

Marginal stability in chemical systems and its relevance in the origin of life

Doriano Brogioli*

Dipartimento di Medicina Sperimentale, Università degli Studi di Milano - Bicocca Via Cadore 48, Monza (MI) I-20052, Italy

(Received 14 July 2011; published 29 September 2011)

Concentration fluctuations are always present in solutions; it has been noticed that, in chemical systems, they can lead to deviations from what is expected from mass-action equations. I recently described the class of the “marginally stable” chemical systems; namely, a system that have an infinity of stationary states forming a continuous curve, and I showed that they present such deviations, which appear as a drift along the stationary-state curve [*Phys. Rev. Lett.* **105**, 058102 (2010)]. Here I describe various marginally stable chemical reaction networks, including replicating molecules, and I present numerical calculations based on reaction-diffusion master equations, showing that the thermodynamic fluctuations induce a drift. This drift can be interpreted in terms of evolution toward a more efficiently replicating system and is analogous to a Darwinian evolution. The concentration fluctuations observed during the drift are scale invariant. Relevance of this phenomenon to the origin of life is discussed. I propose that marginal stability is the mathematical property defining chemical reaction networks potentially involved in the origin of life.

DOI: [10.1103/PhysRevE.84.031931](https://doi.org/10.1103/PhysRevE.84.031931)

PACS number(s): 87.23.Kg, 82.40.Qt, 82.39.Rt, 05.70.Ln

I. INTRODUCTION

One of the most fascinating problems of current science concerns the abiogenesis process (i.e., the origin of life from nonliving matter). This notwithstanding, abiogenesis is still deeply enigmatic. The definition of life is not easy, but generally the feature distinguishing entities involved in abiogenesis is their ability to replicate and undergo Darwinian evolution [1]. The presence of this feature ensures that the entity has overcome a given complexity threshold, after which it can give rise to even increasingly complex offsprings, until the complexity of the forms of life that we observe today is reached. In this context, a sharp distinction between living organisms, prebiotic entities, precursors of life, or even causes of abiogenesis is not possible, nor actually necessary.

Many different approaches have been tried in order to formulate a plausible model of abiogenesis, or at least in order to propose phenomena which could be involved in it. Part of the research has been devoted to the study of molecules which can be replicated, including proteins, DNA, RNA, or even organic polymers which are not currently present in biologic entities, like pRNA [2] or PNA [3]. Such molecules are called “informational polymers” (i.e., polymers that carry information coded into their sequence). Other research lines focus on the process of replication. In current living organisms, DNA is replicated by enzymes, the replicases, which are proteins, but the researchers proposed other possible replication mechanisms that could have worked in the early stages of life: self-replication of proteins, DNA, or RNA through spontaneous base pairing and binding [4–6], or through the action of intercalators [7].

In the RNA-world hypothesis [8], the informational polymer is RNA, and replication requires two identical RNA molecules: one acting as a template, and the other as a replicase. This idea comes from the observation that some RNA molecules, called ribozymes, have catalytic activity, including ligation. Directed (artificial) evolution [9] allowed

the synthesis of ribozymes that can self-replicate: in particular, ligases that catalyse their own synthesis from suitable substrates [10–12]. A partial success has been obtained in creating a ribozyme able to replicate arbitrary RNA templates [13]. The merit of this hypothesis is that only a single kind of polymer is needed, which acts both as an informational polymer and as a molecule with catalytic properties. A ribozyme ligase that acts on itself has been created and used in a completely artificial model [14], which mimics Darwinian evolution.

Another line of research is the study of the formation, in plausible conditions, of the monomers of informational polymers [15] (amino acids [16], adenine [17], pyrimidine nucleoside bases [18,19], ribose 2,4-biphosphate [20]), and their polymerization [21].

It has been noticed that life requires not only replication but also metabolism, and thus a classification into “replication-first” and “metabolism-first” models can be done [22–24]. Models focused on autocatalytic cycles [25] fall into the second class. Examples of metabolic networks which can be involved in abiogenesis are the reverse citric acid cycle [26–28] and the formose cycle [29,30]; a discussion can be found in Ref. [31]. It has been proposed that the concentration of different chemical species can behave as a “compositional genome,” replicated by proto-metabolic networks [32]; this model has been recently discussed [33]. Another research line in this field concerns the theoretical study of interesting families of chemical networks, like hypercycles [34], and the evaluation of the likelihood of the spontaneous presence of autocatalytic sets in random chemical networks [35–37].

Most of the above-described models assume the presence of a membrane, delimiting a small volume (the prebiotic entity) from the environment. The compositional genome cannot even be defined in absence of the membrane. In the case of RNA world, the need for membranes has a more subtle reason [38]. An RNA template is replicated by another RNA molecule, acting as a replicase ribozyme; if the template itself is a replicase of the same species, self-replication takes place. If a superior mutant of the replicase is present, it replicates the other molecules at a higher rate, but it has

*dbrogioli@gmail.com

no advantage in being replicated, and thus its concentration does not increase; in such a system, replication is present, but Darwinian evolution does not take place. In general, some physical effect is needed, besides chemistry, in order to obtain evolution. In Ref. [38], it is proposed that the physical effect can be the presence of volumes delimited by membranes, defining entities in competition with each other [39–43]. In other words, a protocell including a more active mutant replicase will replicate faster, and thus the total number of the mutant replicases will increase. The difficulty of this model is that the self-replicating polymers must emerge simultaneously with the membranes that contain them.

Recently I showed that a simpler physical effect can lead to evolution in a chemical system; namely, the thermodynamic fluctuations of concentration of the polymers [44]. Fluctuations are always present in every solution: in the simplest cases, they arise from the Poissonian distribution of the number of molecules in a given small volume. Local differences in molecule concentration play the role of the volumes delimited by membranes, leading to the increase of the concentration of the superior mutants. Anyhow, no individual living entities are present—evolution involves the whole solution. In this sense, I propose that the first stages of abiogenesis took place extremely slowly, on macroscopic length scales, and later evolution led to the development of membranes, compartmentalization and, finally, individual microscopic living entities, which were selected due to their ability to evolve faster. In other, more dramatic words, there has been no “first living cell,” but a “living ocean,” which later split into protocells.

It has already been noticed that fluctuations can induce “deviant nonclassical effects” on particular biochemical systems [45,46]; namely, deviations from mass-action equations. The above-described evolution in replicating systems is an example of such a deviant nonclassical effect originating from fluctuations. The replicating systems I will describe in Sec. III share the key feature of “marginal stability” (i.e., the presence of a manifold formed by marginally stable states). In particular, they present a stationary-state curve (i.e., a one-dimensional manifold of marginally stable stationary states). In Sec. II, I will describe the numerical method for the calculation of the effect of fluctuations. The results show that the chemical system first follows the mass-action equations and approaches a state that is stationary with respect to the mass-action equations. Then, under the effect of fluctuations, the system moves between the adjacent marginally stable states, violating the mass-action equations. From the numerical calculations, it can be noticed that the effect of fluctuations is not a random walk, but a drift along the stationary-state curve, pointing toward the increase of the most active replicase: that drift represents the Darwinian evolution. In Sec. IV I will discuss quantitatively the validity of the numerical method, and I will show that the fluctuations are scale invariant.

Summarizing the above concepts, thermodynamic fluctuations can provide a mechanism for evolution, alternative to the presence of membranes, in marginally stable chemical systems. I propose to consider marginal stability as the feature defining chemical reaction networks potentially involved in abiogenesis. This approach has the advantage that it can

be applied automatically to large sets of reaction networks and has the potential to discover completely new chemical models. This allows us to find candidates of chemical networks connected with abiogenesis, also if they apparently do not involve replication, as in the case of metabolism-first models. A similar screening has been applied in the case of recognition of self-catalytic cycles [35–37]. Another advantage of this approach is that marginal stability can be assessed without the need to specify the physical structure associated with the chemical system.

II. NUMERICAL METHOD

Chemical kinetics is usually studied by means of mass-action equations. The state is represented by a vector (c_1, c_2, c_3, \dots) , where c_S is the concentration of species S . The kinetics is defined by means of a first-order ordinary differential equation of the form

$$\begin{aligned} \dot{c}_1(t) &= F_1[c_1(t), c_2(t), c_3(t), \dots], \\ \dot{c}_2(t) &= F_2[c_1(t), c_2(t), c_3(t), \dots], \\ &\dots, \end{aligned} \quad (1)$$

where the functions F_S describe the mass-action law and have the form

$$F_1[c_1, c_2, c_3, \dots] = k_1 c_1^\alpha c_2^\beta c_3^\gamma + \dots \quad (2)$$

In order to describe nonstirred extended systems, diffusion must be taken into account. In principle, this could be done by using reaction-diffusion equations of the form

$$\begin{aligned} \dot{c}_1(\vec{x}, t) &= F_1[c_1(\vec{x}, t), c_2(\vec{x}, t), \dots] + D_1 \nabla^2 c_1(\vec{x}, t), \\ \dot{c}_2(\vec{x}, t) &= F_2[c_1(\vec{x}, t), c_2(\vec{x}, t), \dots] + D_2 \nabla^2 c_2(\vec{x}, t), \\ &\dots, \end{aligned} \quad (3)$$

where the concentrations $c_S(\vec{x}, t)$ are functions of both point \vec{x} and time t , and D_S is the diffusion coefficient of species S .

This approach neglects concentration fluctuations, which are always present in any solution. In the simplest case, they coincide with the Poisson fluctuations of the number of solved molecules in a given volume of solvent, which changes over time due to Brownian motions of molecules entering and exiting the volume. In Landau’s fluctuating hydrodynamics [47], fluctuations are taken into account by adding a white noise term to the partial differential equations describing diffusion and advection. Unfortunately, the same procedure is much more difficult when chemical reactions are taken into account and does not work at all in some cases [48]; the present case falls among them.

I will follow a different approach, based on the reaction-diffusion master equations (RDME) [49–51], whose solutions are calculated through the Gillespie algorithm [52]. The reaction volume is divided into $\mathcal{N} \times \mathcal{N} \times \mathcal{N}$ cubic cells, with side l . The dynamic variables $n_S(i, j, k)$ represent the number of molecules of the species S in the cell with integer indices i, j, k . I define a constant \mathcal{C} so that $c_S(i, j, k) = (\mathcal{C}/l^3)n_S(i, j, k)$. The random variables $n_S(i, j, k)$ change as the result of a continuous-time Markov process, defined by the probabilities of events, which include chemical reactions, modeled as “birth and death” processes, and diffusion, modeled as a random

walk. The probability of the occurrence of a given event e in a time interval dt is defined as $\mathcal{P}_e dt$; the values \mathcal{P}_e are called “transition rates” or “event rates.”


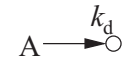
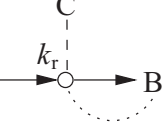
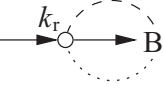
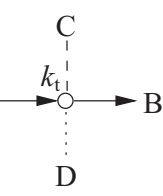
The rates \mathcal{P}_e of the chemical reaction events are calculated on the basis of the reaction constants of the mass-action equations and a combinatorial factor [52]. For example, the generation of molecule A in a cell with indices (i, j, k) corresponds to a transition $n_A(i, j, k) \rightarrow n_A(i, j, k) + 1$, with a transition rate $\mathcal{P}_e = k_g l^3 / C$, where k_g is the mass-action reaction coefficient of the zeroth-order generation reaction.

The random walk between cells is included in the model in order to account for the effect of diffusion. The probability that a molecule moves from one cell to a neighboring cell, during a time interval dt , is dtD/l^2 . The rate of such an event is thus $\mathcal{P}_e = D/l^2$. Periodic conditions are imposed, so that $n_S(i, j, k) = n_S(i + \mathcal{N}, j, k) = n_S(i, j + \mathcal{N}, k) = n_S(i, j, k + \mathcal{N})$.

Table I shows the transition rates \mathcal{P}_e for the reactions of interest. The reactions are represented as directed graphs, as is usual in the literature about catalytic chemical networks [53, 54]. A circle represents a reaction, incoming arrows represent reactants, outgoing arrows represent reaction products. Dashed lines represent catalytic activity (i.e., they are a shortcut for representing a couple of incoming and outgoing lines). I use dotted lines to represent a particular class of catalytic activity; namely, templating activity, which is the effect of a molecule acting as a source of information for creating an identical copy (replication case) or a different molecule which carries the same information (translation). The reaction constant labels k_c are placed near the circle representing the corresponding reaction.

In the present work, as often in the literature about RDME, the name “master equations” is used also if the master equations are not actually solved by analytic methods, nor explicitly formulated at all. Instead, the solution of the RDME is obtained numerically by means of the Gillespie algorithm

TABLE I. Transition rates \mathcal{P}_e of chemical reaction events and diffusion events. In the first column, reactions are represented by a circle. An outgoing arrow points to the product and an ingoing arrow comes from the reactant. Dashed lines connect the circle to the catalyzer molecules and dotted lines represent the particular catalytic effect named “templating action.” The labels k_c near the circle representing the reaction show the reaction constant. The chemical reactions include generation and degradation reactions: replication of a polymer B performed by a replicase C , by the polymerization of a monomer set A , self-replication of a ribozyme B , and translation of a gene D into a protein B by means of the translation molecular machinery C . Diffusion events are also considered.

Event e	Description	Mass-Action Term	Transition	\mathcal{P}_e
	Generation	k_g	$n_A(i, j, k) \rightarrow n_A(i, j, k) + 1$	$k_g l^3 / C$
	Degradation	$k_d A$	$n_A(i, j, k) \rightarrow n_A(i, j, k) - 1$	$k_d n_A(i, j, k)$
	Replication	$k_r ABC$	$n_A(i, j, k) \rightarrow n_A(i, j, k) - 1$ $n_B(i, j, k) \rightarrow n_B(i, j, k) + 1$	$\frac{k_r C^2}{j^6} n_A(i, j, k)$ $n_B(i, j, k) n_C(i, j, k)$
	Self-Replication	$k_r AB^2$	$n_A(i, j, k) \rightarrow n_A(i, j, k) - 1$ $n_B(i, j, k) \rightarrow n_B(i, j, k) + 1$	$\frac{k_r C^2}{j^6} n_A(i, j, k)$ $n_B(i, j, k)$ $[n_B(i, j, k) - 1]$
	Translation	$k_t ACD$	$n_A(i, j, k) \rightarrow n_A(i, j, k) - 1$ $n_B(i, j, k) \rightarrow n_B(i, j, k) + 1$	$\frac{k_t C^2}{j^6} n_A(i, j, k)$ $n_C(i, j, k) n_D(i, j, k)$
	Diffusion Along x Axis	$D_A \frac{\partial^2}{\partial x^2} A$	$n_A(i \pm 1, j, k) \rightarrow n_A(i, j, k) + 1$ $n_A(i, j, k) \rightarrow n_A(i, j, k) - 1$	$\frac{D}{l^2} n_A(i, j, k)$
	Diffusion Along y axis	$D_A \frac{\partial^2}{\partial y^2} A$	$n_A(i, j \pm 1, k) \rightarrow n_A(i, j, k) + 1$ $n_A(i, j, k) \rightarrow n_A(i, j, k) - 1$	$\frac{D}{l^2} n_A(i, j, k)$
	Diffusion Along z Axis	$D_A \frac{\partial^2}{\partial z^2} A$	$n_A(i, j, k \pm 1) \rightarrow n_A(i, j, k) + 1$ $n_A(i, j, k) \rightarrow n_A(i, j, k) - 1$	$\frac{D}{l^2} n_A(i, j, k)$

[52], which is based on a Monte Carlo approach. Actually, the method can be seen as a simulation of the chemical system, which is able to give also the kinetics.

The cell size l is not a physical parameter of the system, but only a parameter of the mathematical model (the continuous-time Markov process) and must be chosen so that the model closely mirrors the physical system. For $l \rightarrow +\infty$, the solution of the RDME approaches the solution of the mass-action equations. For finite, decreasing values of l , the behavior changes, due to the development of concentration fluctuations. Generally there is a value l_{\max} , such that, for $l \ll l_{\max}$, the results are independent of l itself. It is well known that, unfortunately, the results do not converge in the limit of small cell size $l \rightarrow 0$: in this sense, the RDME are not the discretization of a continuous model. In the blessed case, there is a range of cell size $l_{\min} \ll l \ll l_{\max}$ for which the results are independent of l itself; in that range, the model can be considered representative of the so-called mesoscopic behaviour of the system. In this context, “mesoscopic” refers to the fact that the model gives results concerning small length scales, but it relies only on the “macroscopic” parameters such as the reaction constants k_j and the diffusion constants D_S ; the only needed knowledge of the “microscopic” behavior is the number of molecules per unit volume per unit concentration, explicitly accounted for by the constant C ; no hypothesis on the microscopic interactions between the molecules is needed.

Up to now, general conditions for the existence of the range $l_{\min} \ll l \ll l_{\max}$ have not been found. For a given reaction network, with some choice of values of reaction and diffusion constants, it is possible that the results are dependent of l in every range. I argue that, in this case, the fluctuations still have an effect, and the results are qualitatively significant, but the system cannot be fully described in terms of the macroscopic parameters; that is, the knowledge of the microscopic behavior of the reacting molecules is necessary in order to accurately simulate the system.

The choice of the cell size l must be discussed separately for every system. Now I will give an account of the discussion about the choice of l done in the literature.

Bernstein [51] cites just one condition under which the RDME results are representative of the physical system: “the diffusion time across each element for every species should be much less than the typical reaction time.” This condition can be expressed as

$$\mathcal{R}_{c,S} \ll \mathcal{R}_{d,S}, \quad (4)$$

where $\mathcal{R}_{c,S}$ and $\mathcal{R}_{d,S}$ are the molecular rates of the chemical reactions c and of diffusion events of the molecules of species S . The molecular rate $\mathcal{R}_{e,S}$ of an event e for a species S is defined as

$$\mathcal{R}_{e,S} = \frac{\langle \mathcal{P}_e(i,j,k) \rangle}{\langle n_S(i,j,k) \rangle}, \quad (5)$$

so that $\mathcal{R}_{e,S} dt$ is the probability that a molecule of the species S undergoes an event e in a time interval dt .

For example, considering a trimolecular reaction $E + F + G \rightarrow H$, if the molecules are randomly distributed in the cells, the molecular reaction rate of the species E is

$$\mathcal{R}_{c,E} \approx k_c c_F c_G, \quad (6)$$

where c_F and c_G are the concentrations of the two molecular species, and k_c is the reaction constant. The molecular rate of diffusion along all directions is

$$\mathcal{R}_{d,S} = 6D/l^2. \quad (7)$$

In this example, Eq. (4) imposes an upper bound on the value of l :

$$l \ll \sqrt{\frac{6D}{k_c c_F c_G}}. \quad (8)$$

A similar calculation will be performed in Sec. IV for the trimolecular reactions of replication and translation.

Accurate comparisons between microscopic models and RDME have been performed [49,55] for the case of annihilation reactions. For the validity of RDME, Baras and Mansour [49] list a set of three conditions:

- (i) The system is in local thermal equilibrium.
- (ii) The cell size is longer than the mean-free path.
- (iii) The cell size is smaller than the correlation length.

Condition (iii) means that the space must be discretized in steps fine enough to well represent the correlation function. In turn, the correlation length is related to the reactive mean-free path, defined as the average distance traveled by a diffusing molecule before it undergoes a reactive event. Thus, condition (iii) can be expressed as

$$l \ll \sqrt{\frac{6D}{\mathcal{R}_{c,S}}}, \quad (9)$$

which exactly corresponds to the condition expressed by Eq. (4).

The condition (i) is guaranteed if each cell contains a large number of molecules, of the order of a few hundred [49]. This can be expressed as $l \gg \lambda$, where the length λ is defined as the side of a cubic volume that contains one molecule at each time, on average. If values $l < \lambda$ are used in the calculations, the results concerning length scales shorter than λ must be considered not reliable. Average concentrations and information on length scales longer than λ can still be valid; in particular, if the binning of adjacent cells gives results similar to the ones obtained with a longer l . The value l_{\min} is defined so that discrepancies occur for $l < l_{\min}$.

In Sec. IV I will provide a discussion of the range of validity of RDME in the studied cases. I will show that, in the case discussed in this work, there is a range of cell size l for which the cross spectra of the concentrations do not depend on l . Moreover, I will show that the fluctuations are scale invariant, the correlation function is a power law, and thus a correlation length cannot be defined. This situation is typical of the so-called self-organized critical systems [56]: the mesoscopic behavior of the system can be described as scale invariant.

III. EXAMPLES OF marginally STABLE CHEMICAL SYSTEMS

A. Ribozyme self-replication

Figure 1 shows a toy model where a hypothetical replicase R copies a template T by polymerizing the monomers B .

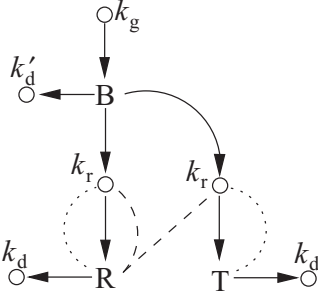


FIG. 1. Chemical reaction network representing a model of a replication reaction. A ribozyme R (a replicase) catalyzes the formation of a copy of a template molecule T , by catalyzing the polymerization of the substrates B . The dashed lines represent the catalysis and the dotted lines represent the particular kind of catalytic action in which the molecule acts as a template. The hypothetical ribozyme R copies any generic RNA template T , including other molecules R .

Moreover, the replicase R also copies other replicase molecules R by polymerizing the monomers B . More precisely, the ribozyme R can act both as a replicase and as a template, but the same molecule cannot play both the roles simultaneously. The production of a new molecule R thus requires two molecules R : one as a replicase, and one as a template. This is the case of a ribozyme that can copy any RNA template. The replication reaction is catalyzed by the replicase R , but the template T itself acts as a catalyst by performing a “templating” action (i.e., by templating its own polymerization). The ribozyme replicase is the basic element of the RNA-world hypothesis; such a ribozyme is not present in current living organisms, but a partial success has been obtained in artificially creating a ribozyme able to replicate arbitrary RNA templates [13].

A source of monomers B is included in the model along with a catabolic reaction which removes both the monomers B and the polymers R and T . Due to the catabolic term, the chemical reaction network is manifestly open, since it exchanges molecules with the environment. In general, the nonequilibrium can be represented by an energy flow; for example, in the form of nucleoside triphosphates [10–13]. The catabolic term is needed to avoid the “death” of the reaction network due to the accumulation of polymers.

The reaction constants of the two replication reactions performed by R are assumed to be equal to k_r . This is the case if the two polymers have the same length, since the polymerization takes place at a constant speed in terms of the number of bases added to the polymer per unit time. Since the polymers R and T have the same length, I also assume that they undergo the catabolic reaction with the same reaction constant k_d .

The mass-action equations corresponding to the reaction network in Fig. 1 are

$$\begin{aligned}\dot{T} &= k_r T R B - k_d T, \\ \dot{R} &= k_r R^2 B - k_d R, \\ \dot{B} &= -k_r (T + R) R B - k_d B + k_g,\end{aligned}\quad (10)$$

where \dot{X} represents the temporal derivative of the variable X and k_j are the reaction constants. Moreover, I assume

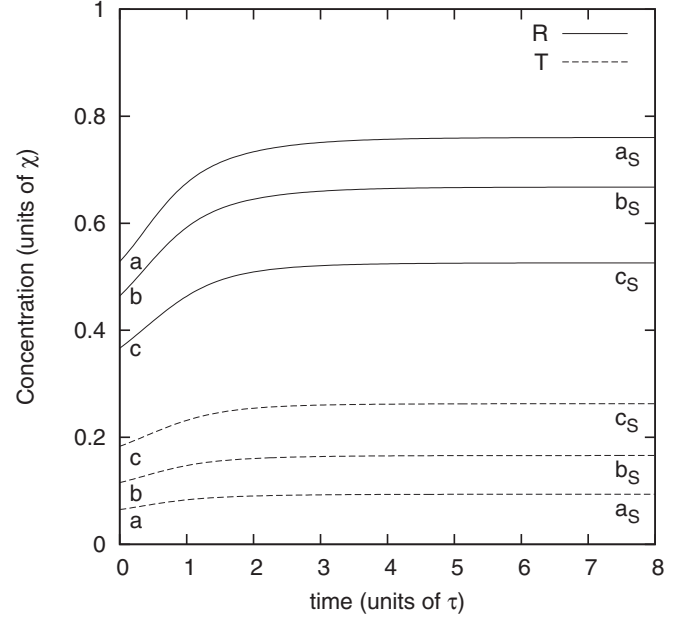


FIG. 2. Kinetics of the chemical reaction network sketched in Fig. 1. The concentrations of the replicase R and of the template T are shown. The results are calculated through the mass-action equations (11). The reaction constant is $k_r = 9/(\tau \chi^2)$; the values of k_d and k_g are irrelevant under the used adimensionalization. Three sets of initial values are shown; they differ for the value of the concentrations R and T (points a , b , and c). The concentrations reach a different stationary value, depending on the initial concentrations a_s , b_s , and c_s , respectively.

$k'_d = k_d$. A unit of B does not represent a single molecule of the monomer, but the whole set of monomers needed to produce a polymer T or R .

A suitable adimensionalization is performed by using $\tau = 1/k_d$ as the unit of time, $\chi = k_g/k_d$ as the unit of concentration, and $\lambda = \sqrt[3]{C/\chi}$ as the unit of length. In these units, the only relevant parameters of the system are k_r , the values of the diffusion coefficients D_R , D_T , D_B , and l . With this choice, the average number of molecules contained in a cell with size λ is 1.

Figure 2 shows the kinetics calculated through the mass-action equations (11). Three sets of initial values are shown; they differ for the values of the concentrations T and R (points a , b , and c). The concentrations reach a different stationary value, depending on the initial concentrations a_s , b_s , and c_s , respectively. This shows that there are many different stationary states.

The stationary states can be found by looking for the concentrations \tilde{T} , \tilde{R} , \tilde{B} for which $\dot{T} = \dot{R} = \dot{B} = 0$ in Eq. (11). The condition for the existence of different stationary states is $k_r > 4/(\tau \chi^2)$. Under this condition, the stationary states form a curve, which can be expressed as

$$\begin{aligned}\tilde{T}(s) &= \chi \left(1 - \frac{1}{\tau \chi^2 k_r s} - s \right), \\ \tilde{R}(s) &= \chi \frac{1}{\tau \chi^2 k_r s}, \\ \tilde{B}(s) &= \chi s,\end{aligned}\quad (11)$$

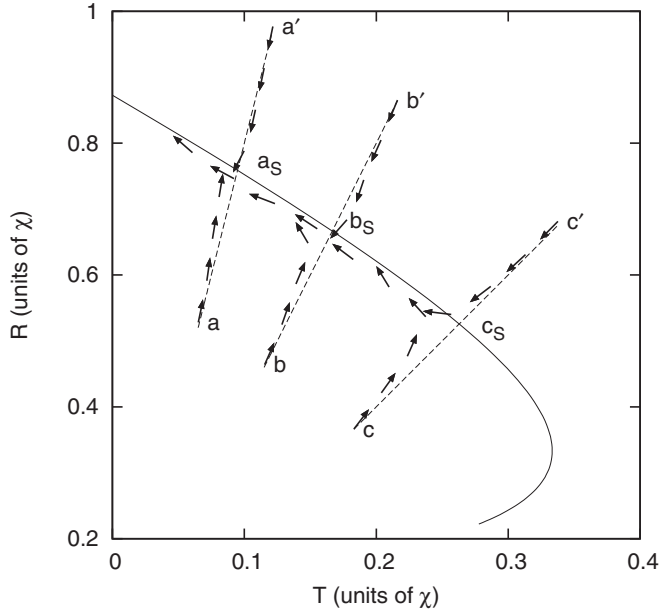


FIG. 3. Kinetics of the chemical reaction network sketched in Fig. 1. The graph shows the trajectories projected in the T - R plane. The dashed lines represent the trajectories calculated by means of the mass-action equations (11), starting from different initial states. Each trajectory converges to different stationary states. The solid line represents the curve of the stationary states of the mass-action equations expressed by Eq. (11). The trajectories represented by the arrows are calculated by means of the reaction-diffusion master equations. The parameters are $k_r = 9/(\tau\chi^2)$, $D_T = D_R = 4\lambda^2/\tau$, $D_B = 20\lambda^2/\tau$, $l = 2\lambda$, and $\mathcal{N} = 20$. The values of k_d , k_g , and C are irrelevant under the adimensionalization used.

where s is the curvilinear abscissa. By linearizing the equations around the stationary states, it can be shown that the stationary states are marginally stable for the values of the parameter s :

$$s < \frac{1}{2}. \tag{12}$$

This part of the curve represents the marginally stable states.

Figure 3 shows a projection on the R - T plane of the stationary-state curve. The kinetics obtained by integrating the mass-action equations is also shown as dashed lines. Starting from initial states $a, b, c, a', b',$ and c' , different stationary states $a_s, b_s,$ and c_s are reached. The presence of many different stationary states is the kinetic counterpart of “inheritance:” the newly produced molecules R and T are of the same species as their ancestors, and the two populations can coexist in different proportions. This conservation is mathematically expressed by the fact that the ratio $\varphi = R/T$ is a constant with respect to the kinetics of the mass-action equations. The conservation of φ is a consequence of the linear dependence of the first two equations of Eqs. which, in turn, is the expression of the replication law.

In mechanics, an example of marginally stable system is a marble on a horizontal track. The marble can remain at any point of the track, since any point is a marginally stable equilibrium state. On the other hand, a disturbance will result in a displacement of the marble along the track. The marble will not return to the original position, but it will reach a different stationary point since there is no restoring force.

By analogy, the presence of the stationary-state curve of Fig. 3 allows the chemical system to move smoothly under the effect of spontaneous concentration fluctuations, passing from a stationary state to one of the surrounding, without being called back. This is not possible in a system that presents a single stable stationary state, because the system eventually comes back to the stationary attractive state, after having been displaced.

The trajectories obtained with RDME are reported in Fig. 3 as arrows. At first the system follows the trajectory of the mass-action equations (11) (dashed lines) and approaches the stationary-state curve, but then motion along the stationary-state curve is observed. Quite surprisingly, this motion is not a random walk, but a drift pointing toward the increase of the replicate R .

This drift has a straightforward explanation: in volumes with higher concentrations of R , more polymers will be replicated but, in turn, a bigger fraction of them will be R . The analogy of this effect with Darwinian evolution will be discussed in more detail in the following subsection. It is worth noting that the directionality of the drift implies that the stationary states cannot be thermodynamic equilibrium states.

B. Ribozyme self-replication: case of two replicases

Selection, in the context of Darwinian evolution, generally refers to the presence of an evolutive advantage for one of the two mutants of the same entity, which show different efficiencies. In the model presented in the previous section, the two entities are RNA molecules, but one is a ribozyme replicase, while the other is completely inactive. In this subsection, I will consider two replicases R_1 and R_2 with different activities. The chemical reaction network is shown in Fig. 4. Both the replicases R_1 and R_2 are able to copy other molecules by polymerizing the monomers B . I assume that the replicase R_2 is a superior mutant; it can be generated by a mutation during the replication.

As it can be noticed in Fig. 4, the various replication reactions are four. In each reaction, one molecule acts as a replicase, and another molecule acts as a template. Both the replicase and the template can be R_1 or R_2 , leading to the four different reactions. The reaction constant depends on the efficiency of the replicase, but not on the features of the template, which is a consequence of the fact that the replication

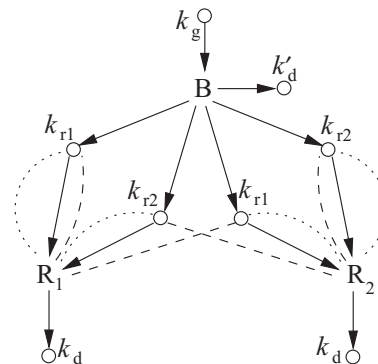


FIG. 4. Chemical reaction network including two ribozymes R_1 and R_2 , acting as replicases, able to copy each other and themselves.

takes place progressively by addition of single monomers. The reaction constants of the two replications in which R_1 acts as a replicase and R_1 or R_2 as a template are equal (k_{r1}), and the reaction constants of the two replications in which R_2 acts as a replicase, and R_1 or R_2 as a template, are equal (k_{r2}).

The mass-action equations are

$$\begin{aligned} \dot{R}_1 &= k_{r1}R_1^2B + k_{r2}R_1R_2B - k_dR_1, \\ \dot{R}_2 &= k_{r1}R_2R_1B + k_{r2}R_2^2B - k_dR_2, \\ \dot{B} &= -(k_{r1}R_1 + k_{r2}R_2)(R_1 + R_2)B - k_dB + k_g. \end{aligned} \tag{13}$$

The replicase R_2 is more active than R_1 thus, $k_{r2} > k_{r1}$. Moreover, I assume $k'_d = k_d$.

A suitable adimensionalization is performed by using $\tau = 1/k_d$ as the unit of time, $\chi = k_g/k_d$ as the unit of concentration, and $\lambda = \sqrt[3]{C/\chi}$ as the unit of length. In these units, the only relevant parameters of the system are k_{r1} , k_{r2} , and the values of the diffusion coefficients D_{R1} , D_{R2} , D_B , and l .

A straightforward calculation shows that there is a stationary-state curve that can be expressed as

$$\begin{aligned} \tilde{R}_1(s) &= \chi \frac{1/s + \tau\chi^2k_{r2}(s-1)}{\tau\chi^2(k_{r1} - k_{r2})}, \\ \tilde{R}_2(s) &= \chi \frac{1/s + \tau\chi^2k_{r1}(s-1)}{\tau\chi^2(k_{r2} - k_{r1})}, \\ \tilde{B}(s) &= \chi s, \end{aligned} \tag{14}$$

where s is the curvilinear abscissa. The stationary-state curve is present if $k_{r2} > 4/(\tau\chi^2)$. Also in this case, the stationary states are marginally stable for

$$s < \frac{1}{2}. \tag{15}$$

Figure 5 shows a projection on the plane R_1 - R_2 of the stationary-state curve (solid line) of the trajectories calculated by means of the mass-action equations (dashed lines) and by means of the RDME (arrows). The trajectories start from the states a, b, c, a', b', c' . Following the mass-action equations, the trajectories (dashed lines) approach the marginally stable states a_s, b_s, c_s , respectively. The RDME trajectories (arrows) at first follow the mass-action equations but, when the stationary state is approached, a drift along the stationary-state curve can be observed, directed toward the increase of the most active replicase R_2 .

The observed drift represents a phase of evolution: the concentration of the more efficient mutant replicase R_2 increases over R_1 and becomes dominant. After this, a third, more active replicase R_3 can arise due to random mutations, and its concentration will increase, and so on, in analogy with Darwinian evolution.

C. Model of replication with DNA and enzymes

Currently, no ribozyme able to replicate RNA templates is found in living organisms. Replicases actually exist, but they are peptides, which replicate DNA molecules (not other enzymes). The DNA replication mechanism of current living organisms is sketched very roughly in Fig. 6. The species R and T represent DNA molecules, genes, or sets of genes. The genes are replicated by an enzyme R' , which polymerizes the

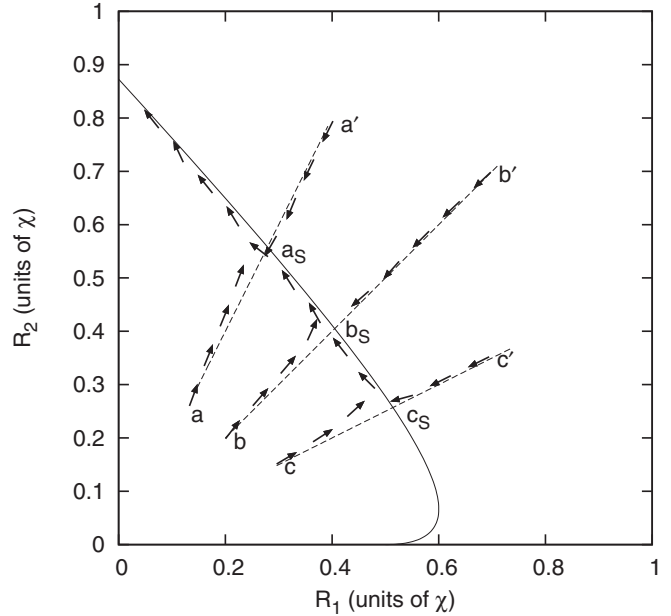


FIG. 5. Kinetics of the chemical reaction network described in Fig. 4 with two replicases: R_1 and the more active R_2 . The solid line represents the stationary-state curve of the mass-action equations, expressed by Eqs. (15). The dashed lines represent the trajectories calculated by means of the mass-action equations (14). The trajectories shown by the arrows are calculated by means of the reaction-diffusion master equations. The parameters are $k_{r1} = 4/(\tau\chi^2)$, $k_{r2} = 9/(\tau\chi^2)$, $D_{R1} = D_{R2} = 4\lambda^2/\tau$, $D_B = 20\lambda^2/\tau$, $l = 2\lambda$, and $\mathcal{N} = 20$. The values of k_d , k_g , and C are irrelevant under the adimensionalization used.

nucleotides B . The information for building the enzyme R' is contained in the gene R , which is said to “code” for it. The operation of building an enzyme (e.g., R') by polymerizing the amino acid A , following the instructions contained in a gene (e.g., R), is called “translation.” This operation is performed by an extremely complex molecular machinery T' , which involves not only enzymes, but also RNA, through the intermediate step called “transcription.” The translation is performed by the ribosome, which is an extremely complex multimolecular cellular structure. Modelling this machinery is out of the scope of the present work. Although it is extremely

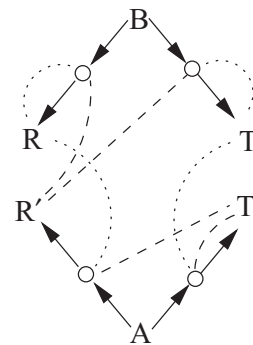


FIG. 6. Sketch of the replication mechanism of current biological entities, including DNA and enzymes. The gene sets R and T are replicated by the enzyme R' and are translated into the active species R' and T' by the translation machinery T' .

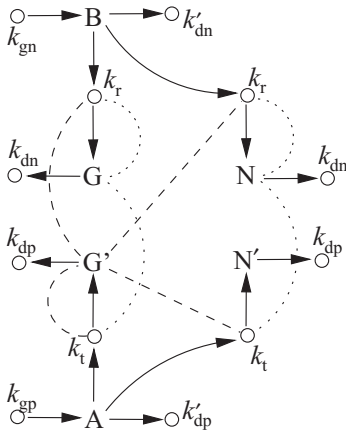


FIG. 7. Reaction network representing the competition between two sets of genes G and N , coding respectively for the active species G' and inactive species N' . Each unit of B and A represent the set of monomers needed to build the corresponding polymer.

complex, the recipe for building it is contained in a set of DNA molecules: the set of genes T .

The genes R and T only have the role of carrying information while the active species are R' and T' . In this extremely simplified view, only T' generates entities identical to itself, but it performs this task by reading the recipe from a different species (i.e., T).

In this section I will consider a chemical reaction network inspired by the above-described mechanism. I will consider the species G , representing the genes R and T as a whole, and G' , representing the active species R' and T' , both of which are able to replicate G and generate other G' by translating G . Moreover, I will consider the competition between the set of genes G , which codes for the efficient active species G' , and another set of genes N , which codes for an inactive species N' . The reaction network is shown in Fig. 7. The polymers G , N , G' , and N' are generated by polymerization of the monomer sets B and A (a naïve representation of nucleotides or amino acids). Each unit of B and A contains the set of monomers needed to create the corresponding polymer. As a further simplification, the model does not consider the energy supply that, in living organisms, is provided by ATP.

The mass-action equations are

$$\begin{aligned}
 \dot{G} &= k_r G' G B - k_{dn} G, \\
 \dot{N} &= k_r G' N B - k_{dn} N, \\
 \dot{B} &= -k_r G' B (G + N) - k_{dn} B + k_{gn}, \\
 \dot{G}' &= k_t G' G A - k_{dp} G', \\
 \dot{N}' &= k_t G' N A - k_{dp} N', \\
 \dot{A} &= -k_t G' A (G + N) - k_{dp} A + k_{gp}.
 \end{aligned} \tag{16}$$

As in the previous sections, I assume that the replication speed is the same for the two sets of genes G and N . The same applies to the translation; namely, the speed of the translation of G into G' is equal to the speed of the translation of N into N' . Moreover, I assume $k'_{dn} = k_{dn}$ and $k'_{dp} = k_{dp}$.

The system is adimensionalized by taking $\tau = 1/k_{dn}$ as the unit of time, $\chi = k_{gn}/k_{dn}$ as the unit of concentration

for G , N , and B , $\chi' = k_{gp}/k_{dn}$ as the unit of concentration for G' , N' , and A , and $\lambda = \sqrt[3]{C/\chi}$ as the unit of length. In these units, the only relevant parameters of the system are k_r , k_t , k_{dp} , and the values of the diffusion coefficients D_G , D_N , $D_{G'}$, $D_{N'}$, D_B , and D_A for the various molecules. A cell with size λ contains on average one molecule of the species G , N , or B .

The stationary-state curve can be parametrized as follows:

$$\begin{aligned}
 \tilde{G}(s) &= \frac{k_{dp}}{s k_t \chi'}, \\
 \tilde{N}(s) &= \frac{[s(s k_r \tau \chi' \chi + 1) k_t + k_r] k_{dp} - s k_r k_{dp}^2 \tau - s k_r k_t \chi' \chi}{s^2 k_r k_t k_{dp} \chi' \tau - s^2 k_t^2 \chi' - s k_r k_t \chi'}, \\
 \tilde{B}(s) &= \frac{s k_t \chi}{s k_t + k_r - s k_r k_{dp} \tau}, \\
 \tilde{G}'(s) &= \frac{s k_t + k_r - s k_r k_{dp} \tau}{s k_r k_t \chi \tau}, \\
 \tilde{N}'(s) &= \frac{s k_r k_{dp}^2 \tau + s k_r k_t \chi' \chi - [(s^2 k_r \tau \chi' \chi + s) k_t + k_r] k_{dp}}{s k_r k_t k_{dp} \chi' \tau}, \\
 \tilde{A}(s) &= \chi' s,
 \end{aligned} \tag{17}$$

Figure 8 shows a projection on the G - N plane of the trajectories. As in the above-described models, the

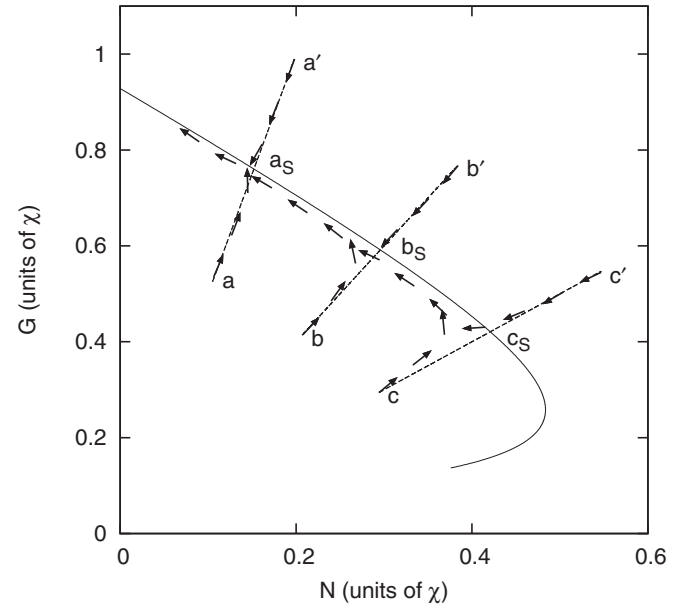


FIG. 8. Kinetics of the reaction network sketched in Fig. 7, representing the competition between two informational polymers replicated by an enzyme. The coordinates represent the concentration of the sets of genes G and N , which code respectively for active and inactive species G' and N' . The dashed lines represent the trajectories calculated by means of the mass-action equations. The solid line represents the curve of the stationary states of the mass-action equations (17), expressed by Eqs. (18). The trajectory shown by the arrows is calculated by means of the reaction-diffusion master equations. The parameters are $k_r = 15/(\tau \chi \chi')$, $k_t = 15/(\tau \chi \chi')$, $k_{dp} = 1/\tau$, $D_R = D_N = D_B = D_{R'} = D_{N'} = D_A = 4\lambda^2/\tau$, $l = 2\lambda$, and $\mathcal{N} = 20$. The values of k_{dn} , k_{gn} , k_{gp} , and C are irrelevant under the adimensionalization used.

stationary-state curve is represented by a solid line, and the trajectories calculated by means of the mass-action equations are represented by dashed lines. Also in this case, the trajectories are straight lines crossing the origin; this is a consequence of the conservation of the ratio G/N . Moreover, also in the present case, the trajectories obtained with the RDME (arrows) starting from the states $a, b, c, a', b',$ and c' at first follow the mass-action equation results (dashed lines) and approach the marginally stable states $a_S, b_S,$ and c_S ; then a drift along the stationary-state curve can be observed, directed toward the increase of the set of genes G coding for the active species G' . Also in this case, the drift represents a step of evolution toward a more efficient self-replicating system.

IV. DISCUSSION OF VALIDITY OF NUMERICAL METHOD.

In order to quantitatively evaluate the speed at which the growth of the most active molecules takes place, the dependence of the numerical model on the linear dimension of the cell, l , must be discussed. The range of l for which the RDME are valid will be discussed in this section. I will start discussing the reaction network sketched in Fig. 7.

The graph of Fig. 9 shows the dependence of the drift speed dG/dt on l , for various values of diffusion coefficients. Two regimes can be easily observed.

At short l , a power-law $dG/dt = A_1(D)l^{-\alpha}$ decrease of the speed is observed. The numeric factor $A_1(D)$ of the power law decreases when the diffusion coefficient D increases. The best fitting value for the exponent is $\alpha = 1.3$.

At longer l , all the curves roughly collapse into a single, steeper curve, compatible with a power law $dG/dt = A_2l^{-3}$. The collapse can be easily explained observing that, in the limit of large cells, diffusion plays no role, so that A_2 does not depend on D . The drift is a second-order effect of the number fluctuations [see Eq. (23) and the following comments] and thus is proportional to the average number of molecules in a cell, which is in turn proportional to the volume l^3 . Based on this argument, I will assume that the exponent is exactly -3 (although the data do not allow an accurate evaluation) and that A_2 does not depend on D .

For l shorter than a given value, dependent on D , the solution of the RDME suddenly changes. The trajectories in the concentration space are no longer attracted by the stationary-state curve—on the contrary, they fall to the origin. The shaded area of Fig. 9 represents the region for which the drift along the stationary-state curve is not found.

Now I will compare the results of Fig. 9 with the range of validity of the RDME, $l_{\min} \ll l \ll l_{\max}$, described in Sec. II. One of the conditions for the validity of the RDME is Eq. (4), which must hold for replication and translation events. Numerical calculations show that, for the conditions used for the graph in Fig. 9, the molecular rates of replication and translation are approximately equal: $\mathcal{R}_{r,S} \approx \mathcal{R}_{t,S}$. For diffusion, $\mathcal{R}_{d,S} = 6D/l^2$, which is a power law with exponent 2. For replication, $\mathcal{R}_{r,G}$ is a constant: $\mathcal{R}_{r,G} = k_r c_G' c_B$. The condition $\mathcal{R}_{r,G} \ll \mathcal{R}_{d,G}$ thus defines an upper

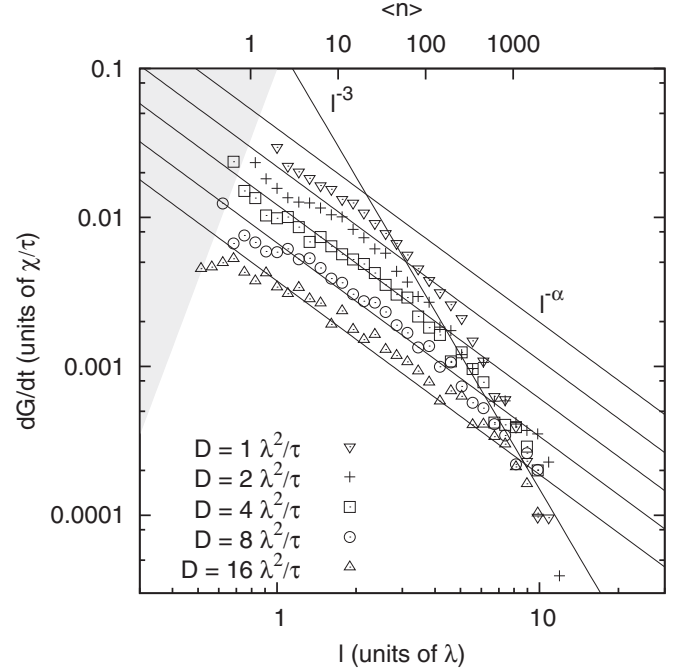


FIG. 9. Dependence of the drift speed dG/dt on the cell size l . Results refer to the reaction network of Fig. 7. The concentration of the set of genes G which codes for the active species G' increases with time. The values of the parameters are $k_r = 15/(\tau \chi \chi')$, $k_t = 15/(\tau \chi \chi')$, and $k_{dp} = 1/\tau$, while the values of $k_{dn}, k_{gn}, k_{gp},$ and C are irrelevant under the adimensionalization used. The value of \mathcal{N} ranges from 15 to 35, decreasing for increasing l , in order to perform the calculations in a reasonable time. The average number of molecules per cell, $\langle n \rangle$, is also shown. The starting values of the concentrations are chosen on the stationary-state curve, with the condition $G = N$. The various data sets refer to different diffusion coefficients D , which are equal for all the molecules.

bound:

$$l_{\max} \propto D^{1/2}. \quad (18)$$

In Fig. 9, the border of the shaded area can be considered the lower bound l_{\min} . For $l \lesssim l_{\min}$, the results are no longer valid; the numerical calculations show that the trajectories of the RDME equations fall to the origin.

By comparing the range $l_{\min} \ll l \ll l_{\max}$ with the values of l shown in Fig. 9, it can be noticed that it corresponds to the range where $dG/dt = A_1(D)l^{-\alpha}$.

The roll-off between the power laws with exponent -3 and $-\alpha$ can be identified with l_{\max} . This identification allows for calculation of the scaling law for $A_1(D)$:

$$A_1(D) \propto D^{(\alpha-3)/2}. \quad (19)$$

This dependence has been used to draw the lines with exponent $-\alpha$ in Fig. 9.

The scaling law $dG/dt \propto D^{(\alpha-3)/2}$ is explicitly shown in Fig. 10. The abscissa of the graph, D_P , represents the diffusion coefficient of the polymers $G, N, G',$ and N' . Two data sets are shown, with two different values of the diffusion coefficient D_M of the monomers B and A . It can be noticed that the drift speed does not depend on D_M . This means that the kinetics of

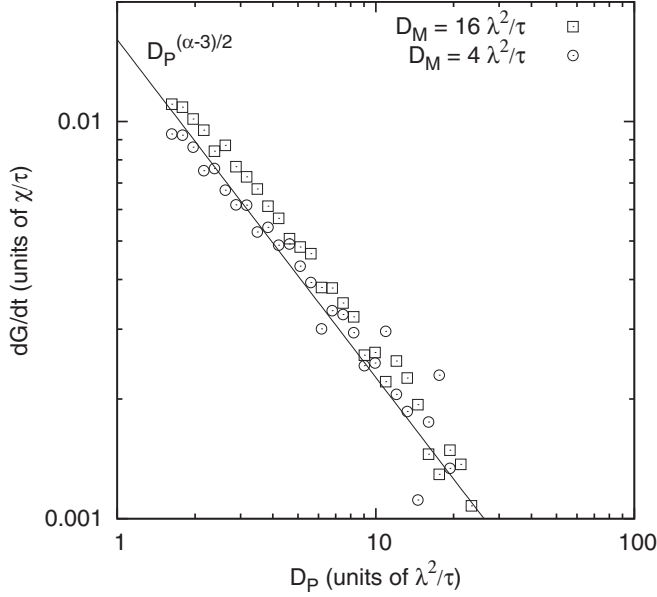


FIG. 10. Dependence of the drift speed dG/dt on the diffusion coefficient. Results refer to the reaction network of Fig. 7. Parameters are as in Fig. 9. The abscissa D_P represents the diffusion coefficient of the polymers; namely, $D_P = D_G = D_N = D_{G'} = D_{N'}$. The two data sets refer to different diffusion coefficients D_M of the monomers; namely, $D_M = D_B = D_A$. A power-law dependence can be observed with an exponent compatible with $(\alpha - 3)/2$. Changing the diffusion coefficient of the monomers does not affect the results.

the monomers is dominated by the chemical reactions inside cells, and not by diffusion.

If a mesoscopic regime is present, the results should be independent of l in the range $l_{\min} \ll l \ll l_{\max}$. For example, in Landau fluctuating hydrodynamics [47] this independence actually takes place in the mesoscopic regime. The naïve argument is the following: In the mesoscopic regime, concentration fluctuations are present. Their size can be roughly quantified by the correlation length ξ of the concentration. For $l \ll \xi$, the fluid inside each cell can be considered completely uniform, reducing even more the cell size should not change the results. The present case shows a different behavior; in particular, a plateau is not present in the graph of Fig. 9. This leads to the suggestion that no characteristic length scale is present; this means that the concentration fluctuations are scale invariant. A similar argument for deducing the scale invariance is typical of self-organized criticality theory (see, e.g., [56]).

In order to investigate this point, I evaluated the spatial cross spectra $\mathcal{S}_{S,S'}(\vec{q})$ between the concentrations of the species S and S' , defined as follows:

$$\mathcal{S}_{S,S'}(\vec{q}) = \langle \Re[c_S(\vec{q})c_{S'}^*(\vec{q})] \rangle, \quad (20)$$

where $c_S(\vec{q})$ are the Fourier transforms of the concentrations:

$$c_S(\vec{q}) = \frac{\mathcal{C}}{\sqrt{(lN)^3}} \sum_{i,j,k} n_S(i,j,k) e^{i\vec{q} \cdot [i,j,k]}. \quad (21)$$

Figure 11 shows the spatial cross spectrum of the concentrations of G and G' , as a function of the spatial wavelength q . It can be noticed that the cross spectrum follows a power law with exponent $-\beta \approx -1.7$. This confirms the scale invariance

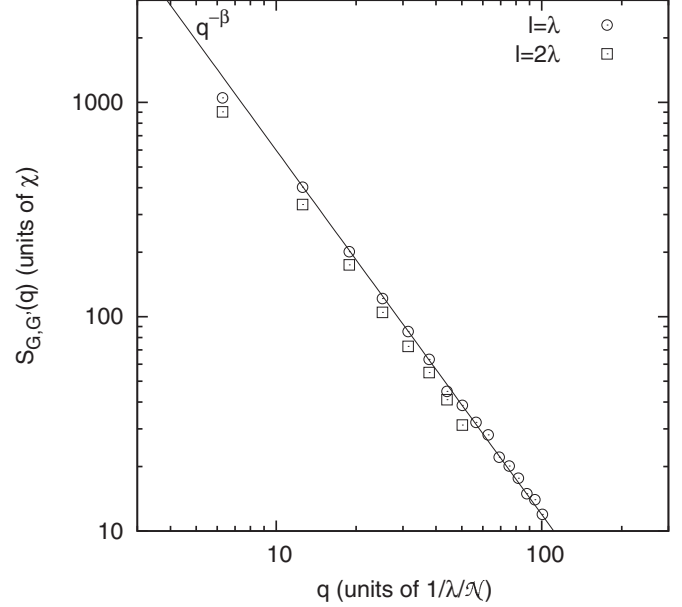


FIG. 11. Spatial cross spectrum $S_{G,G'}(k)$ of the concentrations of G and G' [see Eq. (20) for the definition]. Results refer to the reaction network of Fig. 7. Parameters are as in Fig. 9. A correction for the periodic boundary conditions has been applied. The two data sets refer to $l = \lambda$ ($\mathcal{N} = 32$) and $l = 2\lambda$ ($\mathcal{N} = 16$). A power-law decay can be observed with an exponent compatible with $-\beta = -1.7$.

of the concentration in the accessible range of spatial scales. Moreover, the cross spectrum does not change when l is changed, which confirms that the model is actually describing mesoscopic properties of the physical systems and not only artifacts due to the chosen value of l .

It is possible to relate the power law exponent $-\alpha$ of the drift speed dG/dt to the power law exponent $-\beta$ of the cross-spectrum of the concentrations of G and G' . In the following, I will evaluate $d(G/N)/dt$, which is proportional to dG/dt . The calculations are easier because G/N is a constant of motion of the mass-action equations [while the derivative $d(G/N)/dt$ is calculated with respect to the RDME]. By using the values of the transition rates shown in Table I, the drift speed can be evaluated:

$$\frac{dG}{dtN} = \frac{d\langle n_G \rangle}{dt\langle n_N \rangle} = \frac{k_r \mathcal{C}^2}{l^6} \frac{1}{\langle n_N \rangle^2} (\langle n_G n_G n_B \rangle \langle n_N \rangle - \langle n_G' n_N n_B \rangle \langle n_G \rangle). \quad (22)$$

Now I rewrite this expression in terms of the average number of molecules $\bar{n}_S = \langle n_S \rangle$ and of the displacements $\delta n_S = n_S - \bar{n}_S$:

$$\begin{aligned} \frac{dG}{dtN} = \frac{k_r \mathcal{C}^2}{l^6} \frac{1}{\bar{n}^2} [& \langle \delta n_G \delta n_G \rangle \bar{n}_B \bar{n}_N \\ & + \langle \delta n_G \delta n_B \rangle \bar{n}_G \bar{n}_N - \langle \delta n_G' \delta n_N \rangle \bar{n}_B \bar{n}_G \\ & - \langle \delta n_N \delta n_B \rangle \bar{n}_G \bar{n}_G + \langle \delta n_G' \delta n_G \delta n_B \rangle \bar{n}_N \\ & - \langle \delta n_G' \delta n_N \delta n_B \rangle \bar{n}_G]. \end{aligned} \quad (23)$$

For $l \gg \lambda$, $\bar{n}_S \gg 1$, and $\delta n_S \ll \bar{n}_S$ the third-order terms in δn_S can be neglected compared with the second-order terms. Numerical calculations show that this approximation holds for

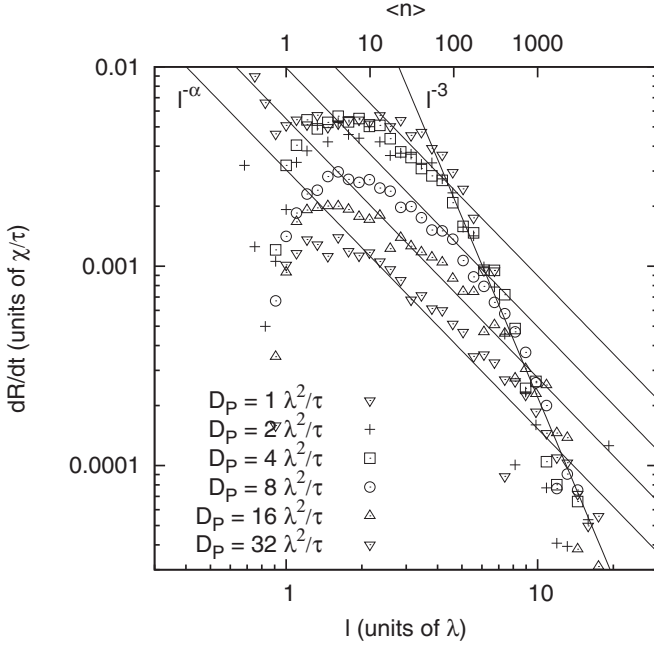


FIG. 12. Dependence of the drift speed dR/dt on cell size l . Results refer to the reaction network of Fig. 1. The concentration of the replicase R increases with time. The value of the replication reaction constant is $k_r = 9/(\tau\chi^2)$, while the values of k_d, k_g , and \mathcal{C} are irrelevant under the adimensionalization used. The value of \mathcal{N} ranges from 15 to 35, decreasing for increasing l in order to perform the calculations in a reasonable time. The average number of molecules per cell, $\langle n \rangle$, is also shown. The starting values of the concentrations are chosen on the stationary-state curve with the condition $R = 2T$. The various data sets refer to different diffusion coefficients. The diffusion coefficient of the polymers is $D_P = D_R = D_T$, while the diffusion coefficient of the monomer is $D_M = D_B = 5D_P$.

all the values of l shown in the graph of Fig. 9. Moreover, the largest term is $\langle \delta n_G \delta n_G \rangle$:

$$\frac{dG}{dt} \frac{dG}{dN} = \frac{k_r \mathcal{C}^2}{l^6} \frac{B}{N} \langle \delta n_G \delta n_G \rangle. \quad (24)$$

This quantity can be evaluated from the integral of the cross-spectrum shown in Fig. 11:

$$\frac{\langle \delta n_G \delta n_G \rangle \mathcal{C}^2}{l^6} \propto \int \mathcal{S}_{G,G}(\vec{q}) d^3q. \quad (25)$$

The effect of the discretization with steps l is that the integral must be limited to values $|\vec{q}| < \pi/l$. Assuming that the cross spectrum follows a power law with exponent $-\beta$,

$$\frac{dG}{dt} \propto l^{\beta-3}, \quad (26)$$

which leads to the relation $\beta - 3 = -\alpha$ between the exponent $-\beta$ of the correlation function and the exponent $-\alpha$ of the drift speed. This equality is confirmed by the graph of Fig. 11: the slope of the curve is compatible with the exponent $-\beta = \alpha - 3 \approx -1.7$.

The case of the reaction network sketched in Fig. 1 (the replicase that can self-replicate) is more complex. Figure 12 shows the dependence of the drift speed dR/dt on the cell size l . For $l \gtrsim 2\lambda$, the graph can be described as before; namely,

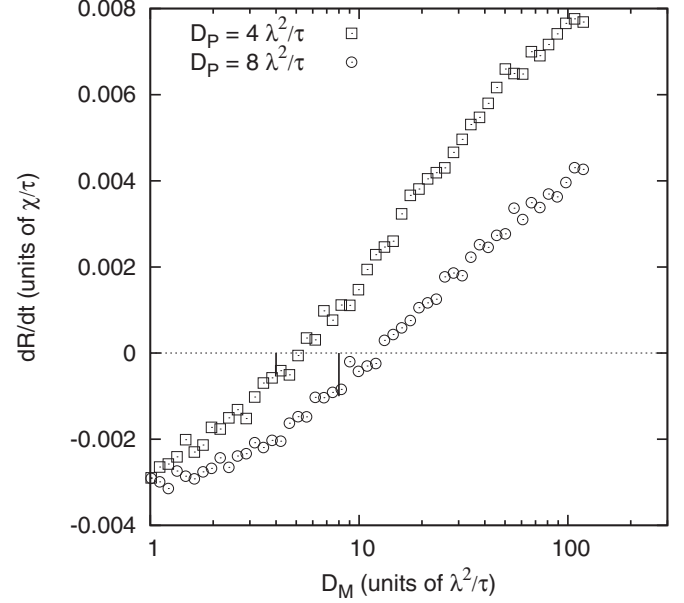


FIG. 13. Dependence of the drift speed dR/dt on the diffusion coefficient. Results refer to the reaction network of Fig. 1. Parameters are as in Fig. 12 and $l = 3\lambda$. The abscissa D_M represents the diffusion coefficient of the monomers; namely, $D_M = D_B$. The two data sets refer to two different diffusion coefficients D_P of the polymers; namely, $D_P = D_R = D_T$. The drift becomes negative for the unrealistic condition $D_M < D_P$.

as two power laws with exponents -3 and $-\alpha$. For shorter l , a plateau is observed. For the lower values of the diffusion coefficient, the power law $l^{-\alpha}$ cannot be observed at all.

Another difference from the previous case of Fig. 7 is that, in this case, the drift speed depends also on the diffusion coefficient D_M of the monomers, as shown in Fig. 13. It can be noticed that the drift speed decreases as the diffusion coefficient of the monomers D_M decreases and becomes negative approximately when $D_M < D_P$. In real cases, $D_M > D_P$; anyhow, the calculations show that, when this condition is not met, the drift even inverts the direction.

The observed behavior is explained by the following calculations: As in the previous case, the temporal derivative of the ratio R/T is expressed in terms of transition rates:

$$\frac{dR}{dt} \frac{1}{T} = \frac{d}{dt} \frac{\langle n_R \rangle}{\langle n_T \rangle} = \frac{k_r \mathcal{C}^2}{l^6} \frac{1}{\langle n_T \rangle^2} [\langle n_R (n_R - 1) n_B \rangle \langle n_T \rangle - \langle n_R n_T n_B \rangle \langle n_R \rangle], \quad (27)$$

and the equations are expressed in terms of the average number of molecules $\bar{n}_S = \langle n_S \rangle$ and of the of the displacements $\delta n_S = n_S - \bar{n}_S$:

$$\begin{aligned} \frac{dR}{dt} \frac{1}{T} = & \frac{k_r \mathcal{C}^2}{l^6} \frac{1}{\bar{n}_T^2} [\langle \delta n_R^2 \rangle \bar{n}_B \bar{n}_T \\ & + \langle \delta n_R \delta n_B \rangle \bar{n}_R \bar{n}_T - \langle \delta n_R \delta n_T \rangle \bar{n}_B \bar{n}_R \\ & - \langle \delta n_T \delta n_B \rangle \bar{n}_R^2 + \langle \delta n_R^2 \delta n_B \rangle \bar{n}_T \\ & - \langle \delta n_R \delta n_T \delta n_B \rangle \bar{n}_R - \langle \delta n_R \delta n_B \rangle \bar{n}_T - \bar{n}_R \bar{n}_B \bar{n}_T]. \end{aligned} \quad (28)$$

The second-order terms in δn_S are much larger than the third-order terms in the limit $l \gg \lambda$. The numerical calculations show that the second-order terms in δn_S are

dominant for $l \gtrsim 2\lambda$, while the plateau at $l \lesssim 2\lambda$ is due to the effect of third-order terms. Anyhow, in the region of the plateau the average number of molecules $\langle n_S \rangle$ is not enough to guarantee the validity of the RDME [49]. For $l \gg 2\lambda$, the numerical calculations show that the largest terms are the following:

$$\frac{d}{dt} \frac{G}{N} = \frac{k_r \mathcal{C}^2}{l^6} \frac{1}{T^2} [(\langle \delta n_R^2 \rangle - \langle \delta n_R \rangle) BT + \langle \delta n_R \delta n_B \rangle RT]. \quad (29)$$

The term $\langle \delta n_R^2 \rangle - \langle \delta n_R \rangle$ vanishes for the Poissonian distribution and represents the enhancement of the variance of the concentration due to the development of long-range correlations. Also in this case, the power spectrum of the fluctuations is a power law with exponent $-\alpha = -1.3$. The term $\langle \delta n_R \delta n_B \rangle$ is negative (i.e., it reduces the drift speed). In the limit of infinite diffusion coefficient D_M , the number of monomers B in a cell is random, and is not correlated with other concentrations: $\langle \delta n_R \delta n_B \rangle$ vanishes as the diffusion of the monomers increases. This explains the increase of the drift speed observed when the diffusion coefficient D_B is increased.

V. CONCLUSIONS

I showed that the chemical reaction networks presented in Sect. III are marginally stable. The marginal stability comes from replication and, in particular, from inheritance; that is, from the fact that the generated molecule (the offspring) inherits from the replicated molecule (the parent) the property of belonging to a given chemical species.

Marginal stability allows drift along the stationary-state curve under the effect of concentration fluctuations. This phenomenon can be classified as a deviant nonclassical effect [45,46]. The drift can be interpreted as the evolution toward a more efficiently replicating system.

One of the open problems in abiogenesis is the identification of the mechanism through which a molecular species, which

is more efficient in replicating other molecules, increases its concentration. The effect of concentration fluctuations in the described marginally stable systems is a possible solution which is more simple than the already-proposed solutions [38].

On the other hand, the chemical reaction networks proposed in various models of abiogenesis require marginal stability, if this has not been noted before. For example, the RNA-world hypothesis actually relies on a reaction network similar to the one sketched in Fig. 4, which is marginally stable. For the case of the “compositional genome” [32], it has been noticed that the presence of a single stationary state of the mass-action equations prevents evolution [33], which means that marginal stability is a necessary condition for evolution in this model.

Summarizing, I propose that the marginal stability is the necessary mathematical property needed by chemical reaction networks in order to undergo evolution. Such a formal property can be used in order to find candidates of chemical reaction networks potentially involved in abiogenesis; in particular, in the case of metabolism-first models.

I showed that the fluctuations generating the drift along the stationary-state curve are self-similar and scale invariant and that their properties are a genuine mesoscopic phenomenon, which does not depend on the value of the parameter l of the mathematical method. Unfortunately the speed of the drift depends on the parameter l , acting as a high-frequency cutoff, which means that the actual speed depends on the microscopic molecular interactions and not only on the macroscopic parameters such as the reaction constants and the diffusion coefficients. Anyhow, the speed of the drift is of the order of some percent of the replication speed. Whenever a marginally stable chemical network will be setup *in vitro*, the drift will be easily observable.

ACKNOWLEDGMENTS

I am grateful to Francesco Mantegazza, Domenico Salerno, Valeria Cassina, and Roberto Ziano for stimulating discussions.

-
- [1] M. Tessera, *International Journal of Molecular Sciences* **12**, 3445 (2011).
 - [2] A. Eschenmoser, *Origins Life Evol. Biosphere* **27**, 535 (1997).
 - [3] P. P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm, and B. Nordén, *Nature (London)* **368**, 561 (1994).
 - [4] D. H. Lee, J. R. Granja, J. A. Martínez, K. Severin, and M. R. Ghadiri, *Nature (London)* **382**, 525 (1996).
 - [5] G. von Kiedrowski, *Angew. Chem.* **25**, 932 (1986).
 - [6] W. S. Zielinski and L. E. Orgel, *Nature (London)* **327**, 346 (1987).
 - [7] N. V. Hud and F. A. L. Anet, *J. Theor. Biol.* **205**, 543 (2000).
 - [8] R. F. Gesteland, T. R. Cech, and J. F. Atkins, *The RNA World* (Cold Spring Harbor Lab. Press, Plainview, 1999).
 - [9] O. Kuchner and F. H. Arnold, *Trends Biotechnol.* **15**, 523 (1997).
 - [10] N. Paul and J. F. Joyce, *PNAS* **99**, 12733 (2002).
 - [11] T. A. Lincoln and G. F. Joyce, *Science* **323**, 1229 (2009).
 - [12] D. E. Kim and J. F. Joyce, *Chem. Biol.* **11**, 1505 (2004).
 - [13] W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, and D. P. Bartel, *Science* **292**, 1319 (2001).
 - [14] S. B. Voytek and G. F. Joyce, *Proc. Natl. Acad. Sci. USA* **104**, 15288 (2007).
 - [15] J. W. Szostak, *Nature (London)* **459**, 171 (2009).
 - [16] S. L. Miller, *Science* **117**, 528 (1953).
 - [17] J. Oró and A. P. Kimball, *Biochem. Biophys. Res. Commun.* **2**, 407 (1960).
 - [18] J. P. Ferris, A. Sanchez, and L. E. Orgel, *J. Mol. Biol.* **33**, 639 (1968).
 - [19] M. W. Powner, B. Gerland, and J. D. Sutherland, *Nature (London)* **459**, 239 (2009).
 - [20] D. Müller, S. Pitsch, A. Kittaka, E. Wagner, C. E. Wintner, and A. Eschenmoser, *Helv. Chim. Acta* **73**, 1410 (1990).
 - [21] J. P. Ferris, A. R. jr Hill, R. Liu, and L. E. Orgel, *Nature (London)* **381**, 59 (1996).
 - [22] S. A. Kauffman, *Self Organization and Selection in Evolution* (Oxford University Press, New York, 1993).

- [23] R. Shapiro, *Origins of life and evolution of biospheres* **39**, 346 (2009).
- [24] F. A. L. Anet, *Curr. Opin. Chem. Biol.* **8**, 654 (2004).
- [25] R. Shapiro, *Quarterly Review of Biology* **81**, 105 (2006).
- [26] G. Wächtershäuser, *Microbiol. Rev.* **52**, 452 (1988).
- [27] C. Huber and G. Wächtershäuser, *Science* **276**, 245 (1997).
- [28] H. J. Morowitz, J. D. Kostelnik, J. Yang, and G. D. Cody, *Proc. Natl. Acad. Sci. USA* **97**, 7704 (2000).
- [29] R. Breslow, *Tetrahedron Lett.* **21**, 22 (1959).
- [30] A. Ricardo, M. A. Carrigan, A. N. Olcott, and S. A. Benner, *Science* **303**, 196 (2004).
- [31] L. E. Orgel, *Proc. Natl. Acad. Sci. USA* **97**, 12503 (2000).
- [32] D. Segré, D. Ben-Eli, and D. Lancet, *Proc. Natl. Acad. Sci. USA* **97**, 4112 (2000).
- [33] V. Vasas, E. Szathmáry, and M. Santos, *PNAS* **107**, 1470 (2010).
- [34] M. Eigen, *Naturwissenschaften* **65**, 7 (1978).
- [35] W. Hordijk, J. Hein, and M. Steel, *Entropy* **12**, 1733 (2010).
- [36] E. Mossel and M. Steel, *J. Theor. Biol.* **233**, 327 (2005).
- [37] W. Hordijk and M. Steel, *J. Theor. Biol.* **227**, 451 (2004).
- [38] J. W. Szostak, D. P. Bartel, and P. L. Luisi, *Nature (London)* **409**, 387 (2001).
- [39] M. M. Hanczyc, S. M. Fujikawa, and J. W. Szostak, *Science* **302**, 618 (2003).
- [40] I. A. Chen, R. W. Roberts, and J. W. Szostak, *Science* **305**, 1474 (2004).
- [41] I. A. Chen, *Science* **314**, 1558 (2006).
- [42] D. W. Deamer, *Nature (London)* **454**, 37 (2008).
- [43] S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobé, D. A. Treco, and J. W. Szostak, *Nature (London)* **454**, 122 (2008).
- [44] D. Brogioli, *Phys. Rev. Lett.* **105**, 058102 (2010).
- [45] M. S. Samoilov and A. P. Arkin, *Nat. Biotechnol.* **24**, 1235 (2006).
- [46] H. Qian, S. Saffarian, and E. L. Elson, *PNAS* **99**, 10376 (2002).
- [47] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1959).
- [48] M. M. Mansour, C. V. D. Broek, G. Nicolis, and J. W. Turner, *Ann. Phys. (NY)* **131**, 283 (1981).
- [49] F. Baras and M. M. Mansour, *Phys. Rev. E* **54**, 6139 (1996).
- [50] M. M. Mansour, J. Dethier, and F. Baras, *J. Chem. Phys.* **114**, 9265 (2001).
- [51] D. Bernstein, *Phys. Rev. E* **71**, 041103 (2005).
- [52] D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- [53] J. D. Lohn, S. P. Comombano, J. Scargle, D. Stassinopoulos, and G. L. Haith, in *Proceedings of the 1998 IEEE International Conference on Evolutionary Computation* (Anchorage, AK, USA, 1998), pp. 487–492.
- [54] D. Bonchev and O. Mekenyan, *Graph Theoretical Approaches to Chemical Reactivity* (Kluwer Academic Publishers, Dordrecht, 1994).
- [55] S. A. Isaacson and D. Isaacson, *Phys. Rev. E* **80**, 066106 (2009).
- [56] P. Bak, C. Tang, and K. Wiesenfeld, *Phys. Rev. Lett.* **59**, 381 (1987).