Cellular automata model for gene networks

J. A. de Sales, M. L. Martins, and D. A. Stariolo
Departamento de Física, Universidade Federal de Viçosa, 36571-000, Viçosa, Minas Gerais, Brazil
(Received 1 May 1996; revised manuscript received 17 October 1996)

In order to study the overall behavior of gene networks, we propose a simple cellular automata (CA) model in which each binary gene is connected to K other inputs (including itself) interacting through asymmetric short- and long-range couplings. Using numerical simulations and mean-field calculations, collective dynamical properties of this CA model were investigated. It is shown that the CA exhibits three different dynamical regimes: a frozen, a marginal, and a chaotic phase, where an initial damage vanishes, remains limited, and grows to a finite fraction of the lattice sites, respectively. The results presented are also consistent with the observed biological scaling laws for the number of differentiated cells and cell length cycles as a function of the number of genes in an organism. [S1063-651X(97)03303-5]

PACS number(s): 87.10.+e, 0.5.50.+q, 64.60.Cn

I. INTRODUCTION

It is the activity of specific structural genes, which are controlled by associated regulatory genes, that generates the vital phenomena of proliferation, differentiation, development, and persistence of spatial and functional ordered patterns in the life span of each organism. Thus one of the most challenging problems in modern biology is the understanding of the complex set of biochemical interactions responsible for differences in the rate of synthesis of various proteins by differentiated cells in eukaryotes, the so-called gene control mechanism [1].

This mechanism regulates the repertoire of different types of mRNA molecules, the abundance of each mRNA, and the number of times each mRNA is used before it is destroyed. Consequently, it determines the kind and the amount of enzymes and structural proteins, the products of gene activity, contained in each cell. In this lies the essence of cell differentiation: it is the protein content of one cell that makes it different from another one. The gene control mechanism regulates the concentration or abundance of proteins over a range from one to two orders of magnitude and very few genes are subjected to absolute ''on-off'' regulation. Therefore, although cell specific proteins can be possible among those that determine the cell specific character, the major differences among cell types are in the regulatory genes, as proposed by Wilson [2] and King and Wilson [3], who found few discrepancies in the structural proteins of chimpanzee and human species.

At the transcriptional level, the most frequent process of gene control, the changes in the rate of synthesis of a particular protein are regulated, primarily, by increasing or decreasing the synthesis of a primary RNA transcript in the cell nucleus. A simplified model for the transcriptional control [4] shown in Fig. 1 can essentially model genes transcribed by RNA polymerase I, II, and III. At any instant each gene is either active or inactive for transcription, with or without a transcription complex bound to it, respectively. Finally, the target of these regulatory molecules can be either a single or a set, of perhaps ten or more, genes.

One of the first models for genetic regulation and cell differentiation has been proposed by Kauffman [6]. It consists of N binary genes, all of them connected to K other fixed genes or inputs randomly chosen among the remaining ones. The state σ_i of each gene i, $i=1,2,\ldots,N$, is updated by a transition function f_i that represents the gene interactions via the proteins, chosen at random among all Boolean functions with K variables.

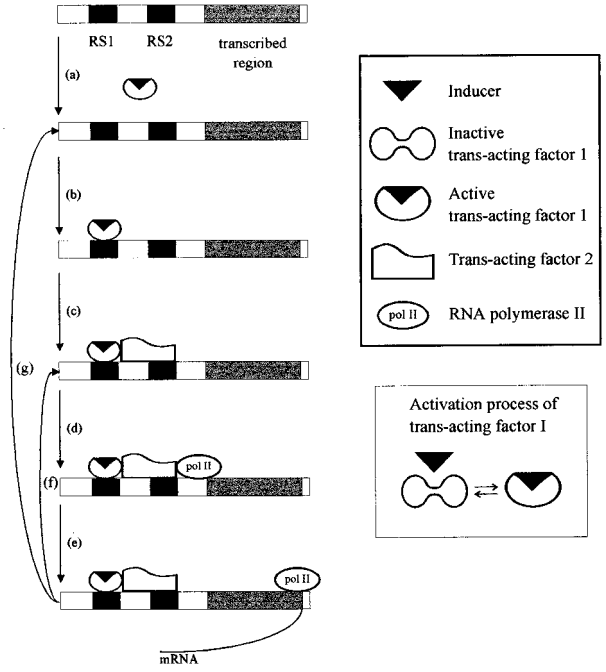


FIG. 1. Simplified model of gene induction. The steps are the following: (a) the inactive transactive factor 1 is activated by binding an inducer molecule; (b) the active transactive factor 1 binds to a regulatory sequence RS_1; (c) a second transactive factor 2 binds, perhaps facilitated by protein-protein interactions with the bounded transactive factor 1, to the regulatory sequence RS_2; (d) RNA polymerase recognize this transcription complex (transactive factors bounded to enhancer-promoter regions); (e) RNA synthesis initiates; (f) while the transcription complex remains, RNA polymerase repeatedly recognizes it; and (d) RNA synthesis continues (e); (g) the dissociation of transcription complex from the enhancer-promoter sequences initializes the whole process. From Ref. [5].

The dynamical properties of these random Boolean networks have been studied by several authors [7–12] and, in brief, the main results follow: for networks with connectivity $K>2$ the attractors are chaotic with low stability to minimal perturbations, cycle lengths increase exponentially with N (undesirable aspects in biological modeling), and the average number of alternative cycles is proportional to N. In contrast, nets with $K=2$ show striking spontaneous order. The expected length of state cycles is only $N^{1/2}$, the number of alternative attractors is also $N^{1/2}$, and each cycle is stable to almost all minimal perturbations. All of these results are discussed in Ref. [13].

Recently, Bastolla and Parisi have investigated analytically [14] and by computer simulations [15] the distribution of cycle lengths, average number of attractors, and distribution of attraction basins for the Kauffman model. They found that all systems on the critical line exhibit the same behavior—cycle lengths and number of attractors increase as $N^{1/2}$ in an annealed approximation—as the $K=2$ original Kauffman proposal.

In this paper we study, by numerical simulations and mean-field analytic calculations, the dynamical properties of a cellular automata (CA) model which incorporates both long-range interactions (as in the original Kauffman model) and short-range gene interactions (next and next-nearest couplings, as in the lattice version of Kauffman's model). In Sec. II the CA model is described. In Sec. III we present our simulation results, concerning average periods, number of different attractors, and stability against mutations. A mean-field study of this last property is also done. In Sec. IV we discuss our results in terms of the biological data currently available. Finally, we conclude in Sec. IV.

II. CELLULAR AUTOMATA MODEL

The activity of gene induction is not a single event occurring within the cell. On the contrary, multiple gene interactions are frequently observed, i.e., induction or repression of certain gene sets. The final effect or cellular response is obtained from all these interactions, initially triggered by a specific extracellular stimulus. This process reminds us of lattice models in which any ''site'' is capable of influence upon other ''contact sites,'' ''positively'' or ''negatively.''

In the CA model presented in this paper the genome is represented by a set of N binary genes σ_i, $i=1,2,\ldots,N$. When $\sigma_i=1$ the gene is active for transcription and the specific enzymes or structural proteins it codifies are produced. On the other hand, when $\sigma_i=0$ the gene is inactive and the products it codifies are not synthesized. The network state at a given time t is specified by the activity pattern $\sigma_1(t),\sigma_2(t),\ldots,\sigma_N(t)$. Each gene i is regulated by $K-1$ other genes and by itself, through a function of the previous state of its regulatory elements. In analogy with neural network models [16] the gene activity state at the next time step is given by

$$\sigma_i(t+1)=\mathrm{sgn}\left(J_{ii}\sigma_i(t)+\sum_{l=1}^{K-1}J_{ij_l(i)}\sigma_{j_l(i)}(t)\right), \qquad (1)$$

where $J_{ij_l(i)}$ is the coupling constant representing the regulatory action of the $j_l(i)$ $(l=1,2,\ldots,K-1)$ input on gene i and J_{ii} is the autogenic regulation. $\mathrm{sgn}(x)=0$ if $x\leqslant 0$ and $\mathrm{sgn}(x)=1$ if $x>0$. All the gene states are simultaneously updated. In order to accomplish this, a given gene evaluates the present stimulus from all its regulatory genes, including itself. If the overall stimulus it receives at time t is positive, the gene activates, or stays active if it was already active; otherwise it turns inactive or stays inactive.

The coupling constants J_{ij} model the extremely complex and partially unknown set of biochemical interactions briefly discussed in the Introduction. Our choice of the J_{ij}'s takes into account the following biological features. (i) The products of a determined gene can activate, inhibit, or not affect the transcription of another gene. In our model all the activatory interactions will assume the same value $+J$ and the inhibitory ones $-J$. When the gene j does not influence the expression of a different gene i, the coupling constant is $J_{ij}=0$, corresponding to a diluted bond. (ii) The gene interactions are asymmetric, i.e., $J_{ij}\neq J_{ji}$. The case in which a given gene i activates another gene j that, in turn, inhibits i, is biologically frequent. (iii) Autogenic or self-regulation gene control is frequent in living organisms. In the present CA model the self-control is provided by the J_{ii} coupling constants.

Since the molecular biologists have elucidated only partially the real connectivity matrix among genes, we have chosen a random distribution of nonsymmetrical J_{ij} (valid also for the self-interactions J_{ii}) described by

$$P(J_{ij})=\frac{(1-p_1)}{2}\left[\delta(J_{ij}-J)+\delta(J_{ij}+J)\right]+p_1\delta(J_{ij}), \qquad (2)$$

where $\delta(x)$ is Dirac's delta function and $J=1$. Therefore, for a particular gene network, each bond J_{ij} is activatory $(+1)$ or inhibitory (-1), with probability $(1-p_1)/2$, or diluted $(J=0)$ with probability p_1. As noticed by Weisbuch [17], a random interaction matrix is a positive choice once one is looking for generic properties independent in any critical manner on a particular interaction structure, which probably varies from one organism to the other.

Also, since almost all known regulated genes in prokaryotes and eukaryotes are directly controlled by up to six or ten gene products, our CA model involves $K=9$ regulatory inputs per gene, including itself. Of them $K-1$ inputs are either chosen at random among all the other remaining genes, with probability p_2, or are its neighbor genes with probability $1-p_2$. Thus the $p_2=1$ limit corresponds to an infinite-range model with connectivity $K=9$ (including the self-interaction J_{ii}), whereas the $p_2=0$ limit corresponds to a square lattice in which each site has a Moore neighborhood defined by its eight nearest and next-nearest neighbors. For any other p_2 values the simultaneous presence of short- and long-range couplings reflects the biological fact that a given gene can be regulated by either its nearest neighbors or distant DNA sequences, whose proteins, produced in the cytoplasm, diffuse towards the cell nucleus. Therefore, as Keller, Thomas, and Pohley [18] observed, spatial distances are frequently an irrelevant feature for functional biological networks. Once the K inputs of each gene and the corresponding interactions J_{ij} are chosen at the beginning, the CA

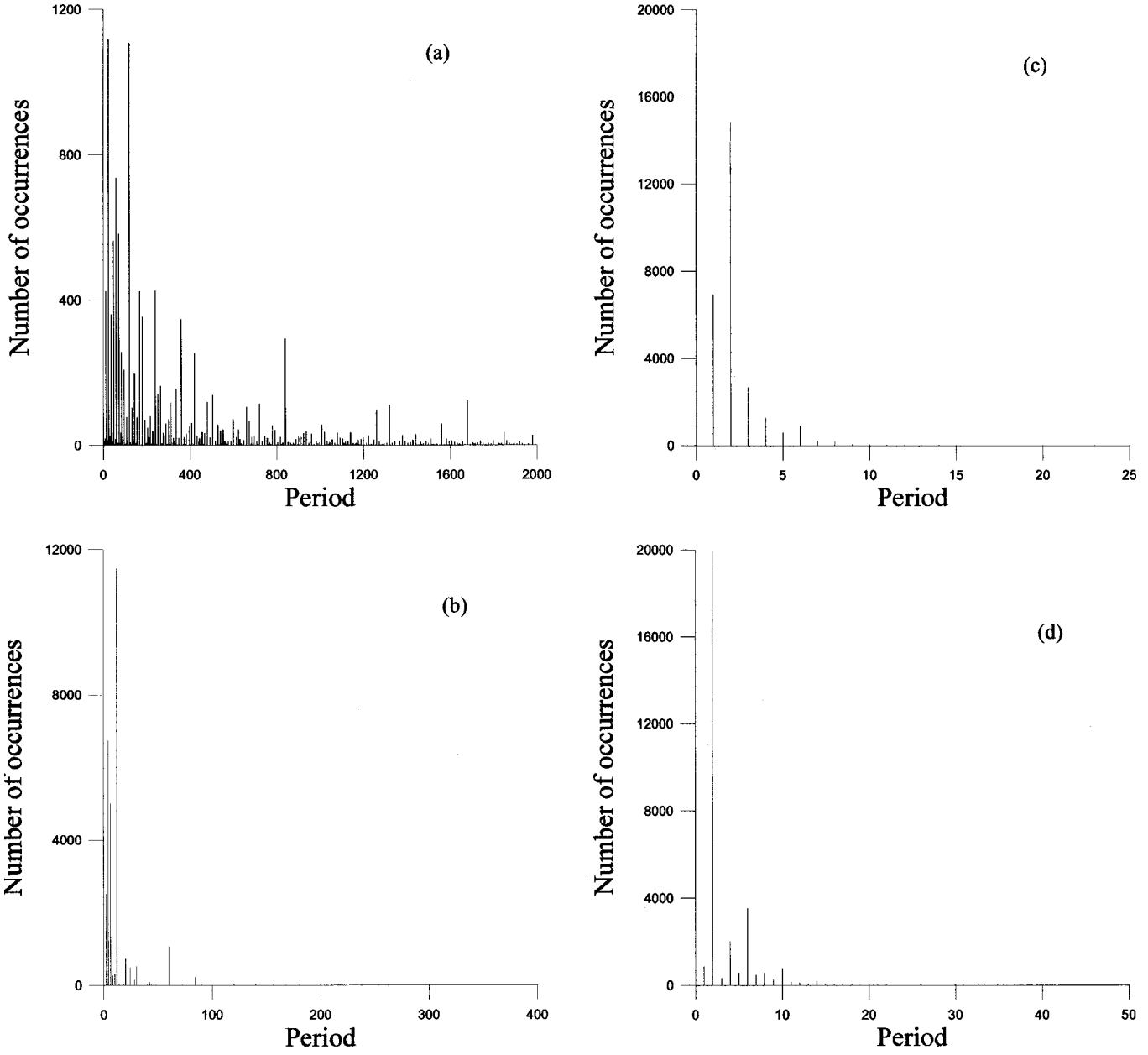


FIG. 2. Distribution of periods for (a) $p_2=p_1=0$ (local, nondiluted); (b) $p_1=0.80$, $p_2=0$ (diluted, local); (c) $p_1=0$, $p_2=0.70$ (nondiluted, nonlocal); and (d) $p_1=0.20$, $p_2=0.75$ (diluted, nonlocal) gene networks. The data correspond to simulations for 10 000 different nets with $N=400$ genes, each one tested with three random initial states.

structure is fixed forever. The quenched CA defined by Eq. (1), where there is a zero threshold, corresponds to a maximally disordered system.

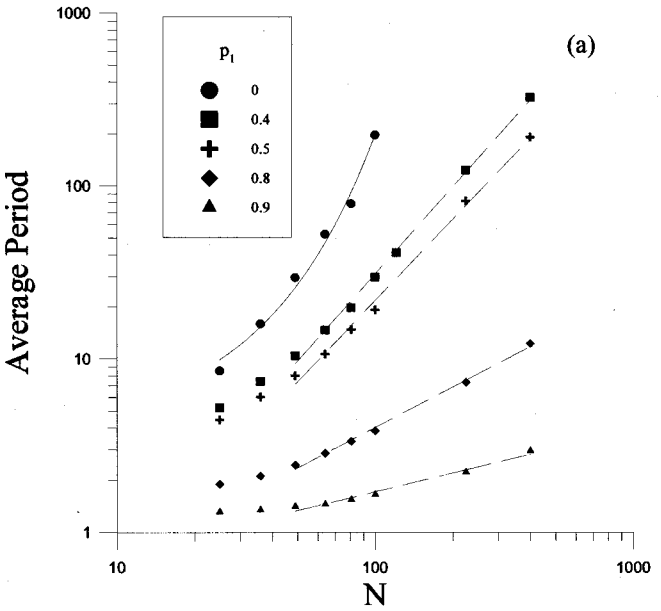
III. RESULTS

In the present CA model there are, except in the $p_2=1$ limit, geometrically correlated (nearest-neighbor) couplings which have made an analytical solution [16] unfeasible. Therefore the dynamical properties of the CA were studied mainly through computer simulations. In all the simulations, any initial states of the genes were equally probable and the largest genome size used was $N=625$ genes. In Sec. III D we compared the simulation results with analytic calculations on a simplified mean-field model.

A. Periods and number of attractors

Since the phase space of a finite CA contains only 2^N different configurations, its deterministic dynamics finally will drive the system towards an attractor, either a limit cycle or a fixed point.

Figure 2(a) shows the period distribution for the cycles in the $p_2=0$ limit, corresponding to a CA with only local (nearest and next-nearest neighbor) interactions. The presence of dilution, Fig. 2(b), affected the former period distribution by decreasing the frequencies of long period cycles. Also, the average period of a nondiluted net can be reduced, Fig. 2(c), if the probability p_2 of long-range interactions is increased. Therefore the combined effects of dilution and long-range couplings, shown in Fig. 2(d), could be important to design gene nets with limit cycles of low average period.



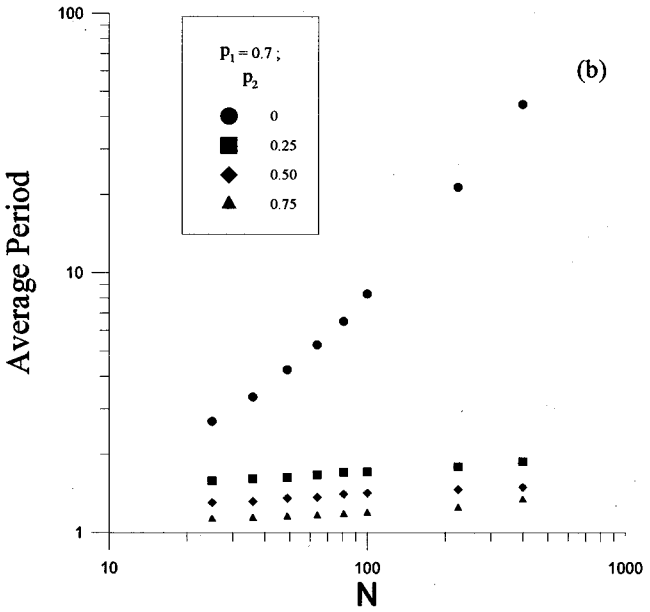


FIG. 3. (a) Average period of attractors as a function of N for local ($p_2=0$) gene networks. A transition between an exponential and a power law scaling of the periods with N occurs near $p_{1c}=0.40$. The solid (dashed) lines correspond to exponential (power) fittings. (b) The effect of long-range couplings, shown in a log-log plot, for $p_1=0.70$ and various p_2 values. The data correspond to 10 000 different nets, each one tested with five random initial states.

Figure 3(a) shows the average period of the attractors for the diluted, short-range CA model. It was found numerically that the average period increases exponentially with the number N of genes for $p_1<0.40$ and as a power law for $0.40\leqslant p_1\leqslant 1$. The critical value of $p_1\approx 0.40$ corresponds to an effective average connectivity $K_{\text{eff}}\approx 5.4$, greater than the critical value $K=4$ for the local model considered by Kurten [16]. The introduction of long-range couplings decreases the average periods as shown in Fig. 3(b). In the local ($p_2=0$) case the exponents range from 1.68, for $p_1=0.40$, to 0.39, for

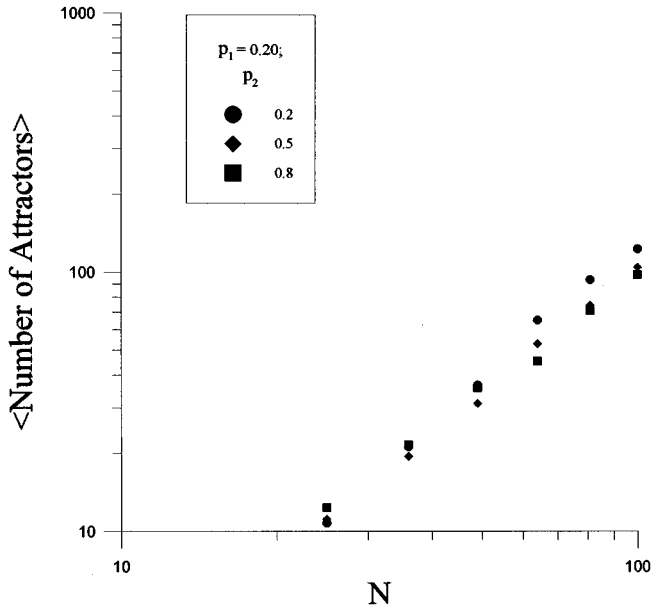


FIG. 4. A typical log-log plot of the number of different attractors as a function of the genome size N for diluted ($p_1=0.20$) and nonlocal ($p_2\neq 0$) gene networks. The exponents range from 1.80, for $p_2=0.20$, to 1.47 for $p_2=0.80$. The data correspond to 4.000 different nets, each one tested with ten random initial states.

$p_1=0.90$. However, the analysis of the average period as a function of the number N of genes is very difficult, especially in the critical region and in the low-diluted regime, where the periods are very long.

Another important quantity is the number of distinct cycles, which provides information about the structure of the attractors in the CA phase space. The numerical simulations carried out have revealed that this number increases as a power law of the number of genes in both regimes. A typical log-log plot of the number of cycles versus N is shown in Fig. 4. The exponents vary from 1.80, for $p_2=0.20$, to 1.47, for $p_2=0.80$, and fixed $p_1=0.20$.

B. Stability of the attractors

Even in the regime where the period of cycles grows exponentially as N increases, it does not mean that the flow on these attractors is divergent or exhibits extreme sensitivity to initial conditions.

In order to investigate the stability of the attractors the damage spread throughout the CA have been studied. To do this, first the automaton was simulated during a given transient. Then a replica of the system was made where an ''initial damage'' was created by flipping randomly a fraction p either of the genes or of the regulatory connections. As time evolves, the initial damage spreads through a damaged region where the genes in the two systems have different values. This damage is measured by a normalized Hamming distance Ψ defined by

$$\Psi(t)=\frac{1}{N}\sum_{i=1}^{N}|\sigma_i'(t)-\sigma_i(t)|, \tag{3}$$

i.e., the fraction of genes (σ_i') in the replica system that differ from their counterparts (σ_i) in the original system.

The CA is in its ordered phase if, for large N and sufficiently long time t, an arbitrarily small initial Hamming distance $\Psi(0)$ vanishes or does not grow. On the contrary, the CA is in its chaotic phase if Ψ approaches a finite non-null value for $\Psi(0)\rightarrow 0$. Consequently, small perturbations introduced in the CA initial state or connectivity structure grow into differences of a large size. Therefore in the chaotic phase the attractors become unstable and unable to return homeostatically, after small perturbations, to its previous limit cycle. This means, in summary, that the CA dynamical evolution exhibits extreme sensitivity to initial conditions.

Typical behaviors of the Hamming distance are shown in Fig. 5, for the diluted ($p_1 \neq 0$) and local ($p_2=0$) case. The complete dynamical phase diagram in parameter space (p_1, p_2) is shown in Fig. 6. To get the limit $\Psi(0)\rightarrow 0$ properly, in the numerical simulations the following trick [19] has been used: consider three initial states σ_1, σ_2, and σ_3 with $\Psi_{12}(0)=\Psi_{23}(0)=\frac{1}{2}\Psi_{13}(0)=s \ll 1$, where s is a fixed number. Here Ψ_{12} denotes the Hamming distance between the configurations σ_1 and σ_2, and so on. Then

$$\Psi(t)=\Psi_{12}(t)+\Psi_{23}(t)-\Psi_{13}(t) \tag{4}$$

is a reliable extrapolation to $\Psi(0)\rightarrow 0$. The phase diagram, shown in Fig. 6, presents a frozen and a chaotic phase for which a small initial damage vanishes or attains a finite value, respectively. In addition, there is a third phase localized in a narrow band between the chaotic and frozen regions, for which the final damage remains at the same size of the initial damage. *In this marginal phase the attractors are stable since a small initial damage neither vanishes nor grows.* The existence of this marginal phase, which, as will be seen in a next subsection, is due to the simultaneous presence of local and long-range interactions in the genomic networks, is the central result obtained by our simulations. In contrast, Derrida and Pomeau [8] using an annealed approximation have shown that, for an infinite system, the Hamming distance in the Kauffman model attains a finite nonzero value in the chaotic phase and zero elsewhere. Moreover, the annealed approach for the Hamming distance evolution is exact, in the limit of large systems, until times of the order of $\ln N$ [8,20].

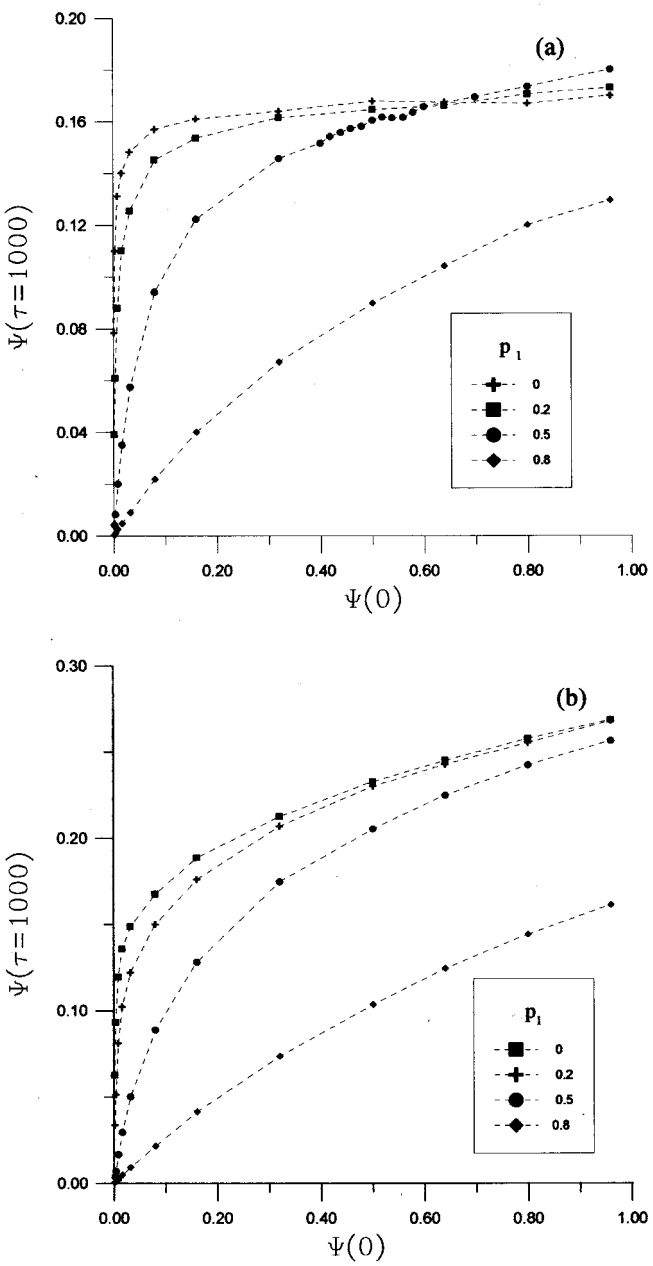


FIG. 5. Long time Hamming distance Ψ as a function of the initial damage $\Psi(0)$ for diluted ($p_1 \neq 0$) and local ($p_2=0$) CA. In (a) the initial damage was introduced in the gene activities and in (b) only the gene interactions were damaged. The data correspond to 2.000 different nets with $N=625$ genes, each one tested with five random initial states.

C. Percolation of frozen components and dynamical order

As is shown in Fig. 7, the CA dynamics crystallizes a subset of genes in fixed active or inactive states, which constitute a frozen component. The structure of this frozen core is determined by the control parameters (p_1, p_2). In the frozen phase a frozen core percolates through the gene net and isolates usually several subsets of genes oscillating in complex patterns. Since each oscillating subset has its own period, the global period cycle is the product of all such subset periods. The marginal phase corresponds to the boundary region where a frozen core begins to percolate and the subset of oscillating genes is just splitting in separated islands. This region marks the phase transition between order and chaos. In contrast, in the chaotic phase a frozen core does not percolate and the great majority of genes oscillate in complex cycles. Only small islands of frozen genes form and therefore a damage introduced in one site can propagate via its coupled elements to a finite fraction of genes in the network. In this phase the genome exhibits sensitivity to initial conditions. So, as has been noticed by Kauffman [13] and Weisbuch and Stauffer [12], the appearance of a percolating frozen core seems to be a sufficient condition for an ordered behavior in our CA model.

Figure 8 shows the typical behavior of the fraction of oscillating genes as a function of the genome size N. In the frozen phase this fraction decreases with N and appears to attain an asymptotic value in which only a few percent of genes oscillate. In contrast, the average fraction of oscillating genes increases with N in the chaotic phase, but again it is very difficult to analyze this regime since the periods are

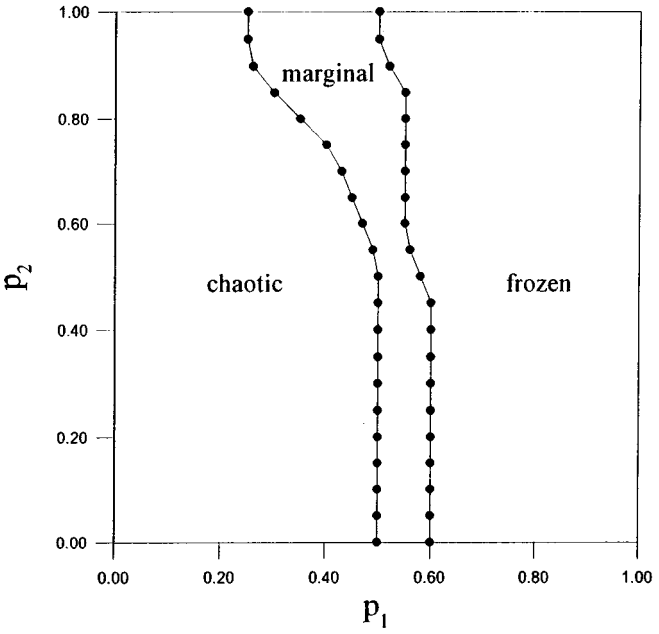


FIG. 6. Phase diagram in the unit square (p_1,p_2) considering the CA sensitivity to initial conditions. The system presents three phases: frozen, marginal, and chaotic, depending on whether the $t\to\infty$ Hamming distance Ψ vanishes, remains of the same size, or approaches an independent finite value for almost all $\Psi(0)\to 0$, respectively. The data correspond to 1.000 different nets with $N=400$ genes. Each net was tested with one random initial state.

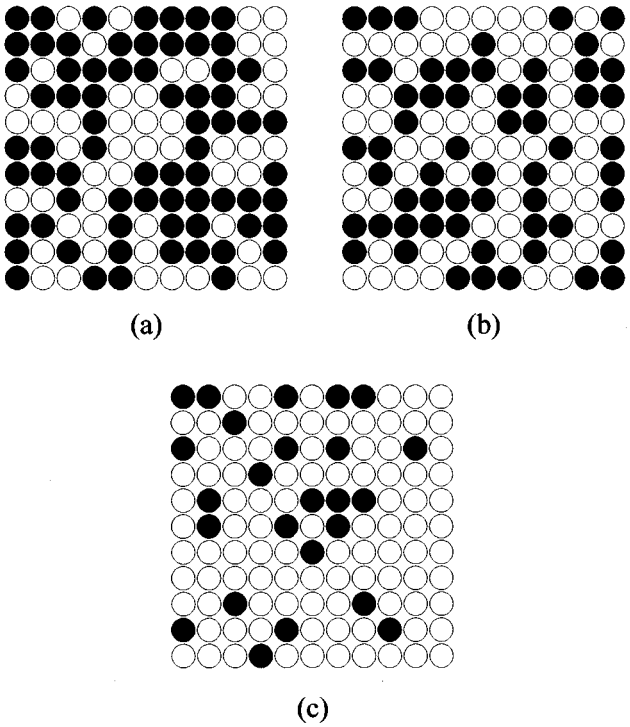


FIG. 7. Typical subsets of oscillating (●) and frozen (○) components in networks with $N=121$ genes in the (a) chaotic ($p_1=0.25$), (b) marginal ($p_1=0.50$), and (c) frozen ($p_1=0.75$) phases, with fixed $p_2=0.75$. In the frozen regime the oscillating genes form small isolated islands, contrary to the chaotic phase in which the majority of genes oscillate in complex cycles. The marginal regime corresponds to the boundary region where a frozen core begins to percolate.

very long. On the other hand, the fraction of oscillating genes is quasi-independent of N in the marginal phase and fluctuates around 20%, therefore predicting a core of frozen genes comprising, on average, 80% of them. Again this is a central result since Flyvbjerg [21] has shown that the frozen core in the infinite Kauffman model exhibits only two regimes. In the chaotic phase the fraction of frozen genes tends to a value less than 1, while in the frozen phase and on the critical line this fraction tends to 1. Therefore the ordered phase in the Kauffman model seems to be excessively rigid as compared with the marginal regime presently found.

Also, our simulations reveal that, with fixed p_2 (the probability of long-range couplings), the fraction of oscillating genes decreases with the increase of the dilution probability p_1, since the overall stimulus received by a gene typically decreases [see Eq. (1)]. Therefore the number of genes frozen in the inactive state grows. On the other hand, fixing p_1, the fraction of oscillating genes slightly decreases with the increase of p_2 in the nonchaotic phases. To understand this result we consider a given element at the border of an oscillating island. If p_2 increases, so does the chance of this element to receive an input from a distant gene, most probably from a frozen one since in these phases the majority of the gene activities are fixed. Consequently the chance of this gene to enter into a cycle decreases, and the size of nonfrozen islands and the fraction of oscillating genes also decrease.

D. A mean-field calculation

In this section analytical results of a mean-field version for the dynamics of the model introduced in Sec. II are presented. The main simplification refers to the connectivity structure of the CA. Here it is assumed that each gene presents connections with K other genes chosen at random from the total N, itself included. Then if K remains finite as N grows to infinity (as is the case we are interested in), it can be proved [22] that the dynamics of individual genes are not correlated and can be solved analytically. The calculations will be restricted to this case. The complete model, with long- as well as short-range interactions, in which the spatial structure of the neighborhood of the genes begins to be important, is very difficult to study analytically because correlations develop rapidly as time passes and the expressions for the dynamical evolution become intractable.

Concretely, we calculated the time evolution of the Hamming distance, Eq. (3), between two replicas of the CA, with an initial fraction of damaged genes in one of them. We were able to introduce the effects of random noise, or ''temperature'' in the dynamics that is defined by

$$P(\sigma_i)=\frac{1}{1+\exp(-2\sigma_i h_i/T)}, \tag{5}$$

with T being the temperature of the system or level of noise in the network dynamics and $h_i=\sum_{j=1}^{K}J_{ij}\sigma_j+h$ represents the effect on gene i from its K neighbors, h being an external input. Note that in the limit $T\to 0$ the dynamics Eq. (5) re-

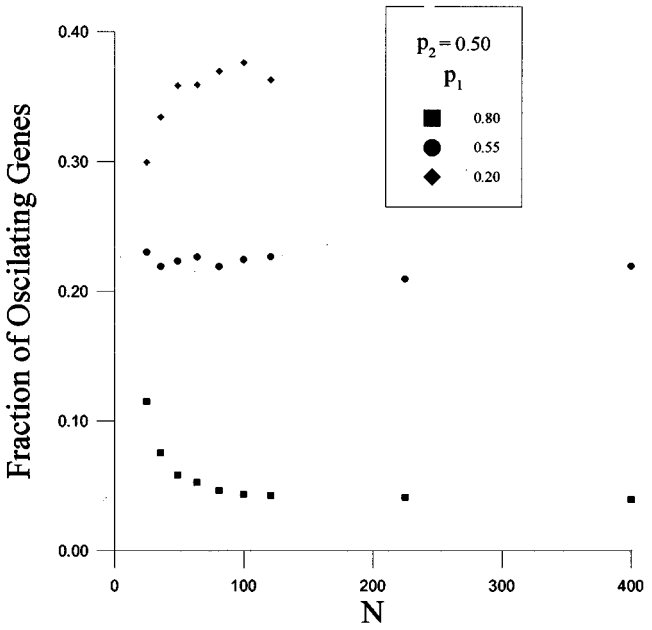


FIG. 8. Fraction of oscillating genes as a function of genome size N and fixed $p_2=0.50$. In the frozen ($p_1=0.80$) phase only a few percent of the genes oscillate, while, in contrast, the fraction of oscillating genes increases in the chaotic ($p_1=0.20$) phase. In the marginal ($p_1=0.55$) regime this fraction fluctuates around 20% of the genes and predicts a core of frozen genes comprising, on average, 80% of them.

duces to Eq. (1) with a nonzero threshold h. Following the lines of Ref. [23] the following map for the time evolution of $\Psi(t)$ was obtained:

$$\Psi(t+1)=\frac{1}{2^K}\sum_{p=0}^{K}\sum_{n=0}^{K-p}\sum_{m=0}^{p}[\Psi(t)]^p[1-\Psi(t)]^{K-p}\binom{K}{p}\times\binom{K-p}{n}\binom{p}{m}\left\{\frac{1}{1+e^{-2(X_n+|Y_m|+h)/T}}-\frac{1}{1+e^{-2(X_n-|Y_m|+h)/T}}\right\},\tag{6}$$

with $X_n=[2n-(K-p)]J$ and $Y_m=(2m-p)J$.

Before analyzing the solutions of this equation we note that there are four free parameters: the connectivity K, the noise level T, the external input h, and the mean strength of the interactions J. The effect of K and T will be analyzed fixing $J=1$ and $h=0$ throughout. The long time behavior of $\Psi(t)$ was studied for connectivities K ranging from 1 to 9 as a function of the temperature T. The main result is shown in Fig. 9(a): for each $K\geqslant 3$ there is a phase transition at a finite T from a chaotic phase where $\Psi(\infty)>0$ to an ordered or frozen phase with $\Psi(\infty)=0$. An important point to stress is that the asymptotic values of $\Psi(t)$ are *independent of the initial damage* $\Psi(0)$. This is in contrast with the results found in the simulations and suggests that spatial structure and short-range interactions may be relevant for the presence of the marginal phase in which the Hamming distance stays small for small initial damage or ''mutations.'' In Fig. 9(b) it can be seen, for a typical case with $K=5$, how the final

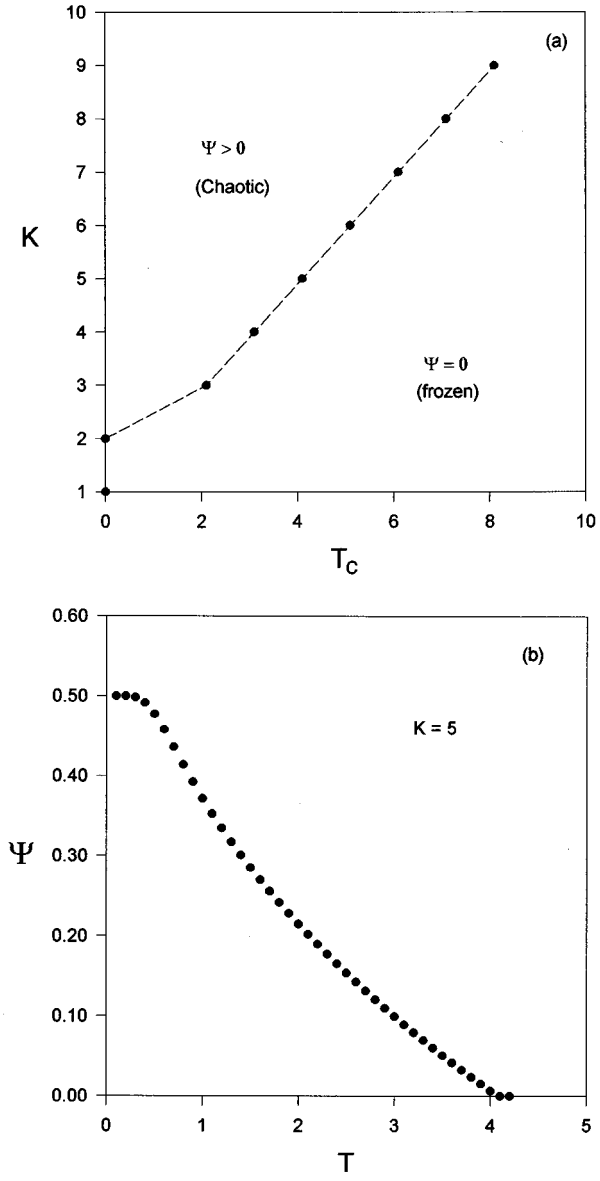


FIG. 9. (a) Mean-field phase diagram in which only the chaotic ($\Psi>0$) and frozen ($\Psi=0$) phases appear, independently of the initial damage. The parameters $J=1$ and $h=0$ are fixed. (b) Typical behavior of the order parameter Ψ as a function of the external noise T for a fixed connectivity $K=5$, characterizing a continuous phase transition. Again, $J=1$ and $h=0$ are fixed.

damage behaves with increasing T. It presents a continuous transition to $\Psi=0$ at a critical temperature T_c which increases with increasing K.

IV. DISCUSSION

In the context of cell differentiation Kauffman interprets the cell types as stable cyclic patterns of gene expression emerging from the interactions among genes and their products. Consequently, the total number of cyclic attractors represents the highly limited number of differentiated cell types in a living organism. Also, the periods of these attractors are related to the various differentiated cell cycles [24,25]. Bearing in mind this biological interpretation, our numerical

simulations are consistent with the observed data, since the following items are true.

(i) Always in the period distribution, shown in Fig. 2, the majority of cycles has short periods and a few examples have very long ones, as is the case for the biological observed mean cycle time for each level of genomic size (or complexity) [25].

(ii) Even in the local case the exponents controlling the power law increase of the average periods, ranging from 1.68 to 0.39, are consistent with the biological evidence that the average mitotic period increases either as the square root of the DNA content per cell or linearly with (perhaps slightly faster than) the number of transcribed genes across many phyla [25]. Also, these periods increase exponentially only above an effective connectivity of $K_{\mathrm{eff}} \approx 5.4$, close to the average number of regulatory sites per gene in a multicellular organism suggested as being, at least, five [26].

(iii) The observed number of cell types does not increase as the square or as an exponential function of the gene number in organisms across many phyla [25]. Again, in both CA regimes (ordered and chaotic), the number of different attractors increases as a power law of the genomic size, with exponents in the biologically expected range.

(iv) Finally, our CA exhibits a marginal regime in which the gene expression patterns are stable against almost all perturbations and, in addition, limited mutations are permitted. Therefore in this regime the Darwinian adaptive evolution that occurs by gradual accumulation of useful minor mutations is possible. Also, in the marginal phase the fraction of oscillating genes fluctuates around 20%, which means that on average 80% of the genes comprise a frozen core. This result is consistent with biological data, since it is observed that 70% or more of the genes are transcribed into heterogeneous nuclear RNA among all cell types of an organism.

In summary, our very simple model can exhibit some of the properties observed in living organisms. Moreover, it reinforces the hypothesis, raised by Langton [27] in the context of complex systems, dynamics, and computation, that life occurs in a marginal region at the edge of chaos, in which the cell types are stable and endowed with the necessary flexibility to allow mutations and, consequently, diversity and natural evolution, basic features of life.

V. CONCLUSION

In the present paper we have investigated numerically a simple CA model for gene control, which consists of binary genes interacting through asymmetric short- and long-range activatory, inhibitory, or diluted coupling constants chosen from a specified probability distribution. This model with a zero threshold corresponds to a maximally disordered CA version. We found that the CA parameter space is partitioned by attractors (limit cycles) which are either sensitive or not to the initial conditions. In the chaotic regime the average period of the attractors increases exponentially with the genome size, while in the ordered phase we observed a power law increase of their average period. The stability analysis of these attractors demonstrated that within the ordered phase the cycles are stable against damage spreading. Indeed, the ordered regime is divided into two regions: a frozen one, in which the final Hamming distance is always zero; and a marginal one, in which the final Hamming distance is of the same order of the initial one. The marginal region, localized between order and chaos, has stable attractors or cell types endowed with the necessary flexibility to allow mutations and therefore natural evolution, a basic feature of life. The existence of this marginal phase is the central result obtained by our simulations. In contrast, in the chaotic regime even a small initial damage spreads to a finite fraction of genes. Also, the marginal regime is a result of the simultaneous presence of local and long-range interactions in the genomic networks. This conclusion is supported by the absence of this marginal regime in a simplified model with only long-range interactions. The presence of noise in the simplified model did not destroy the basic phase transition between the chaotic and the frozen phases in the CA dynamics but affects the value of the critical connectivity for which a chaotic phase develops for each noise level.

Finally, the observed power laws in the ordered regime for both the average period cycles and number of distinct attractors are in the range suggested by the biological data concerning the cell cycle length and number of differentiated cells in an organism. However, more extensive simulations involving genome sizes of 10^3 or 10^4 genes and with nonzero thresholds are necessary in order to investigate if the marginal regime, mainly, persists even in the infinite size limit, since in Ref. [15] it is suggested that the number of cycles and their lengths increase very fast with random Boolean net size, which seriously prejudices the biological interpretation.

ACKNOWLEDGMENTS

A stimulating discussion with Marina L. Martins is acknowledged. The authors would also like to acknowledge partial support by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais—FAPEMIG—and Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq.

[1] J. Darnel, Jr., H. Lobish, and D. Baltimore, *Molecular Cell Biology* (Scientific American Books, New York, 1990).
[2] A. C. Wilson, Stadler Symp. **7**, 117 (1975).
[3] M. C. King and A. C. Wilson, Science **188**, 107 (1975).
[4] M. Ptashne and A. A. F. Gann, Nature (London) **346**, 329 (1990).
[5] M. S. H. Ko, J. Theor. Biol. **153**, 181 (1991).
[6] S. A. Kauffman, J. Theor. Biol. **22**, 437 (1969).
[7] B. Derrida and H. Flyvbjerg, J. Phys. A **19**, L1003 (1986).
[8] B. Derrida and Y. Pomeau, Biophys. Lett. **1**, 45 (1986).
[9] B. Derrida and H. Flyvbjerg, J. Phys. (Paris) **48**, 971 (1987).
[10] B. Derrida and H. Flyvbjerg, J. Phys. A **20**, L1107 (1987).
[11] S. A. Kauffman, Physica D **10**, 145 (1984).
[12] G. Weisbuch and D. Stauffer, J. Phys. (France) **48**, 11 (1987).
[13] S. A. Kauffman, Physica D **42**, 135 (1990).
[14] U. Bastolla and G. Parisi, Physica D **98**, 1 (1996).

[15] U. Bastolla and G. Parisi (unpublished).
[16] K. E. Kurten, J. Phys. (France) **59**, 2313 (1989).
[17] G. Weisbuch, J. Theor. Biol. **143**, 507 (1990).
[18] U. Keller, R. Thomas, and H. J. Pohley, J. Stat. Phys. **52**, 1129 (1988).
[19] H. E. Stanley, D. Stauffer, J. Kertész, and H. J. Herrmann, Phys. Rev. Lett. **59**, 2326 (1987).
[20] H. J. Hilhorst and M. Nijmajer, J. Phys. (France) **48**, 185 (1987).
[21] H. Flyvbjerg, J. Phys. A **21**, L955 (1988).
[22] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).
[23] B. Derrida, J. Phys. A **20**, L721 (1987).
[24] S. A. Kauffman, Sci. Am. **64** (August) (1991).
[25] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press, New York, 1993).
[26] E. H. Davidson, Development **108**, 365 (1990).
[27] C. G. Langton, Physica D **42**, 12 (1990).