Spontaneous emergence of self-replication in chemical reaction systems

Yu Liu[a,1] and David Sumpter[a]

[a] Department of Mathematics, Uppsala University, 75105 Uppsala, Sweden

Explaining origins of life requires us to explain how self-replication arises. Specifically, how can self-replicating entities develop spontaneously from chemical reaction systems in which no reaction is self-replicating? Previous models either supply a framework for minimal living systems or only consider catalyzed reactions, and thus fail to provide a comprehensive theory. We establish a general model for chemical reaction systems that properly accounts for energetics, kinetics and conservation laws. We find that (1) some systems are collectively-catalytic (e.g., the citric acid cycle), while others self-replicate as a whole (e.g., the formose reaction); (2) side reactions do not always inhibit such systems; (3) many alternative chemical universes often contain one or more such systems; (4) in some self-replicating systems, the entropy of certain parts spontaneously decreases; (5) complex self-replicating molecules emerge spontaneously from simple reaction systems through a sequence of transitions. Together these results start to explain origins of prebiotic evolution.

Origin of life | Self-replication | Collectively-catalytic | Prebiotic evolution | Biological complexity

$\mathbf{S}$elf-replication is one of the central properties of life (1), and to explain life's origins we need to explain how self-replication arose. It is widely accepted that before DNA, life replicated through RNA molecules (2). However, it remains unclear how the building blocks of RNA, such as nucleotides, became available on the primitive Earth, and even if these building blocks were abundant, it is unclear how they were assembled into the first RNA (3, 4). It is plausible that self-replication did not originate from a single complex independent self-replicating molecule. In the early stage of evolution, the "precursor life" could be very different from what we see today (3). For example, in Wächtershäuser's iron-sulfur world hypothesis, the precursor life does not have nucleic acids but consists of a self-replicating (or autocatalytic, in his words) metabolic network (5). Another proposal, by Szathmáry, is that life evolved from "holistic limited hereditary replicators" such as the formose reaction—in which sugar is replicated from formaldehyde—to "modular unlimited hereditary replicators" such as RNA and today's DNA (6–8).

There are lots of biological examples of self-replicating systems that do not rely on a single complex template molecule (as is the case for DNA and RNA molecules). These include: the malic acid cycle (a metabolic path of some bacteria and plants for synthesis of malates); the Calvin cycle in photosynthesis (9); the reductive citric acid cycle for a certain group of chemoautotrophs (10, 11); the metabolic pathways of ATP (as well as some other coenzymes such as NAD^+ and CoA) found in many different living organisms (12); and the whole metabolic reaction network of *E. coli* (13). Self-replication has also been identified in non-living systems such as the formose reaction (9, 14) and experiments in labs (15–17), including self-replication of nucleotide-based oligomers (18).

Several theoreticians have tried to explain the origin of self-replication in terms of a system of coupled chemical reactions (19). For example, the chemoton model describes a system composed of three coupled quasi-self-replicating subsystems (metabolism, membrane, and template), which as a whole is able to self-replicate (9). The chemoton can be considered as a model of a minimal living system, but cannot explain how this system spontaneously develops from a soup of simple molecules. The RAF (reflexively autocatalytic and food-generated) theory (20–22), extended from Kauffman's autocatalytic sets theory (23), is also a step in the direction of understanding origins of self-replication. A set of chemical reactions is RAF if (1) every reaction in this set is catalyzed by at least one molecule involved in this set, and (2) every molecule involved in this set can be produced from a small food molecule set. RAF sets are shown to be able to readily emerge from a set of chemical reactions, and always consist of smaller RAF sets, demonstrating the capability to evolve (24, 25). Other similar models also contributed to this theory (26–29), and many of the biological observations mentioned in the previous paragraph can be put into the framework of RAF theory (13, 15–18).

Although RAF theory is interesting, it has limitations as an explanation of self-replication. Firstly, the theory stipulates that every reaction in the set is catalyzed. Even though catalyzed reactions are very common in living systems, not every reaction involves a catalyst (e.g., condensation reactions (7, 30)). In the early stages of biological evolution, probably no reaction required sophisticated biotic catalysts (e.g., enzymes) (10, 13), so there is a strong motivation not to include these enzymes (18, 31, 32) or even catalysts as given in a model of the origins of self-replication. Uncatalyzed reactions have significant effects on the dynamics of the whole system, e.g., "innovating" new species of molecules and then triggering other RAF sets (33, 34). The second concern with RAF theory is that it is a purely graph-theoretic approach. As a result, there is no constraint on how chemical reactions are constructed and coupled (although it gives the theory much freedom, the construction of reaction systems is too arbitrary to investigate the model systematically). Extra assumptions need to be made about chemical kinetics in order to investigate the dynamics of how populations of molecules change over time (35). Here we see another reason to relax the assumption of studying only catalyzed reaction: catalyzed reactions are never elementary reactions, so the kinetics cannot be simply calculated from the reaction equations (36).

Any theoretical approach to the origin of self-replication should explicitly include energetics (13, 37), an aspect which

[1] To whom correspondence should be addressed. E-mail: yu.ernest.liu@hotmail.com

is missing from all these models and theories above. There are several reasons for this. Firstly, energetics (e.g., Gibbs energy) determines whether a chemical reaction is spontaneous, so to investigate the spontaneity of the emergence of life, it has to be considered. Secondly, in order to concretely discuss the issue—famously put forward by Schrödinger—that life maintains its order by feeding on "negative entropy" (38), energetics has to be explicitly taken into account: entropy is negatively related to the thermodynamic free energy (which is Gibbs energy in the scenario of constant pressure and temperature). It should be noted that the relationship between life and entropy is investigated in different ways and contexts (39–41). For example, Branscomb explained the specific mechanisms to increase thermodynamic free energy in two real-world biochemical scenarios (40): the hypothesized "alkaline hydrothermal vent" (42) on the prebiotic Earth, and the system where the Ferredoxin I protein translocates protons. In the context of statistical physics, a lower limit was derived for the amount of heat generated in a non-equilibrium system where a process of self-replication occurs (41).

In this paper, we set up a general model for chemical reaction systems that properly accounts for energetics, kinetics and the conservation law. Catalysts are not explicitly included in the model, but we later find that catalysis can emerge, along with self-replication and potentially complex molecules, in our system.

Model

Chemical reaction system. We model a well-mixed soup of molecules, each of which is defined by its integer mass, i. A molecule's type is thus denoted $\bar{i}$. Only two types of reaction are possible: synthesis of two molecules to create a molecule of greater mass (e.g., $\bar{2} + \bar{4} \rightarrow \bar{6}$) and decomposition into two molecules to create two molecules of lower mass (e.g., $\bar{6} \rightarrow \bar{1} + \bar{5}$). All reactions that conserve mass—the total mass on the left-hand side of the equation adds up to those on the right-hand side—can occur. For convenience, we define a reaction pair to be a reaction and its corresponding reverse reaction.

Each type of molecule $\bar{i}$ has its own standard Gibbs energy of formation G_i°. Then, as illustrated in Fig. 1, a reaction is
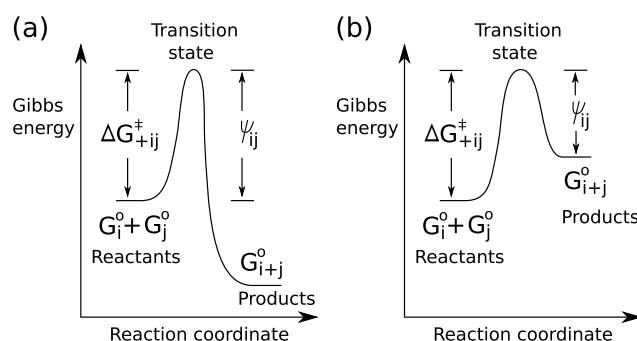

Fig. 1. Diagram of Gibbs energy for a synthesis reaction $\bar{i} + \bar{j} \rightarrow \overline{i+j}$. (a) For the case that the synthesis reaction is spontaneous, i.e., $G_i^\circ + G_j^\circ > G_{i+j}^\circ$. (b) For the case that it is non-spontaneous, i.e., $G_i^\circ + G_j^\circ \leqslant G_{i+j}^\circ$.

either spontaneous—meaning that the total standard Gibbs energy of formation of the reactants is greater than that of

the products—i.e., $G_i^\circ + G_j^\circ > G_{i+j}^\circ$, or non-spontaneous, i.e., $G_i^\circ + G_j^\circ \leqslant G_{i+j}^\circ$. If one reaction in a reaction pair is spontaneous, the other is non-spontaneous, and vice versa.

According to transition state theory (36), the reactants have to overcome the Gibbs energy of activation (namely $\Delta G_{+ij}^{\ddagger}$ in Fig. 1) in order for the reaction to occur. In the model, any reaction pair is either *low-barrier*—the energy barrier (corresponding to ψ_{ij} in Fig. 1) is low and the reaction rate is thus high—or *high-barrier*. In our system all reactions are possible, but to define a specific chemical reaction system we write a list of low-barrier (and spontaneous) reactions only. For example,

$$\begin{cases} \bar{5} \rightarrow \bar{1} + \bar{4} \\ \bar{6} \rightarrow \bar{1} + \bar{5} \\ \bar{2} + \bar{4} \rightarrow \bar{6} \end{cases} \qquad [1]$$
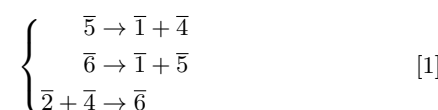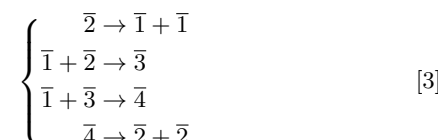
is a model of the citric acid cycle in cellular respiration, simplified so that only carbon-changing reactions are included (9). Specifically, molecule $\bar{1}$ stands for carbon dioxide, $\bar{2}$ for acetyl-CoA, $\bar{4}$ for oxaloacetic acid, $\bar{5}$ for α-Ketoglutaric acid, and $\bar{6}$ for citric acid, respectively.

It is also possible to model the formose reaction (9) which involves the formation of sugars from formaldehyde,

$$\begin{cases} \bar{1} + \bar{2} \rightarrow \bar{3} \\ \bar{1} + \bar{3} \rightarrow \bar{4} \\ \bar{4} \rightarrow \bar{2} + \bar{2} \end{cases} \qquad [2]$$
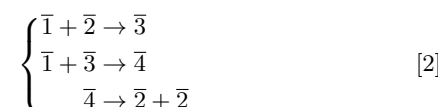
Again, only carbon-changing reactions are considered. Specifically, molecule $\bar{1}$ stands for formaldehyde, $\bar{2}$ for glycolaldehyde, $\bar{3}$ for glyceraldehyde, and $\bar{4}$ for tetrose, respectively. These two reaction systems will serve as special cases in our study, but we will cover a wide range of different systems.

Note that some chemical reaction systems are not physically possible, e.g.,

$$\begin{cases} \bar{2} \rightarrow \bar{1} + \bar{1} \\ \bar{1} + \bar{2} \rightarrow \bar{3} \\ \bar{1} + \bar{3} \rightarrow \bar{4} \\ \bar{4} \rightarrow \bar{2} + \bar{2} \end{cases} \qquad [3]$$

where by adding up these low-barrier reactions, all molecules are cancelled out. As a consequence, one cannot find proper Gibbs energy for each molecule such that each reaction above is spontaneous. We are only interested in physically possible systems.

In order to be as simple as possible, a strong assumption we make is that molecules with the same mass are identical. In real chemistry, molecules with the same mass may have different atomic compositions (e.g., CO and N_2), and even when molecules have the same atomic compositions, they can be different isomers (e.g., ethanol and methoxymethane). Similarly, certain chemical transformations are impossible, e.g., organometallic compounds can never be produced by a chemical reaction system involved with only carbohydrates. Our model currently excludes these real-world possibilities and limitations. However, we emphasise that the "integer molecule mass" in our model does not necessarily correspond to the physical mass, but should be thought of as a rough ordering of the relative complexity of molecules.

Kinetics. We now specify a general model for the kinetics of our chemical system, under the following assumptions: (1) All molecules are ideally gaseous; (2) the whole system is kept at constant pressure and temperature; (3) every possible reaction is elementary. The derivation follows the rate law and transition state theory (36). Here we cover the key points, and a full derivation is given in Supporting Information (SI) section S1.

For any synthesis reaction $\bar{i} + \bar{j} \to \overline{i+j}$, the reaction rate in unit s^{-1} is

$$\gamma_{+ij} = \beta \exp(-\kappa \Delta G^{\ddagger}_{+ij}) \cdot N_i \cdot N_j/(S+N) \qquad [4]$$

where the subscript $_{+ij}$ stands for the synthesis reaction, β and κ are constants, N_i is the number of molecules $\bar{i}$ in the system, N_j is the number of $\bar{j}$, S is the number of solvent molecules, determining the global rate at which molecules interact, N is the number of all the molecules except for solvent molecules, and $\Delta G^{\ddagger}_{+ij}$, as shown in Fig. 1, is defined as

$$\Delta G^{\ddagger}_{+ij} = \begin{cases} \psi_{ij}, & \text{if } G_i^{\circ} + G_j^{\circ} > G_{i+j}^{\circ} \\ \psi_{ij} + G_{i+j}^{\circ} - (G_i^{\circ} + G_j^{\circ}), & \text{if } G_i^{\circ} + G_j^{\circ} \leqslant G_{i+j}^{\circ} \end{cases}$$

When implementing the model, we set values of ψ_{ij} (positive and finite) for each reaction pair and G_i° for each molecule. Together, these give a unique value for $\Delta G^{\ddagger}_{+ij}$.

Likewise, for any decomposition reaction $\overline{i+j} \to \bar{i} + \bar{j}$, the reaction rate in unit s^{-1} is

$$\gamma_{-ij} = \beta \exp\left(-\kappa \Delta G^{\ddagger}_{-ij}\right) \cdot N_{i+j} \qquad [5]$$

where the subscript $_{-ij}$ stands for this decomposition reaction, and

$$\Delta G^{\ddagger}_{-ij} = \begin{cases} \psi_{ij}, & \text{if } G_{i+j}^{\circ} > G_i^{\circ} + G_j^{\circ}, \\ \psi_{ij} + (G_i^{\circ} + G_j^{\circ}) - G_{i+j}^{\circ}, & \text{if } G_{i+j}^{\circ} \leqslant G_i^{\circ} + G_j^{\circ}. \end{cases}$$

Due to the fact that the transition state of a reaction pair $\bar{i} + \bar{j} \to \overline{i+j}$ and $\overline{i+j} \to \bar{i} + \bar{j}$ is identical, by setting ψ_{ij} both $\Delta G^{\ddagger}_{+ij}$ and $\Delta G^{\ddagger}_{-ij}$ are uniquely determined. In addition, by setting ψ_{ij} large or small, we can easily make the reaction pair low-barrier or high-barrier.

Simulation of dynamics. We take the citric acid cycle Eq. (1) as an example to illustrate how to set up the simulation experiment.

(1) Set up all the constants. Assume that the constant pressure the system is kept at is 100 kPa and the constant temperature is 298.15 K, so we obtain that $\beta \approx 6.21 \times 10^{12}$ s^{-1} and $\kappa \approx 0.403$ mol/kJ.

(2) Set G_i° for each type of molecule (up to $\bar{6}$ in this case), to make sure that the three reactions are spontaneous, that is, $G_5^{\circ} > G_1^{\circ} + G_4^{\circ}$, $G_6^{\circ} > G_1^{\circ} + G_5^{\circ}$ and $G_2^{\circ} + G_4^{\circ} > G_6^{\circ}$. Here we set $G_1^{\circ} = -780$, $G_2^{\circ} = -500$, $G_3^{\circ} = -490$, $G_4^{\circ} = -190$, $G_5^{\circ} = -830$ and $G_6^{\circ} = -900$, all in unit kJ/mol. These values are set in the range of normal chemical substances' Gibbs energy of formation (36), roughly in the range $(-1500, 300)$. Note that the choice is not unique and a wide range of choices can be made to allow the system to work.

(3) Set $\psi_{14} = \psi_{15} = \psi_{24} = 10$ for these three low-barrier reactions, and $\psi_{ij} = 100$ for all other reaction pairs (all in units kJ/mol).

(4) Assume that there is an unlimited reservoir of resource molecule $\bar{2}$. That is, whenever a molecule $\bar{2}$ is consumed or produced, it is replenished or removed so that the number of $\bar{2}$ always keeps as a constant $Q = 1000$. This setting makes biological sense if we consider a system separated from the unlimited reservoir by a "wall". As long as some resource molecules are consumed, more will enter the system driven by the chemical gradient.

(5) We denote the number of molecule $\bar{i}$ at time t as $N_i(t)$. Initially, $N_4(0) = 1$, and $N_1(0) = N_3(0) = N_5(0) = N_6(0) = 0$. Molecule $\bar{4}$ triggers the chemical reaction system. We also set $S = 1 \times 10^6$, which is much larger than Q, so that we are in the dilute limit.

We then use the standard Gillespie Algorithm to simulate the dynamics (see details in SI section S2). In addition, we also construct ordinary differential equations (ODEs) to describe the mean-field dynamics (see details in SI section S3).

Results

Collectively-catalytic system. We first simulate the citric acid cycle Eq. (1) and observe that $N_1(t)$ increases linearly. Molecules $\bar{4}, \bar{5}$ and $\bar{6}$ are involved in a cycle of reactions, but the total number is constant. The SI section S4 shows this dynamics and ODEs solutions in more detail. We also observe that each reaction in Eq. (1) occurs approximately the same number of times. We can add up these reactions, cancel out the molecules appeared on both sides of the reaction (in this case, molecule $\bar{4}$, $\bar{5}$ and $\bar{6}$), and then obtain the overall reaction of the system $\bar{2} \to \bar{1} + \bar{1}$. In itself, the reaction $\bar{2} \to \bar{1} + \bar{1}$ is high-barrier, so its reaction rate is extremely low, but through the whole system the actual rate of the overall reaction is several billion times larger (SI section S4). We call system Eq. (1) *collectively-catalytic system*, since the overall reaction $\bar{2} \to \bar{1} + \bar{1}$ is catalyzed by molecule $\bar{4}$, $\bar{5}$ and $\bar{6}$. Note that this outcome is consistent with the biological observation that the citric acid cycle consumes acetyl-CoA and produces carbon dioxide as a waste product.

The linear growth of molecules $\bar{1}$ is because: (1) $\bar{1}$ is an end product which cannot be used by these low-barrier reactions; (2) the number of resource molecules $\bar{2}$ is constant; and (3) no additional molecule $\bar{4}$, $\bar{5}$ and $\bar{6}$ can be produced through these low-barrier reactions (by noting that the number of $\bar{4}$, $\bar{5}$ and $\bar{6}$ on the right-hand side of the reaction system Eq. (1) is the same as the number on the left-hand side). We can use these observations to give a rigorous set of criteria for collectively-catalytic systems. We start by defining an *intermediate molecule* to be any molecule that appears on both the reactant side and the product side. In general, the following stoichiometric criteria are sufficient (but not necessary) to show that a physically possible chemical reaction system, given supplies of resource molecules, is collectively-catalytic: (1) For every low-barrier reaction, at least one type of its reactants comes from the products of other low-barrier reactions (called the criterion for self-driven); (2) by adding up all the low-barrier reactions, for every type of intermediate molecule, the number of times it appears on the reactant side and on the product side is the same (called the criterion for balanced-cancelling).

The citric acid cycle Eq. (1) thus satisfies all these criteria, but we also find that other systems fulfill these criteria too (in fact, any single catalytic reaction can be written as a collectively-catalytic system: see SI section S5 for details). The criteria above give us a way of discerning whether or not

a system is collectively-catalytic based on stoichiometry alone, without the need to investigate its dynamics.

Self-replicating system. We now look at the dynamics of the formose reaction Eq. (2) given resource molecule $\bar{1}$. The result of a simulation is shown in Fig. 2. All three intermediate
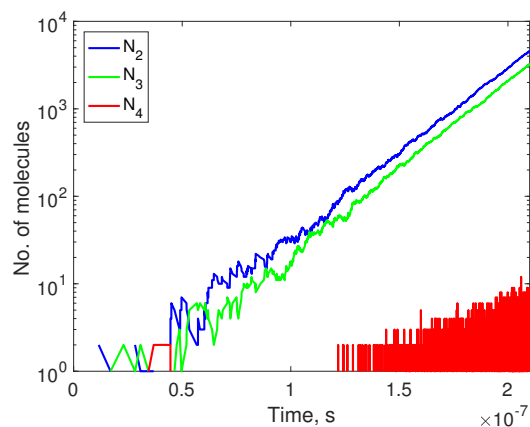

Fig. 2. Dynamics of the formose reaction Eq. (2) in log-normal scale, i.e., x-axis is in normal scale and y-axis is in logarithmic scale. It is not so clear that $N_4(t)$ grows exponentially because $N_4(t)$ is always small. But in solutions of ODEs, we see it clearly (SI Fig. S3a). Note that $N_4(t)$ fluctuates frequently between 0 and small numbers, so the curve looks like a block. We set $G_1^\circ = 220$, $G_2^\circ = -760$, $G_3^\circ = -970$, $G_4^\circ = -1160$, $N_1(t) = Q$, $N_2(0) = 1$ and $N_3(0) = N_4(0) = 0$.

molecules ($N_2(t)$, $N_3(t)$ and $N_4(t)$) increase exponentially. The solutions of the corresponding ODEs are consistent with the simulations (SI section S6).
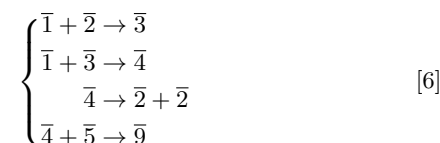
The fact that $\bar{2}$ grows exponentially can be seen directly by adding up the three low-barrier reactions to obtain $\bar{1}+\bar{1}+\bar{2} \to \bar{2}+\bar{2}$. It indicates that if one $\bar{2}$ is present beforehand, one extra $\bar{2}$ can be produced, by transforming two of $\bar{1}$. Then, the additional $\bar{2}$ is further used by the system, and more $\bar{2}$s are produced. Although the molecules $\bar{3}$ and $\bar{4}$ are canceled out when we add up the reactions, they also grow exponentially (with $\bar{4}$ increasing much slower than the other two). The reason is that the actual reaction rate of each low-barrier reaction is not the same (see SI section S6). This observation is important since it illustrates that it is not just one type of molecule that grows exponentially in such systems, but all the intermediate molecules.

We define a *self-replicating system*, of which Eq. (2) is an example, to be a system in which at least one type of molecule is replicated. By investigating various self-replicating systems, we find that not only exponential but also super-exponential growth is observed (see examples in SI section S7). The dynamics indicates that the reactions in self-replicating systems become faster and faster. This is a very special property compared to the collectively-catalytic system, where the overall reaction rate keeps constant, e.g., the citric acid cycle Eq. (1).

In general, the following stoichiometric criteria are sufficient (but not necessary) to show that a physically possible chemical reaction system, given supplies of resource molecules, is self-replicating: (1) The criterion for self-driven (mentioned in last section) is satisfied; (2) there are some types of intermediate molecules, and the number of times it appears on the reactant side is less than that on the product side (called the criterion for

overproduction); (3) there is no type of intermediate molecules that the number of times it appears on the reactant side is larger than that on the product side (called the criterion for no-overintake). The formose reaction Eq. (2) satisfies all these criteria.
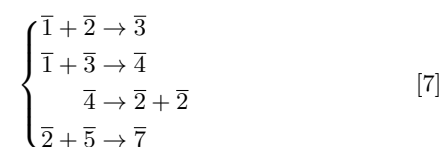
Effect of side reactions on self-replicating systems. Let us consider the following system Eq. (6), which is the formose reaction coupled with an additional reaction that transforms $\bar{4}$ and $\bar{5}$ to $\bar{9}$,

$$\begin{cases} \bar{1}+\bar{2} \to \bar{3} \\ \bar{1}+\bar{3} \to \bar{4} \\ \quad\ \bar{4} \to \bar{2}+\bar{2} \\ \bar{4}+\bar{5} \to \bar{9} \end{cases} \quad [6]$$

The last reaction can be thought of as a side reaction, that consumes the intermediate molecules.

When there is an infinite reservoir of molecule $\bar{1}$, the dynamics of this system are exactly the same as in Fig. 2, because the side reaction $\bar{4}+\bar{5} \to \bar{9}$ cannot proceed without $\bar{5}$. Moreover, even with an infinite reservoir of both molecules $\bar{1}$ and $\bar{5}$, the dynamics of this system are unaffected, because reaction $\bar{4} \to \bar{2}+\bar{2}$ occurs so fast that the side reaction $\bar{4}+\bar{5} \to \bar{9}$ cannot obtain molecule $\bar{4}$ to proceed. In general, decomposition reactions, such as $\bar{4} \to \bar{2}+\bar{2}$, occur faster than synthesis reactions because the reaction rate for synthesis, γ_{+ij}, has an extra term $N_j/(S+N)$ (comparing with γ_{-ij} which is always smaller than 1 (see Eq. (4) and Eq. (5)). In our experimental setting, S is very large, so this extra term is much smaller than 1. In such settings, the self-replicating system outcompetes the side reaction.

Nonetheless, self-replication can be inhibited by some side reactions. Consider the following system Eq. (7),

$$\begin{cases} \bar{1}+\bar{2} \to \bar{3} \\ \bar{1}+\bar{3} \to \bar{4} \\ \quad\ \bar{4} \to \bar{2}+\bar{2} \\ \bar{2}+\bar{5} \to \bar{7} \end{cases} \quad [7]$$

Given infinite reservoirs of both molecule $\bar{1}$ and $\bar{5}$, we no longer observe self-replication, i.e., $N_2(t)$ and $N_3(t)$ do not grow exponentially, because this side reaction consumes the newly-produced molecule $\bar{2}$ before it can be used by reaction $\bar{1}+\bar{2} \to \bar{3}$.

Note that we always assumed $\psi_{ij} = 10$ for every low-barrier reaction (referring to Section "Model"), which means that the reaction rate constant $\beta \exp\left(-\kappa\Delta G_{+ij}^\ddagger\right)$, or $\beta \exp\left(-\kappa\Delta G_{-ij}^\ddagger\right)$, for every low-barrier reaction is identical, denoted as ω. We relax this assumption for now. Let the reaction rate constants for the first three reactions in Eq. (7) be identical, ω, and the rate constant for the side reaction be $\eta\omega$ where $\eta \geqslant 0$. (1) When $\eta = 0$, system Eq. (7) goes back to the original formose reaction. (2) When the side reaction occurs at low rate, i.e., $0 < \eta < 1$, the onset of self-replication is delayed but still occurs eventually (the larger η is, the more delayed the onset of self-replication becomes). (3) When the side reaction has a sufficiently high rate, i.e., $\eta \geqslant 1$, it consumes the newly-produced molecule $\bar{2}$ before it can be used by reaction $\bar{1}+\bar{2} \to \bar{3}$, so self-replication is completely inhibited. $\eta = 1$ is the critical value.

In the real chemical universe we are living in, the formose reaction is often inhibited by side reactions. Many products

further react into a "browning tar" so that the formose reaction has a very low yield of sugars (43). However, this does not preclude the existence of a similar reaction system in the prebiotic world, because, in general, the stability of a self-replicating system depends on the existence and reaction rates of other low-barrier reactions. Whether self-replication is inhibited depends on how side reactions are coupled with the system and what the specific condition is. Also note that all the arguments above apply to collectively-catalytic systems.

Collectively-catalytic and self-replicating systems are common. Another natural question to ask is how common these reaction systems are, such as the citric acid cycle and the formose reaction. Specifically, if we construct alternative chemical universes, where we arbitrarily choose which reactions are low-barrier, we can ask how many of the resulting chemical universes contain collectively-catalytic or self-replicating systems. To answer this question we start by using the criterion shared by both of them, that they are self-driven. Table 1 shows the numbers for different L, which is set to be the mass of the largest molecule in the chemical universe in question, in order to have a measurement of the number of all chemical reaction systems. For example, when $L = 6$, there are in total

Table 1. Number of physically possible universes that contain self-driven, collectively-catalytic or self-replicating systems

L	No. physically possible uni. [1]	No. uni. cont. self-driven [2]	Lower bound uni. cont. CC [3]	Lower bound uni. cont. SR [4]
4	79	8 (10%)	0 (0‰)	2 (25.3‰)
5	681	152 (22%)	5 (7.3‰)	10 (14.7‰)
6	16,825	6,886 (41%)	21 (1.2‰)	74 (4.4‰)
7	401,445	232,552 (58%)	184 (0.5‰)	642 (1.6‰)

[1] Number of physically possible universes. [2] Number of physically possible universes that contain self-driven systems. [3] Lower bound on the number of physically possible universes that contain collectively-catalytic systems. [4] Lower bound on the number of physically possible universes that contain self-replicating systems. All the percentages are calculated with respect to the number of all physically possible universes.

$\overline{1}$, $\overline{2}$, $\cdots$, $\overline{6}$ six types of molecules, and thus 9 reaction pairs. By choosing which reaction is low-barrier, we can construct $\sum_{l=0}^{9} \binom{9}{l} \cdot 2^l = 19683$ alternative chemical universes, 16825 of which turn out to be physically possible. Using the criterion for self-driven given above, we find that 6886 (41%) of all the physically possible universes contain self-driven systems. This percentage increases with L, which indicates that self-driven systems are common, and more common in systems involving more types of molecules.
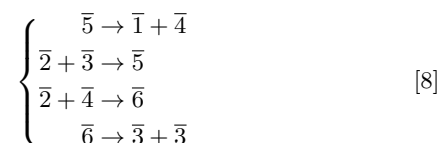
However, we cannot be sure that these self-driven systems are collectively-catalytic or self-replicating, and there is a third type of self-driven system that is non-sustaining system (see SI section S8 for details of non-sustaining system and section S9 for more details of classification for self-driven system). Nevertheless, we can use the stoichiometric criteria mentioned above (note that these criteria are sufficient but not necessary) to give a lower bound on the number of chemical universes that contain collectively-catalytic or self-replicating systems. That is, for a self-driven system, a system is collectively-catalytic if it satisfies the criterion for balanced-canceling, or self-replicating if it satisfies both the criteria for overproduction and no-overintake.

Using the stoichiometric criteria we find that the number of chemical universes containing collectively-catalytic or self-

replicating systems increases with L, although the percentage decreases. This is a lower bound, so does not mean that the actual number of chemical universes containing collectively-catalytic or self-replicating systems decreases. Establishing a firm relationship between the number of chemicals (L) and self-replication will involve simulating dynamics of all systems.

How can life maintain low entropy?. According to the second law of thermodynamics, the total entropy of an isolated system never spontaneously decreases over time. Life, thought of as an open system as opposed to an isolated one, is able to maintain order, i.e., maintain a relatively low entropy level. Schrödinger suggested that this is achieved by life "feeding on negative entropy" (38). His question is how this can happen spontaneously. But before we answer this we first need a way to discuss this question concretely and more quantitively.

Under the framework of our model, if we simply consider life as some self-replicating entity, we should then ask: Is it possible that a self-replicating system spontaneously increases its Gibbs energy or at least keeps it unchanged (we first note that in the scenario of constant pressure and temperature, the decrease of entropy corresponds to the increase of Gibbs energy as $G = H - TS$ where G is Gibbs energy, H is enthalpy, T is temperature and S is entropy)? Let us consider the self-replicating system Eq. (8), given the resource molecule $\overline{2}$:

$$\begin{cases} \overline{5} \to \overline{1} + \overline{4} \\ \overline{2} + \overline{3} \to \overline{5} \\ \overline{2} + \overline{4} \to \overline{6} \\ \overline{6} \to \overline{3} + \overline{3} \end{cases} \qquad [8]$$

The simulation shows that $N_1(t)$, $N_3(t)$ and $N_4(t)$ increase exponentially, as well as $N_5(t)$ and $N_6(t)$ (although they are very small, see SI section S10). Molecule $\overline{1}$ is the end product, while $\overline{3}$, $\overline{4}$, $\overline{5}$ and $\overline{6}$ are replicated through the whole system. We then consider that the "living" system consists of the self-replicating part (namely all of molecules $\overline{3}$, $\overline{4}$, $\overline{5}$ and $\overline{6}$) and the resource molecules in the system.

Now we investigate how Gibbs energy of the system changes. We set Gibbs energy of the initial system to zero (as the reference point), since only relative quantity matters. Therefore, Gibbs energy of the self-replicating part is $G_{replicating}(t) = \sum_{i=3}^{6} N_i(t) \cdot G_i^\circ$. Gibbs energy of the resource molecules in the system is $G_{resource}(t) = -F_2(t) \cdot G_2^\circ$ where $F_2(t)$ is the number of resource molecule $\overline{2}$ ever consumed till time t. Gibbs energy of the waste is $G_{waste}(t) = N_1(t) \cdot G_1^\circ$. Then, Gibbs energy of the living system is $G_{living}(t) = G_{replicating}(t) + G_{resource}(t)$. As shown in Fig. 3, $G_{living}(t)$ increases while $G_{total}(t) = G_{living}(t) + G_{waste}(t)$ decreases.

We have thus given an explicit example of a self-replicating "living" system that spontaneously consumes the resources to increase its own Gibbs energy. Note that our system, as defined, is a well-mixed gas system. So the waste molecules $\overline{1}$ are not automatically separated from other molecules, and Gibbs energy of the gas-mixing process is neglected in the calculation above. However, the contribution of the fact of well-mixed gas is relatively small (see details in SI section S10). So it is still possible for a self-replicating system to spontaneously increase Gibbs energy or at least keep it unchanged.

Spontaneous evolution from simple towards complex. Nature provides many examples of the spontaneous evolution from
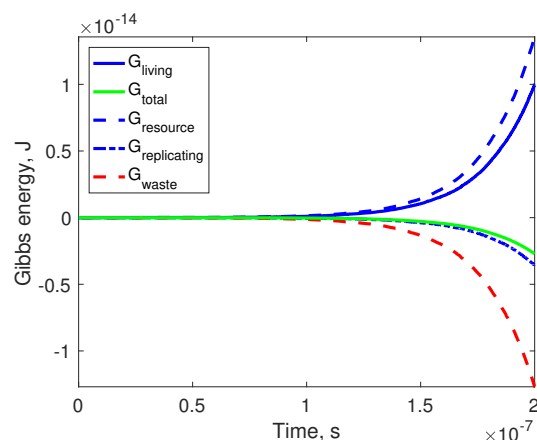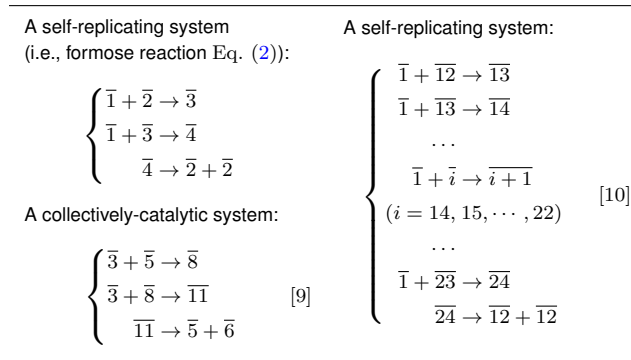
Fig. 3. Time series of Gibbs energy of the self-replicating system Eq. (8). We set $G_1^\circ = -800$, $G_2^\circ = -500$, $G_3^\circ = -400$, $G_4^\circ = -200$, $G_5^\circ = -950$, $G_6^\circ = -750$, $N_2(t) = Q$, $N_3(0) = 1$ and initially other molecules none.

simple towards complex. Is it possible to construct a system showing the similar process? Imagine there is a chemical reaction system composed of the following low-barrier reactions, given an infinite reservoir of only resource molecules $\overline{1}$. The

A self-replicating system
(i.e., formose reaction Eq. (2)):

$$\begin{cases} \overline{1} + \overline{2} \to \overline{3} \\ \overline{1} + \overline{3} \to \overline{4} \\ \qquad \overline{4} \to \overline{2} + \overline{2} \end{cases}$$

A collectively-catalytic system:

$$\begin{cases} \overline{3} + \overline{5} \to \overline{8} \\ \overline{3} + \overline{8} \to \overline{11} \\ \qquad \overline{11} \to \overline{5} + \overline{6} \end{cases} \quad [9]$$

A self-replicating system:

$$\begin{cases} \overline{1} + \overline{12} \to \overline{13} \\ \overline{1} + \overline{13} \to \overline{14} \\ \qquad \cdots \\ \overline{1} + \overline{i} \to \overline{i+1} \\ (i = 14, 15, \cdots, 22) \\ \qquad \cdots \\ \overline{1} + \overline{23} \to \overline{24} \\ \qquad \overline{24} \to \overline{12} + \overline{12} \end{cases} \quad [10]$$

first three reactions constitute the formose reaction Eq. (2), given the resource molecule $\overline{1}$. The three reactions in Eq. (9) constitute a collectively-catalytic system, given the resource $\overline{3}$. The thirteen reactions in Eq. (10) constitute a self-replicating system, given the resource $\overline{1}$. Is it possible that lots of complex molecules, such as $\overline{12}$, $\overline{13}$ and $\overline{14}$, are produced in the end?
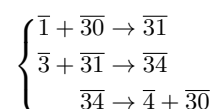
The answer in this case is yes. If the first $\overline{12}$ is produced, the self-replicating system Eq. (10) will be triggered, and consequently N_{12} will grow exponentially, as well as N_{13}, N_{14}, $\cdots$, N_{23}. But how is the first $\overline{12}$ produced? There are three stages: (1) Initially when there are only lots of $\overline{1}$ but nothing else, the system stays almost unchanged for a very long time since no low-barrier reaction could occur. Occasionally, by the high-barrier reaction $\overline{1} + \overline{1} \to \overline{2}$, one molecule $\overline{2}$ is produced. The self-replicating system Eq. (2) is triggered, and then N_2 and N_3 grow exponentially. Very quickly there are lots of $\overline{2}$ and $\overline{3}$. (2) After a relatively long "boring" period, the first $\overline{5}$ is produced by the high-barrier reaction $\overline{2} + \overline{3} \to \overline{5}$. Then the collectively-catalytic system Eq. (9) is triggered, and N_6 grows. Very soon there are lots of $\overline{6}$. (3) After that, occasionally one $\overline{12}$ is produced by the high-barrier reaction $\overline{6} + \overline{6} \to \overline{12}$.

One might naively believe that the first $\overline{12}$ can be produced by other reactions without the need for self-replicating and

collectively-catalytic systems. This is not the case. Despite an abundance of molecules $\overline{1}$ which could be used to "assemble" an initial $\overline{12}$, the production of the first $\overline{12}$ requires high-barrier reactions, e.g., $\overline{6} + \overline{6} \to \overline{12}$. It is only when both the other self-replicating and collective-catalytic systems are producing $\overline{6}$ in sufficient numbers that one such reaction is sufficiently likely to occur. At this point a $\overline{6} + \overline{6} \to \overline{12}$ can occur, despite it being a high-barrier reaction, because of the abundance of $\overline{6}$ molecules.

The previous stage-by-stage procedure could be a general model of how chemical reaction systems evolve towards complex: a relatively simple innovation triggers some self-replicating or collectively-catalytic systems and then a large number of new types of molecules are produced, paving the way for other innovations.

Meanwhile, the more types of molecules, the more probable to have reactions which are high-barrier before becoming practically low-barrier. For example, in the formose reaction Eq. (2), if the reaction $\overline{1} + \overline{3} \to \overline{4}$ is high-barrier, the system is not self-replicating. But imagine that the following three reactions are low-barrier,

$$\begin{cases} \overline{1} + \overline{30} \to \overline{31} \\ \overline{3} + \overline{31} \to \overline{34} \\ \qquad \overline{34} \to \overline{4} + \overline{30} \end{cases}$$

They constitute a collectively-catalytic system. Then the reaction $\overline{1} + \overline{3} \to \overline{4}$ which is the overall reaction of the three can still be considered as low-barrier, and the formose reaction system is still self-replicating. The only question is that we have to wait for the complex molecule $\overline{30}$ to appear.

Discussion

We have set up a general model for chemical reaction systems that properly accounts for energetics, kinetics and the conservation law. Although our model did not explicitly include catalysts, as other models did (20–22), catalysis and autocatalysis emerge in a number of systems.

We found three distinct types of self-driven system, i.e., systems which "feed" themselves. Both collectively-catalytic and self-replicating systems are vital in biology, while the third (non-sustaining) system appears less important. In terms of generating complexity, the self-replicating system plays a more important role, since it is able to replicate innovations. In the self-replicating formose reaction Eq. (2), after the first molecule $\overline{2}$ is produced by a high-barrier reaction, more $\overline{2}$s are easily replicated. In a biological setting, if this molecule spread to other places, it can trigger more self-replicating system. In contrast, in the collectively-catalytic citric acid cycle Eq. (1) for example, after the innovation (the first molecule $\overline{5}$), the second $\overline{5}$ will not appear until the responsible high-barrier reaction occurs once again.

In biology, most metabolic reactions are catalyzed by sophisticated and highly specific enzymes, and thus avoid disturbance from undesired side reactions. In the very early stage of life, however, probably no reaction required enzymes (10, 13). We found that whether side reactions prevent a system from self-replicating depends on how they are coupled with the system and the specific conditions. There appear to exist examples where self-replication can be sustained in the presence of side reactions. However, how sophisticated enzymes develop *de*

novo so that side reactions are avoided is another question, which deserves further investigations.

By arbitrarily constructing alternative chemical universes, we found that lots of them contain self-driven systems, and the lower bounds on the number of chemical universes containing collectively-catalytic or self-replicating systems increase with more types of molecules. This result suggests that in a random chemical universe, it would not be too surprising to observe the emergence of self-replication, one of the central properties of life (1). Although it is not the first theory to propose that self-replication is relatively easy to emerge, as the RAF theory did (23–25), it is the first one requiring no catalyst.

We provided a general model explicitly showing that high thermodynamic free energy molecules can be produced exponentially from low free energy molecules, while specific mechanisms in specific real-world scenarios have been investigated before (39, 40). The example system we showed is a metaphor of why high free energy ATP molecules are constantly produced in organisms (12). In addition, as our model takes energetics (corresponding to entropy) into account, it provides a more concrete way to discuss the issue—famously put forward by Schrödinger (38)—why life is able to spontaneously maintain a relatively low entropy level, although it cannot give the full answer: Answering the questions of how the molecules in living systems can be placed in an ordered structure (namely, a low entropy state) would require extending our model to include spatial effects.

Our model explicitly shows that complexity evolves from extreme simplicity stage by stage. It gives insights into three issues related to the origin of life. Firstly, the first RNA molecule is much more likely to be produced *de novo* by this stage-by-stage procedure, rather than a magic event (3). It provides theoretical support to "metabolism-first" theories (3), such as Wächtershäuser's iron-sulfur world hypothesis (5) and Szathmáry's theory (6–8). Secondly, before life, Earth should have gone through many stages in which different self-replicating systems existed and consequently Earth's compositions were different in each stage. The raw materials for life we should look for are those for the first self-replicating system, rather than those for the extant life (5). That is why, in the current theoretical framework, the raw materials for life (e.g., nucleotides) seem not to be available on the primordial Earth (3, 4). Thirdly, collectively-catalytic and self-replicating systems generate more types of new molecules, and in return, more types of molecules make more reactions feasible (in the form of catalysis and autocatalysis). This could explain why metabolic reactions in extant life always require sophisticated enzymes (13), while no reaction is expected to involve catalysts in the very early stage of life (18, 31, 32).

As we noted when introducing the model, we currently do not fully account for different isomers or for the impossibility of certain chemical transformations. We made our assumptions in order to ensure a form of "mass" conservation. This allows us to automatically exclude physically impossible systems such as system Eq. (3). The current model can be considered as a simpler version of a more general model, with a more intricate way of "ordering" the complexity of molecules, e.g., accounting for different isomers.

Our model provides a convenient platform to construct alternative chemical universes and investigate general properties of chemical reaction systems. It may provide a theoretical

guideline for systematically searching for other chemical paths towards life (or at least self-replicating entities), as pursued in astrobiology (44, 45) and xenobiology (46) for example.

1. Eörs Szathmáry. The origin of replicators and reproducers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1474):1761–1776, Oct 2006.
2. Walter Gilbert. Origin of life: The RNA world. *Nature*, 319(6055):618–618, Feb 1986.
3. M. P. Robertson and G. F. Joyce. The origins of the RNA world. *Cold Spring Harbor Perspectives in Biology*, 4(5):a003608–a003608, Apr 2012.
4. Hyo-Joong Kim, Alonso Ricardo, Heshan I. Illangkoon, Myong Jung Kim, Matthew A. Carrigan, Fabianne Frye, and Steven A. Benner. Synthesis of carbohydrates in mineral-guided prebiotic cycles. *Journal of the American Chemical Society*, 133(24):9457–9468, Jun 2011.
5. G Wächtershäuser. Before enzymes and templates: theory of surface metabolism. *Microbiological Reviews*, 52(4):452–484, Dec 1988.
6. Eörs Szathmáry. *The first replicators*, pages 31–52. Princeton University Press, New Jersey, USA, 1999.
7. Eörs Szathmáry. A classification of replicators and Lambda-Calculus models of biological organization. *Proceedings of the Royal Society of London B: Biological Sciences*, 260(1359):279–286, 1995.
8. John Maynard Smith and Eörs Szathmáry. *The Major Transitions in Evolution*. Oxford University Press, Oxford, UK, 1995.
9. Tibor Gánti. *The Principles of Life*. Oxford University Press, Oxford, UK, 2003.
10. H. J. Morowitz, J. D. Kostelnik, J. Yang, and G. D. Cody. The origin of intermediary metabolism. *Proceedings of the National Academy of Sciences*, 97(14):7704–7708, Jun 2000.
11. G Wächtershäuser. Evolution of the first metabolic cycles. *Proceedings of the National Academy of Sciences*, 87(1):200–204, Jan 1990.
12. Ádám Kun, Balázs Papp, and Eörs Szathmáry. Computational identification of obligatorily autocatalytic replicators embedded in metabolic networks. *Genome Biology*, 9(3):R51–R51, 2008.
13. Filipa L Sousa, Wim Hordijk, Mike Steel, and William F Martin. Autocatalytic sets in E. coli metabolism. *Journal of Systems Chemistry*, 6(4), Apr 2015.
14. Eörs Szathmáry, Mauro Santos, and Chrisantha Fernando. *Evolutionary potential and requirements for minimal protocells*, pages 167–211. Springer, Berlin, Heidelberg, 2005.
15. Nilesh Vaidya, Michael L. Manapat, Irene A. Chen, Ramon Xulvi-Brunet, Eric J. Hayden, and Niles Lehman. Spontaneous network formation among cooperative RNA replicators. *Nature*, 491(7422):72–77, Oct 2012.
16. Tracey A. Lincoln and Gerald F. Joyce. Self-sustained replication of an RNA enzyme. *Science*, 323(5918):1229–1232, 2009.
17. G. Ashkenasy, R. Jagasia, M. Yadav, and M. R. Ghadiri. Design of a directed molecular network. *Proceedings of the National Academy of Sciences*, 101(30):10872–10877, Jul 2004.
18. D. Sievers and G. von Kiedrowski. Self-replication of complementary nucleotide-based oligomers. *Nature*, 369(6477):221–224, May 1994.
19. Philippe Nghe, Wim Hordijk, Stuart A. Kauffman, Sara I. Walker, Francis J. Schmidt, Harry Kemble, Jessica A. M. Yeates, and Niles Lehman. Prebiotic network evolution: six key parameters. *Mol. BioSyst.*, 11(12):3206–3217, 2015.
20. Wim Hordijk and Mike Steel. Chasing the tail: The emergence of autocatalytic networks. *Biosystems*, 152:1 – 10, 2017.
21. Wim Hordijk and Mike Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461, Apr 2004.
22. M. Steel. The emergence of a self-catalysing structure in abstract origin-of-life models. *Applied Mathematics Letters*, 13(3):91 – 95, 2000.
23. Stuart A. Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119(1):1 – 24, 1986.
24. Wim Hordijk. Autocatalytic sets: from the origin of life to the economy. *BioScience*, 63(11):877–881, Nov 2013.
25. Wim Hordijk, Mike Steel, and Stuart Kauffman. The structure of autocatalytic sets: evolvability, enablement, and emergence. *Acta Biotheoretica*, 60(4):379–392, Sep 2012.
26. Wim Hordijk. Evolution of autocatalytic sets in computational models of chemical reaction networks. *Origins of Life and Evolution of Biospheres*, 46(2):233–245, 2016.
27. Vera Vasas, Chrisantha Fernando, Mauro Santos, Stuart Kauffman, and Eörs Szathmáry. Evolution before genes. *Biology Direct*, 7(1):1, 2012.
28. S. Jain and S. Krishna. A model for the emergence of cooperation, interdependence, and structure in evolving networks. *Proceedings of the National Academy of Sciences*, 98(2):543–547, Jan 2001.
29. Peter Wills and Leah Henderson. Self-organisation and information-carrying capacity of collectively autocatalytic sets of polymers: Ligation systems. In *Proceedings from the International Conference on Complex Systems on Unifying Themes in Complex Systems*, pages 613–623, Cambridge, MA, USA, 2000. Perseus Books.
30. Walter Fontana and Leo W. Buss. "the arrival of the fittest": Toward a theory of biological organization. *Bulletin of Mathematical Biology*, 56(1):1 – 64, 1994.
31. Steen Rasmussen, Adi Constantinescu, and Carsten Svaneborg. Generating minimal living systems from non-living materials and increasing their evolutionary abilities. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 371(1701), 2016.
32. Matthew Levy and Andrew D. Ellington. Exponential growth by cross-catalytic cleavage of deoxyribozymogens. *Proceedings of the National Academy of Sciences*, 100(11):6416–6421, May 2003.

33. Wim Hordijk and Mike Steel. Autocatalytic sets in polymer networks with variable catalysis distributions. *Journal of Mathematical Chemistry*, 54(10):1997–2021, Jul 2016.

34. J.Doyne Farmer, Stuart A Kauffman, and Norman H Packard. Autocatalytic replication of polymers. *Physica D: Nonlinear Phenomena*, 22(1):50 – 67, 1986.

35. Joshua I. Smith, Mike Steel, and Wim Hordijk. Autocatalytic sets in a partitioned biochemical network. *Journal of Systems Chemistry*, 5(1):2, 2014.

36. Peter Atkins and Julio de Paula. *Atkins' Physical Chemistry*. Oxford University Press, 10th edition, 2014.

37. William F. Martin, Filipa L. Sousa, and Nick Lane. Energy at life's origin. *Science*, 344(6188):1092–1093, 2014.

38. Erwin Schrödinger. *What Is Life? The Physical Aspect of The Living Cell.* Cambridge University Press, Cambridge, UK, 1944.

39. Jan P. Amend, Douglas E. LaRowe, Thomas M. McCollom, and Everett L. Shock. The energetics of organic synthesis inside and outside the cell. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 368(1622):20120255, 2013.

40. Elbert Branscomb and Michael J. Russell. Turnstiles and bifurcators: The disequilibrium converting engines that put metabolism on the road. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1827(2):62–78, 2013.

41. Jeremy L. England. Statistical physics of self-replication. *The Journal of Chemical Physics*, 139(12):121923, Sep 2013.

42. Michael J. Russell, Roy M. Daniel, and Allan J. Hall. On the emergence of life via catalytic iron-sulphide membranes. *Terra Nova*, 5(4):343–347, 1993.

43. A. Ricardo, M. A. Carrigan, A. N. Olcott, and S. A. Benner. Borate minerals stabilize ribose. *Science*, 303(5655):196–196, Jan 2004.

44. Ricardo Cavicchioli. Extremophiles and the search for extraterrestrial life. *Astrobiology*, 2(3):281–292, 2002.

45. Linda Billings, Vyllinniskii Cameron, Mark Claire, Greg J. Dick, Shawn D. Domagal-Goldman, Emmanuelle J. Javaux, Orion J. Johnson, Chris Laws, Margaret S. Race, Jon Rask, John D. Rummel, Rachel T. Schelble, and Steve Vance. The astrobiology primer: An outline of general knowledge—version 1, 2006. *Astrobiology*, 6(5):735–813, 2006.

46. Markus Schmidt. Xenobiology: A new form of life as the ultimate biosafety tool. *BioEssays*, 32(4):322–331, Mar 2010.