Recursiveness, switching, and fluctuations in a replicating catalytic network

Kunihiko Kaneko
Department of Basic Science, College of Arts and Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153, Japan
(Received 13 February 2003; published 18 September 2003)

A protocell model consisting of mutually catalyzing molecules is studied, in order to investigate how chemical compositions are transferred recursively through cell divisions under replication errors. Depending on the numbers of molecules and species, and the path rate, three phases are found: fast switching state without recursive production, recursive production, and itinerancy between the above two states. The number distributions of the molecules in the recursive states are shown to be log-normal except for those species that form a core hypercycle, and are explained with the help of a heuristic argument.

DOI: 10.1103/PhysRevE.68.031909 PACS number(s): 87.17.Aa, 87.80.Vt, 89.75.Fb

I. INTRODUCTION

In a cell, a huge number of chemicals is synthesized by mutual catalyzation leading to replication of molecules that allow a cell to grow until it is large enough to divide into two. In spite of complexity and diversity in components, a cell sustaining similar chemical compositions is reproduced successively, which is called ''recursive'' production here. The question how such recursive production of cells continues has been discussed in considering the origin of life [1–4], following Oparin's pioneering work. Eigen and Schuster proposed the hypercycle [2] as a mechanism to overcome an inevitable loss in the catalytic activities through mutations. Although the hypercycle itself may be destroyed by parasitic molecules, i.e., molecules which are catalyzed by the hypercycle species but themselves do not catalyze other molecules, it was later shown that compartmentalization by a cell structure or localized patterns can suppress the invasion of parasitic molecules [5–7]. On the other hand, how a cell with a collection of chemicals sustains catalytic activity for reproduction is discussed by Dyson [3] with the use of Ising model, and Kauffman [4], by using autocatalytic reaction network. Achievement of recursive production is shown by Segré *et al.* [8], by introducing some local structure in a random autocatalytic network.

Besides the reproduction, a cell must have ability to evolve to a faster and more robust reproduction. How such evolvability and recursive production are compatible is an important question. Here, by taking a simple model with mutually catalytic, random reaction network, we study their ability for recursive reproduction. Besides fixed recursive production states, we find fast switching states and itinerancy over several quasirecursive states. The latter allows for evolution, since several recursive reproduction states are visited [9]. Finally, we study the characteristics of the number distributions of the molecular species in these replicating cells, to show the importance of the log-normal distribution.

These behaviors, as well as the log-normal distribution, is shown to be universal as long as a cell grows through replications of molecules with catalytic networks. Although a finely-tuned machinery for replication is further added to the present cell, the basic condition mentioned above is still there. Considering the universality of our result, we expect that the distribution law is also satisfied in the present cell, as will be discussed. Indeed, the strategy to predict some universal characteristics for the present cell by a simple model is shown to be valid recently [11].

II. MODEL

We envision a (proto)cell containing k molecular species with some of the species possibly having a zero population. A chemical species can catalyze the synthesis of some other chemical species as

$$[i]+[j]\rightarrow[i]+2[j], \tag{1}$$

with $i,j=1,\ldots,k$ according to a randomly chosen reaction network, where the reaction is set at far from equilibrium. In Eq. (1), the molecule i works as a catalyst for the synthesis of the molecule j, while the reverse reaction is neglected as discussed in the hypercycle model. For each chemical the rate for the path of catalytic reaction (1) is given by ρ, i.e., each species has about $k\rho$ possible reactions. The rate is kept fixed throughout each simulation. Considering catalytic reaction dynamics, the reverse reaction process is neglected, and reactions $i\leftrightarrow j$ are not included. (Here we investigated the case without direct mutual connections, i.e., $i\rightarrow j$ was excluded as a possibility when there was a path $j\rightarrow i$, although this condition is not essential for the results to be discussed.) Furthermore, each molecular species i has a randomly chosen catalytic ability $c_i\in[0,1]$ [12], i.e., the above reaction occurs with the rate c_i. Assuming an environment with an ample supply of chemicals available to the cell, the molecules then replicate leading to an increase in their numbers within a cell. Our main concern here is the dynamics of these molecule numbers N_i of the species i under replication.

During the replication process, structural changes (e.g., alternation of a sequence in a polymer) may occur, which alter the catalytic activities of the molecules. Therefore, the activities of the replicated molecule species can differ from those of the mother species. The rate of such structural changes is given by the replication ''error rate'' μ. As a simplest case, we assume that this error leads to all other molecule species with equal probability, [i.e., with the rate $\mu/(k-1)$]. In reality of course, even after a structural change, the replicated molecule will keep some similarity with the original molecule, and this equal rate of transition to

other molecule species is a drastic simplification. Some simulations where the errors in replication only lead to a limited range of molecule species, however, show that the simplification does not affect the basic conclusions presented here. Hence we use the simplest case for most simulations.

The model is simulated as follows. At each step, a pair of molecules (say, i and j) is chosen randomly. If there is a reaction path between species i and j, and i (j) catalyzes j (i), one molecule of the species j (i) is added with probability c_i (c_j), respectively. The molecule is then changed to another randomly chosen species with the probability of the replication error rate μ. When the total number of molecules exceeds a given threshold (denoted as N), the cell divides into two, such that each daughter cell inherits half ($N/2$) of the molecules of the mother cell, chosen randomly [10]. In order to take the importance of the discreteness in the molecule numbers into account, we adopted a stochastic [13] rather than the usual differential equations approach (see also Ref. [8]).

III. PHASES

The cell state at the nth division is characterized by the molecule numbers of the chemical species $\{N_1^n, N_2^n, \ldots, N_k^n\}$ (with $\Sigma_j N_j^n = N$), Ref. [14], while there are four basic parameters; N, k, μ, and ρ. By investigating the dynamics of 1000 randomly chosen networks, and changing the four parameters, we have found that the behaviors of the system can be classified into just the following three types (dependence on the parameters will be discussed later [15]): (A) Fast switching states without recursiveness, (B) fixed recursive states, and (C) itinerancy over several quasirecursive states.

In phase (A), even though each generation has some dominating species as with regards to the molecule numbers, the dominating species change every few generations. Information regarding the previously dominating species is totally lost often to the point that its population drops to zero. Here no stable mutual catalytic relationships are formed. Indeed, by autocatalytic reactions, the population of one dominant species can be amplified, but soon it is replaced by another chemical that is catalyzed by it [for example, see Fig. 1(a)].

In phase (B), a recursive state is established where the chemical composition is stable enough to withstand the division process. Once reached, this state lasts all over the generations [$O(10^4)$] in the simulation [see Fig. 1(a)].

The recursive state (''attractor'') here is not necessarily a fixed point, and the molecule numbers may oscillate in time. Nevertheless, the overall chemical compositions remain within certain ranges: for example, the major species (i.e., those synthesized by themselves, not by an error) are not altered over the generations. Generally, all the observed recursive states consist of 5–12 species [16], except for those species which exist only as a result of replication errors. (See also Ref. [8] for recursive transmission of state in a network model with some structure.)

For example, in the recursive state depicted in Fig. 1(b), 11 species remain in existence throughout the simulation. As shown, three species have much higher populations than others, which form a hypercycle as $109 \rightarrow 11 \rightarrow 13 \rightarrow 109$. (The numbers 11,13, . . . are indices of chemical species, initially assigned arbitrarily.) The hypercycle sustains the replication of the molecules, and is called ''core hypercycle.'' The catalytic activities of the species satisfy $c_{13} > c_{109} > c_{11}$, and accordingly, the respective populations satisfy $N_{11} > N_{109} > N_{13}$. This relationship is natural, since molecules with higher catalytic activities result in the synthesis of more molecules other than themselves, thus suppressing their own population fraction [17]. Here, through mutual catalyzation, molecules with higher catalytic activities are catalyzed by molecules with lower activities but larger populations. To destroy such a network of mutual support, large fluctuations in molecule numbers are required, which are rare for large N.

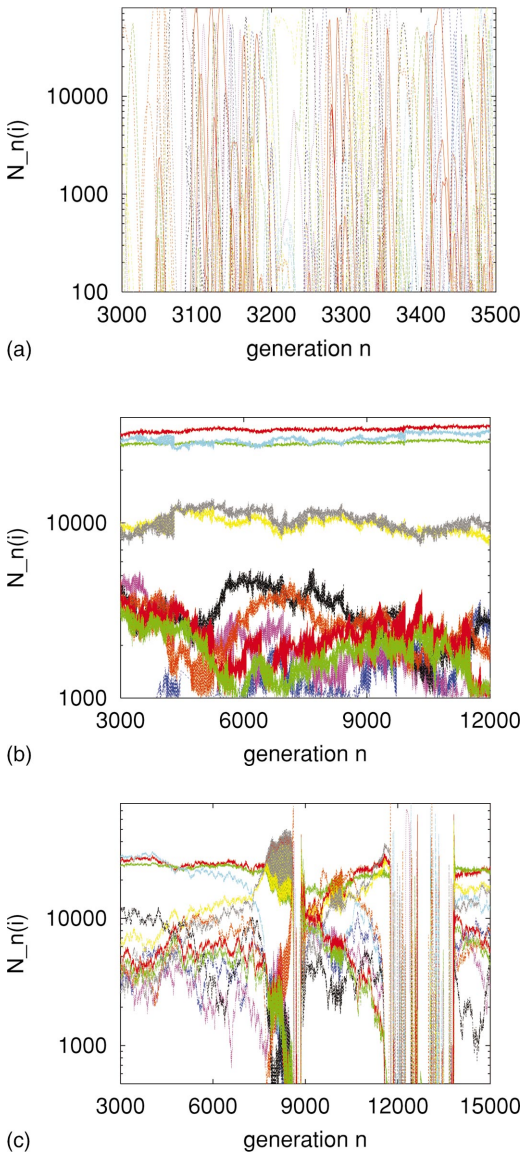

FIG. 1. (Color) The number of molecules, N_i^n, for the species i is plotted as a function of the generation n, i.e., at each successive division event n. $\rho = 0.1$, $k = 200$, and $N = 64\,000$. (a) $\mu = 0.01$, and $k = 500$, (b) $\mu = 0.01$ and $k = 200$, and (c) $\mu = 0.1$ and $k = 200$. For (b) and (c), the same network is adopted. Different colors correspond to different species, while only some species (whose population becomes large during some generation) are plotted.

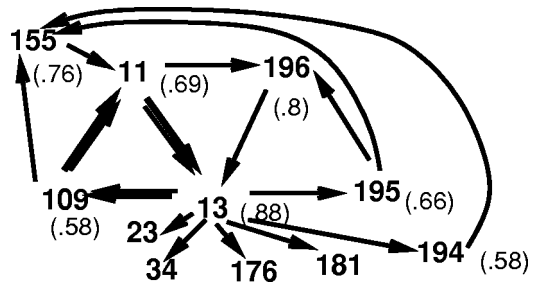

FIG. 2. The catalytic network of the species that constitute the recursive state. The numbers in () denote c_i of the species. The three species connected by thick arrows are the top three species in Fig. 1(a).

Hence parasitic molecule species cannot easily invade the core.

In phase (C), the system alternates between quasirecursive states similar to phase (B), which last for many generations, and fast switching states similar to phase (A). (Existence of this state is rather common, independent of specific choices of models. See also Ref. [8].) The quasirecursive state itself can be subjected to switches between core hypercycles, as can be seen in Fig. 1(c) where a switch occurs from an initial core hypercycle (109,11,13) to the next core hypercycle (11,13,195,155) around the 8500th generation. Subsequently, around the 12 000th generation, the core network is taken over by parasites to enter the phase (A) like fast switching state which in turn gives way for a new quasirecursive state around the 14 000th generation. The former is the competition among core networks, while the latter is the invasion of parasitic molecules. These two types of switches are generally observed, while the latter, which destroys the quasistable recursive state, is more frequent, Once this occurs, successive take-over of parasitic molecules occurs, and the fast switching state (A) appears, accompanied by the decrease in chemical diversity. When N is not so large, the molecule number of the species with the highest catalytic activity in the core hypercycle can become small due to fluctuations, and subsequently succumb to parasitic molecules. (Note the species catalyzes other molecules better, and the number is smaller [18].) Then, the core hypercycle loses its main catalyst, resulting in its collapse and giving way to a fast switching state that in turn will allow the formation of a new core hypercycle (which may not be identical to the previous one).

Which one of the phases (A)–(C) appears, of course, depends on the parameters and the specific structure of the network. By choosing a variety of networks, however, we find a clear dependence of the fraction of the networks on the parameters, leading to a rough sketch of the phase diagram. Generally, the fraction of phase (B) increases and the fraction of phase (A) decreases for increasing N, or for decreasing k, ρ, or μ. For example, the fraction of phase (A) [or phase (C)] gets larger as k is decreased from $k \lesssim 300$ for $N = 50\,000$ (with $\rho = 0.1$ and $\mu = 0.01$), while dependence on ρ will be discussed below.

For a more systematic investigation, it is useful to classify the phases by the similarity of the chemical compositions between two cell division events [8]. This can be done by defining a k-dimensional vector $\vec{V}_n = [p_n(1), \ldots, p_n(k)]$, with $p_n(i) = N_n(i)/N$, and measuring the similarity between ℓ successive generations with the help of the inner product $H_\ell = \vec{V}_n \cdot \vec{V}_{n+\ell} / (|V_n||V_{n+\ell}|)$. In Fig. 2, the average similarity $\overline{H}_{20}$ and the average division time are plotted for 50 randomly chosen reaction networks as a function of the path probability ρ. Roughly speaking, the networks with $\overline{H}_{20} > 0.9$ belong to phase (C), and those with $\overline{H}_{20} < 0.4$ to phase (A), empirically. Hence, for $\rho > 0.2$, phase (A) is observed for nearly all the networks (e.g. 48/50), while for lower path rates, the fraction of phase (C) [with a few of phase (B)] increases. The value $\rho \sim 0.2$ is the phase boundary in this case.

In general, we have found a positive correlation between the growth speed of a cell, the similarity H, and the diversity of the molecules. (In Fig. 2, networks with larger H have smaller division times.) The recursive states maintain higher growth speeds since they effectively suppress parasitic molecules. In Fig. 2, for decreasing path rates, the variations in the division speeds of the networks become larger, and some networks that reach recursive states have higher division speeds than networks with larger ρ. On the other hand, when the path rate is too low, the protocells generally cannot grow since the probability to have useful connections in the network is nearly zero. Indeed, an optimal path rate seems to exist (e.g., around 0.05 for $k = 200$, $N = 12\,800$), for which some networks have high growth speeds. Consequently, in an environment that necessitates competition for growth, protocells having such optimal networks will be more successful than those with suboptimal networks. Existence of optimal path rate for recursive production is also discussed in other models [3,4,8].

IV. FLUCTUATION

Finally, we investigate the fluctuations of the molecule numbers of each of the species. Since the number of molecules is not very large, the fluctuations over the generations can possibly have a significant impact on the dynamics of the system. In order to quantify the sizes of these fluctuations, we have measured the distribution $P(N_i)$ for each molecule species i, by sampling over division events. Our numerical results are summarized as follows:

(i) For the fast switching states, the distribution $P(N_i)$ satisfies the power law

$$P(N_i) \approx N_i^{-\alpha}, \tag{2}$$

with $\alpha \approx 2$.

(ii) For recursive states, the fluctuations in the core network (i.e., 13,11,109 in Fig. 3) are typically small.

(iii) On the other hand, species that are peripheral to but catalyzed by the core hypercycle have log-normal distributions

$$P(N_i) \approx \exp\left(- \frac{(\ln N_i - \overline{\ln N_i})^2}{2\sigma} \right), \tag{3}$$

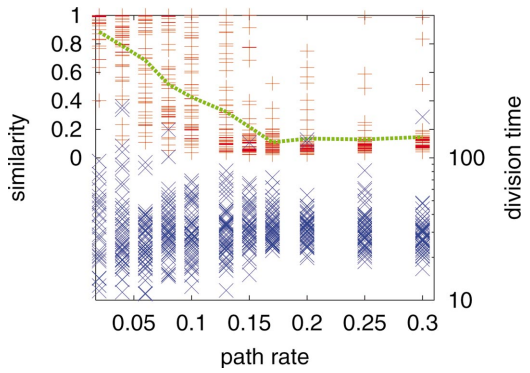

FIG. 3. (Color) The average similarity $\overline{H}_{20}$ (red) and the average division time (blue) are plotted as a function of the path rate ρ. For each ρ, data from 50 randomly chosen networks are plotted. The average is taken over 600 division events. The green line indicates the average similarity $\overline{H}_{20}$ over the 50 networks for each ρ. For $\rho > 0.2$, networks over 98% have $H < 0.4$, and they show fast switching. Around $\rho \approx 0.05$, division time ($\times$) depends largely on network, but some take very small values (i.e., can grow faster). At $\rho = 0.02$, 25 out of 50 networks cannot support cell growth, and 4 cannot support cell growth at $\rho = 0.04$.

as shown in Fig. 4. We have also plotted the variance $\overline{(N_i - \overline{N}_i)^2}$ [$\overline{\cdots}$ is the average of the distribution $P(N_i)$] and the deviation between the peak of $P(N_i)$ and $\overline{N}_i$, divided by the average $\overline{N}_i$. The deviation is positive since the distribution $P(N)$ has a tail to larger value, while by taking $\ln N$, the deviation is decreased drastically for the species except the three core-hypercycle species 11, 109, and 13, which have three largest average $\overline{N}$.

As can be seen in Fig. 5, the variance and the fluctuations in the core network are small, especially for the minority species (i.e., 13). For molecule species that do not belong to the core hypercycle, the variance scaled by the average increases as the average decreases. Furthermore, there is a distinct deviation between the peak and the average (except for the core species), since the distribution has a tail for larger sizes. On the other hand, if we use the variable $\ln N_i$ when plotting the distribution, it is closer to a Gaussian, and the difference between the peak and the average is suppressed.

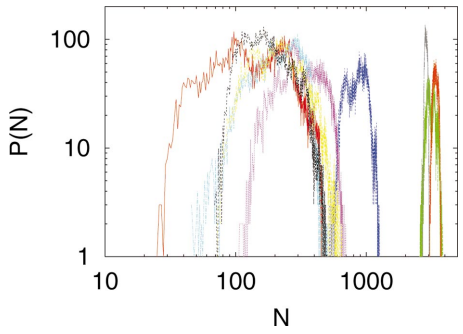

FIG. 4. (Color) The number distribution of the molecules corresponding to the network in Fig. 3. The distribution is sampled from 1000 division events. From right to left, the plotted species are 11,109,13,155,176,181,195,196,23; log-log plot.

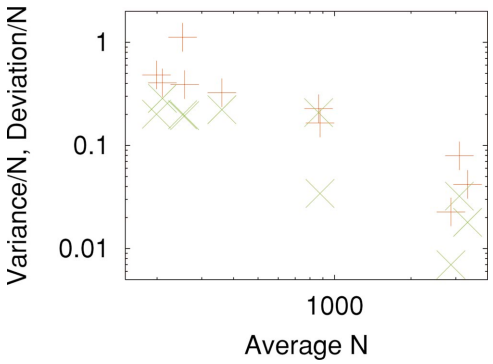

FIG. 5. (Color) Scaled variance and deviation. $\times$ denotes the variance of the molecule number divided by its average, while $+$ shows the difference between the average and the peak of the distribution divided by the average $\overline{N}_i$. From the largest to the smaller, the species 11 (the largest $\overline{N}_i$), 109 (the second largest), 13,155,194,176,195,181,196,23, 34 (smallest $\overline{N}_i$) are plotted; computed in the same way as Fig. 4.

The origin of the log-normal distributions here can be understood by the following rough argument: for a replicating system, the growth of the molecule number N_m of the species m is given by $dN_m/dt = AN_m$, where A is the average effect of all the molecules that catalyze m. We can then obtain the estimate

$$d \ln N_m / dt = \overline{a} + \eta(t) \tag{4}$$

by replacing A with its temporal average $\overline{a}$ plus fluctuations $\eta(t)$ around it [19]. If $\eta(t)$ is approximated by a Gaussian noise, the log-normal distribution for $P(N_m)$ is suggested (this argument is valid if $\overline{a} > 0$). For the fast switching state the growth of each molecule species is close to zero on the average and in this case, by considering the Langevin equation with boundary conditions, the power law follows as discussed in Ref. [20].

If several molecules mutually catalyze each other, however, one would expect their distributions to be close to Gaussian, due to the central limit theorem, and this is indeed the case for the three core species.

By studying a variety of networks, the observed distributions of the molecule numbers can be generally summarized as follows. (1) Distribution close to Gaussian form, with relatively small variances in the core (hypercycle) of the network. (2) Distribution close to log-normal, with larger fluctuations for a peripheral part of the network. (3) Power-law distributions for parasitic molecules that appear intermittently [21].

V. SUMMARY AND DISCUSSION

To sum up, features of a protocell with catalytic reactions and divisions are classified into three phases. Besides the establishment of recursive growth at the phase (B), switching state (C) is found, which makes both recursive growth and evolution compatible, since novel quasirecursive states with different chemical compositions are visited successively,

triggered by extinctions of minority molecules in the core hypercycle networks [9,17].

Previously, we pointed out the relevance of minority molecules to heredity [9,17] in mutually catalyzing systems for making recursive growth and evolution compatible. Indeed, phase (C) satisfies both the features, since novel quasirecursive states with different chemical compositions are visited successively,

To see this evolvability, we have also studied a model with two modifications; first the catalytic activity is set as $c_i = i/k$, i.e., the activity is monotonically increasing with the species index, and instead of global change to any molecule species by replication error, the error is assumed to change the species index only within a given range $i_0 \ll k$ (i.e., when the molecule species j is synthesized, with the error rate μ, the molecule $j+j'$, with j' a random number over $[-i_0, i_0]$, is synthesized). Starting from species with $i < i_{ini}$, one can examine whether the evolution having higher catalytic activity progresses or not. At the phase (C), after one recursive state (consisting of species within the width of the order $2i_0$) lasts for some time, switching to a new recursive state occurs to increase the catalytic activity (i.e., to increase the indices of species i). Once the minority species in the core (that has higher catalytic activity) decreases its population, switching to a new state occurs, which has higher catalytic activities. Hence, evolution from a rather primitive cell with low catalytic activities to those with higher activities is possible at the phase (C) with the aid of minority molecules.

Finally, we showed suppression of the fluctuation of molecules at a core hypercycle network and ubiquity of log-normal distribution of those at a peripheral network. Now, it is important to check if these characteristics are preserved in the present cell that has much finer machinery for reproduction. Here, it is interesting to note that the distributions of the abundances of fluorescent proteins, measured by flow cytometry, generally show log-normal, rather than Gaussian, distributions [22,23], while further supports from accurate experiments and from theoretical study of realistic cell models will be reported elsewhere [23]. Furthermore, it is also known that the size of bacteria and some cells in blood [24] (as well as human body weight) obey log-normal distributions.

ACKNOWLEDGMENTS

I would like to thank C. Furusawa, K. Sato, and T. Yomo for useful discussions, and F. Willeboordse for critical reading of the manuscript. This research was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan (11CE2006).

[1] See, e.g., D. Segré and D. Lancet, EMBO Rep. **1**, 217 (2000), and papers cited therein.
[2] M. Eigen and P. Schuster, *The Hypercycle* (Springer, New York, 1979).
[3] F. Dyson, *Origins of Life* (Cambridge University Press, Cambridge, England, 1985).
[4] S. A. Kauffman, *The Origin of Order* (Oxford University Press, New York, 1993).
[5] M. Boerlijst and P. Hogeweg, Physica D **48**, 17 (1991).
[6] S. Altmeyer and J.S. McCaskill, Phys. Rev. Lett. **86**, 5819 (2001).
[7] E. Szathmary and J. Maynard Smith, J. Theor. Biol. **187**, 555 (1997).
[8] D. Segré, D. Ben-Eli, and D. Lancet, Proc. Natl. Acad. Sci. U.S.A. **97**, 4112 (2000); D. Segré *et al.*, J. Theor. Biol. **213**, 481 (2001).
[9] K. Kaneko, J. Biol. Phys. **28**, 781 (2002).
[10] K. Kaneko and T. Yomo, Bull. Math. Biol. **59**, 139 (1997).
[11] C. Furusawa and K. Kaneko, Phys. Rev. Lett. **90**, 088102 (2003).
[12] In real reactions, the rate also depends on each path. Studies on a model with parameters c_{ij} show qualitatively same behaviors.
[13] We use a stochastic approach, since discreteness in molecule numbers is sometimes found to be important; Y. Togashi and K. Kaneko, Phys. Rev. Lett. **86**, 2459 (2001); see also B. Hess and A.S. Mikhailov, Science **264**, 223 (1994); J. Theor. Biol. **176**, 181 (1995).
[14] The results do not depend on at what stage of the molecule numbers (between $N/2$ and N) the cell state is observed.
[15] The behavior discussed here hardly depends on the choice of initial chemical compositions.
[16] This number does not depend on N, k, ρ, μ, although with the decrease of μ and ρ, there is a tendency of its decrease toward 3, the minimal number for a hypercycle here.
[17] The population sizes of just three species A, B, C, with the catalytic relationship $A \to B \to C \to A$ are estimated by taking the continuum limit $N \to \infty$ and obtaining a fixed point solution of the rate equation for the concentrations of the chemicals as discussed in Ref. [2]. From a straightforward calculation, we have $N_A : N_B : N_C = c_A^{-1} : c_B^{-1} : c_C^{-1}$.
[18] See K. Kaneko and T. Yomo, J. Theor. Biol. **214**, 563 (2002) for relevance of minority molecules to heredity.
[19] As such this equation diverges with time, but here, the cell divides into two before the divergence becomes significant. The finiteness of the total number N means that the asymptotic distribution as $N \to \infty$ is not available, but the argument on the distribution form is valid as long as N is sufficiently large.
[20] A. S. Mikhailov and V. Calenbuhr, *From Cells to Societies* (Springer, New York, 2002); D. Sornette, *Critical Phenomena in Natural Science* (Springer, New York, 2002).
[21] For some networks, the distributions may sometimes be intermediate between log-normal and Gaussian, and even have double peaks.
[22] V. Shivashnakar *et al.* (private communication); see also W.J. Blake *et al.*, Nature (London) **422**, 633 (2003).
[23] C. Furusawa *et al.* (unpublished).
[24] J. Facquet-Danis *et al.*, Biomedicine **25**, 294 (1970); C.E. Claren *et al.*, Biometrics **47**, 607 (1991).