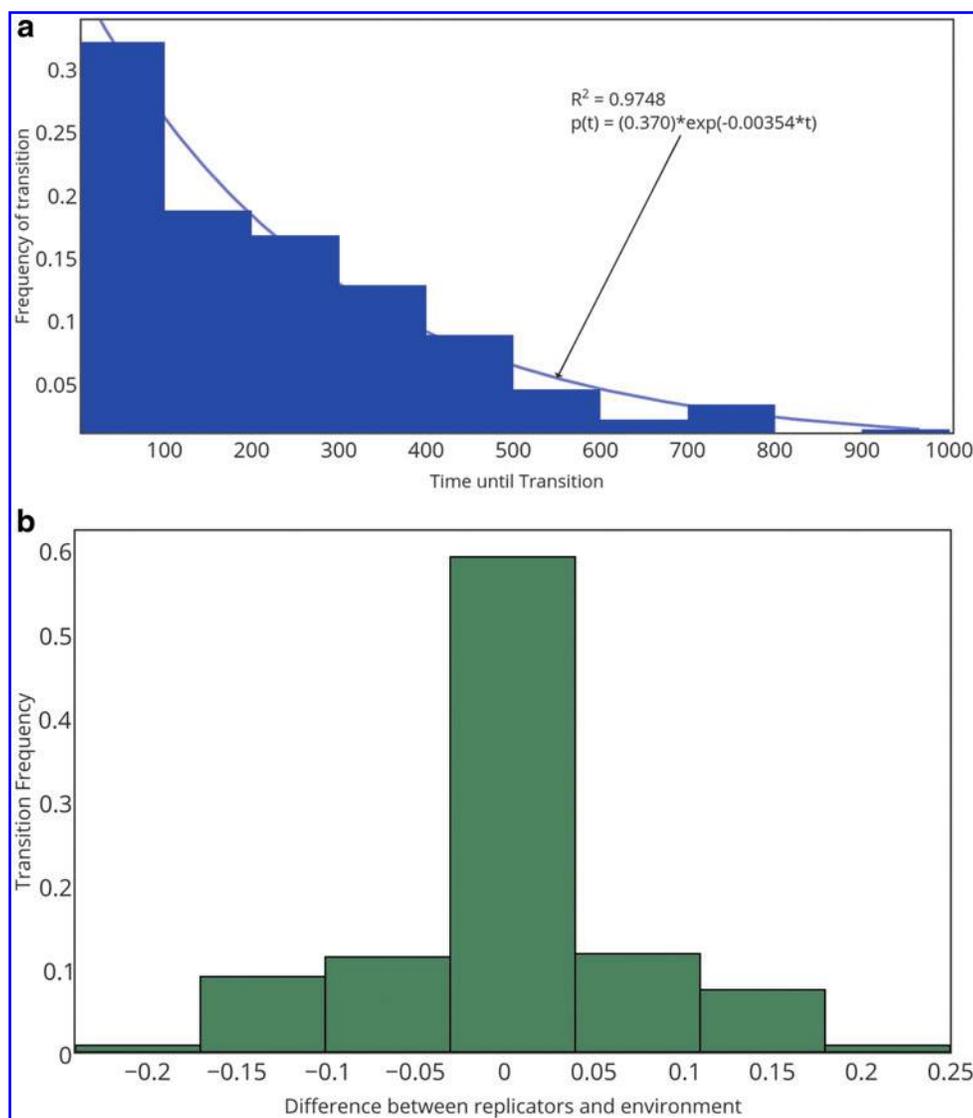


FIG. 2. Panel **a**: The distribution of waiting times until the phase transition occurs for $\mathcal{L}I$ is shown. It follows an exponential distribution, indicative of a first-order phase transition due to large fluctuations. Panel **b**: The frequency of successful phase transitions for $\mathcal{L}I$ as a function of the difference between the composition of extant replicators and their environment. Data shown is for an ensemble statistic of 256 simulations. Simulation parameters were set as $k_p = 0.0005$, $k_d = 0.5000$, and $k_r = 0.0050$ for both figures.

consider the dynamics to be indicative of a genuine first-order phase transition. We note that, similar to other first-order transitions, there are often many frustrated transitions prior to a successful phase transition (see Fig. 1, top left), which can occur when lack of selection on fit sequences leads to a failure in the transition to run to completion (see also Fig. 1, bottom right).

3.2. The dynamics of the transition from nonlife to life

In the nonlife phase, long sequences are exponentially rare, and the majority of system mass is in monomers and dimers (not shown). Sequences of all lengths have relatively similar composition, as shown in the top panel of Fig. 3. The composition of extant polymers is reflective of the combined effects of the abiotic availability of resources and, for $\mathcal{L}I$ – $\mathcal{L}III$, the stability landscape established by Eq. 1.

Replicators can exist in the nonlife phase, albeit at exponentially low abundance. These typically have compositions reflective of the abiotic distribution of resources and form via polymerization.

In the life phase, the composition of replicators need not reflect the bulk composition of the environment; instead, replicator composition is determined by *selection* of the fittest sequences. This can in turn lead to restructuring of the distribution of resources in the environment, as shown in Fig. 3 for $\mathcal{L}I$, where selected replicators are primarily homogeneous 1s or 0s. Due to resource constraints, selection on replicators drives a transition in the composition of shorter sequences. In the life phase, short sequences obtain the opposite compositional signature to that of replicators (bottom panel, Fig. 3). The compositional reversal is seen only below $L=6$. Although $L=6$ sequences cannot replicate, they are formed primarily via degradation of $L=7$

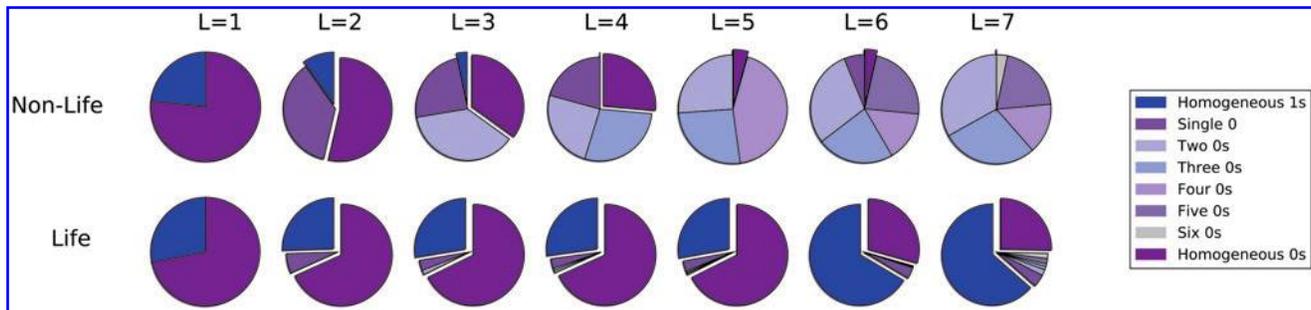


FIG. 3. Ensemble averaged compositions of all sequences with $L \leq 7$ for \mathcal{LI} . The distributions in the top panel characterize the nonlife phase (no selection on replicators) and in the bottom panel characterize the life phase (selection on replicators). Data is averaged over 100 simulations, and simulation parameters are $k_p = 0.0005$, $k_d = 0.5000$, and $k_r = 0.0050$.

replicators; thus their formation is dominated by self-replication (via formation, then degradation of $L=7$ sequences). In the life phase, replicators are selected based on their *intrinsic* fitness and not strictly how well their composition matches the environment. The defining feature of the life phase is therefore not necessarily the presence of replicators, which exist in both phases. Instead, the defining characteristic of “life” in this model is that the distribution of resources is dictated by *selection* on the properties of replicators, and that selection only operates in the life phase.

Figure 3 demonstrates that, due to resource constraints, the selection of replicators can coincide with dynamic restructuring of the entire system [including both monomer and nonreplicating ($L < 7$) sequence populations]. Figure 4 shows an ensemble averaged phase space trajectory through this restructuring for \mathcal{LI} . The phase transition moving from nonlife to life phases is highly unstable and dominated by degradation. In both the nonlife and life phases, polymer formation rates (polymerization and replication) balance rates for polymer degradation, with ratios of formation/

degradation ~ 1 . However, the life and nonlife phases are clearly distinguished in phase space by very different values for the mutual information [here $\mathcal{I}(R;E) \sim 3.0$ for nonlife and $\mathcal{I}(R;E) \sim 0.25$ for life, for results in Fig. 4].

The rampant degradation observed through the phase transition results in a rapid and dramatic restructuring of the extant polymer population and a steep slope in the rate of sequence exploration, as observed in Fig. 5. This is characteristic of the phase transition from nonlife to life, independent of the replicative fitness landscape. This restructuring arises as a result of reallocation of mass from shorter sequences to replicators, which must occur via degradation to monomers that can then be consumed via replication. The extant diversity and the rate of introduction of new sequences are both higher in the life phase than the nonlife phase (Fig. 5), which is attributable to the higher turnover rate of resources in the life phase (due to the higher assembly rate of polymers via replication).

Shown in Fig. 6 is an example time series for the evolution of all sequences with $L=7$ for \mathcal{LI} , binned by sequence

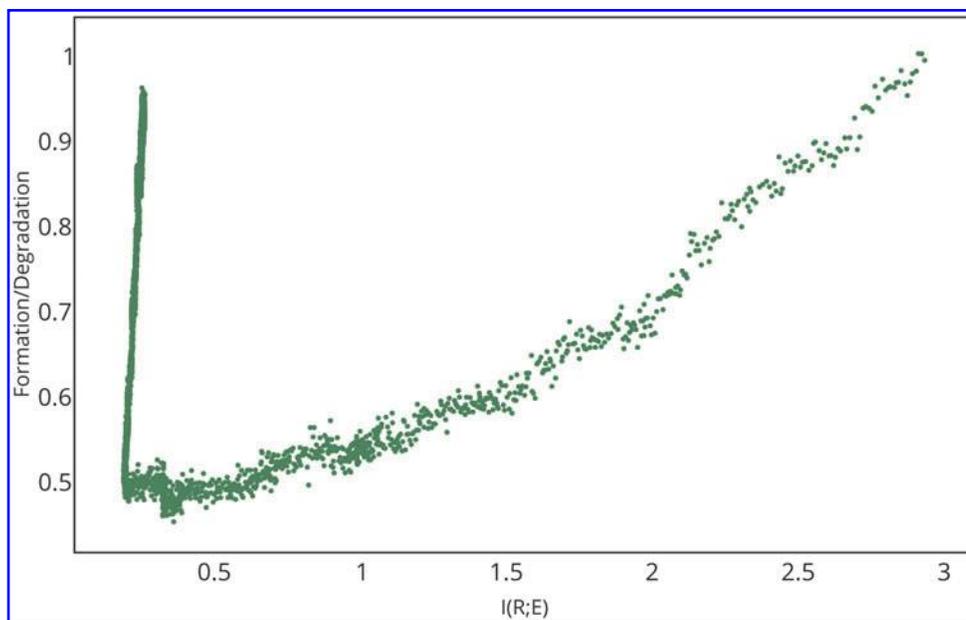


FIG. 4. Phase trajectory for an ensemble of 100 systems transitioning from nonlife to life (plotted vs. time, the system would move from left to right). Axes are the mutual information $\mathcal{I}(R;E)$ between replicators and environment (x axis) and the ratio of formation (polymerization and replication) to degradation rates (y axis). Simulation parameters are $k_p = 0.0005$, $k_d = 0.5000$, and $k_r = 0.0050$.

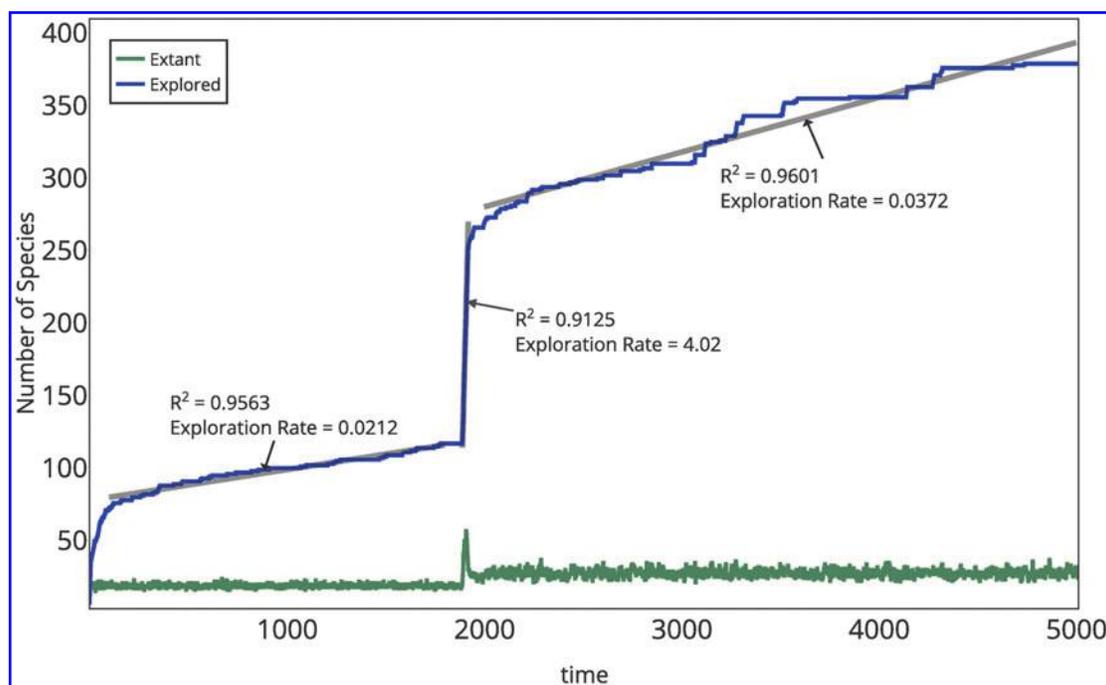


FIG. 5. Exemplary time series for the extant species population size and total number of sequences explored by the system. Linear fits to the explored species are shown. The exploration rate is 75% faster during the life phase compared to the nonlife phase and is 2 orders of magnitude larger during the transition. Simulation parameters are $k_p=0.0005$, $k_d=0.5000$, and $k_r=0.0050$.

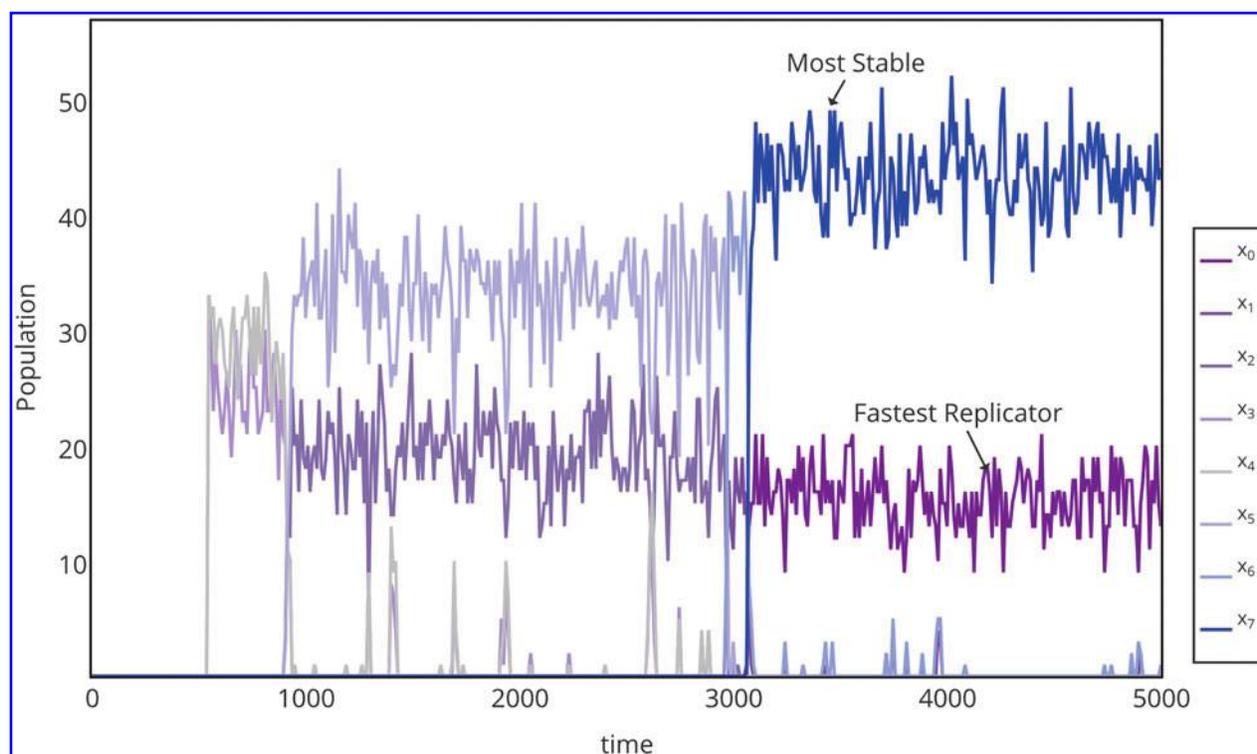


FIG. 6. Series of transitions in the selection of fit, homogeneous “0” and “1” length $L=7$ replicators for \mathcal{LI} . Here, the subscript denotes the number of “1” monomers in the sequence (e.g., x_0 contains no “1s,” x_1 bins all polymers with a single “1” monomer, and x_7 contains all “1s”). Simulation parameters are $k_p=0.0005$, $k_d=0.9000$, and $k_r=1.000$.

composition, for a set of parameters where the transition is prolonged enough to resolve details of the restructuring. Resource constraints enforce selection of sequences in complementary pairs that maintain the symmetry of the bulk resource distribution of the environment (50% 0s and 50% 1s). The system subsequently undergoes a series of abrupt transitions associated with increasing sequence homogeneity, where replicator composition increasingly departs from that of the bulk environment.

3.3. The timescale for life's emergence

The phase transition from nonlife to life described here is a robust feature of the dynamics, observed for different fitness landscapes with qualitatively similar features. Here, we demonstrate that the observed features are also qualitatively robust over a large range of parameter values. Quantitative differences arise in the final abundances of replicators and in the timescale for the transition to occur, which are both sensitive to the specific details of the prebiotic chemistry under consideration. Figure 7 shows the average time to complete the phase transition as a function of the degradation and replication rate constants, k_d and k_r , for $\mathcal{L}I$. For the results presented, the transition was identified as complete when 75% of the total replicating mass was allocated in homogeneous (fit) sequences.

One might *a priori* expect the transition to be most rapid (favored) for fast replication (high k_r) and slow degradation (low k_d); however, this is not always observed. For high degradation rate $k_d = 5.0$, the time to the transition is largely independent of k_r (Fig. 7). Lowering the degradation rate ($k_d = 1.0$ and $k_d = 0.5$, Fig. 7) increases the dependence of the transition time on k_r , which, on average, occurs most rapidly for relatively low k_r . This counterintuitive behavior arises as

a result of the resource constraints. For high degradation rates, there is a high rate of turnover that increases the likelihood of discovering functionally fit sequences, but the probability of survival is low, so the transition time is long regardless of replicative efficiency. For lower degradation rates, high replication rates lock resources in less fit sequences, frustrating the system's restructuring and leading to long transition times.

The rate of degradative recycling seems to be the primary factor in determining the transition timescale. Figure 8 shows the transition time observed for different abiotic resource abundances, quantified by the ratio R of the total number of 1 monomers to total system mass. The transition timescale is not expected to be symmetric with respect to the relative abundance of 0 and 1 monomers. For large values of R (environments rich in 0 monomers that confer stability), where recycling is inherently slower, the average transition time may be much longer than in environments with fewer stable polymers. Our data support this expectation, although the variation in transition times is large. These features suggest that environments that engender degradative recycling at a moderate rate may be the most conducive to nucleating the origin of life under resource-limited conditions.

4. Discussion

We have demonstrated the existence of a spontaneous transition from nonlife to life, which arises due to explicit incorporation of environmental feedback and displays many features in common with first-order physical phase transitions. It might be argued that the dynamics reported here do not represent a true phase transition. In the study of equilibrium physical systems, free energy is the quantity that is minimized to determine the state of the system (Goldenfeld,

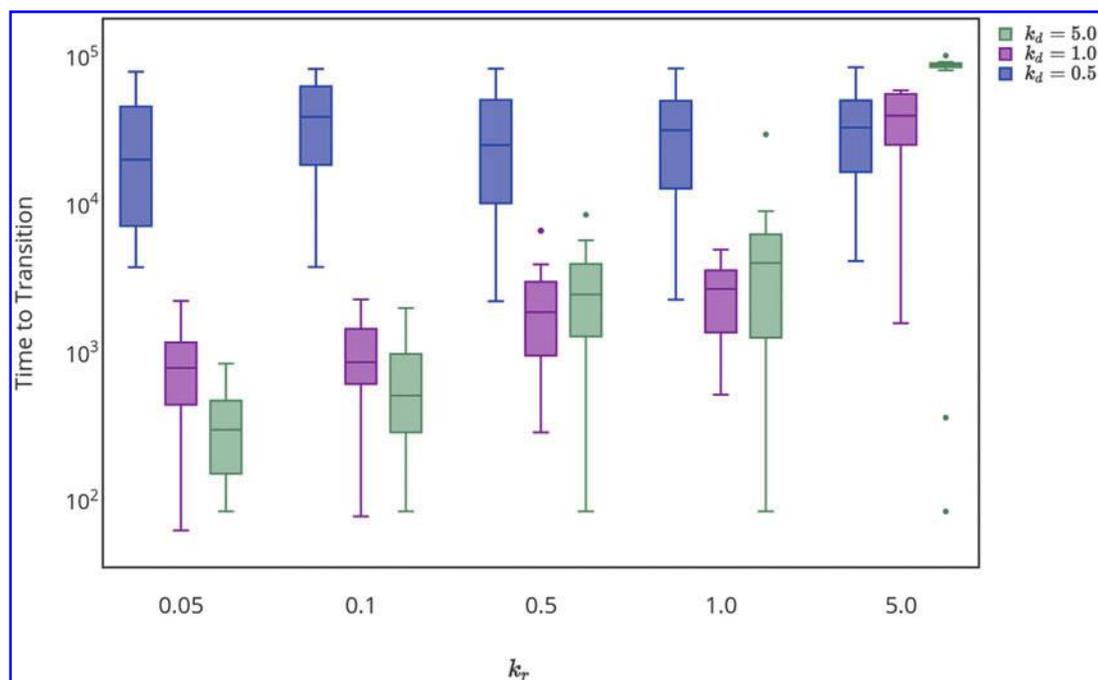


FIG. 7. Timescale for completing the phase transition as a function of reaction rate constants for replication k_r . Data from 25 simulations is shown; all data points are included in the box and whisker plots. The center line for each distribution is the median, the boxes contain half the data points, and the bars show the range. Three values of the degradation rate constant k_d are shown: 5.0 (blue), 1.0 (purple), 0.5 (green). The polymerization rate constant was fixed at $k_p = 0.0005$.

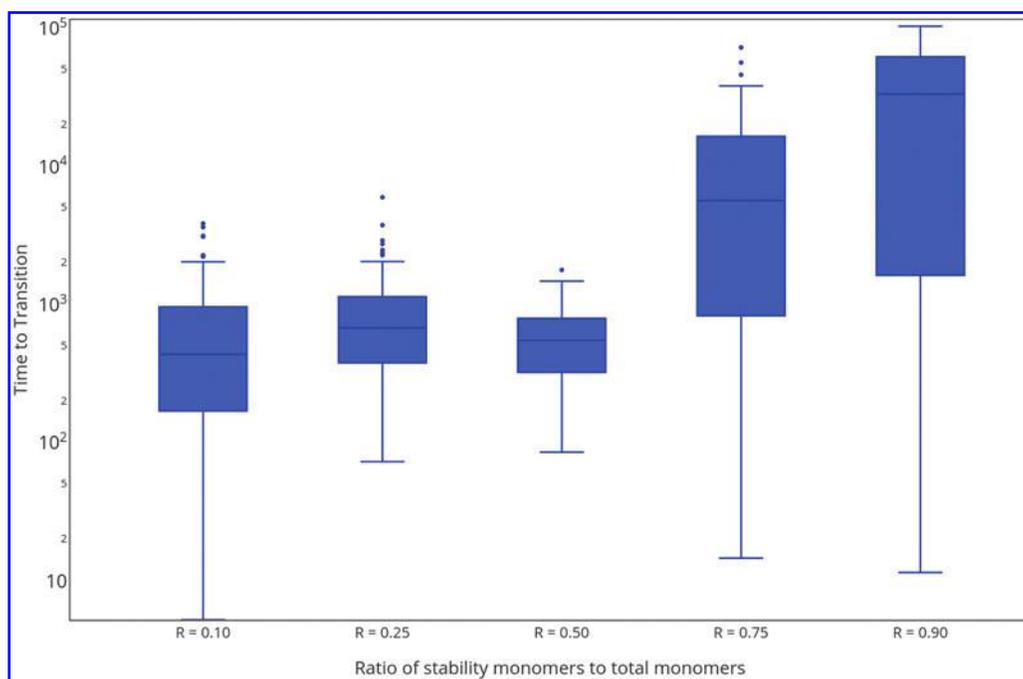


FIG. 8. Timescale for completing the phase transition as a function of the abiotic distribution of resources. Here, the parameter R is the ratio of “1” monomers (which confer stability) to total system mass. Data from over 100 simulations is shown; all data points are included in the box and whisker plots. The center line for each distribution is the median, the boxes contain half the data points, and the bars show the range. Parameters: $k_p=0.0005$, $k_d=0.5000$, and $k_r=0.0050$.

1992). Typically, this involves a payoff between minimizing total energy and maximizing entropy. When these two favor different results, a system is expected to exhibit a first-order phase transition from order to disorder. Here, in our dynamical scenario, a similar tradeoff happens between two processes that consume and try to minimize the number of free monomers [which may be related to the minimization of free energy (Amend *et al.*, 2013)]. These two different ways—*viz.*, maximizing the number of bonds via polymerization or maximizing the number of polymers via replication—yield distinct results with a sharp boundary between them, which motivates the classification of the observed dynamics as a phase transition. Future work will detail whether this is merely a useful analogy or indicative of a deeper connection.

Importantly, the most distinguishing feature of the life phase in our model is not the presence of replicators, since these can also exist in nonlife. Instead it is *selection* on the properties of replicators (such as replicative efficiency and stability in the examples presented here). Selection in turn necessitates a redistribution of matter due to limited resource availability. This restructuring is coincident with a sharp transition in the mutual information shared by replicators and their environment. Previous work connecting information theory to life’s origins has reported that the probability to discover a self-replicator by chance should depend exponentially on the availability of monomers composing it (Adami, 2015). Our results demonstrate an additional necessary feature: in the case of resource-constrained replication, replicators and environment share a similar composition (*e.g.*, have high mutual information). This enables exponential growth of the replicator population based on high dynamic fitness, which in turn enables selection

on the properties of new replicators discovered. When the fittest replicators do not match the bulk composition of the system, they force a redistribution of resources to accommodate their selection. We further note that very few measures have been proposed to explicitly quantify the origin-of-life transition. Here, mutual information between replicators and environment accurately measures the progress of the phase transition reported (perhaps acting as an order parameter), independent of the specific attributes of the replicator selected. Future work should elucidate the relationship between the fitness landscape and the system dynamics and magnitude of the mutual information. This will help identify how broadly applicable this approach is and perhaps provide insights to other candidate scenarios for the origin of life, such as in the formation of autocatalytic sets (Nghe *et al.*, 2015).

While we have nominally identified selection on replicators with “life” in this simple model, we note that the presence of replication is perhaps a necessary, though not sufficient, criterion to define life (see, *e.g.*, Walker and Davies, 2012), which remains an important open philosophical and scientific question (Mix, 2015). The information-theoretic characterization of this transition is consistent with proposals that life is most defined by its informational properties (Walker and Davies, 2012; Adami, 2015) [here, replicators might be interpreted as driving the dynamics of the entire system in a “top-down” manner due to adaptive selection (Ellis, 2012)]. The life phase may be interpreted as a state where the kinetics of individual replicators (*e.g.*, as quantified by their replicative efficiency and stability) dictate the behavior of the entire system, which is consistent with the notion that life is a kinetically driven state of matter (Pross, 2005). Although our motivation is to understand the origin of life utilizing this model system, we note that the

model is sufficiently general to capture features that may be universal to a broader class of evolutionary transitions. In particular, the dynamics could be universally characteristic of the discovery of novel, selectable patterns in the distribution of resources among replicating populations. For example, the abrupt nature of the transition shares features in common with punctuated equilibrium (Gould and Eldredge, 1977). The dynamics of this phase transition also demonstrate behavior that may be characteristic to niche construction and/or mass extinctions. In particular, the system's restructuring necessitates a period of instability driven by rampant destruction of extant diversity (extinction), which is followed by an explosion in novel diversity. The relationship to the phase transition reported here could be tested, for example, by analyzing the connection between resource distribution patterns and abrupt evolutionary transitions in the evolutionary record of life on Earth.

Interestingly, the features most characteristic of the phase transition reported are heavily dependent on degradative recycling of finite resources, which mediates selection on fit sequences by recycling less fit ones and yields an abrupt transition due to rapid resource reallocation. This suggests new perspectives regarding the role of degradation in the origin of life, which is typically viewed as an impediment in prebiotic chemistry, rather than a process central to early evolution (Atkins *et al.*, 2005). Cast under new light in the resource-constrained dynamics observed here, it is perhaps not a coincidence that RNA, as a biopolymer that played a prominent role in early evolution, is highly susceptible to hydrolysis, perhaps resolving an apparent paradox in the origin of life (Benner, 2014). The properties of this phase transition are in principle testable in the laboratory in experimental systems that permit recycling of biopolymers, for example, as reported in the work of Vaidya *et al.* (2013). In particular, the observed dynamics should place further constraints on the kinds of chemistries (defined by relative rates k_d , k_r , and k_p and relative resource abundances) that are most conducive to mediating the transition from nonlife to life (see also, *e.g.*, Walker *et al.*, 2012).

Due to the explicit coupling between replicators and environment, the transition reported here displays many features one might expect for a newly emergent biosphere that are not observable in open-flow reactor models. In particular, restructuring during the phase transition drives a vast increase in extant diversity and in the rate of exploration of novel diversity. This indicates that the emergence of life should coincide with an explosive growth of novelty in resource-limited systems. Concomitantly, during the transition, the system is dramatically restructured, indicating that the emergence of life should have significantly altered the environment of early Earth. It is well known that biology alters its environment over evolutionary and geological timescales and that the presence of life defines many features of the Earth system. Our results indicate that this may be a universal characteristic of life, from the very first appearance of replicators, and is most dramatic in cases where life is composed of sequences rarely produced abiotically.

The model includes the possibility of many frustrated trials before life first emerged (see, *e.g.*, Fig. 2), with success entailing a transformation of the environment as a necessary component of the process of biogenesis [perhaps consistent with the notion of a "Gaian bottleneck" (Chopra and

Lineweaver, 2016)]. These features indicate that it should be difficult to retrace the precise history of the origin of life: the replicators that are ultimately selected will, in general, neither be reflective of the ancestral planetary environment from which life first emerged, nor will they be representative of the replicators that first nucleated the origin of life. Thus, as is often suggested, here we see an explicit example that the conditions favoring the emergence of life may not be the same as those favoring its subsequent evolution.

Finally, we point out that simple replicators such as those presented here may not be the most effective architecture for a self-reproducing system. In this model, the total composition of the system remains fixed; what life does is restructure the distribution of matter within the system, due to the propagation of selectable replicating resource allocation patterns. An interesting open question is how this phase transition might play out for more lifelike replicative systems, such as those with the architecture of a von Neumann self-reproducing automata (von Neumann, 1966; Walker and Davies, 2012; Marletto, 2015), a subject we leave to future work.

A. Appendix

Herein, we explicitly measure the mutual information between two variables *as a time-series variable* itself to track the progress of the phase transition from nonlife to life. To generate a time series for mutual information, we use the *pointwise mutual information*, \mathcal{P} . Given two random variables $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_m\}$, \mathcal{P} is quantified as

$$\mathcal{P}(x_i : y_i) = \log \frac{p(x_i, y_i)}{p(x_i)p(y_i)} \quad (\text{A1})$$

(Csiszar and Korner, 2011), where $p(x_i)$ and $p(y_i)$ are the probabilities of observing the event where X is in state x_i and Y is in state y_i , respectively, and $p(x_i, y_i)$ is the joint probability of this event occurring. We generated probability distributions by counting the frequency of a given event (*e.g.*, abundance of 0 and 1 monomers and of replicators of a given sequence composition) in our time series data. In the results presented here, the distributions were generated using time series data from an ensemble of 100 experimental runs over 10,000 time steps each. To ensure that the frequency-based probability distributions were not biased by counting states from different phases of the system (see below), the frequencies were generated from data that sampled equally from both phases. The probabilities of different states therefore represent ensemble statistics that do not depend on time, while the particular ordering of states in a time series is used to determine the time series \mathcal{P} . In stochastic systems, such as ours, \mathcal{P} will fluctuate rapidly in time and is unlikely to yield useful insights. We therefore sum \mathcal{P} over a fixed time window to yield the mutual information for that window. Explicitly, for a window size of w , the mutual information at time t is defined as the average of $\mathcal{P}(x_i : y_i)$ and is given by

$$\mathcal{I}(X(t) : Y(t)) = \sum_{i=t-(w/2)}^{t+(w/2)} p(x_i, y_i) \log \frac{p(x_i, y_i)}{p(x_i)p(y_i)} \quad (\text{A2})$$

(Csiszar and Korner, 2011). This value will depend on time, not because the probabilities of different states will depend

on time, but rather the realization of different states is time ordered. Determining an appropriate size for w is important. If w is too large, the entire measurement collapses into one value yielding no insights into how the system is evolving in time. By contrast, if w is too small, fluctuations wash out interesting larger scale structure. We chose w heuristically, such that the value of $\mathcal{I}(R;E)$, tracking the mutual information between replicators and environment, was relatively constant, but large fluctuations could still be resolved. For the results presented $w=100\Delta t$, where $\Delta t=0.1k_h^{-1}$ is the resolution of the time series data in natural units. We note that different values of w change the results quantitatively but not qualitatively: the system still maintains a nonzero value of the mutual information in the nonlife phase which tends toward zero in the life phase.

Acknowledgments

This project/publication was made possible through support of a grant from Templeton World Charity Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of Templeton World Charity Foundation. The authors wish to thank Paul C.W. Davies and Nigel Goldenfeld for constructive conversations on this work and the Aspen Center for Physics (supported in part by the National Science Foundation under grant no. PHY-1066293) for hosting S.I.W. and T.B., where the initial seeds of the idea that nucleated this project were matched with the right environment.

References

Adami, C. (2015) Information-theoretic considerations concerning the origin of life. *Orig Life Evol Biosph* 45:309–317.

Amend, J.P., LaRowe, D.E., McCollom, T.M., and Shock, E.L. (2013) The energetics of organic synthesis inside and outside the cell. *Philos Trans R Soc Lond B Biol Sci* 368, doi:10.1098/rstb.2012.0255.

Atkins, J.F., Gesteland, R.F., and Cech, T.R., editors. (2005) *The RNA World*, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Benner, S.A. (2014) Paradoxes in the origin of life. *Orig Life Evol Biosph* 44:339–343.

Chopra, A. and Lineweaver, C.H. (2016) The case for a Gaian bottleneck: the biology of habitability. *Astrobiology* 16:7–22.

Csiszar, I. and Korner, J. (2011) *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Cambridge University Press, Cambridge, UK.

Eigen, M. (2000) Natural selection: a phase transition? *Biophys Chem* 85:101–123.

Ellis, G.F.R. (2012) Top-down causation and emergence: some comments on mechanisms. *Interface Focus* 2:126–140.

Gillespie, D.T. (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 22:403–434.

Gillespie, D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361.

Goldenfeld, N. (1992) *Lectures on Phase Transitions and the Renormalization Group*, Addison-Wesley, Advanced Book Program, Reading, MA.

Gould, S.J. and Eldredge, N. (1977) Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3:115–151.

Manapat, M., Ohtsuki, H., Bürger, R., and Nowak, M.A. (2009) Originator dynamics. *J Theor Biol* 256:586–595.

Marletto, C. (2015) Constructor theory of life. *J R Soc Interface* 12, doi:10.1098/rsif.2014.1226.

Mix, L.J. (2015) Defending definitions of life. *Astrobiology* 15:15–19.

Nghe, P., Hordijk, W., Kauffman, S.A., Walker, S.I., Schmidt, F.J., Kemble, H., Yeates, J.A., and Lehman, N. (2015) Prebiotic network evolution: six key parameters. *Mol Biosyst* 11:3206–3217.

Nowak, M.A. and Ohtsuki, H. (2008) Prevolutionary dynamics and the origin of evolution. *Proc Natl Acad Sci USA* 105:14924–14927.

Ohtsuki, H. and Nowak, M.A. (2009) Pre-life catalysts and replicators. *Proc R Soc Lond B Biol Sci* 276:3783–3790.

Paul, N. and Joyce, G.F. (2004) Minimal self-replicating systems. *Curr Opin Chem Biol* 8:634–639.

Pross, A. (2005) On the emergence of biological complexity: life as a kinetic state of matter. *Orig Life Evol Biosph* 35:151–166.

Schirrmeister, B.E., Gugger, M., and Donoghue, P.C.J. (2015) Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology* 58:769–785.

Smith, E. and Morowitz, H.J. (2016) *The Origin and Nature of Life on Earth: The Emergence of the Fourth Geosphere*, Cambridge University Press, Cambridge, UK.

Szabó, P., Scheuring, I., Czárán, T., and Szathmáry, E. (2002) *In silico* simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* 420:340–343.

Szathmáry, E. and Maynard Smith, J. (1997) From replicators to reproducers: the first major transitions leading to life. *J Theor Biol* 187:555–571.

Vaidya, N., Walker, S.I., and Lehman, N. (2013) Recycling of informational units leads to selection of replicators in a prebiotic soup. *Chem Biol* 20:241–252.

von Kiedrowski, G. (1986) A self-replicating hexadeoxy-nucleotide. *Angew Chem Int Ed Engl* 25:932–935.

von Kiedrowski, G. (1993) Minimal replicator theory I: Parabolic versus exponential growth. In *Bioorganic Chemistry Frontiers*, edited by H. Dugas and F.P. Schmidtchen, Springer, Berlin, pp. 113–146.

von Neumann, J. (1966) *Theory of Self-Reproducing Automata*, edited by A.W. Burks, University of Illinois Press, Urbana, IL.

Walker, S.I. and Davies, P.C.W. (2012) The algorithmic origins of life. *J R Soc Interface* 10, doi:10.1098/rsif.2012.0869.

Walker, S.I., Grover, M.A., and Hud, N.V. (2012) Universal sequence replication, reversible polymerization and early functional biopolymers: a model for the initiation of prebiotic sequence evolution. *PLoS One* 7, doi:10.1371/journal.pone.0034166.

Wu, M. and Higgs, P.G. (2009) Origin of self-replicating biopolymers: autocatalytic feedback can jump-start the RNA world. *J Mol Evol* 69:541–554.

Wu, M. and Higgs, P.G. (2012) The origin of life is a spatially localized stochastic transition. *Biol Direct* 7, doi:10.1186/1745-6150-7-42.

Address correspondence to:

Sara I. Walker
School of Earth and Space Exploration
Arizona State University
PO Box 871404
Tempe, AZ 85287

E-mail: sara.i.walker@asu.edu

Submitted 9 February 2016

Accepted 22 September 2016

This article has been cited by:

1. Neveu Marc, Hays Lindsay E., Voytek Mary A., New Michael H., Schulte Mitchell D.. 2018. The Ladder of Life Detection. *Astrobiology* **18**:11, 1375-1402. [[Abstract](#)] [[Full Text](#)] [[PDF](#)] [[PDF Plus](#)]
2. William B. Miller, John S. Torday, František Baluška. 2018. Biological evolution as defense of 'self'. *Progress in Biophysics and Molecular Biology* . [[Crossref](#)]
3. Gloria A. A. Saracino, Federico Fontana, Shehrazade Jekhmane, João Medeiros Silva, Markus Weingarth, Fabrizio Gelain. 2018. Elucidating Self-Assembling Peptide Aggregation via Morphoscanner: A New Tool for Protein-Peptide Structural Characterization. *Advanced Science* **5**:8, 1800471. [[Crossref](#)]
4. Pamela Knoll, Oliver Steinbock. 2018. Inorganic Reactions Self-organize Life-like Microstructures Far from Equilibrium. *Israel Journal of Chemistry* **58**:6-7, 682-692. [[Crossref](#)]
5. William B. Miller. 2018. Biological information systems: Evolution as cognition-based information management. *Progress in Biophysics and Molecular Biology* **134**, 1-26. [[Crossref](#)]
6. Sara Imari Walker, Cole Mathis. Network Theory in Prebiotic Evolution 263-291. [[Crossref](#)]
7. Douglas Moore, Sara I Walker, Michael Levin. 2017. Cancer as a disorder of patterning information: computational and biophysical perspectives on the cancer problem. *Convergent Science Physical Oncology* **3**:4, 043001. [[Crossref](#)]
8. Sara Imari Walker. 2017. Origins of life: a problem for physics, a key issues review. *Reports on Progress in Physics* **80**:9, 092601. [[Crossref](#)]