Thresholds in Origin of Life Scenarios

Cyrille Jeancolas, Christophe Malaterre, Philippe Nghe

Thresholds in origin of life scenarios

Cyrille Jeancolas[1,2], Christophe Malaterre[3], Philippe Nghe[1*]

[1], Laboratoire de Biochimie, CBI, ESPCI Paris, Université PSL, CNRS, 75005 Paris, France

[2], Laboratoire d'Anthropologie Sociale, Collège de France, 52 rue du Cardinal Lemoine, 75005 Paris, France

[3], Département de philosophie and Centre de recherche interuniversitaire sur la science et la technologie (CIRST), Université du Québec à Montréal, 455, Boulevard René-Lévesque Est Montréal (Québec) H3C 3P8, Canada

Summary

Thresholds are widespread in origin of life scenarios, from the emergence of chirality, to the appearance of vesicles, of autocatalysis, all the way up to Darwinian evolution. Here, we analyze the "error threshold" — which poses a condition for sustaining polymer replication — and generalize the threshold approach to other properties of prebiotic systems. Thresholds provide theoretical predictions, prescribe experimental tests and integrate interdisciplinary knowledge. The coupling between systems and their environment determines how thresholds can be crossed, leading to different categories of prebiotic transitions. Articulating multiple thresholds reveals evolutionary properties in prebiotic scenarios. Overall, thresholds indicate how to assess, revise and compare origin of life scenarios.

Keywords: thresholds, abiogenesis, origin of life, major transitions, prebiotic systems, evolution

Introduction

The transition from non-living to living matter is usually formulated in a historical manner, aiming to address the origin of life on Earth (Oparin, 1938; de Duve, 1991). However, the sparsity of historical evidence suggests that the "origin of life cannot be discovered; it has to be reinvented" (Eschenmoser and Loewenthal, 1991). This view, together with increasing experimental possibilities and the question of life on exoplanets (Rimmer *et al.*, 2018) broadens the field of investigation to synthetic systems with lifelike properties (Solé, 2016), such as metabolism, reproduction and evolution (Attwater and Holliger, 2014).

* Correspondence:
philippe.nghe@espci.psl.eu (P.N.), Lead Contact
cyrille.jeancolas@espci.psl.eu (C.J)
malaterre.christophe@uqam.ca (C.M.)

Origin of life scenarios consist of successions of transitions. Major transitions have been outlined (Smith and Szathmáry, 1995; Szathmáry, 2015) but there are so far few empirical validations. Indeed, experiments may face huge parameter spaces and waiting times. Furthermore, studying transitions requires inputs from many disciplines, including chemistry, biochemistry, molecular biology, evolutionary biology, physics, astrophysics, geology, or geochemistry among others (Preiner *et al.*, 2019). Thus, it remains a challenge to decompose scenarios at a sufficiently fine-grained scale for experimental tests to be possible, while allowing articulation of elementary steps and overall plausibility estimates.

Here, we propose thresholds to be an operational notion that can be used to decompose origin of life scenarios, articulate theory and experiment, and assess scenario plausibility by coordinating interdisciplinary efforts. A threshold can be defined as a major qualitative change undergone by a physical-chemical system upon relatively minor changes in the values of systemic or environmental control parameters. One of the most studied thresholds is Eigen's "error threshold" which constrains replication-based scenarios of early genetic polymers (Eigen, 1971; Kun *et al.*, 2015; Takeuchi, Hogeweg and Kaneko, 2017). A number of other prebiotic transitions invoke thresholds explicitly or implicitly: the appearance of homochirality (Hawbaker and Blackmond, 2019), replication (Kauffman, 1986; Szathmáry, 2006), compartmentalization (Hargreaves, Mulvihill and Deamer, 1977), Darwinian evolution (Goldenfeld, Biancalani and Jafarpour, 2017).

Below, we first sample the diversity of thresholds found in origin of life research. Thresholds are encountered at different stages and levels of organization. In the second section, as a case study, we examine the original formulation of the "error threshold" for template-based replication, highlighting its interdisciplinary nature. In the third section, we analyze revisions to the error threshold that aim at establishing more precise or more plausible models. In the fourth section, we generalize the threshold approach to properties of prebiotic systems other than replication. Introducing a phase diagram representation allows us to categorize different types of threshold transitions as a function of the relationship between systems and their environment. In the fifth section, we investigate how to articulate multiple thresholds toward building complete scenarios. In turn, we determine qualitative relationships between transitions, such as entrenchment, contingency or transience. These relationships are typically used in evolutionary biology but are more generally applicable to prebiotic threshold transitions. Finally, we discuss how thresholds can be used to coordinate research efforts to study the origin of life.

1. The diversity of thresholds in origin of life

In this section, we illustrate the diversity of threshold-type transitions in origin of life research with a non-exhaustive list of examples taken from chemistry and thermodynamics all the way up to evolutionary and Darwinian dynamics.

The prebiotic synthesis of the building blocks of life (i.e. amino acids, nucleic acids, sugars, fatty acids) requires reactants to reach sufficient concentrations for the reactions to take place, hence the existence of concentration thresholds. Concentration thresholds can be overcome by acting on physical-chemical conditions resulting from enclosure into vesicles (Luisi, Stano and de Souza, 2014), wet-dry cycles (Nelson *et al.*, 2001), freeze-thaw cycles (Monnard, Kanavarioti and Deamer, 2003; Trinks, Schröder and Biebricher, 2005), adsorption on mineral surfaces (Lambert, 2008) or

thermophoresis effects in hydrothermal vents (Baaske *et al.*, 2007). Eutectic ice and meteoritic impacts have been pointed out to overcome a concentration threshold of hydrogen cyanide for its polymerization into purines (Miyakawa, Cleaves and Miller, 2002; Parkos *et al.*, 2018) (figure 1.a). This pathway to purines includes yet another threshold which concerns free-energy barriers for the formation of intermediate compounds, which could be overcome through exposition to light (Boulanger *et al.*, 2013) (figure 1.b). Since the famous Miller's experiment (Miller, 1953), lightning is often considered to overcome such energy thresholds as well. The subsequent polymerization of building blocks (e.g. nucleotides) must also overcome energy thresholds (Dickson, Burns and Richardson, 2000), through chemical activation (Wachowius and Holliger, 2019) or adsorption on mineral surfaces (Hazen and Sverjensky, 2010). More generally, when it comes to forming the chemical bonds of biomolecules, a kinetic barrier has been estimated to be around 100 kJ/mol at moderate temperature (Pascal, 2012). Another threshold transition in prebiotic chemistry concerns the emergence of homochirality among biomolecules, i.e. the fact that in extant life amino acids are left-handed and sugars right-handed (Hawbaker and Blackmond, 2019).

The emergence of prebiotic compartments can be studied as a phase transition in the thermodynamic sense. Several thresholds that separate phases of dissociated components from self-assembled structures have been identified. Self-assembly thresholds controlling compartments formation can be overcome by varying physical-chemical parameters such as pH, temperature, or molecular crowding. It has been studied for the formation of vesicles (Bachmann, Luigi and Lang, 1992), coacervates (Jiang *et al.*, 2015) or microdomains in liquid crystals (Todisco *et al.*, 2018) (figure 1.c). For finite-size structures, such as peptide nanospheres (Carny and Gazit, 2005), mathematical models have established a threshold separating the self-assembled state from a "yield catastrophe phase" where constituents nucleate without stabilizing to final structures, controlled by activation and dimerization rates (Gartner *et al.*, 2020).

Transitions toward self-organization in reaction networks require the spontaneous formation of autocatalytic networks (Kauffman, 1993) (figure 1.d). A first threshold concerns catalytic closure, initially formulated for a model of networks of ligation and fragmentation reactions between peptides (Kauffman, 1986). Catalytic closure implies the existence of a set of peptides such that the formation of any member of the set is catalyzed by another member of the set. In this model, the transition happens when $\frac{E}{N} > \frac{1}{2}$, with E being the number of catalyzed reactions and N the number of peptides (respectively edges and nodes of the network) (Cohen, 1988). Other models have simulated autocatalytic transitions in the context of template-based replication (Chen and Nowak, 2012; Mathis, Bhattacharya and Walker, 2017). Wu et al. for instance examined how variations in catalytic feedback efficiency and spontaneous polymerization rate could induce the crossing of a threshold delimiting a regime of slow spontaneous RNA synthesis from a regime of rapid autocatalytic RNA synthesis (figure 1.e). This transition is stochastic, meaning that the systems explore random chemical compositions in a fixed environment before undergoing the transition.

Once formed, autocatalytic networks need to be sustained. For this, catalysts within autocatalytic networks should exhibit high enough specificity towards the reactions of the network to overcome degradation and losses due to side reactions. Given an autocatalytic set of p reactions, each catalyzed with a specificity s_i (where $i = 0,..,p$), disappearance can be avoided provided $\prod_{i=1...p} s_i > \frac{1}{2}$, thereupon revealing a "decay threshold" (Szathmáry, 2006; Vasas *et al.*, 2012; Blokhuis, Lacoste and Nghe, 2020). Furthermore, for survival to be possible, the concentration of feedstock compounds must stay above a specific value most of the time. For a single autocatalytic reaction, this condition is

fulfilled if $R > \frac{a}{k}$, with R being the concentration of reactants, a the degradation kinetic coefficient of the autocatalyst and k the kinetic rate of autocatalyst production (King, 1977).

Later stages in origin of life concern the emergence of template-based replication of genetic polymers, which is discussed in detail in the next section, and the emergence of protocells. Early compartments and genetic polymers may not have been strictly associated in lineages. Indeed, compartments provide a way for multilevel selection (Poole, 2009) even when genes are submitted to pooling and mixing (Matsumura *et al.*, 2016). The "progenote stage" has been described as a stage where genes heavily exchange between dividing protocells via horizontal gene transfer (HGT) (Woese and Fox, 1977). The transition from dominant HGT to lineages where genetic polymers and compartments are strongly correlated has been coined the "Darwinian threshold" (Woese, 2002) (figure 1.f). Mathematical models (Arnoldt, Strogatz and Timme, 2015; Goldenfeld, Biancalani and Jafarpour, 2017) show that such a threshold can be crossed when the progenotes HGT rate decreases below a specific value. This decrease might result from mutations and environmental changes such as protocell density, alkaline shifts, or nutrient limitation (Claverys, Prudhomme and Martin, 2006; Kovács *et al.*, 2009). In contrast, other authors posit the advent of Darwinian evolution before the appearance of protocells. This is the case of Eigen's theory of template-based replication of genetic polymers (Eigen, 1971; Eigen and Schuster, 1977; Eigen, McCaskill and Schuster, 1988), which we detail in the coming section.

2. The error threshold and the replicase

The error-threshold, introduced by Eigen (1971), has certainly been the most studied threshold in origin of life theories. We draw here on the main findings to make a case study and highlight the usefulness of studying thresholds, the replicase scenario being reviewed elsewhere (Kun *et al.*, 2015; Takeuchi, Hogeweg and Kaneko, 2017). In the RNA world hypothesis, an RNA capable of catalyzing the polymerization of its own sequence as well as others (Cheng and Unrau, 2010), a generalist *self-replicase*, is regarded as a primordial replicator capable of Darwinian evolution. The appearance and maintenance of such a replicase regime—a process that we will refer to as the "replicase scenario"—is constrained by the error threshold which imposes upper bounds on the length of the copied reference sequences given an error-rate during replication. More specifically, to avoid the disappearance of the sequence to be copied across rounds of replication, called the "error catastrophe", the following condition must be fulfilled:

$$l \leq \frac{ln\,(s)}{e} \qquad (1)$$

where e is the replicase copying error rate and s the selective advantage, defined as the ratio of the replication rate of the reference sequence to that of the erroneous sequences. Note that formula (1) was initially derived by Eigen et al. assuming that a polymerase is present and non-limiting or that replication is non-enzymatic (see next section for more precise models). Formula (1) is valid for $s > 1$, otherwise this model forbids any reference sequence, including the replicase, to be sustainable. From (1), we deduce a function $e_{max} = \frac{ln(s)}{l}$ (red curve on figure 2.a). Due to the minimum structure complexity required for replicase activity, replicases of a given length l cannot reach perfect fidelity and their error rate must be above a certain value $e_{min}(l)$ (blue curve on figure 2.a). Bounds e_{min}

and e_{max} delineate a domain in the space parametrized by (l, e) so that above e_{min} there exist replicases, and under e_{max}, replicase copying can be sustained through faithful replication. If these domains overlap, i.e. if there exists an l such that $e_{max}(l) \geq e_{min}(l)$, then the replicase regime is possible in the intersection domain (green zone in figures 2.a and 2.b; note that the sheer possibility of the replicase regime does not imply its plausibility). Otherwise, $e_{max}(l) < e_{min}(l)$ for all l, and the replicase regime has no existence domain (figure 2.c). The latter case corresponds to the so-called "Eigen's paradox": no long replicases without replication fidelity, but no fidelity without long replicases.

If the replicase regime is possible, the next step is to study the plausibility of emergence of a sustainable replicase. First needed is the number N_l of sequences of length l that have a replicase activity with an error rate e. The higher this number, the greater the chances of reaching at least one such replicase by a pre-existing generative process. The distribution $f(e, l)$ of these sequences is depicted in figure 2.d for a given length l. Second, the pre-existing generative process also determines the plausibility of the transition. A typical model for this regime is random polymerization, where every sequence is generated with equal probability for a given length. There, the fraction of generated polymers of a certain length $g(l)$ in the polymer population decreases rapidly with l. From the distributions $f(e, l)$ and $g(l)$, we estimate the probability for a sustainable replicase to appear to be:

$$P_r = \sum_l \; g(l) \int_{e_{min}(l)}^{e_{max}(l)} f(e, l) de \qquad (2)$$

This probability can be used in the context of an environmental scenario supplying estimates of the total number N of random oligomers present at any moment, their global renewal rate ρ resulting from sequence generation and destruction dynamics. The typical time t required for one replicase to appear then obeys:

$$\rho \; t \; N \; P_r \sim 1 \qquad (3)$$

The parameters found in (2) and (3) highlight the highly interdisciplinary nature of the question: the threshold e_{max} is determined by replication dynamics studied in theoretical biology; the distribution f and threshold e_{min} arise from genotype-to-function relationships obtained from biochemical considerations; g depends on the chemistry of spontaneous polymerization; ρ and N depend on geochemical processes, and t should be compared with estimates from geology and planetary system stability analysis.

The analysis leading to (3) is only part of the answer for the replicase scenario: for the replication dynamics to actually start, a replicase needs to encounter either a copy of its own sequence or a complementary sequence. This should probably take place in a compartmentalized setting, where only replicase sequences are present, since otherwise the copying of other sequences would immediately take over the population. Furthermore, the analysis presented so far may be questioned, given that formula (1-3) rely on a number of strong assumptions. We examine some of these considerations in the coming section.

3. Refining and revising the replicase scenario

The error threshold of equation (1) relies on a number of simplifying assumptions that have been refined in subsequent models presented here. Such models may lead to lower e_{max}, thus reducing the size of the green zone in figure 3, which goes to the detriment of the replicase scenario. For instance, Eigen's original model ignores the effects of changes in replicase concentration, whereas the latter is impacted by replication. Including this aspect greatly lowers e_{max} (Obermayer and Frey, 2009). Experimental data about RNA replication highlight another required model revision, because replication errors typically lead to shorter sequences rather than to point mutations (Ichihashi and Yomo, 2016). As a result, mutants tend to replicate faster than the original sequence, which means $s < 1$ by definition of the selective advantage, so that replicase regimes cannot be sustained (see Section 2). Note however that it is still unclear whether results obtained with protein replicases used in these experiments extrapolate to hypothetical RNA replicases. In contrast, other model refinements tend to increase e_{max}, thus broadening the range of viable replicases. This is the case when considering the abundance of neutral mutations in the neighborhood of the replicase sequence, leading to a "relaxed" error threshold which is significantly lower than Eigen's (Kun, Santos and Szathmáry, 2005; Takeuchi, Poorthuis and Hogeweg, 2005).

Other modifications are revisions of the scenario. So-called "hypercycles" were first proposed as a way to increase the robustness of the system to replication errors (Eigen and Schuster, 1977). This model is based on a network with multiple molecular species, each species being replicated in a template-based manner while helping the replication of other members of the network. The resulting cooperative dynamics can raise the error threshold, thereby increasing the replicase regime zone (Takeuchi, Hogeweg and Kaneko, 2017). Other scenarios introduce compartments and more generally spatial structure. The Stochastic Corrector Model (SCM) considers lineages of compartments which grow, are selected, and divide (Szathmáry and Demeter, 1987; Grey, Hutson and Szathmary, 1995). Transient Compartmentalization (TC) assumes simpler cycles consisting of compartmentalization, selection, and pooling (Matsumura *et al.*, 2016; Blokhuis *et al.*, 2020). Unlike the original model and its equation (1), SCM and TC authorize the presence of parasite sequences that replicate faster than the reference sequence. This was experimentally demonstrated for TC with replicated RNA (Matsumura *et al.*, 2016). Indeed, in both cases, erroneous copies only locally invade compartments and are removed from the population by compartment-level selection. Surface clusters are an even less constrained setting that allows the maintenance of replication, although less efficiently than non-permeable compartments (Szabó *et al.*, 2002; Shah *et al.*, 2019). These mechanisms increase e_{max}, thus also increase the plausibility of the replicase scenario. However, this comes at the cost of assuming the existence of spatial structures and introducing additional parameters (number of molecules per compartment, compartment lifetime, selection process, etc.).

Once the error threshold is determined, the next question is to estimate the fraction of viable replicases in the sequence space. This corresponds to determining the distribution f in equation (2), which depends on the relationship between sequence and catalytic function, including the minimum possible error rate e_{min} given a certain replicase length. This problem cannot be solved computationally yet, and heavily relies on experimental data. Experiments on synthetic replicases show that processivity — the ability to copy long templates — is a limiting factor even before considering fidelity. Processivity is notably hampered when copying folded RNAs, which is the case for replicase templates. Consistently, replicases synthesized in the laboratory can copy other RNAs as long as themselves but cannot copy themselves (Horning and Joyce, 2016; Attwater *et al.*, 2018).

Another issue is that the sequence space cannot be covered experimentally due to its astronomical size (~10^{114} for a typical 190 nucleotides long replicase (Johnston *et al.*, 2001)). Consequently, experimental data need to be combined with theory and computation in order to estimate f. For instance, the diversity of RNA secondary structures is predicted to increase as $1.4848 \times n^{-\frac{3}{2}}(1.8488)^n$ where n is the sequence length (Schuster *et al.*, 1994). This is consistently smaller than the 4^n possible sequences. Taking structure similarity as a proxy for function indicates an average redundancy in sequences with similar catalytic properties, potentially reducing the size of the space to characterize. A complementary approach is to estimate the tail of distribution f between e_{min} and e_{max} (figure 2.d) using laws of extreme statistics (Gumbel, 1958). Indeed, replicases are rare sequences for which statistics are expected to follow extreme value theory. The latter restricts distributions of rare properties to functional forms (among the Weibull, Gumbel and Fréchet laws) with few parameters. Such approach has been tested experimentally for proteins (Boyer *et al.*, 2016) and ribozymes (Pressman *et al.*, 2017). Despite the complexity of the genotype-to-phenotype relationship, the limited number of parameters involved in these laws may allow estimates of f from a restricted sample of replicases.

Finally, the transition scenario toward existence of a sustainable replicase may be revised by considering different polymer generation processes anterior to the replicase regime. Polymer generation could have been enhanced by catalysis of polymerization (Hazen and Sverjensky, 2010), ligation (Mutschler *et al.*, 2018), or recombination (Blokhuis and Lacoste, 2017), as well as by size selection (Mizuuchi *et al.*, 2019). An alternative is that autocatalytic sets could have preceded template-based replication and biased polymer populations toward longer and more functional sequences (Vasas *et al.*, 2012). All these processes can add weight to the distribution g in equation (2) in the length categories corresponding to replicases and contribute to the plausibility of their emergence.

4. Generalizing the threshold approach

We have seen that the error threshold sets a number of requirements on polymerases for them to sustain replication. We now generalize the threshold approach to other properties of prebiotic systems. In figure 3, the axes correspond to relevant parameters of the system and its environment, and regions α and β respectively correspond to systems without and with a property of interest (regions separated by a threshold line). We call this representation a phase diagram, by analogy with equilibrium thermodynamics, extending its use to non-equilibrium regimes as well. In the replicase scenario, region α corresponds to a regime of randomly produced RNAs and region β to sustainable template-based RNA replication, the boundary between the two being the error threshold. In contrast with figure 2 where the axes stand for system parameters only, the x-axis in figure 3 captures system parameters (e.g. replicase error rate) and the y-axis environment variables (e.g. temperature). Note that it is not always obvious to distinguish prebiotic systems from their environment, and choosing a variable as being either part of the system or of the environment may be a matter of choice. Note also that the x-axis represents chosen control parameters of the system; other parameters of the system may lie in other dimensions, in particular as encoded by the green shading of β in figure 3, which indicates the presence of a specific property of the system. A point in the diagram then represents a system with certain fixed properties in a given environment. Such phase diagrams allow us to classify modes by which a system overcomes a threshold and acquires a novel property, i.e. moves from region α to region β. We classify transitions into four modes:

1. *Extrinsic transition*: the acquisition of a new property is here solely driven by the environment. In figure 3a, the system is initially located at point A in region α and environment E_1; a change from E_1 to E_2 brings the system to B in region β. Since this transition is not caused by any change in the system control parameters, there is no variation along the x-axis. Transitions in prebiotic chemistry are typically framed in the context of extrinsic transitions (Kawamura and Maurel, 2017; Kitadai and Maruyama, 2018; Benner *et al.*, 2019), where concentration thresholds are overcome during wet-dry cycles in ponds (Campbell *et al.*, 2019), where steps of nucleobase synthesis occur at the junction of interconnecting streams (Patel *et al.*, 2015), or where hydrogen cyanide is synthesized upon meteorite impacts (Parkos *et al.*, 2018). This mode can also be applied to the appearance of vesicles triggered by CO_2 (Bachmann, Luigi and Lang, 1992).

2. *Intrinsic transition*: the novel property here results from changes in parameters of the system while the environment remains unchanged. In figure 3.b, the system is initially located at point A in α with environment E_1, and moves to B in β without any environmental change, simply moving along the x-axis. This is the case when a replicase appears from random polymerization as described in section 3 (Mathis, Bhattacharya and Walker, 2017). Another example is the appearance of autocatalytic sets as modelled by Jain and Krishna (1998). This model assumes that catalytic species only are limiting and follow a dynamic where the least fit species tend to disappear and be randomly replaced by novel catalytic species. Under these assumptions, an autocatalytic set inevitably emerges and fixes. Note that in the examples cited above, extrinsic transitions are deterministic while intrinsic transitions are stochastic, but this need not always be the case. For a discussion on contingency versus determinism in the origin of life see Luisi (2003).

3. *Scaffolding transition*: the environment changes only transiently, and fixing a novel property requires to combine the environmental change with a system-driven change. Here, the environment plays the role of a scaffold in the sense that it transiently supports the emergence of a property which is then internalized, i.e. becomes an intrinsic property of the system (Caporael, Griesemer and Wimsatt, 2014). In figure 3c, the system starts at point A in region α in environment E_1, then moves to point B in region β via an extrinsic transition triggered by an environmental change to E_2. At this stage, returning to E_1 would bring the system back into α. However, if the property acquired in β allows exploration of neighboring areas along the x-axis, the system may reach B' in β while in environment E_2. Then, an environmental change from E_2 back to E_1 would bring the system from B' to B", yet still within β. Such scaffolding transition could be observed when a dilute solution of polymers in a pond (system in A) is submitted to partial evaporation (from E_1 to E_2), thereby reaching a concentration threshold above which autocatalysis becomes possible (system in B), leading it to accumulate enough catalysts (from B to B') so that once back to wet conditions, autocatalysis is maintained (from B' to B").

4. *Symbiotic transition*: two distinct systems aggregate into a new system thereby acquiring a novel property. In figure 3.d, systems A and B both in α merge into a new system C located in β in a constant environment. Such transition is hypothesized when compartments meet autocatalytic chemistries (Hordijk *et al.*, 2018; Joyce and Szostak, 2018). Both may pre-exist separately, but evolution by natural selection requires autocatalytic chemistries to be compartmentalized in order to provide a collective level of selection (Vasas *et al.*, 2012).

5. Articulating series of thresholds

In previous sections, we analyzed single threshold transitions. However, origin of life scenarios combine multiple transitions (Szathmáry, 2015; Solé, 2016). There are to date only a few attempts to assemble complete detailed scenarios (Martin *et al.*, 2003; Damer and Deamer, 2020). Research efforts have focused on combining transitions for specific stages, such as the synthesis of building blocks (Patel *et al.*, 2015; Kitadai and Maruyama, 2018; Becker *et al.*, 2019), the emergence of functional RNAs from random ones (Briones, Stich and Manrubia, 2009; Higgs and Lehman, 2015) or the emergence of evolution from catalytic micelles (Lancet, Zidovetzki and Markovitch, 2018), to cite a few. Below, we apply the phase diagram representation to multiple thresholds. This analysis reveals different types of articulations between threshold transitions: *accumulation*, *entrenchment*, *contingency*, *transience*, *facilitation*. These concepts are borrowed from evolutionary biology but are used here to analyze systems that do not necessarily evolve in a Darwinian manner. We indeed consider these concepts to be generally applicable to successions of threshold transitions, heritable transitions through Darwinian evolution only being a particular case.

We first consider the crossing of two successive thresholds. In figures 4.a-c, the green and blue regions β and γ indicate systems which possess properties p_β and p_γ, respectively; in region α, the system does not possess any of them. Here, the x- and y-axes depict relevant variables that can be environmental, systemic or both. For *accumulation* of properties p_β and p_γ to be possible, regions β and γ must intersect. Figures 4a-c depict systems that start from region α, then cross the threshold leading to β, followed by a second threshold-crossing to region γ. A notion stronger than accumulation is *entrenchment*, where any property p_γ posterior to p_β requires it and fixes it irreversibly. This is illustrated in figure 4a, where region γ is entirely included in region β. For example, p_β could stand for the existence of catalytic RNAs and p_γ of autocatalytic RNAs. Another threshold articulation is when β and γ intersect without inclusion. Such a situation is characterized by *contingency* in the sense that p_γ may exist independently of p_β (and vice-versa) so that any of the two properties may appear first (figure 4.b, from points A to B or points A to b). This would be the case with the appearance of RNAs (p_γ) together with other genetic polymers (p_β) (Cleaves *et al.,* 2019). Alternative polymers further illustrate *transience*, where a property is acquired, and then lost (figure 4.b, any path from point A to point D). For instance, TNA may have appeared, coexisted with RNA, then disappeared (Yu, Zhang and Chaput, 2012). Finally, a threshold transition can make a subsequent transition more plausible when acquiring p_β lowers the threshold to acquire p_γ (figure 4.c); this is referred to as *facilitation*. For example, compartmentalization (p_β) lowers the error threshold that delineates sustainable replication (p_γ) (Matsumura *et al.*, 2016).

We now consider the articulation between more than two thresholds as depicted in figure 4.d as a Venn diagram, where thresholds delineate regions corresponding to different properties. For example, a chemical system may start in region α where it has the ability to synthetize random RNAs. By reaching β, it gains the ability to synthetize random peptides as well. Next, reaching γ allows autocatalysis among sets of peptides. Furthermore, in these examples, the system is allowed to *accumulate* the properties that define α, β and **γ** because they all intersect. An example of *entrenchment* here is RNA oligomer synthesis (α) that is a necessary condition for the existence of autocatalysis in RNA sets (δ is fully included in α). *Transience* happens if the system loses its ability to synthetize random RNAs (α) but acquires protein RNA-replicases (φ), as it allows replication of long RNAs from single nucleotides.

The notion of *entrenchment* can be further refined. As we have seen, it occurs when a system cannot get rid of a feature because too many features have evolved on that basis. As a result, the entrenched feature can narrow down the access to other properties as well as open other possibilities. The former is known as "contingent irreversibility" (Smith and Szathmáry, 1995) and the latter as "generative entrenchment" (Schank and Wimsatt, 1986) or "enabling constraint" (Kauffman, 2014). In figure 4.d, reaching region δ allows transition to region ε but not φ. The advent of homochirality illustrates these notions. Indeed, a racemic mixture of oligonucleotides (region β in figure 4.d) could break its symmetry because of a triggered small enantiomeric excess amplified by an autocatalytic reaction leading either to L-ribose RNA only (region γ) either to D-ribose RNA only (region δ) (Hawbaker and Blackmond, 2019). Entrenchment would follow with peptide synthesis: L-ribose RNA systems could imply D-amino acids peptides only (region φ) whereas D-ribose RNA systems could imply L-amino acids peptides only (region ε) (Illangasekare *et al.*, 2010).

Finally, we relate the diagram representations to the concept of *lifeness* (also referred as *aliveness* or *life index*) proposed by several authors (Bruylants, Bartik and Reisse, 2010; Sutherland, 2017; Malaterre and Chartier, 2019). Lifeness stands for a scale where so-called "infrabiological systems" (Szathmáry, 2005) are positioned between a non-living state and a living state. It remains unclear which quantities or measures best account for lifeness. For instance, Bedau (2012) proposes a discrete scale from 0 to 9 depending on the number of interactions between three subsystems: container, metabolism and program, while Malaterre and Chartier (2019) propose a multidimensional gradual scale that integrates system and environment-related functions. In any case, projecting the paths of figure 4d on any such measure would make threshold transitions appear as jumps, the directionality of paths towards a living state corresponding to increasing lifeness (figure 4.e). In this view, we propose that lifeness corresponds, at first order, to the number of thresholds crossed relative to the total number of thresholds in a given scenario. This suggests a way toward formalizing the notion of chemical evolution: early prebiotic systems may not yet evolve in a canonically Darwinian manner, but in the sense of acquiring and sustaining qualitatively novel properties by crossing thresholds, thereby increasing their lifeness before getting a chance to increasing their fitness.

6. Discussion

Thresholds define conditions of existence for particular states along the path from inanimate matter to life. Although not all prebiotic transitions are threshold transitions, thresholds pose well-defined problems that can structure interdisciplinary efforts to understand the origin of life (Fig. 5). Thresholds may initially be deduced from apparent paradoxes, so-called "catch-22" or "chicken-and-egg" situations stated as: "without X, no Y, and without Y, no X" (Benner, 2018). For instance, from a bird's-eye view of life, metabolism, genetics and compartments appear so intertwined that there is no obvious scenario for their gradual emergence. As we have seen, thresholds are also detected when taking a closer look at molecular processes, such as template-based replication, or from computer simulations and experiments. Their signatures are discontinuities, hysteresis and waiting time distributions, as observed with first order phase transitions in physics.

Quantitatively characterizing a threshold requires back-and-forth adjustments between theory and experiment. Interestingly, thresholds pose conditions of existence of a given state without referring to an anterior nor a posterior state. Thus, their study is independent of a particular transition or scenario

in the first place. The starting point is typically a simplified model establishing relationships between parameters. Such a "toy-model" has been highly productive in the case of the error threshold, notably pointing to the disciplinary diversity of its parameters. From this point, thorough theoretical investigations of the parameter space and regimes help elaborating hypotheses testable with synthetic experiments in chemistry, systems chemistry, physical-chemistry, soft matter physics, etc. (Solé, 2016; Preiner *et al.*, 2019). In turn, models may be refined until convergence. A possible outcome is that the proposed regime is impossible due to multiple incompatible thresholds, resulting in a true paradox and the revision of the scenario with alternative regimes. Otherwise, the regime is possible and then comes the question of its plausibility.

To assess plausibility, transition from an anterior state must be accounted for (Orgel, 2008; Schwartz, 2013). Theory, possibly again with back-and-forth adjustment with experiments, estimates the probability of crossing the threshold given the anterior state and environmental conditions as determined by geology, geochemistry, or planetary dynamics (Stüeken *et al.*, 2013; Sasselov, Grotzinger and Sutherland, 2020). Comparing the crossing probability from different anterior states and environments selects between transitions (e.g. random polymerization or autocatalytic sets before to template-based replication). The plausibility of scenarios finally results from the probability of successions of such transitions. If this total probability is deemed too low, alternative scenarios may lead to consider novel thresholds (e.g. error thresholds with compartment instead of without).

So far, no one agrees on whether any scenario is devoid of true paradoxes (impossibilities). However, the threshold approach outlined above can establish firm building blocks on the way to constructing scenarios. For this to be effective, studies on the origin of life ought to be explicit with the thresholds they address, the parameters involved, and the relationships between these parameters, following the example of the error threshold.

Resource availability

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact Philippe Nghe (philippe.nghe@espci.psl.eu)

Materials Availability

This study did not use or generate any reagents

Data and Code Availability

This study did not use any dataset or code

Acknowledgments

The authors thank the Origines et Conditions d'Apparition de la Vie (OCAV) IRIS initiative at PSL Research University for stimulating discussions. The manuscript also benefited from the comments of two anonymous reviewers for *iScience*. CJ acknowledges financial support by Université de Paris and the École Doctorale FIRE - Programme Bettencourt. This work has received the support of "Institut Pierre-Gilles de Gennes" (laboratoire d'excellence, "Investissements d'avenir" program ANR-10-IDEX-0001-02 PSL and ANR-10-LABX-31). PN acknowledges funding from the Human Frontier Science Program (Grant RGY0077/2019). CM acknowledges funding from an ESPCI Joliot Chair, Canada Social Sciences and Humanities Research Council [Grant 430-2018-00899] and Canada Research Chairs [CRC-950-230795].

Author Contributions

All authors contributed equally to this study.

Declaration of Interest

The authors declare no competing interests.

References

Arnoldt, H., Strogatz, S. H. and Timme, M. (2015) 'Toward the Darwinian transition: Switching between distributed and speciated states in a simple model of early life', *Physical Review E*, 92(5), pp. 1–9. doi: 10.1103/PhysRevE.92.052909.

Attwater, J. *et al.* (2018) 'Ribozyme-catalysed RNA synthesis using triplet building blocks', *eLife*, 7, pp. 1–25. doi: 10.7554/eLife.35255.

Attwater, J. and Holliger, P. (2014) 'A synthetic approach to abiogenesis', *Nature Methods*. Nature Publishing Group, 11(5), pp. 495–498. doi: 10.1038/nmeth.2893.

Baaske, P. *et al.* (2007) 'Extreme accumulation of nucleotides in simulated hydrothermal pore systems', *Proceedings of the National Academy of Sciences of the United States of America*, 104(22), pp. 9346–9351. doi: 10.1073/pnas.0609592104.

Bachmann, P. A., Luigi, P. and Lang, J. (1992) 'Micelles As Models for Prebiotic Structures', *Nature*, 357(May), pp. 1013–1015.

Becker, S. *et al.* (2019) 'Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides', *Science*, 366(6461), pp. 76–82. doi: 10.1126/science.aax2747.

Bedau, M. A. (2012) 'A functional account of degrees of minimal chemical life', *Synthese*, 185(1), pp. 73–88. doi: 10.1007/s11229-011-9876-x.

Benner, S. A. (2018) 'Prebiotic plausibility and networks of paradox-resolving independent models', *Nature Communications*. Springer US, 9(1), pp. 9–11. doi: 10.1038/s41467-018-07274-y.

Benner, S. A. *et al.* (2019) 'When Did Life Likely Emerge on Earth in an RNA-First Process?', *ChemSystemsChem*, 1900035. doi: 10.1002/syst.201900035.

Blokhuis, A. *et al.* (2020) 'The generality of transient compartmentalization and its associated error thresholds', *Journal of Theoretical Biology*. Elsevier Ltd, 487, p. 110110. doi: 10.1016/j.jtbi.2019.110110.

Blokhuis, A. and Lacoste, D. (2017) 'Length and sequence relaxation of copolymers under recombination reactions', *The Journal of Chemical Physics*. AIP Publishing LLC, 147(9), p. 094905. doi: 10.1063/1.5001021.

Blokhuis, A., Lacoste, D. and Nghe, P. (2020) 'Universal motifs and the diversity of autocatalytic systems', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.2013527117.

Boulanger, E. *et al.* (2013) 'Photochemical Steps in the Prebiotic Synthesis of Purine Precursors from HCN', *Angewandte Chemie*, 125(31), pp. 8158–8161. doi: 10.1002/ange.201303246.

Boyer, S. *et al.* (2016) 'Hierarchy and extremes in selections from pools of randomized proteins', *Proceedings of the National Academy of Sciences of the United States of America*, 113(13), pp. 3482–3487. doi: 10.1073/pnas.1517813113.

Briones, C., Stich, M. and Manrubia, S. C. (2009) 'The dawn of the RNA World: Toward functional complexity through ligation of random RNA oligomers', *RNA*, 15, pp. 743–749. doi: 10.1261/rna.1488609.Orgel.

Bruylants, G., Bartik, K. and Reisse, J. (2010) 'Prebiotic chemistry: A fuzzy field', *Comptes Rendus Chimie*. Academie des

sciences, 14(4), pp. 388–391. doi: 10.1016/j.crci.2010.04.002.
Budin, I. and Szostak, J. W. (2010) 'Expanding Roles for Diverse Physical Phenomena During the Origin of Life', *Annual Review of Biophysics*, 39(1), pp. 245–263. doi: 10.1146/annurev.biophys.050708.133753.
Campbell, T. D. *et al.* (2019) 'Prebiotic condensation through wet–dry cycling regulated by deliquescence', *Nature Communications*. Springer US, 10, p. 4508. doi: 10.1038/s41467-019-11834-1.
Caporael, L. R., Griesemer, J. R. and Wimsatt, W. C. (2014) 'Developing scaffolds: An introduction', in *Developing scaffolds in evolution, culture, and cognition*, pp. 1–20.
Carny, O. and Gazit, E. (2005) 'A model for the role of short self-assembled peptides in the very early stages of the origin of life', *The FASEB Journal*, 19(9), pp. 1051–1055. doi: 10.1096/fj.04-3256hyp.
Chen, I. A. and Nowak, M. A. (2012) 'From Prelife to Life: How Chemical Kinetics Become Evolutionary Dynamics', *Accounts of Chemical Research*, 45(12). doi: 10.1021/ar2002683.
Cheng, L. K. L. and Unrau, P. J. (2010) 'Closing the circle: replicating RNA with RNA.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 2(10), p. a002204. doi: 10.1101/cshperspect.a002204.
Claverys, J.-P., Prudhomme, M. and Martin, B. (2006) 'Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria', *Annual Review of Microbiology*, 60(1), pp. 451–475. doi: 10.1146/annurev.micro.60.080805.142139.
Cleaves, H. J. *et al.* (2019) 'One among Millions: The Chemical Space of Nucleic Acid-Like Molecules', *Journal of Chemical Information and Modeling*. doi: 10.1021/acs.jcim.9b00632.
Cohen, J. E. (1988) 'Threshold phenomena in random structures', *Discrete Applied Mathematics*, 19(1–3), pp. 113–128. doi: 10.1016/0166-218X(88)90008-X.
Damer, B. and Deamer, D. (2020) 'The hot spring hypothesis for an origin of life', *Astrobiology*, 20(4), pp. 429–452. doi: 10.1089/ast.2019.2045.
Dickson, K. S., Burns, C. M. and Richardson, J. P. (2000) 'Determination of the free-energy change for repair of a DNA phosphodiester bond', *Journal of Biological Chemistry*, 275(21), pp. 15828–15831. doi: 10.1074/jbc.M910044199.
de Duve, C. (1991) *Blueprint for a cell : The nature and origin of life*. Burlington: Patterson.
Eigen, M. (1971) 'Selforganization of matter and the evolution of biological macromolecules', *Die Naturwissenschaften*, 58(10), pp. 465–523. doi: 10.1007/BF00623322.
Eigen, M., McCaskill, J. and Schuster, P. (1988) 'Molecular quasi-species', *Journal of Physical Chemistry*, 92(24), pp. 6881–6891. doi: 10.1021/j100335a010.
Eigen, M. and Schuster, P. (1977) 'The Hypercycle A Principle of Natural Self-Organization', *Die Naturwissenschafte*, 64, pp. 541–565. doi: 10.1007/978-3-642-67247-7_1.
Eschenmoser, A. and Loewenthal, E. (1991) 'Chemistry of Potentially Prebiological Natural Products', *Chemical Society Reviews*, 67(281), pp. 1–16.
Farmer, J. D., Kauffman, S. A. and Packard, N. H. (1986) 'Autocatalytic replication of polymers', *Physica*, 220, pp. 50–67.
Gartner, F. M. *et al.* (2020) 'Stochastic yield catastrophes and robustness in self-assembly', *eLife*, 9. doi: 10.7554/elife.51020.
Goldenfeld, N., Biancalani, T. and Jafarpour, F. (2017) 'Universal biology and the statistical mechanics of early life', *Philosophical Transactions of the Royal Society A*, 375(2109). doi: 10.1098/rsta.2016.0341.
Grey, D., Hutson, V. and Szathmary, E. (1995) 'A re-examination of the stochastic corrector model', *Proceedings of the Royal Society B: Biological Sciences*, 262(1363), pp. 29–35. doi: 10.1098/rspb.1995.0172.
Gumbel, E. J. (1958) *Statistics of Extremes*. New-York: Columbia University Press.
Hargreaves, W. R., Mulvihill, S. J. and Deamer, D. W. (1977) 'Synthesis of phospholipids and membranes in prebiotic conditions', *Nature*, 266(5597), pp. 78–80. doi: 10.1038/266078a0.
Hawbaker, N. A. and Blackmond, D. G. (2019) 'Energy threshold for chiral symmetry breaking in molecular self-replication', *Nature Chemistry*. Springer US, 11(10), pp. 957–962. doi: 10.1038/s41557-019-0321-y.
Hazen, R. M. and Sverjensky, D. A. (2010) 'Mineral Surfaces, Geochemical Complexities, and the Origins of Life', *Cold Spring Harbor Perspectives in Biology*, 2, p. a002162. doi: 10.1016/0169-5347(88)90031-6.
Higgs, P. G. and Lehman, N. (2015) 'The RNA World: Molecular cooperation at the origins of life', *Nature Reviews Genetics*. Nature Publishing Group, 16(1), pp. 7–17. doi: 10.1038/nrg3841.
Hordijk, W. *et al.* (2018) 'Population Dynamics of Autocatalytic Sets in a Compartmentalized Spatial World', *Life*, 8(3), p. 33. doi: 10.3390/life8030033.
Horning, D. P. and Joyce, G. F. (2016) 'Amplification of RNA by an RNA polymerase ribozyme', *Proceedings of the National Academy of Sciences*, 113(35), pp. 9786–9791. doi: 10.1073/pnas.1610103113.
Ichihashi, N. and Yomo, T. (2016) 'Constructive approaches for understanding the origin of self-replication and evolution', *Life*, 6(3), pp. 1–12. doi: 10.3390/life6030026.
Illangasekare, M. *et al.* (2010) 'Chiral histidine selection by D-ribose RNA', *Rna*, 16(12), pp. 2370–2383. doi: 10.1261/rna.2385310.
Jain, S. and Krishna, S. (1998) 'Autocatalytic sets and the growth of complexity in an evolutionary model', *Physical Review Letters*, 81(25), pp. 5684–5687. doi: 10.1103/PhysRevLett.81.5684.
Jiang, H. *et al.* (2015) 'Phase Transition of Spindle-Associated Protein Regulate Spindle Apparatus Assembly', *Cell*. Elsevier Inc., 163(1), pp. 108–122. doi: 10.1016/j.cell.2015.08.010.
Johnston, W. K. *et al.* (2001) 'RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension', *Science*, 292(5520), pp. 1319–1325. doi: 10.1126/science.1060786.
Joyce, G. F. and Szostak, J. W. (2018) 'Protocells and RNA Self-Replication', *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 10(9), p. a034801. doi: 10.1101/CSHPERSPECT.A034801.
Kauffman, S. A. (1986) 'Autocatalytic sets of proteins', *Journal of Theoretical Biology*, 119(1), pp. 1–24. doi: 10.1016/S0022-5193(86)80047-9.

Kauffman, S. A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press.
Kauffman, S. A. (2014) 'Prolegomenon to patterns in evolution', *BioSystems*. Elsevier Ireland Ltd, 123, pp. 3–8. doi: 10.1016/j.biosystems.2014.03.004.
Kawamura, K. and Maurel, M. C. (2017) 'Walking over 4 Gya: Chemical Evolution from Photochemistry to Mineral and Organic Chemistries Leading to an RNA World', *Origins of Life and Evolution of Biospheres*. Origins of Life and Evolution of Biospheres, 47(3), pp. 281–296. doi: 10.1007/s11084-017-9537-2.
King, G. A. M. (1977) 'Symbiosis and the origin of life', *Origins of Life*, 8, pp. 39–53. doi: 10.1201/9780203009918.
Kitadai, N. and Maruyama, S. (2018) 'Origins of building blocks of life: A review', *Geoscience Frontiers*. Elsevier, 9(4), pp. 1117–1153. doi: 10.1016/j.gsf.2017.07.007.
Kovács, Á. T. *et al.* (2009) 'Ubiquitous late competence genes in Bacillus species indicate the presence of functional DNA uptake machineries: Minireview', *Environmental Microbiology*, 11(8), pp. 1911–1922. doi: 10.1111/j.1462-2920.2009.01937.x.
Kun, Á. *et al.* (2015) 'The dynamics of the RNA world: Insights and challenges', *Annals of the New York Academy of Sciences*, 1341(1), pp. 75–95. doi: 10.1111/nyas.12700.
Kun, Á., Santos, M. and Szathmáry, E. (2005) 'Real ribozymes suggest a relaxed error threshold', *Nature Genetics*, 37(9), pp. 1008–1011. doi: 10.1038/ng1621.
Lambert, J. F. (2008) 'Adsorption and polymerization of amino acids on mineral surfaces: A review', *Origins of Life and Evolution of Biospheres*, 38(3), pp. 211–242. doi: 10.1007/s11084-008-9128-3.
Lancet, D., Zidovetzki, R. and Markovitch, O. (2018) 'Systems protobiology: Origin of life in lipid catalytic networks', *Journal of the Royal Society Interface*, 15(144). doi: 10.1098/rsif.2018.0159.
Luisi, P. L. (2003) 'Contingency and determinism', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 361(1807), pp. 1141–1147. doi: 10.1098/rsta.2003.1189.
Luisi, P. L., Stano, P. and de Souza, T. (2014) 'Spontaneous Overcrowding in Liposomes as Possible Origin of Metabolism', *Origins of Life and Evolution of Biospheres*, 44(4), pp. 313–317. doi: 10.1007/s11084-014-9387-0.
Malaterre, C. and Chartier, J.-F. (2019) 'Beyond categorical definitions of life: a data-driven approach to assessing lifeness', *Synthese*. Springer Netherlands. doi: 10.1007/s11229-019-02356-w.
Martin, W. *et al.* (2003) 'On the origins of cells: A hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1429), pp. 59–85. doi: 10.1098/rstb.2002.1183.
Mathis, C., Bhattacharya, T. and Walker, S. I. (2017) 'The Emergence of Life as a First-Order Phase Transition', *Astrobiology*, 17(3), pp. 266–276. doi: 10.1089/ast.2016.1481.
Matsumura, S. *et al.* (2016) 'Transient compartmentalization of RNA replicators prevents extinction due to parasites', *Science*, 354(6317), pp. 1293–1296.
Miller, S. L. (1953) 'A production of amino acids under possible primitive earth conditions', *Science*, 117(3046), pp. 528–529. doi: 10.1126/science.117.3046.528.
Miyakawa, S., Cleaves, H. J. and Miller, S. L. (2002) 'the Cold Origin of Life : Hydrogen Cyanide and Formamide', *Origins of Life and Evolution of the Biosphere*, 32(1986), pp. 195–208.
Mizuuchi, R. *et al.* (2019) 'Mineral surfaces select for longer RNA molecules', *Chemical Communications*. Royal Society of Chemistry, 55(14), pp. 2090–2093. doi: 10.1039/c8cc10319d.
Monnard, P. A., Kanavarioti, A. and Deamer, D. W. (2003) 'Eutectic Phase Polymerization of Activated Ribonucleotide Mixtures Yields Quasi-Equimolar Incorporation of Purine and Pyrimidine Nucleobases', *Journal of the American Chemical Society*, 125(45), pp. 13734–13740. doi: 10.1021/ja036465h.
Mutschler, H. *et al.* (2018) 'Random-sequence genetic oligomer pools display an innate potential for ligation and recombination', *eLife*, 7, pp. 1–26. doi: 10.7554/elife.43022.
Nelson, K. E. *et al.* (2001) 'Concentration by Evaporation and the prebiotic Synthesis of Cytosine', *Origins of Life and Evolution of the Biosphere*, 31(3), pp. 221–229. doi: 10.1023/A:1016582308525.
Obermayer, B. and Frey, E. (2009) 'Escalation of error catastrophe for enzymatic self-replicators', *Europhysics Letters*, 88(4). doi: 10.1209/0295-5075/88/48006.
Oparin, A. I. (1938) *The Origin of Life*. New-York: Macmillan Publishers Limited.
Orgel, L. E. (2008) 'The implausibility of metabolic cycles on the prebiotic earth', *PLoS Biology*, 6(1), pp. 0005–0013. doi: 10.1371/journal.pbio.0060018.
Parkos, D. *et al.* (2018) 'HCN Production via Impact Ejecta Reentry During the Late Heavy Bombardment', *Journal of Geophysical Research: Planets*, 123(4), pp. 892–909. doi: 10.1002/2017JE005393.
Pascal, R. (2012) 'Suitable energetic conditions for dynamic chemical complexity and the living state', *Journal of Systems Chemistry*, 3(1), pp. 1–5. doi: 10.1186/1759-2208-3-3.
Patel, B. H. *et al.* (2015) 'Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism', *Nature Chemistry*. Nature Publishing Group, 7(4), pp. 301–307. doi: 10.1038/nchem.2202.
Poole, A. M. (2009) 'Horizontal gene transfer and the earliest stages of the evolution of life', *Research in Microbiology*. Elsevier Masson SAS, 160(7), pp. 473–480. doi: 10.1016/j.resmic.2009.07.009.
Preiner, M. *et al.* (2019) 'The future of origin of life research: bridging decades-old divisions', *Life*, (January), p. under review. doi: 10.3390/life10030020.
Pressman, A. *et al.* (2017) 'Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences', *Nucleic Acids Research*, 45(14), pp. 8167–8179. doi: 10.1093/nar/gkx540.
Rimmer, P. B. *et al.* (2018) 'The origin of RNA precursors on exoplanets', *Science Advances*, 4(8), pp. 1–12. doi: 10.1126/sciadv.aar3302.
Sasselov, D. D., Grotzinger, J. P. and Sutherland, J. D. (2020) 'The origin of life as a planetary phenomenon', *Science Advances*, 6(6), pp. 1–10. doi: 10.1126/sciadv.aax3419.

Schank, J. C. and Wimsatt, W. C. (1986) 'Generative Entrenchment and Evolution', *Proceedings of the Biennial Meeting of the Philosophy of Science Associatio*, 2, pp. 33–60.
Schuster, P. *et al.* (1994) 'From sequences to shapes and back: A case study in RNA secondary structures', *Proceedings of the Royal Society B: Biological Sciences*, 255(1344), pp. 279–284. doi: 10.1098/rspb.1994.0040.
Schwartz, A. W. (2013) 'Evaluating the plausibility of prebiotic multistage syntheses', *Astrobiology*, 13(8), pp. 784–789. doi: 10.1089/ast.2013.1057.
Shah, V. *et al.* (2019) 'Survival of RNA replicators is much easier in protocells than in surface-based, spatial systems', *Life*, 9(3). doi: 10.3390/life9030065.
Smith, J. M. and Szathmáry, E. (1995) *The Major Transitions in Evolution*. Oxford: Oxford University Press.
Solé, R. (2016) 'Synthetic transitions: Towards a new synthesis', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1701). doi: 10.1098/rstb.2015.0438.
Stüeken, E. E. *et al.* (2013) 'Did life originate from a global chemical reactor?', *Geobiology*, 11(2), pp. 101–126. doi: 10.1111/gbi.12025.
Sutherland, J. D. (2017) 'Opinion: Studies on the origin of life-The end of the beginning', *Nature Reviews Chemistry*. Macmillan Publishers Limited, 1, pp. 1–8. doi: 10.1038/s41570-016-0012.
Szabó, P. *et al.* (2002) 'In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity', *Nature*, 420(6913), pp. 340–343. doi: 10.1038/nature01187.
Szathmáry, E. (2005) 'In search of the simplest cell', *Nature*, 433(February), pp. 469–470.
Szathmáry, E. (2006) 'The origin of replicators and reproducers', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1474), pp. 1761–1776. doi: 10.1098/rstb.2006.1912.
Szathmáry, E. (2015) 'Toward major evolutionary transitions theory 2.0', *Proceedings of the National Academy of Sciences of the United States of America*, 112(33), pp. 10104–10111. doi: 10.1073/pnas.1421398112.
Szathmáry, E. and Demeter, L. (1987) 'Group selection of early replicators and the origin of life', *Journal of Theoretical Biology*, 128(4), pp. 463–486. doi: 10.1016/S0022-5193(87)80191-1.
Takeuchi, N., Hogeweg, P. and Kaneko, K. (2017) 'Conceptualizing the origin of life in terms of evolution', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 375(2109), pp. 1–10. doi: 10.1098/rsta.2016.0346.
Takeuchi, N., Poorthuis, P. H. and Hogeweg, P. (2005) 'Phenotypic error threshold; additivity and epistasis in RNA evolution', *BMC Evolutionary Biology*, 5, pp. 1–9. doi: 10.1186/1471-2148-5-9.
Todisco, M. *et al.* (2018) 'Nonenzymatic Polymerization into Long Linear RNA Templated by Liquid Crystal Self-Assembly', *ACS Nano*, 12(10), pp. 9750–9762. doi: 10.1021/acsnano.8b05821.
Trinks, H., Schröder, W. and Biebricher, C. K. (2005) 'Ice and the origin of life', *Origins of Life and Evolution of the Biosphere*, 35(5), pp. 429–445. doi: 10.1007/s11084-005-5009-1.
Vasas, V. *et al.* (2012) 'Evolution before genes', *Biol. Direct*, 7(1), p. 1. doi: 10.1186/1745-6150-7-1.
Wachowius, F. and Holliger, P. (2019) 'Non-Enzymatic Assembly of a Minimized RNA Polymerase Ribozyme', *ChemSystemsChem*, 1(1–2), pp. 12–15. doi: 10.1002/syst.201900004.
Woese, C. R. (2002) 'On the evolution of cells', *Proceedings of the National Academy of Sciences of the United States of America*, 99(13), pp. 8742–8747. doi: 10.1073/pnas.132266999.
Woese, C. R. and Fox, G. E. (1977) 'The Concept of Cellular Evolution', *Journal of Molecular Evolution*, 10, pp. 1–6.
Wu, M. and Higgs, P. G. (2012) 'The origin of life is a spatially localized stochastic transition', *Biology Direct*. Biology Direct, 7(1), p. 1. doi: 10.1186/1745-6150-7-42.
Yu, H., Zhang, S. and Chaput, J. C. (2012) 'Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor', *Nature Chemistry*. Nature Publishing Group, 4(3), pp. 183–187. doi: 10.1038/nchem.1241.

Main figure titles and legends

Figure 1. A sample of thresholds in origin of life research. (a) Phase diagram representing necessary conditions for a meteoritic impact to deliver a quantity of HCN above a threshold value of 10 mM (required for its polymerization into nucleobases, assuming an oxidizing early Earth atmosphere). Diagram obtained by mathematical modelling; meteorite radius represented on the x-axis; impact velocity on the y-axis; several mixing depths below water surface are represented (from Parkos et al. (2018) with permission). **(b)** Diagram of the photoactivated reactions from cis-DAMN to AICN compounds which are key intermediates in the prebiotic synthesis of purine nucleotides. Ground and excited states respectively represented in solid and dotted lines; dashed arrows indicate backward reactions. This pathway reveals different free-energy barriers of at least 52 kcal/mol that can be overcome thanks to light energy (from Boulanger et al. (2013) with permission). **(c)** Liquid-crystal phase diagram delimiting 3 areas: a homogeneous phase of mixed RNA and polyethylene glycol (PEG) (in blue), a liquid-liquid phase separation of RNA-rich droplets in PEG solution (in orange) and a liquid-liquid crystals (LC) microdomains phase separation (in green). Going from one area to another is made possible by varying temperature or molecular crowding (i.e. PEG concentration) (from Todisco et al. (2018) with permission). **(d)** Graph describing an autocatalytic set of theoretical polymers that has reached catalytic closure

(threshold). The reactions are represented by nodes connecting reactants and products of ligation and cleavage reactions. Dashed lines indicate catalysis and point from catalyst to catalyzed reaction (from Farmer et al. (1986) with permission). **(e)** Phase diagram for an RNA polymerization model as a function of catalyst feedback efficiency k and spontaneous polymerization rate r_0. So called "dead" state characterized by slow spontaneous rate of RNA synthesis, "living" state by rapid autocatalytic rate of RNA synthesis (from Wu and Higgs (2012) with permission). **(f)** Diagram representing the conjectured transition from a state of progenote cells (genetic transfer mainly achieved horizontally) to a state of speciated individuals (lineages resulting from vertical genetic transfer). Passed the Darwinian Threshold, individuals go from a communal evolution to a Darwinian evolution. Such a transition can be triggered by a decreased in competence initiated by environmental parameters or selected mutations (from Goldenfeld, Biancalani and Jafarpour (2017) with permission).

Figure 2. Phase diagrams for the replicase scenario. (a) Phase diagram showing the areas where the replicase regime is possible (green area delimited by the red and blue curves) in the parameter space (e, l), where e is the error rate of the replicase and l the length of the copied sequences, including the replicase itself. The red curve represents the function e_{max} that sets the maximal replicase error rate e_{max} for each given length l of the copied sequence, and for a given selective advantage (s in equation (1)); it gives the maximal length of copied sequences, for each given error rate e. When e is low enough, $e_{max}(l) \sim \frac{1}{l}$; when e is close to 1, equation (1) is no longer valid and l exponentially tends towards 0 (Eigen, 1971). Above the red curve, the copied sequences vanish in the error catastrophe. The blue curve represents the function e_{min} that sets the minimal error rate e_{min} a replicase can reach due to biochemical considerations. As first approximation, higher fidelity needs to be encoded in longer sequences. When e tends towards 1, the blue curve reaches a plateau corresponding to the minimum length that a replicase must have to catalyze a phosphodiester bond between two nucleotides. For a given length, there is no existing replicase under e_{min}. Note that l stands for the length of any copied sequence in the case of the red curve, whereas it only stands for the length of the replicase in the case of the blue curve; the representation of the two lengths on the same graph is possible thanks to the fact that the replicase regime assumes that the replicase is itself subjected to replication as well. The blue and red curves are obtained by merging models and experiments; their shape can therefore change. For example, in **(b)**, the existence zone is wider than in **(a)**. In **(c)**, there is no existence zone at all; as a consequence, the replicase scenario must be revised or abandoned in this case. Figure **(d)** represents the distribution f, i.e. the number of available replicases as a function of their copying error rate (at a given illustrative length of 200 nucleotides, as in **(b)**). The purple zone bounded between e_{min} and e_{max} depicts the total number of possible sustainable replicases.

Figure 3. Different ways of crossing thresholds. The y-axis represents a relevant environmental variable whereas the x-axis represents a variable of the system itself, in **(a)**, **(b)** and **(c)**. The phase diagrams are divided in two: a green region β where the system displays a specific property p and a region α where it does not. **(a)** The system, initially in point A, acquires the property p in point B thanks to an environmental change from E_1 to E_2, i.e. via an extrinsic transition. Vesicle formation from a micellar solution can be so described, where E would be for instance the concentration of dissolved CO_2 (Bachmann, Luigi and Lang, 1992). **(b)** The system undergoes a transition from A to B because of an internal variation of the system, in a given environment, i.e. via an intrinsic transition. Stochastic transitions from random polymerization to replicating polymers fit in this representation, where X could describe the diversity of generated polymers (Mathis, Bhattacharya and Walker, 2017). **(c)** The system acquires p (from A to B) thanks to an environmental change. It keeps this property in B" while the environmental variable goes back to its initial value E_1. This is made possible thanks to a transitory change of a variable of the system itself in environment E_2 (from B to B'). For instance, the property p could be the ability to sustain an autocatalytic network, the transient environmental change the shift between dry and rainy periods, and the transition from B to B' the accumulation of autocatalysts. The y-variable would then be environmental hygrometry and the x-variable catalyst concentration. **(d)** Here, two independent systems in A and B fuse into a single one in C that displays p, in environment E_1. This is the case when autocatalytic networks get encapsulated in compartments, endowing the novel system with the property to evolve through natural selection (Vasas *et al.*, 2012).

Figure 4. Articulating different thresholds. Figures **(a)**, **(b)** and **(c)** represent combinations of two phase-diagrams, with x- and y-axes describing relevant systemic and environmental variables (the representation is simplified to fit a 2D illustration). The green region β and the blue region γ represent two regions where properties p_β and p_γ can be acquired respectively. The arrows show different ways to cross a succession of two thresholds. **(a) Entrenchment.** Region γ is included into region β so that the system must acquire p_β before acquiring p_γ. In this configuration, p_β could stand for the ability to generate random polymers and p_γ for the property of having an autocatalytic set of polymers (Mathis, Bhattacharya and Walker, 2017). **(b) Contingency and transience**. Region γ partially intercepts region β so that the system can combine p_β and p_γ in C by first acquiring either one of the two properties. Here p_β and p_γ could stand for the ability to generate one type of

genetic polymer or the other. The system in C would therefore generate both types of polymers at the same time, such as RNA and TNA. Transition from C to D, through a second threshold, shows that a specific property can be lost across phase transitions. It could be the loss of TNA that is not present in the actual living world (Cleaves *et al.*, 2019). **(c) Facilitation**. Getting p_β first lowers the threshold delineating p_γ. This is the case when compartmentalization (p_β) raises the error-threshold of a replicating system (p_γ) making it more likely to appear (Matsumura *et al.*, 2016). **(d)** This **Venn diagram** shows different ways of crossing a series of thresholds. Colored edges represent thresholds that delimit domains with novel properties. Each domain is denoted with a Greek letter. Points denoted with a capital Latin letter stand for a given system state. Arrows between points represent possible trajectories of a prebiotic system across successive domains. The figure exhibits contingency (the system can reach C or D from B), enabling constraint (the system in D can reach F but not E), accumulation (the system in C has accumulated properties gained in A and B), entrenchment (the property acquired in B remains in all other locations) and transience (the system in F has lost the property acquired in A). **(e)** This graph illustrates how crossing successions of thresholds may account for an increase in lifeness over time.

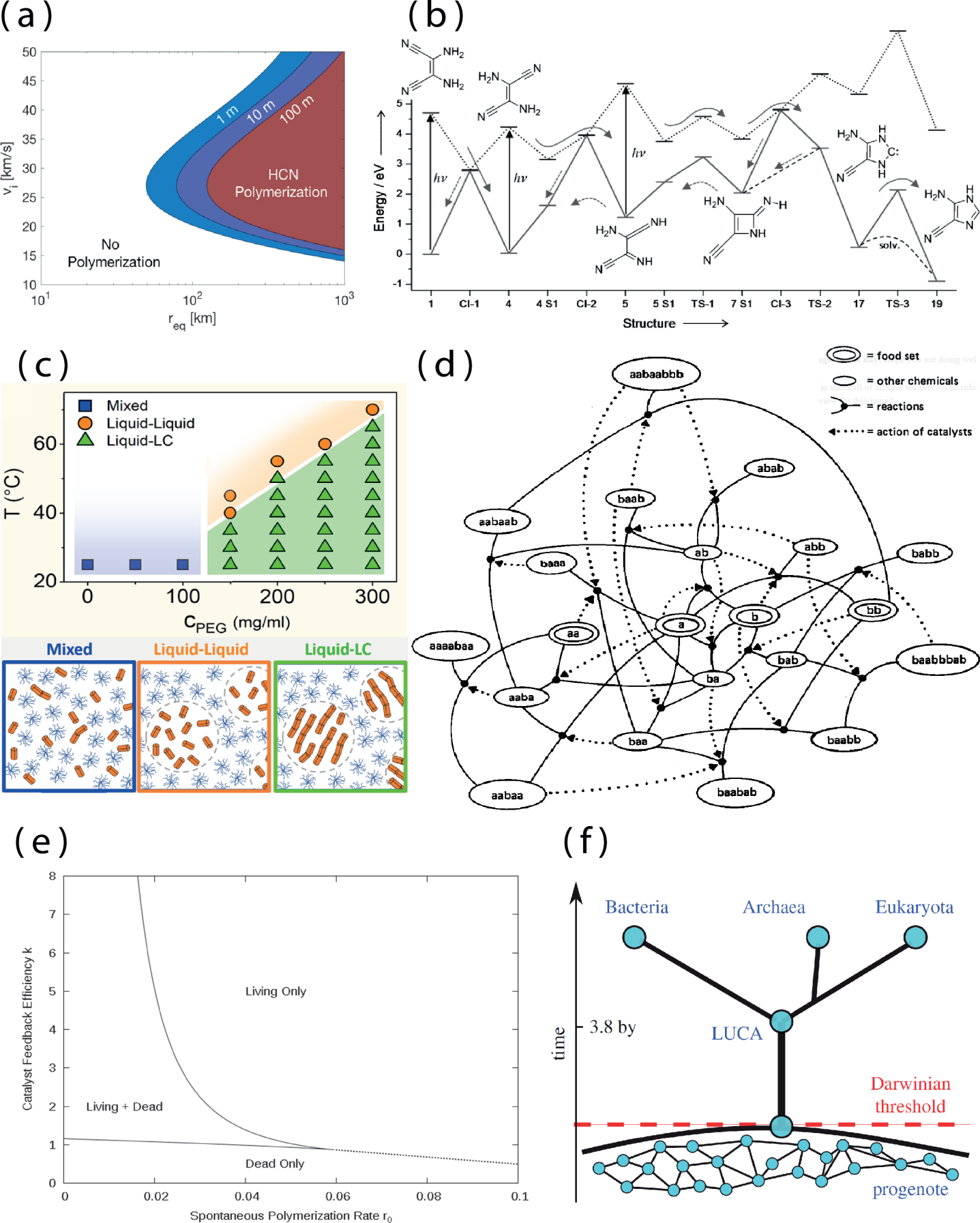
Figure 5. Threshold transitions in origin of life research. Identifying a threshold in an origin of life scenario allows the formalization of a parameterized theoretical model. This model can be refined by the output of appropriate experiments, which reciprocally benefit from a refinement of the theoretical model. These back-and-forth relations make it possible to refine the threshold model. The parameters of the threshold model help to integrate several disciplines among which physics, chemistry or molecular biology. If the threshold is such that the transition from one state to another is possible, it can be used as a basis for modelling such a transition within a broader scenario for the transition from non-living to living matter. Transition possibilities are then constrained by hypothesized environments as determined by Earth sciences, geochemistry, astronomy etc. Plausibility estimates of such transitions allow one to build and estimate multi-stage scenarios articulated around well-identified thresholds. If a scenario reaches a dead-end, it needs to be revised and novel thresholds identified.

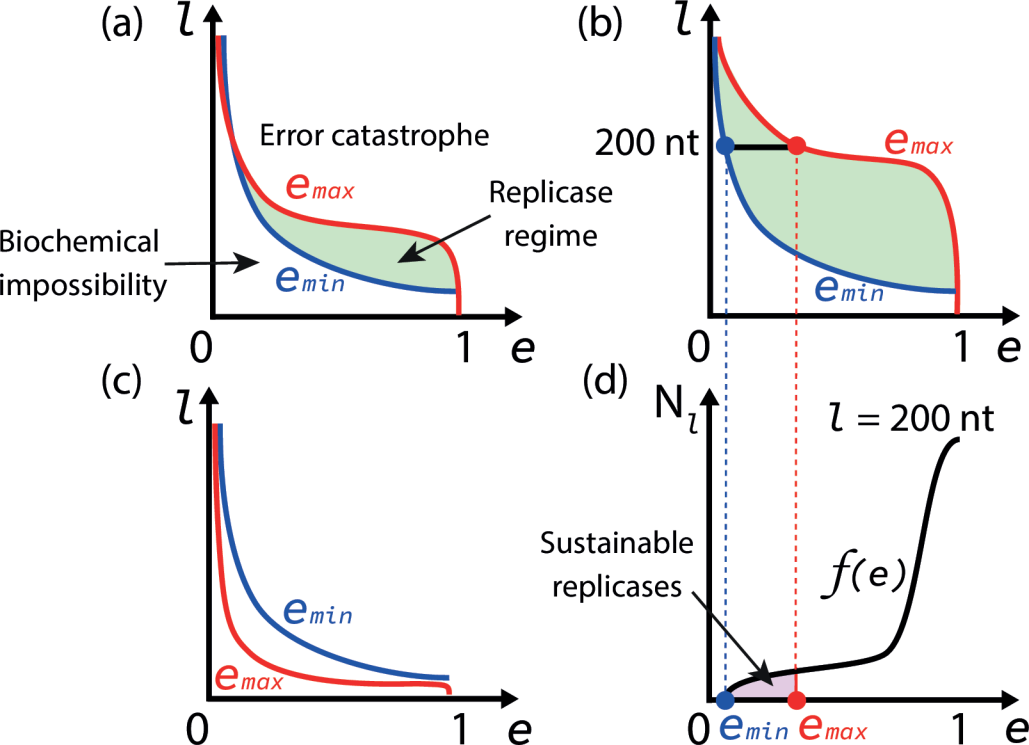
Main tables and legends

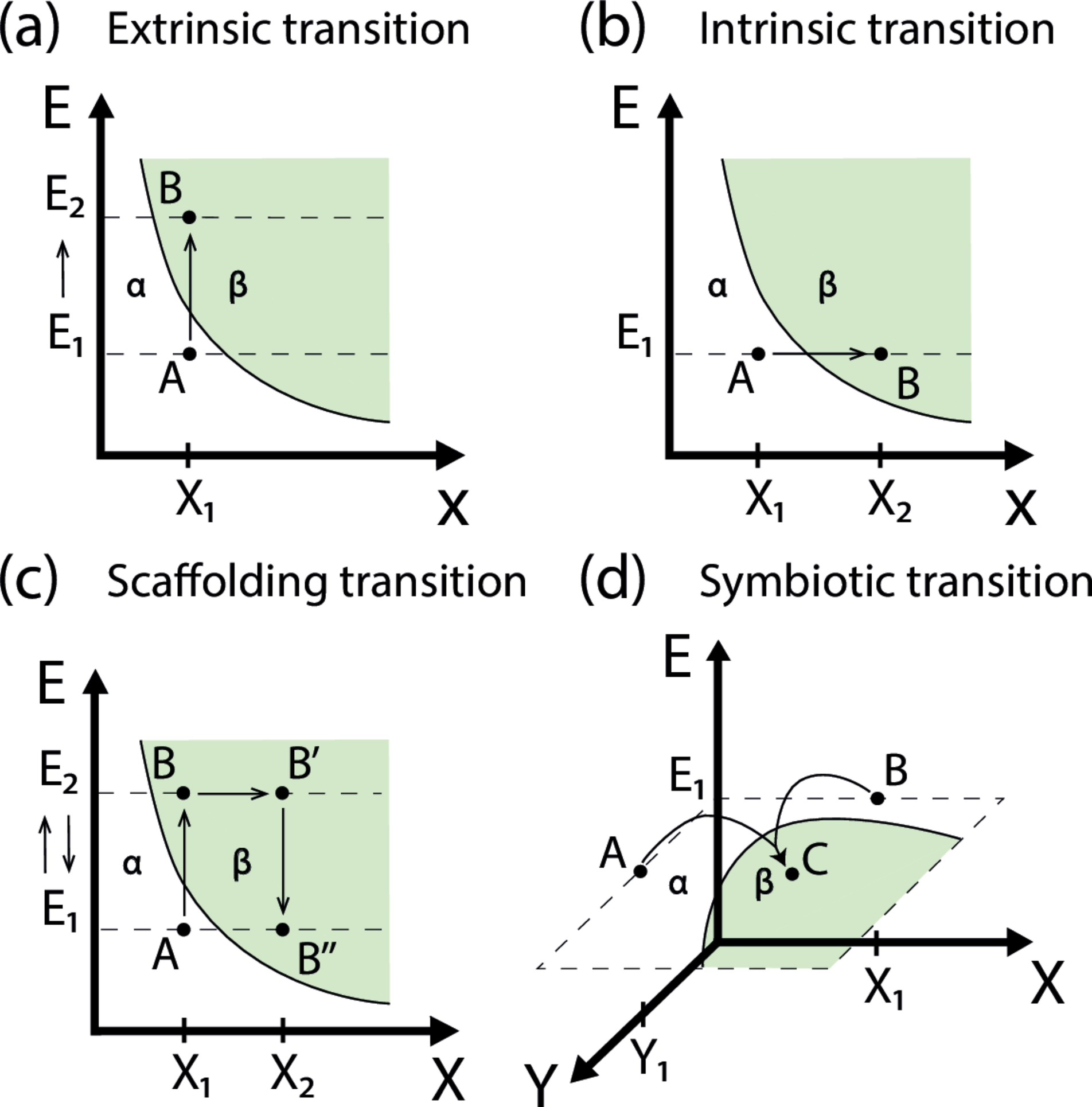
Threshold examples	**System state before threshold**	**System state after threshold**	**Variables triggering the transition (in model or experiment)**	**Naturalized variables posited to have triggered the transition in an origin of life scenario**
Chirality symmetry breaking (Budin and Szostak, 2010; Hawbaker and Blackmond, 2019)	Racemic state of a solution	Homochiral state of a solution	Enantiomeric excess	Presence of circularly polarized light, stereospecific crystallization, isotopic enantioselective initiators, auto-catalysis
Spontaneous polymerization (Monnard, Kanavarioti and Deamer, 2003; Baaske *et al.*, 2007; Lambert, 2008)	Solution of monomers	Solution of polymers	Concentrations of monomers, chemical activation	Ponds evaporation, freeze-thaw cycle, mineral surfaces, thermophoresis in hydrothermal vents, in situ closure in vesicles
Self-assembly of compartments (Bachmann, Luigi and Lang, 1992; Todisco *et al.*, 2018)	Solution of free constituents	Solution of molecular self-assembled compartments	Concentration of the constituents, salinity, pH, temperature, molecular crowding	Increase of CO_2 concentration, day-night temperature variations, wet-dry cycles

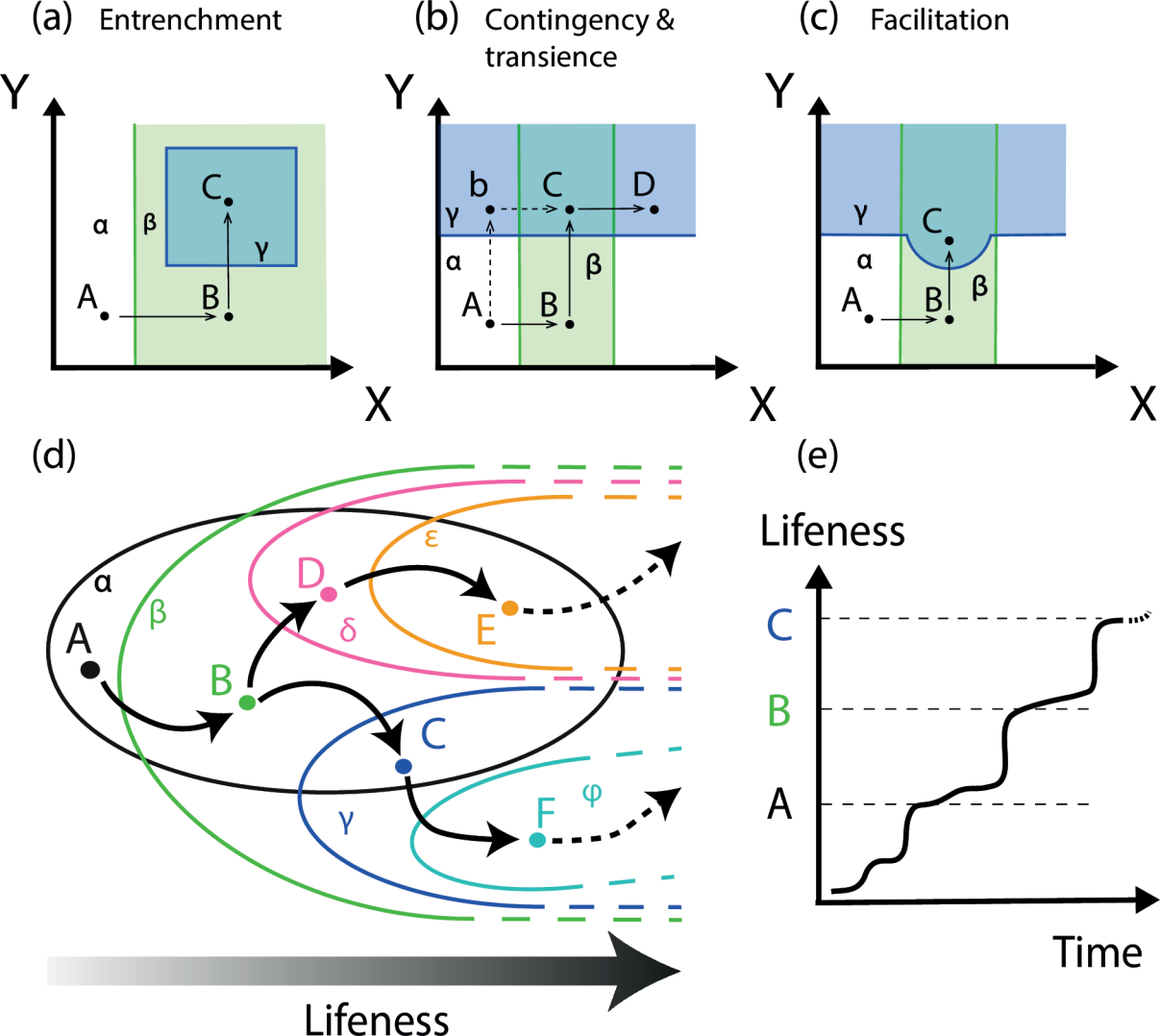
Catalytic closure threshold (Kauffman, 1986)	Solution of polymers with few catalysts	Closed collective autocatalytic sets	Number of catalysts and reactions catalyzed	Spontaneous and effective synthesis of diverse polymers
Error threshold (Eigen, 1971; Takeuchi, Hogeweg and Kaneko, 2017)	Unreplicated polymers	Polymers copied by template-based replication	Selective advantage, copying error rate, polymers length	Selection of compartments, genotype-phenotype redundancy, mineral surfaces
Decay threshold (King, 1977; Szathmáry, 2006; Vasas *et al.*, 2012)	No autocatalysis	Autocatalytic set	Kinetics, network topology, feedstock concentration	Transient depletion in reactants or rare product of pre-existing reactions
Darwinian threshold (Woese, 2002; Goldenfeld, Biancalani and Jafarpour, 2017)	Progenotes with high Horizontal Gene Transfer (HGT)	Speciated individuals with high Vertical Gene Transfer	HGT strength (i.e. competence), fitness	Decrease in cell density, nutrient limitation, alkaline shift, toxic chemicals release

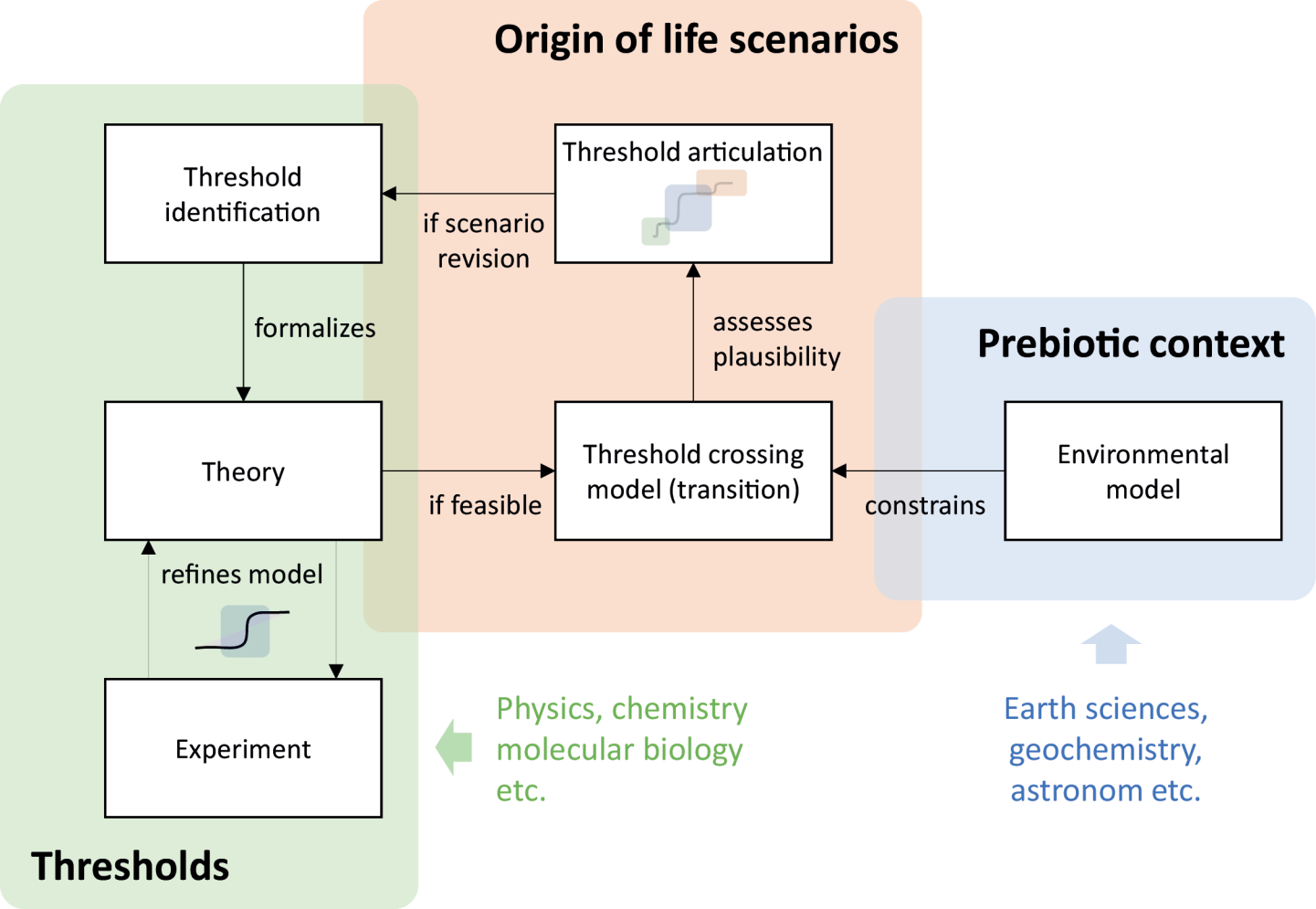
773 **Table 1. Examples of thresholds in origin of life scenarios.** Systems states before and after crossing the
774 threshold are listed, along with corresponding physico-chemical variables and prebiotically relevant phenomena.
775

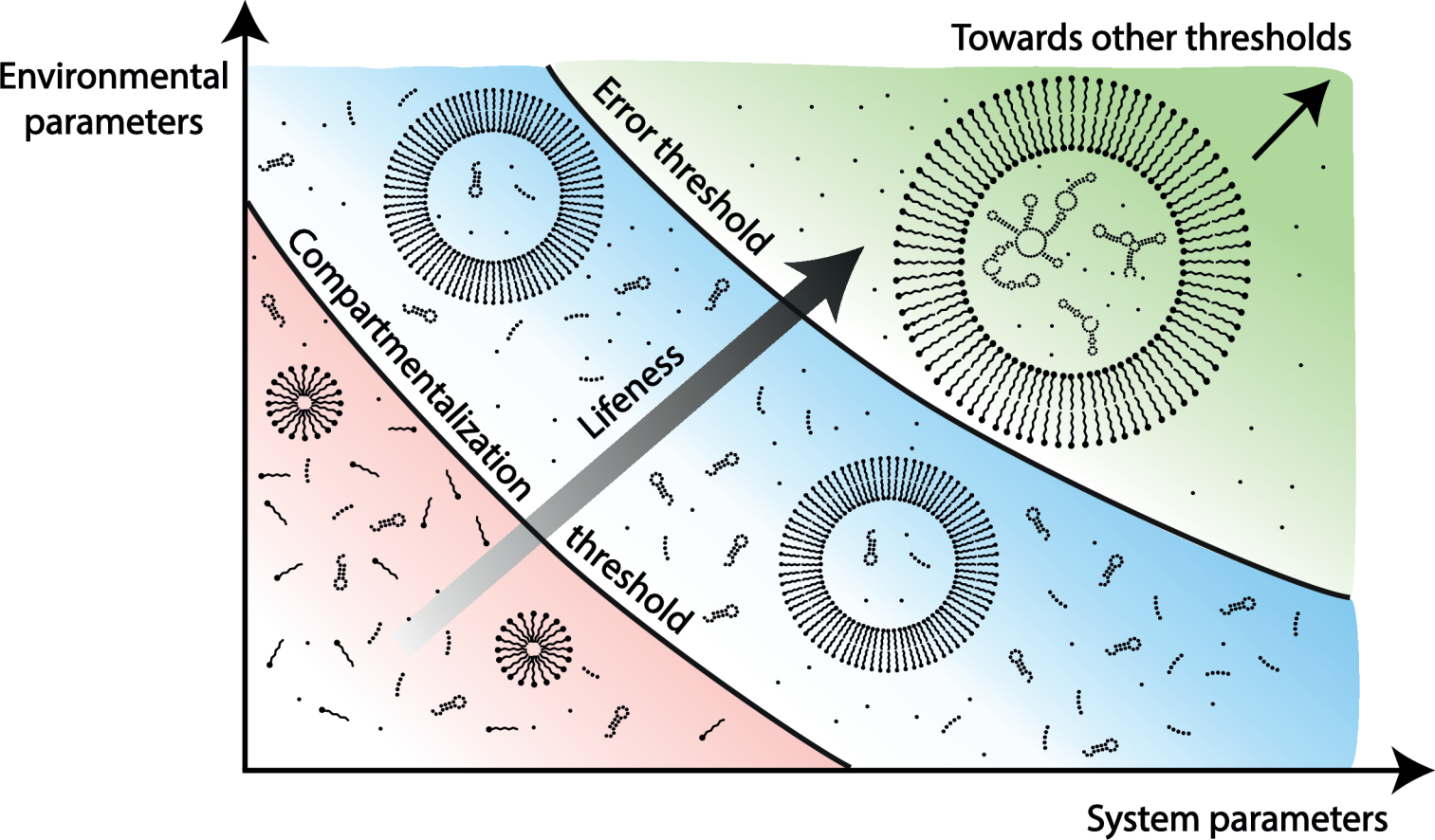

(a)
1 m
10 m
100 m
HCN Polymerization
No Polymerization
v_i [km/s]
r_{eq} [km]
(b)
Energy / eV
Structure
1
CI-1
4
4 S1
CI-2
5
5 S1
TS-1
7 S1
CI-3
TS-2
17
TS-3
19
hν
solv.
(c)
Mixed
Liquid-Liquid
Liquid-LC
T (°C)
C_{PEG} (mg/ml)
Mixed
Liquid-Liquid
Liquid-LC
(d)
= food set
= other chemicals
= reactions
= action of catalysts
aabaabbb
abab
baab
aabaab
ab
abb
babb
baaa
a
b
bb
aa
aaaabaa
bab
baabbbab
ba
aaba
baa
baabb
aabaa
baabab
(e)
Catalyst Feedback Efficiency k
Spontaneous Polymerization Rate r_0
Living Only
Living + Dead
Dead Only
(f)
Bacteria
Archaea
Eukaryota
time
3.8 by
LUCA
Darwinian threshold
progenote


(a)
l
Error catastrophe
e_max
Replicase
regime
Biochemical
impossibility
e_min
0
1
e
(b)
l
200 nt
e_max
e_min
0
1
e
(c)
l
e_min
e_max
0
1
e
(d)
N_l
l = 200 nt
Sustainable
replicases
f(e)
0
e_min
e_max
1
e


(a) Extrinsic transition
E
E2
E1
B
A
α
β
X1
X
(b) Intrinsic transition
E
E1
α
β
A
B
X1
X2
X
(c) Scaffolding transition
E
E2
E1
B
B′
A
B″
α
β
X1
X2
X
(d) Symbiotic transition
E
E1
B
A
C
α
β
X1
X
Y1
Y


(a) Entrenchment
(b) Contingency & transience
(c) Facilitation
Y
X
α
β
γ
A
B
C
b
D
(d)
ε
δ
E
F
φ
Lifeness
(e)
Time


Origin of life scenarios
Threshold identification
Threshold articulation
if scenario revision
formalizes
assesses plausibility
Prebiotic context
Theory
Threshold crossing model (transition)
Environmental model
if feasible
constrains
refines model
Experiment
Physics, chemistry molecular biology etc.
Earth sciences, geochemistry, astronom etc.
Thresholds


Towards other thresholds
Environmental parameters
Error threshold
Compartmentalization threshold
Lifeness
System parameters