Energetics in a model of prebiotic evolution

B. F. Intoy and J. W. Halley*
School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA
(Received 26 July 2017; revised manuscript received 11 October 2017; published 11 December 2017)

Previously we reported [A. Wynveen *et al.*, Phys. Rev. E **89**, 022725 (2014)] that requiring that the systems regarded as lifelike be out of chemical equilibrium in a model of abstracted polymers undergoing ligation and scission first introduced by Kauffman [S. A. Kauffman, *The Origins of Order* (Oxford University Press, New York, 1993), Chap. 7] implied that lifelike systems were most probable when the reaction network was sparse. The model was entirely statistical and took no account of the bond energies or other energetic constraints. Here we report results of an extension of the model to include effects of a finite bonding energy in the model. We studied two conditions: (1) A food set is continuously replenished and the total polymer population is constrained but the system is otherwise isolated and (2) in addition to the constraints in (1) the system is in contact with a finite-temperature heat bath. In each case, detailed balance in the dynamics is guaranteed during the computations by continuous recomputation of a temperature [in case (1)] and of the chemical potential (in both cases) toward which the system is driven by the dynamics. In the isolated case, the probability of reaching a metastable nonequilibrium state in this model depends significantly on the composition of the food set, and the nonequilibrium states satisfying lifelike condition turn out to be at energies and particle numbers consistent with an equilibrium state at high negative temperature. As a function of the sparseness of the reaction network, the lifelike probability is nonmonotonic, as in our previous model, but the maximum probability occurs when the network is less sparse. In the case of contact with a thermal bath at a positive ambient temperature, we identify two types of metastable nonequilibrium states, termed locally and thermally alive, and locally dead and thermally alive, and evaluate their likelihood of appearance, finding maxima at an optimal temperature and an optimal degree of sparseness in the network. We use a Euclidean metric in the space of polymer populations to distinguish these states from one another and from fully equilibrated states. The metric can be used to characterize the degree and type of chemical equilibrium in observed systems, as we illustrate for the proteome of the ribosome.

DOI: 10.1103/PhysRevE.96.062402

I. INTRODUCTION

The central problem of prebiotic evolution arises because the simplest models assuming the formation of a genome at an early stage in the history of life encounter formidable statistical odds, suggesting that initiation of life starting with a genome is extremely unlikely [1] (Eigen's paradox). For that reason, models [1–4] such as the one considered here [5,6] have been formulated in which the initial events do not involve an information carrying genome. Instead, a system of polymers, possibly proteins or a collection of proteins and RNA, is postulated to form a metastable, autocatalytic chemical state which grows and evolves and incorporates an information carrying genome as an evolutionary adaptation at a later stage. Such a picture acquires some support from phylogenetic analysis of protein structures, which suggests that biologically relevant proteins may have been present on earth hundreds of millions of years before the appearance of the ribosome [7].

Using an adaptation of such a model due to Kauffman [3], we recently showed numerically [5,6] that, under certain conditions, one was more likely to find nonequilibrium dynamic steady states in such a Kauffman-like model if the reaction network was sparse. In the model by Kauffman [3] and here, a system of polymers interacting chemically by ligation and scission is simulated stochastically. Starting such simulations from a population of short polymers (the "food set", here dimers and monomers whose population is maintained constant during the simulation), one explores the distribution of steady states achieved at long times. Sorting these final states according to whether they have properties deemed to be lifelike, one then obtains estimates of the likelihood that systems with those properties will emerge. Our work in Refs. [5,6] was distinguished from that of [3] by exploring the likelihood that the resulting final states will be out of chemical equilibrium, whereas the earlier work imposed only the requirement that the system of polymers grows to large populations of lengthy polymers. With this change, which may be regarded as the use of a more restricted definition of "lifelike", we found that the likelihood of such lifelike states was much reduced. In terms of the control parameter p of the model, defined to be the fraction of possible ligations and scissions which actually occur, we found that p was required to be very small (about 0.005) if a substantial likelihood of lifelike states was to be achieved. Small p describes sparse networks of reactions, thus the result gave lifelike systems only for sparse reaction networks. Both models grossly simplify the chemistry relative to real systems, reducing the description of the polymers to strings of digits representing monomers. However, within the model, our result in Ref. [5] suggests that life might be more likely to originate in desertlike conditions or in a very dilute gas such as might be found in the upper atmosphere of a planet, rather than in a pond or ocean trench where the reaction network is dense and many paths to chemical equilibrium are open. We explored a further extension of the model in Ref. [6] in which a spatial dimension was added, but in the present paper there

*Corresponding author: halle001@umn.edu

is no spatial dimension and the reactants are regarded as well mixed.

The model used in Ref. [5] was purely entropic and did not take any account of bond energies. In a sense, that neglect corresponded to assuming an infinite temperature, although that description is imprecise. It is unlikely that an abstracted model of the chemistry like that in Ref. [5] can capture all the relevant qualitative features of the problem of the emergence of lifelike properties in nonbiological chemical systems without including some account of the energetics. Therefore, we here report results from an extension of the model of Refs. [5,6], which includes bond energies. We include only the largest energy in the problem, namely, the bond formation energy associated with the bonds between monomers which are formed during ligation and broken during scission. (If our polymers were proteins, then the associated bond energy would be of the order of 0.1 eV.) We take the bond energies to be independent of monomer type, with the view that the differences between bond energies are usually smaller than the average bond energy so that we are taking account of the biggest energy in the problem and neglecting the rest. Of course, smaller energy scales associated, for example, with hydrogen bonding and folding are known to be enormously important in terrestrial biochemistry, so such effects should eventually be taken into account. In the implementation reported here, the bond energies only enter the computations in the dynamics, where they determine appropriate reaction rates consistent with detailed balance as described in the next section. As in the earlier work, the goal is to determine the nature of the resulting final steady states and to determine the likelihood that they will have various lifelike characteristics. We find here that when bond energies are introduced, several kinds of nonequilibrium states may emerge and we numerically estimate their likelihood.

The introduction of energy into the problem requires specification of the degree of energetic isolation which will be imposed on the system during the simulated dynamics. Our earlier model was already open with respect to polymer and monomer number, because we maintained a constant population of food (monomers and dimers) throughout the simulation and, for numerical reasons, limited the total number of polymers to a maximum. Once we introduce a bonding energy and start the system from a population of only food, one sees that we have started the system near its maximum energy and at a very low entropy, because the entropy is associated with the fact that there is a number of possible polymers which increases exponentially with polymer length. Bonding (ligation) results in lowering of the net bonding energy and scission raises it. At the same time, as bonds are introduced, the polymers grow longer and the entropy rises. As the population grows, more high-energy, low-entropy food is added to maintain a constant food supply until the system reaches an entropically steady state. If that state has maximized the entropy, then we deem it "dead". Otherwise, we call it lifelike, subdivide the lifelike systems into various categories, and determine their frequency of appearance, as discussed below and in [5,6]. Since we are interested in systems not in chemical equilibrium, the polymer systems of interest will not be characterized by a temperature, because temperature is only defined for systems in local thermal equilibrium. However, the rates of reactions (if they are allowed at all by the small p) will act to drive each system toward maximum entropy consistent with its coarse-grained description.

If the system is in thermal contact with a reservoir characterized by a positive temperature, then we assume that the rates of reaction have a dependence on the reservoir temperature given by the usual detailed balance condition as described in more detail in the next section. Under conditions in which such a reservoir does not exist or is sufficiently weakly coupled to the system, then we can also define a maximum entropy state toward which the reactions are driving it, provided that the total energy and polymer number are slowly varying relative to the reaction rate. We term these two conditions thermally connected and thermally isolated. The thermally connected case is intended to roughly model situations, as in aqueous solvent in an ocean trench, in which a thermal bath is strongly coupled to the system. We envision the thermally isolated case as a rough approximation for situations, such as those in the upper atmosphere of early earth, where an ambient temperature might not even be well defined and the system is weakly coupled energetically to its surroundings.

In the thermally connected cases we impose a finite positive temperature in the detailed balance condition, but we compute the chemical potential μ on the fly from the polymer number N, taking account of bond energies in the expression for N. We find that in this thermally connected case, we can define and study three kinds of nonequilibrium (deemed lifelike) steady states and measure their properties and likelihoods within the model as a function of p. If no account is taken of the nature of the collective dynamics in the metastable final state, then the probability of occurrence of nonequilibrium entropically steady-state systems mainly increases with decreasing ambient temperature (increasing β). However, when we select as lifelike only those systems which remain dynamically active by a criterion based on the time Fourier transform of a polymer-polymer time correlation function as defined earlier [5] we find that the probability of producing systems which are both dynamically and entropically lifelike is nonmonotonic in the temperature variable with peaks in the neighborhood of an ambient temperature equal to the bonding energy.

In the thermally isolated case, we determine, at each reaction step, an instantaneous value of the inverse temperature β and chemical potential μ from the instantaneous values of the energy and polymer number and use those to fix the detailed balance conditions for the next step. In those cases, we have often found that the isolated equilibrium (not reached in the states of interest) is characterized by a negative temperature.

In each case, we explored another lifelike property by selecting those systems generated by the model that were growing exponentially in population while remaining dynamically active and out of equilibrium entropically. Imposing that constraint with the others discussed above resulted in a diminution of the number of dynamical states which satisfied all the criteria. However, statistically significant numbers of such states appeared and further study of their nature is left for future work.

We have found in all cases that the likeliest nonequilibrium states are closer to, though separated from, the corresponding

isolated equilibrium state characterized by a negative temperature than from a thermal equilibrium state characterized by a positive temperature of the order of the bonding energy. The same analysis used here to measure the separation of the nonequilibrium states from the corresponding equilibria can also be used to characterize living biological systems of proteins as we illustrate for the ribosome. We discuss the possible usefulness of this method of analysis for studying nonequilibrium systems in extraterrestrial environments to determine if they have lifelike characteristics.

II. DESCRIPTION OF THE MODEL

As in Ref. [5] and elsewhere [3,4], artificial chemistries associated with abstracted polymers consisting of strings of binary digits undergoing scission and ligation are generated. The parameter p controls the probability that, in a given realization, any reaction possible involving polymers up to a maximum length $l_{\max}$ is included in the network. From the resulting chemical networks we select, as we did previously [5], those which are "viable", by which we mean that there is at least one reaction path from a "food set" of small polymers to at least one polymer of maximum length. The probability that a network is viable is then found as the ratio of the number of realizations of the network which are viable divided by the total number of realizations. As in Ref. [5] but differently from the model described by us in Ref. [6], we assume here that the system is well mixed and no effects of spatial diffusion are considered. (The study could be extended to include spatial variations of local temperature as well as population.)

The difference between the model used here and that used in the work reported in Ref. [5] is in the simulation of population dynamics in the generated networks. Previously, we selected fixed rates for each reaction from a uniform distribution between 0 and 1 (in units of the inverse time step) and left them invariant as we implemented the reaction model using the Gillespie algorithm [8] as described in more detail in Ref. [5]. Such a procedure took no account of any difference in energy between reactants and products. Instead, in the work reported here, to any polymer (binary string) of length l we attribute an energy $-(l-1)\Delta$, where Δ is a positive real number which is the bonding energy between two monomers. The total energy E of any population $\{n_m\}$ of polymers in which n_m is the number of polymers of type m is $E = -\sum_{L=1}^{l_{\max}}(L-1)N_L\Delta$. Here the $N_L = \sum_{m \text{ of length } L} n_m$ is the same set of macrovariables used in Refs. [5,6]. The total number of polymers N is $N = \sum_{L=1}^{l_{\max}} N_L$.

We first describe the method used to take account of the energy in the case that the system is isolated except for the addition or removal of polymers to the food set to keep its total population fixed and the removal of polymers to keep the total number below a fixed maximum value. We consider the configurational entropy associated with a coarse-grained prescription of the state given by the number of molecules N_L for each length L between $L = 1$ and $L = l_{\max}$. Using the fact that, in the model, there are 2^L possible molecules of length L, counting possible states for a given state specification $\{N_L\}$ is the same problem that occurs in the boson statistics problem [9] (though of course this is not to imply that this model has any quantum features). The result is

$$S/k_B = \sum_L (\ln[(2^L + N_L - 1)!] - \ln(N_L!) - \ln[(2^L - 1)!]). \tag{1}$$

For fixed E and N, it is a standard textbook exercise to write down the values $\overline{N_L}$ of the populations which maximize this entropy, giving

$$\overline{N_L} = \frac{2^L - 1}{\exp[-\beta(E,N)\mu(E,N) - \beta(E,N)\Delta(L-1)] - 1}, \tag{2}$$

in which the parameters $\beta(E,N)$ and $\mu(E,N)$ are determined from the total energy E and polymer number N by the implicit equations [with (2)]

$$E = -\sum_{L=1}^{l_{\max}} (L-1)\overline{N_L}\Delta \tag{3}$$

and

$$N = \sum_{L=1}^{l_{\max}} \overline{N_L} \tag{4}$$

and Stirling's approximation has been used. (In the similar case in Refs. [5,6] $\Delta = 0$ and the remaining equation for the chemical potential is trivial to solve.) In this model the equilibrium population $\overline{n_m}$ of any polymer of length L is $\overline{N_L}/2^L$.

Though we are interested in dynamical states which do not reach such an equilibrium state, the chemical dynamics will drive any reaction toward it. (In Kauffman-like models, such equilibrium states are not always achieved because many of the reactions have zero rate, so equilibrium distributions cannot always be reached.) To describe how we take account of this, we recall the master equation used earlier to describe our implementation of the polymer dynamics,

$$\begin{aligned} dn_l/dt = \sum_{l',m,e} \big[& v_{l,l',m,e}\left(-k_d n_l n_{l'} n_e + k_d^{-1} n_m n_e\right) \\ & + v_{m,l',l,e}\left(+k_d n_m n_{l'} n_e - k_d^{-1} n_l n_e\right)\big], \end{aligned} \tag{5}$$

where n_l is the number of polymers of species l, $v_{l,l',m,e}$ is proportional to the rate of the reaction $l + l' \xrightarrow{e} m$, e denotes the catalyst, l and l' denote the polymer species combined during ligation or produced during cleavage, and m denotes the product of ligation or the reactant during cleavage. (In terms of this description, kinetic blocking can occur because some of the rates $v_{l,l',m,e}$ and $v_{m,l',l,e}$ are set to zero when the network is formed.)

The parameter k_d was described in Refs. [5,6] as a rough proxy for temperature and was set to 1 in the reported simulations. From the form of Eq. (5) one sees that, in general, k_d parametrizes the difference between the forward and reverse reactions. For example, in the first line of (5) k_d^2 is the ratio of the rate of the reaction $l + l' \to m$ to the rate of the reaction $m \to l + l'$. Here we do not regard k_d as a parameter, but recompute it on the fly at each time step in the dynamical simulation so that the rates implied by the master equation would drive the system to equilibrium at the current values of E and N in the absence of kinetic blocking and if E and N were not changing in time. The latter condition is not trivially

satisfied in our simulations. Both E and N are changing as we start the simulations from a food set of small polymers, so the assumption that they are approximately fixed during many reaction steps is an adiabatic approximation which can be checked as discussed in Appendix A. Given those assumptions, our condition is related to the condition of detailed balance: If the system were to achieve equilibrium, the summand of the first line in Eq. (5), for example, would have the form

$$-k_d \overline{n_l}\,\overline{n_{l'}}\,\overline{n_e} + k_d^{-1}\overline{n_m}\,\overline{n_e}, \tag{6}$$

where the overline on the n's indicates equilibrium values and the standard detailed balance condition (which is sufficient but not always necessary for achievement of equilibrium in the absence of kinetic blocking) requires that this be zero, giving

$$k_d^2 = \overline{n_m}/\overline{n_l}\,\overline{n_{l'}}. \tag{7}$$

Within the framework just described, k_d is therefore a function of the current values of E and N through the solution of the implicit equations (2)–(4) for β and μ as well as of m, l, and l'. The dependence on m, l, and l' is simplified by the fact that the equilibrium populations depend only on polymer lengths. However, the implicit equations are not analytically soluble and we solve them numerically on the fly during the dynamical simulations.

In the thermally isolated case in which the dynamics are driven through Eqs. (3)–(5) and (7) by quantities depending on $\beta(E,N)$ and $\mu(E,N)$, there is only one equilibrium state of interest, namely, that which maximizes the entropy at the system's total energy and polymer number. In this isolated case of a thermally isolated system, a consequence of Eq. (7) is a strong dependence of the results of the dynamics on the detailed composition of the food set which was not present in our previous simulations. Previously, Refs. [5,6], and following [3], the boundary condition on the food set was set to be $N_1 + N_2 = N_f$. (In the single-site simulations reported [5], N_f was set to 500.) However, with the running value for k_d given by (7), if only the sum $N_1 + N_2 = N_f$ in the food set is fixed, then the number of dimers N_2 is rapidly depleted to zero so that there are only monomers in the food set (see Appendix D). Here we report simulations in which the ratio of N_1 to N_2, as well as their sum, is fixed at various values, to explore the dependence of the resulting systems on that effect.

In the thermally connected case, the formulation is the same, but, instead of treating the inverse temperature β as a Lagrange multiplier to be determined from E, we treat β as a (positive) parameter. This corresponds physically to the assumption that the network is in contact with a thermal bath which exchanges energy rapidly, on the reaction time scale, with the polymers in it. (The latter condition can be satisfied even if the total energy E is changing slowly, because the net interchange of energy with the ambient environment can be small even though many small interchanges of both signs are occurring. However, we do not need to assume that E is changing slowly in the second case.)

We have only one implicit equation (4) to solve for the chemical potential in the thermally connected case and it is a function of N and β. However, the equilibrium population depends on the bonding energy and the implicit equation must still be solved numerically for $\mu(\beta,N)$ at the current value of N and the fixed value of β. The results are strikingly different from the isolated case, because we find that, using the usual Kauffman initial condition of a population of food polymers, the assumption of thermal isolation leads in most cases to a dynamics driving the more isolated system not in contact with a thermal bath toward an equilibrium characterized by large negative temperatures (small negative β). (The relevant parameter is actually $\beta\Delta$. If we measure energies and temperatures in units of Δ, then the Δ disappears from the formulation, but we have kept it here.) In the thermally isolated and thermally connected cases, the numerical solution of the implicit equations is carried out on the fly using Newton's method as described in Appendix B.

When the dynamics are driving the system through Eqs. (2)–(7) toward a predetermined temperature in the connected case, there are two macrostates of interest in determining whether the system is behaving in a lifelike manner or not. The first of these occurs when the populations $\{N_L\}$ maximize the entropy in steady state at values of energy E and polymer number N in the steady state. The second state of interest occurs when populations $\{N_L\}$ maximize the entropy associated with the assigned ambient temperature $1/k_B\beta$ and polymer number N. If the system is found in the first state but not in the second, it means that it has attained a local thermal equilibrium different from that it would have if it were in equilibrium with the external thermal bath parametrized by β. We term this locally dead but thermally alive (LDTA). There are two relevant temperatures in that case: one associated with the local equilibrium and the other associated with the ambient bath. As noted earlier, in that case the local temperature can be negative and we find such cases. Once the system is in the equilibrium state associated with the externally applied temperature, the internally equilibrated state will be the same as the externally equilibrated one and we judge such a state both locally and thermally dead (LDTD).

States in which neither type of equilibrium has been reached are termed locally and thermally alive (LATA). In such states, three entropies can be calculated, namely, the (i) instantaneous entropy S computed by inserting the instantaneous values of the quantities $\{N_L\}$ into Eq. (1), (ii) the local equilibrium entropy $S_{\rm L}$ computed from the current N and E by solving (3) and (4) for $\beta\Delta$ and μ and inserting the result into (2) and the resulting values of $\{\overline{N_L}\}$ into (1), and (iii) the thermal equilibrium entropy $S_{\rm T}$ computed in the same way from the fixed β and N and the solution to (4). The relation $S < S_{\rm L}$ is always obeyed, but there are no general inequalities relating S to $S_{\rm T}$ or $S_{\rm T}$ to $S_{\rm L}$. In practice, because the temperature associated with $S_{\rm L}$ is usually large and negative, whereas the temperature associated with $S_{\rm T}$ is chosen to be positive and moderate (e.g., $\beta\Delta \approx 1$), it is frequently the case that $S_{\rm L} > S_{\rm T}$ and S may be larger than $S_{\rm T}$. Negative temperatures associated with $S_{\rm L}$ arise because, with the Kauffman-like initial conditions which we use in the dynamics, the system starts in a state of very high energy and very low entropy, associated with maximum entropy characterized by a negative temperature near infinity. Instead of using entropy ratios, we determine how close each simulated steady state is to each of these possible equilibria by computing the Euclidean distances R_L and R_T of the steady state from the corresponding equilibrium states in

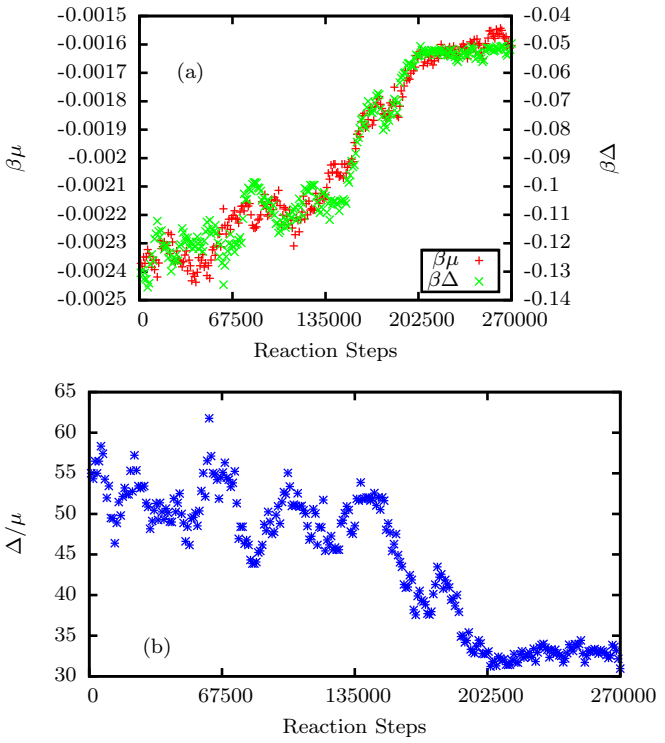

FIG. 1. Values of (a) $\beta\mu$ and $\beta\Delta$ as well as (b) the ratio Δ/μ as computed on the fly during a simulation of the model in which the dynamics were continuously driven toward equilibrium determined by the instantaneous energy E and polymer number N through Eq. (7). Note that the ratio N_1/N_f as well as N_f was fixed during this and similar runs.

the macrospace $\{N_L\}$ as described in the next section [Eqs. (8) and (9)].

To determine whether the states resulting from simulation using the dynamics algorithms we have described have lifelike properties we successively apply the following selection criteria to the simulation results. First, we require that the final state be an entropically steady state. For thermally isolated

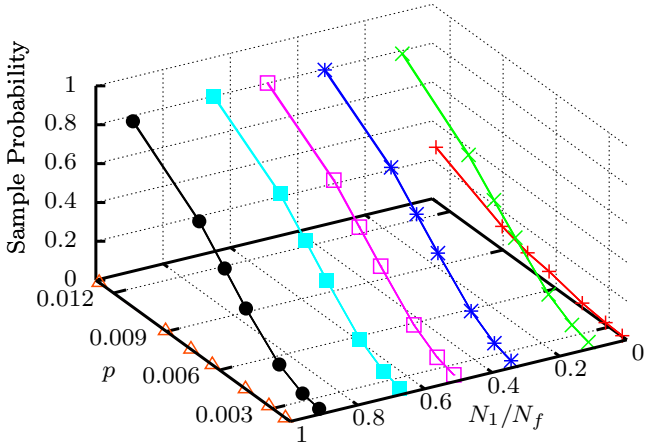

FIG. 2. Dependence of the likelihood of producing any state as a function of p and N_1/N_f. Except very near $N_1/N_f = 0$ and $N_1/N_f = 1$, this essentially maps the probability of having a viable network, with little dependence on N_1/N_f. The data are from at least 809 301 simulations. More details are in Appendix F.

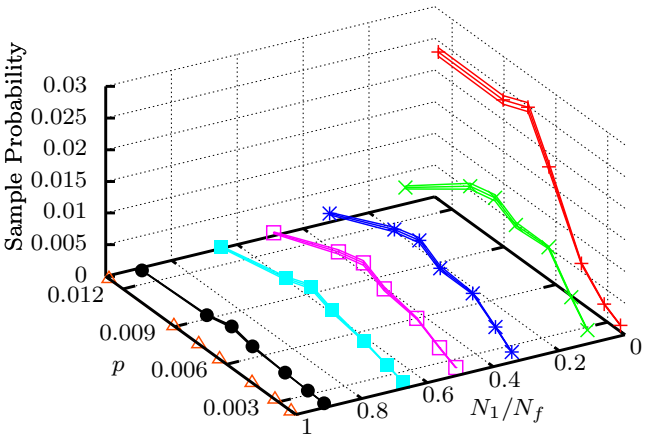

FIG. 3. Dependence on N_1/N_f and p of the likelihood of producing a nonequilibrium state in a thermally isolated system. Here we required that the steady-state entropy at the end of the run be less than 70% of the maximum possible entropy at the final E and N, given that the food population (N_1 and N_f) was fixed. The data are from the same set of simulations which were used to generate Fig. 2.

simulations, instantaneous entropy is computed from Eq. (1) using the instantaneous values of the quantities N_L and the history of that entropy over 10 000 time steps into the past is used to determine if the system is in a steady state as described in Ref. [6]. Second, we determine, for isolated systems, whether the ratio of the average entropy to the maximum entropy [determined by inserting (2) into (1)] is sufficiently below 1 to be deemed lifelike. For the systems simulated with an ambient positive temperature we determine how close they are to the two described equilibrium points by use of the normalized distances R_L and R_T [defined in Eqs. (8) and (9) in Sec. III below] in the macrospace $\{N_L\}$ and classify the state as LATA, LDTA, or LDTD. Third, in both thermally isolated and thermally connected simulations,

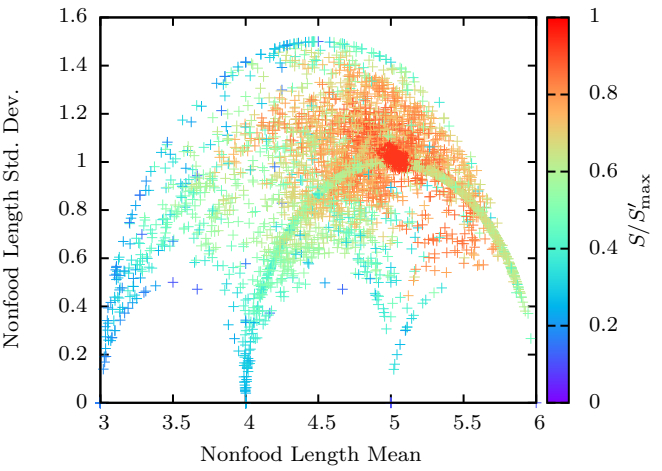

FIG. 4. Scatter plot of values of the mean and standard deviation of the nonfood polymer length produced from 14 753 thermally isolated simulations of the model. Here $p = 0.0064$ and $N_1/N_f = 0.33$.

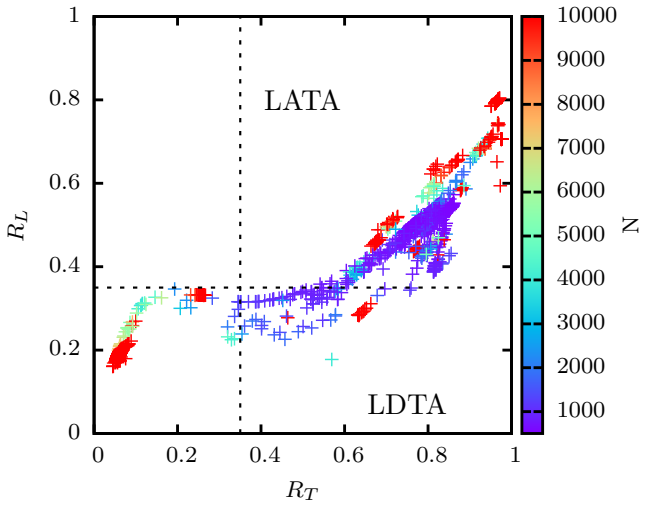

FIG. 5. Sample scatter plot of the distances defined in Eqs. (9) and (8) from the final state of runs at fixed positive ambient temperature. The colors indicate populations. Once sees several examples of the high population cases of LDTD (bottom left corner), LDTA (bottom right corner), and LATA (top right corner) with high populations. Here $\beta\Delta = 50$ corresponds to a relatively low temperature and it is likely that many states are not very dynamically active. (No dynamics cut, as discussed below, has been made). Here $p = 0.009\,05$ and $l_{\max} = 8$. The data are from 17 751 simulations.

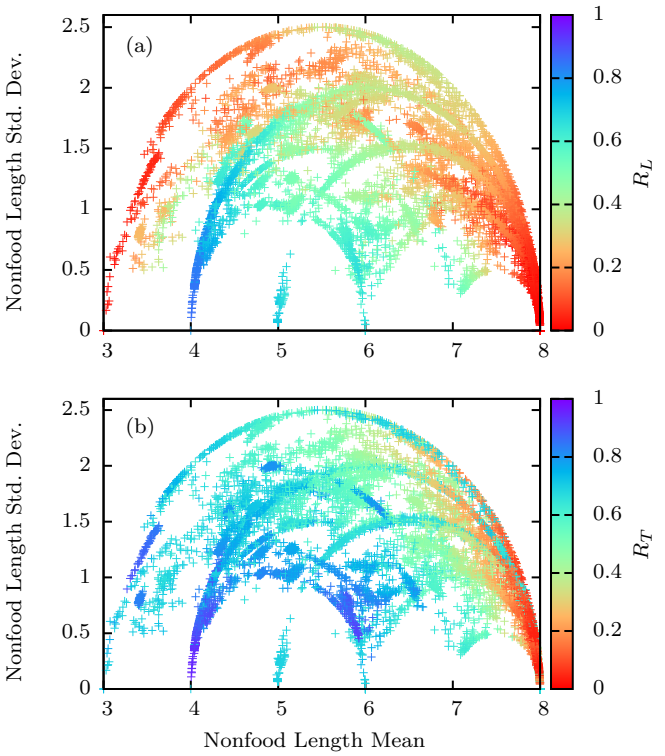

FIG. 6. Scatter plot of values of the mean and standard deviation of the nonfood polymer length produced from simulations driven toward equilibrium at fixed positive temperature for $p = 0.002\,26$, $\Delta\beta = 0.316$, and $l_{\max} = 8$. The colors indicate the values of the distances (a) R_L and (b) R_T. Points which are red in both plots are LDTD systems. Blue in both plots indicates LATA systems. Blue in (b) and red in (a) indicate LDTA systems. The data are from 9116 simulations.

we apply a dynamics criterion to determine whether the steady state is showing any chemical activity over the time scale of the steady-state period. The criterion is based on the time Fourier transform of a population correlation function as described in Ref. [5] and is reviewed in Appendix C. The criterion is parametrized by a frequency ω_m. Large ω_m means that the system is more dynamically active than small ω_m. The dynamics criterion turns out to be significant in excluding states associated with low ambient temperature, because many of those, though out of equilibrium, turn out to be dynamically quite inert and glasslike. Finally, we here introduce a fourth criterion, namely, that the population continues to grow, though the normalized distance of the state from the entropic equilibria is not changing in time as described in Appendix A. We report results of successively imposing these four criteria.

III. SIMULATION DETAILS AND RESULTS

A. Thermally isolated systems

To carry out meaningful simulations in the thermally isolated case described in the preceding section, in which the dynamics is driven toward the equilibrium associated with the current values of the total energy and polymer number, we find that, beginning, as we do, following the literature on these types of models, with a set of monomers and dimers (the food set), the very high initial energy leads, through Eqs. (2)–(4) and (7), to driving the system toward negative temperatures. Because the model's energy is bounded from above (by zero), negative temperatures are well known to be thermodynamically consistent in the system when it is thermally isolated [9]. One may think of this as arising because the thermal distribution of polymers requires an energy near its maximum. We illustrate in Fig. 1 with a display of the instantaneously determined values of $\beta\Delta$ in such a case. The details of the dynamics vary from run to run in such data but the temperatures are often negative. The chemical potentials are also negative but $\beta\Delta/\beta\mu$ is greater than 1 and often 10 or larger.

Note that in addition to fixing $N_f = 500$, we have here fixed N_1/N_f, where N_f is the number of polymers in the food set taken here to be the polymers of length 1 and 2. It turns out to be necessary to fix N_1/N_f in these thermally isolated runs because, if we do not, the dimers are almost immediately all dissociated into monomers and the likelihood of finding a kinetically trapped nonequilibrium state gets small (see also Appendix D). We illustrate this in Figs. 2 and 3. We see there that if we do not select for nonequilibrium states, then there is very little dependence on the ratio N_1/N_f. However, when we do select for nonequilibrium steady states, then there is a strong dependence, with the probability of finding a nonequilibrium steady state going essentially to zero as $N_1/N_f \to 1$.

We get some insight into the nature of the states produced from the scatter plot in Fig. 4, which shows the standard deviation in the polymer length versus the mean polymer length in the steady states arising from this model for a particular value of p. The "scars" in this scatter plot arise from states in which polymers of just two lengths are active, as shown in Appendix E.

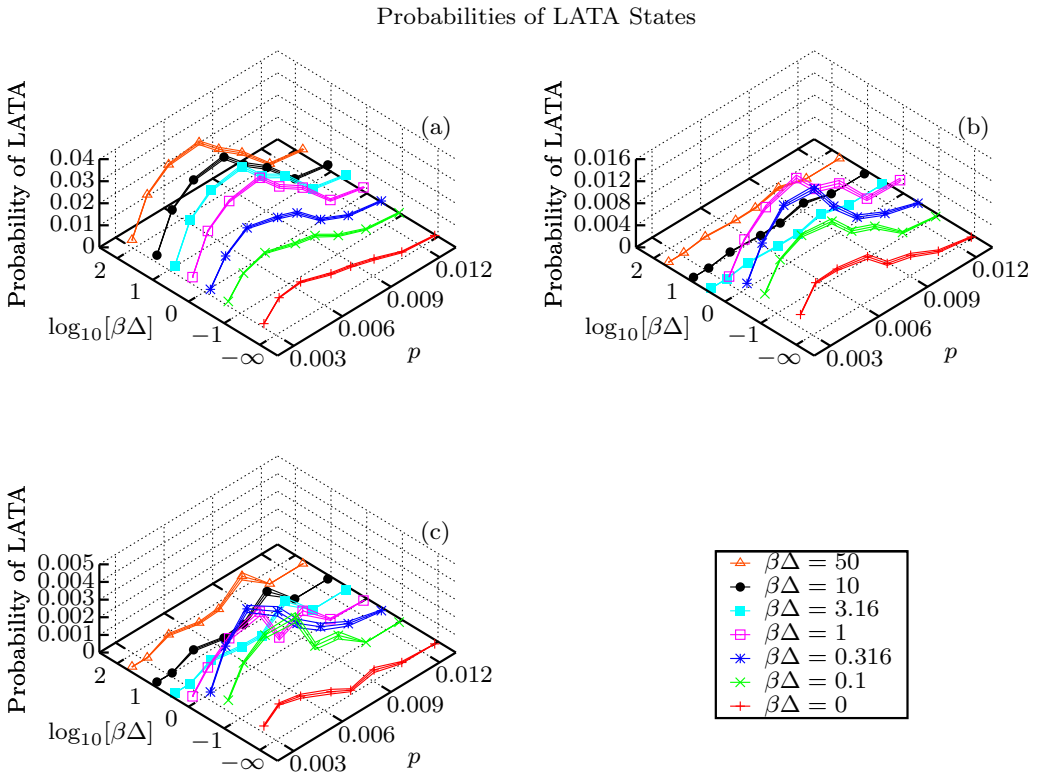

FIG. 7. Probabilities of producing an entropically steady state at the indicated values of the Kauffman parameter p determining the reaction network sparseness as defined in the text and the ambient inverse temperature $\beta\Delta$. Here $l_{\max} = 8$ and states were deemed LATA if R_L and R_T were both greater than 0.35. (a) No cut was made to exclude states with no dynamic activity or which were not growing exponentially in population. (b) Result of subjecting the states counted in the data of (a) to the dynamics cut described in Appendix C. Here we required that $\omega_m > 10(\frac{2\pi}{\delta t_{\text{Gillespie}}})$. (c) Result of requiring that the states counted in (b) also have exponentially growing populations. Here we required that the final states be exponentially growing at a rate of one inverse Gillespie time steps or greater (see Appendix A). All probabilities shown take account of the probability that an artificial chemical network generated by our algorithm as described in Ref. [5] is viable as defined there and in Sec. II.

B. Systems with dynamics driven toward a fixed positive temperature

As discussed in Sec. II, in this case there are two possible equilibrium points of interest. To determine whether the system has achieved either local or thermal equilibrium with the ambient thermal bath we compute the Euclidean distance of the final steady-state population distribution in the macrospace $\{N_L\}$ from the corresponding equilibrium points. Namely, we calculate

$$R_L = \sqrt{\sum_L (N_L - \overline{N_L(\beta(E,N),\mu(E,N))})^2}/\sqrt{2}N \tag{8}$$

for the distance from the locally equilibrated state and

$$R_T = \sqrt{\sum_L (N_L - \overline{N_L(\beta,\mu(\beta,N))})^2}/\sqrt{2}N \tag{9}$$

for the distance from the thermally equilibrated state in the case of simulation at a fixed ambient temperature. The numbers of polymers in the food set (here N_1 and N_2) are fixed and variations in their contributions to the entropy are small and not of interest.

We show scatter plots of values of R_L and R_T for a series of runs at fixed positive ambient temperature in Fig. 5. Scatter plots of standard deviation of the polymer length versus the mean polymer length for this positive temperature ensemble have a structure similar to that shown for the isolated case in Fig. 4. An example is shown in Fig. 6.

We have a large data bank of such systems. Their properties are very diverse and the scatter plots show structure corresponding to classes of states which have not been fully analyzed. However, we note in Fig. 6 that many of the nonequilibrium steady-state systems involve predominantly only two polymer lengths outside the food set as identified (Appendix E) by the scars in the plot. These may be less lifelike than the systems that do not lie on scars.

Here we focus on features which we deem of particular interest for study of prebiotic evolution, namely, dynamical behavior and population growth. We proceed with the analysis in a series of steps. As described in the preceding section, we first sort states into LATA, LDTA, or LDTD categories by introducing cutoffs on the values of R_L and R_T. The cutoffs are somewhat arbitrary, but the results are not qualitatively affected by the values chosen. In Figs. 7 and 8 we show probabilities of finding LATA and LDTA states as a function of p and $\beta\Delta$. Probabilities are calculated as the product of the probability that a network at the given p is viable (as defined in Sec. II and [5]) times the ratio of the number of times a final

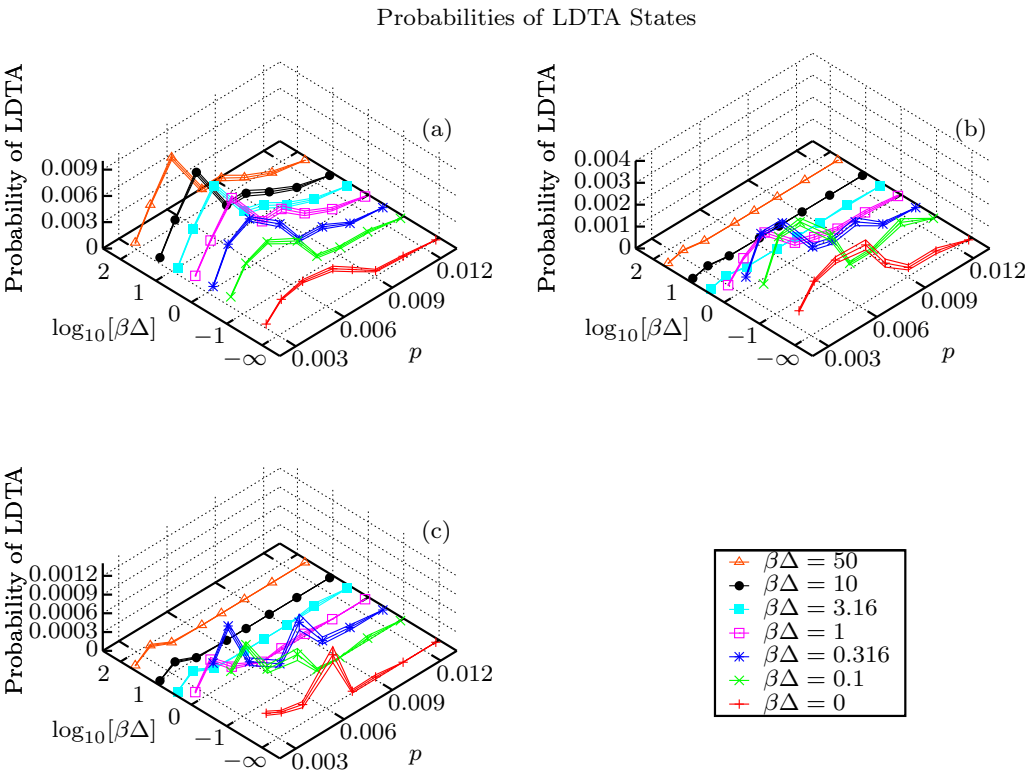

FIG. 8. Same as Fig. 7 except that $R_L < 0.35$ and $R_T > 0.35$ which we here term LDTA states. The data are from at least 1 101 877 simulations. More details are in Appendix F.

entropically steady state of the relevant (here LATA or LDTA) type appears divided by the total number of simulated viable systems. Probabilities of LDTD states are not shown, since they are regarded within our definitions as dead. Locally alive and thermally dead systems do not occur (as one can see, for example, in the scatter plot in Fig. 5) because no distinct local equilibrium can be defined once the system has equilibrated to an external thermal bath. We illustrate what is happening in these simulations with Fig. 9 showing the history of a run at an ambient inverse temperature of $\beta\Delta = 10$, which resulted in a state which passed all four criteria for lifelike.

IV. DISCUSSION AND CONCLUSIONS

We have considered two conditions in this study of a Kauffman-like model modified to take account of bonding energies. In the first condition, the system is energetically isolated (though it is open with respect to polymer number). In that first case, the chemical dynamics drive the system toward the equilibrium associated with its current values of total energy and polymer number. We find in that case, with simulations starting with only polymers in a food set of monomers and dimers, that the temperature equilibrium toward which the dynamics drive the system is negative and large. Further, if the composition of the food set is not controlled, the system contains an instability which prevents any evolution from occurring. Controlling the ratio of monomers to dimers in the food set however, we find evolution to lifelike states by our previous criteria that such states be out of chemical equilibrium, with the probability of occurrence of such states strongly dependent on the fixed value of the monomer to dimer ratio in the food set. Analysis of the nonequilibrium states realized in this case shows that many of them are found to be dynamical states in which polymers of only two lengths outside the food set are active (Fig. 4 and Appendix E). The internal effective temperature toward which the dynamics drive the system remains negative in these nonequilibrium steady states. Applying a dynamics cut to the ensemble of nonequilibrium steady states to select those with more than a minimal amount of dynamical activity reduces the number of surviving states by about an order of magnitude.

In the second condition, the dynamics drive the system toward equilibrium at a fixed positive temperature, modeling a system in energetic contact with a thermally equilibrated environment. In that case, two kinds of nonequilibrium steady states are found. In the first, the system is out of equilibrium both with respect to the temperature associated with its current values of the total energy and total polymer number and also with respect to the equilibrium associated with the fixed external temperature and the current polymer number. We term such states locally and thermally alive. We also find states which are in equilibrium with respect to their internal temperature (as determined from their total energy and polymer number) but out of equilibrium with respect to the externally fixed temperature and current polymer number. We term such states locally dead and thermally alive. The likelihood of LATA and LDTA states, as a function of p and the external inverse temperature $\beta\Delta$, was numerically estimated

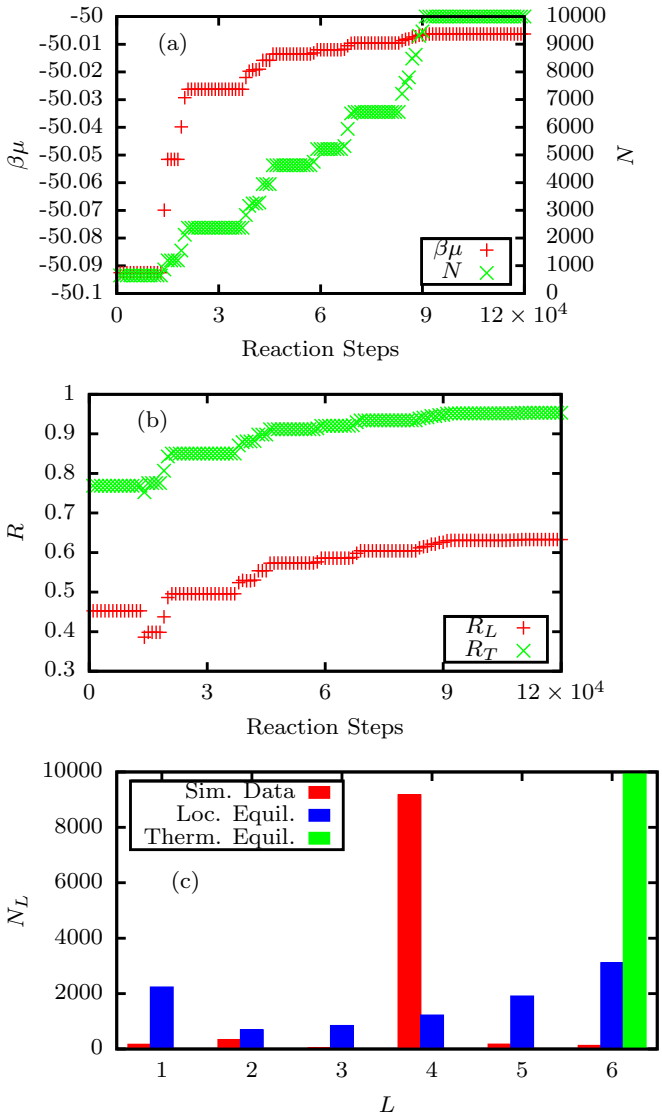

FIG. 9. History of a run which produced a lifelike state according to the criteria applied. (a) Computed value of the chemical potential $\beta\mu$ and the total polymer population N as a function of simulation time. (b) Values of R_L and R_T as a function of simulation time. (c) Macrospace polymer distribution for the final state (red), population distribution of a locally equilibrated state at the same energy and total polymer number (blue), and population distribution of a globally equilibrated state at the ambient temperature used in the dynamics algorithm (green). This was a single run with $l_{\max} = 6$, $p = 0.002\,26$, and $\beta\Delta = 10$.

as shown in Figs. 7 and 8. We found that the likelihood of both kinds of nonequilibrium rose with decreasing external temperature (increasing $\beta\Delta$).

However, when we applied the dynamics cut to exclude systems with little dynamical activity, the ensemble of surviving states was less likely to be lifelike at the lowest temperatures. Thus we find an optimum external temperature for the appearance of lifelike states which is of order $\beta\Delta \approx 1$. In this condition in which the system is driven toward equilibrium at a finite fixed positive temperature we again find that many of the nonequilibrium states contain polymers of just two lengths outside the food set (Fig. 6 and Appendix E). We applied a final cut (Appendix A) to this ensemble of nonequilibrium states to select those exhibiting exponential population growth. This final growth cut further reduces the number of surviving members of the ensemble. However, we find a significant number of such LATA states. The peak in probability as a function of temperature is weakened and the peak as a function of p is quite sharp [Fig. 7(c)]. Locally dead and thermally alive states are found to be much less probable than LATA states (Fig. 8). Overall, these results show that lifelike characteristics are most likely to emerge within the model in an ambient temperature bath when the sparseness of the network, parameterized by p, is small (around 1%, as in our previous work) and the ambient temperature is of the order of the bond energy.

The model provides some insight into the nature of nonequilibrium steady states which are most likely to be achieved in prebiotic evolution. In particular, it is noteworthy that in the thermally isolated case, the finite upper bound on the energy results in dynamics which drive the system toward a large negative temperature. We see in data like that in Fig. 9 that the polymer distributions in LATA states are in fact far from either the population distribution toward which the dynamics are driving the system (shown in green) or the distribution which would maximize the entropy associated with the total energy and polymer number (shown in blue and associated with a negative temperature).

The measures used here to evaluate the degree to which the simulated systems are lifelike can also be determined from data on real living, nonliving, and indeterminate polymer systems. In preliminary work, for example, we found the chemical potential and temperature associated with the isolated equilibrium of a system with the protein distribution found in the ribosome [10] in *E. coli*. We solved Eqs. (2)–(4), suitably modified to take account of the fact that 20 monomers, rather than 2, are available in the proteins. Using the observed fact that there are $N = 52$ proteins in the ribosome, the value of $-E/\Delta = 6561$ obtained from within the model from the observed population distribution, and the observed $l_{\max} = 393$, we found $\beta\mu = -4.09$ and $\beta\Delta \approx -3.00$. Evaluating the Euclidean distance R_L from the local equilibrium point gave $R_L = 0.01$ and an entropy value of $S = 19\,806$ compared to the locally equilibrated entropy of $S_L = 19\,918$. On the other hand, we also evaluated the chemical potential for the thermally equilibrated state associated with a temperature of 20 °C, assuming $\Delta = 0.1$ eV and the same value of N giving $\beta\mu = -2725$, $\beta\Delta = 3.958$, $R_T = 0.70$, and $S_T = 61\,067$. Thus, somewhat unexpectedly, the protein distribution in this ribosome is much closer to isolated chemical equilibrium associated with a high-energy initial state than to chemical equilibrium expected from equilibrium with an ambient temperature of the order to be expected at the surface of the earth. The three population distributions for the ribosome are shown (with a histogram bin size of 10) in Fig. 10. Further studies applying these measures to experimental and natural systems to evaluate their usefulness in determining the extent to which a system of unknown provenance (for example, on an exoplanet) might be regarded as lifelike are left for future work.

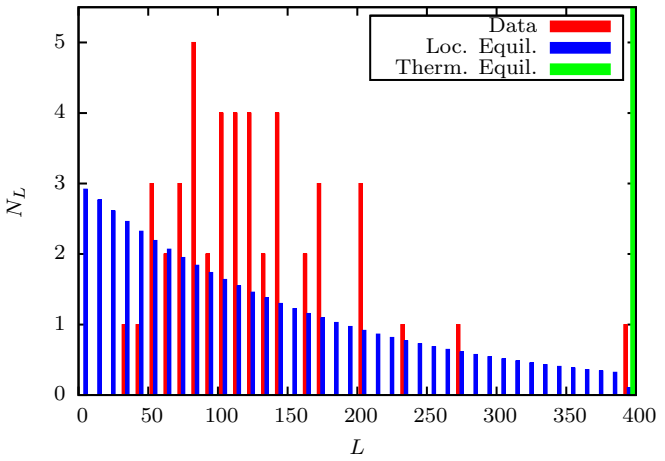

FIG. 10. Polymer length distribution for the proteins in the ribosome of *E. coli* (red) compared with the local equilibrium distribution associated with the energy of those proteins within the model (blue) and with the thermal distribution associated with the same number of polymers at an ambient temperature of the order of that to be expected at the surface of the earth (green). In the thermal distribution, all the 52 polymers are in the largest bin, so the number of monomers is much larger and the height of the green peak is not to scale.

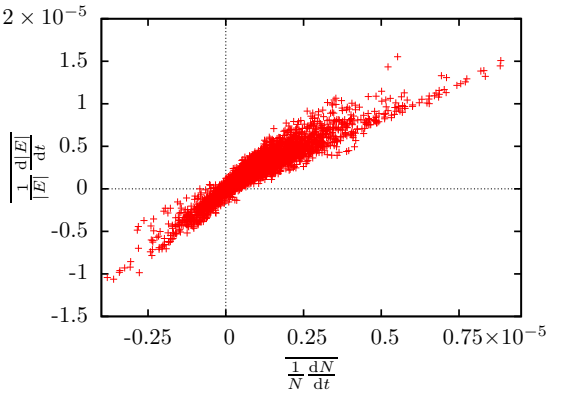

FIG. 11. Logarithmic reaction time derivative of the polymer number and the energy for isolated systems during the entire run with $p = 0.002\,26$ and $N_1/N_f = 0.33$. The data are for 7174 systems.

ACKNOWLEDGMENTS

This work was supported by the United States National Aeronautics and Space Administration (NASA) through Grant No. NNX14AQ05G and used the computational resources of the Minnesota Supercomputing Institute, the Open Science Grid, the University of Minnesota School of Physics and Astronomy Condor cluster, and the NASA Advanced Supercomputing division Pleiades supercomputer. We thank A. Wynveen for many helpful discussions and for reading and helpfully commenting on a preliminary version of the manuscript. We thank Professor Niels Fischer for assistance in locating the data on the ribosome which was used to produce Fig. 10.

APPENDIX A: TEST OF THE ADIABATIC APPROXIMATION AND SELECTION FOR EXPONENTIAL POPULATION GROWTH

The use of detailed balance as expressed in Eq. (7) can only be approximately correct in the isolated case if the rate of change of the total energy E and the total polymer number N is much slower than the rate of individual reactions. (The latter is of the order of the inverse of the time step.) In the case of contact with a thermal bath, the approximation is justified if just the rate of change of N is much slower than the reaction rate. To check if these conditions are satisfied we estimated the logarithmic time derivatives of the polymer number N and system energy E by sampling their values every 1000 simulation steps over the entire run (typically at least 10^5 simulation steps). We calculated $N(t)$ and $E(t)$ and from them the average of the discrete logarithmic time derivative of N and E,

$$\overline{(1/N)dN/dt} = (1/\mathcal{N})\sum_{\tau=1}^{\mathcal{N}} \frac{(N_{\tau+1} - N_\tau)}{(1/2)(N_{\tau+1} + N_\tau)}, \qquad \text{(A1)}$$

$$\overline{(1/E)dE/dt} = (1/\mathcal{N})\sum_{\tau=1}^{\mathcal{N}} \frac{(E_{\tau+1} - E_\tau)}{(1/2)(E_{\tau+1} + E_\tau)}, \qquad \text{(A2)}$$

in which τ labels data every 1000 reaction steps and $\mathcal{N}$ is the number of data points.

Figure 11 is a scatter plot showing the average logarithmic reaction time derivatives for 7174 thermally isolated systems. Note that the fastest change in polymer number or energy is of the order of 10^{-5}, giving a time scale of variation of both quantities of order 10^5 reaction steps, satisfying the adiabatic condition.

Figure 12 is a typical scatter plot of the logarithmic reaction time derivatives of the energy and polymer number for systems coupled to a positive temperature heat bath. As noted, only the logarithmic derivative of N is relevant in this case. It is also generally less than or of order 10^{-5} inverse reaction time steps, again giving a time scale of 10^5 reaction time steps, which is much greater than 1 as required. For both N and E there are large variances about the average values of the logarithmic derivatives with respect to reaction time, indicating that exponential growth is not a good model for these quantities as a function of reaction time. However, these

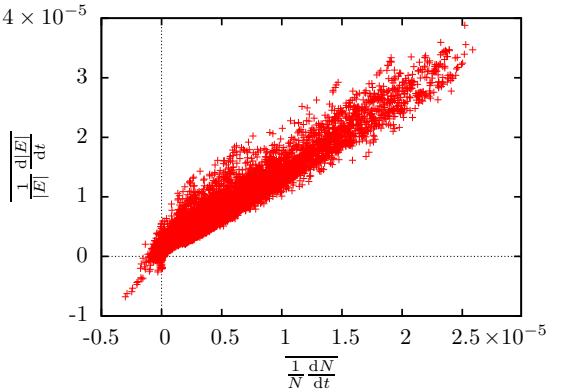

FIG. 12. Logarithmic derivative of the polymer number and the energy for systems coupled to a thermal bath with $p = 0.002\,26$ and $\beta\Delta = 3.16$. The data are for 9343 systems.

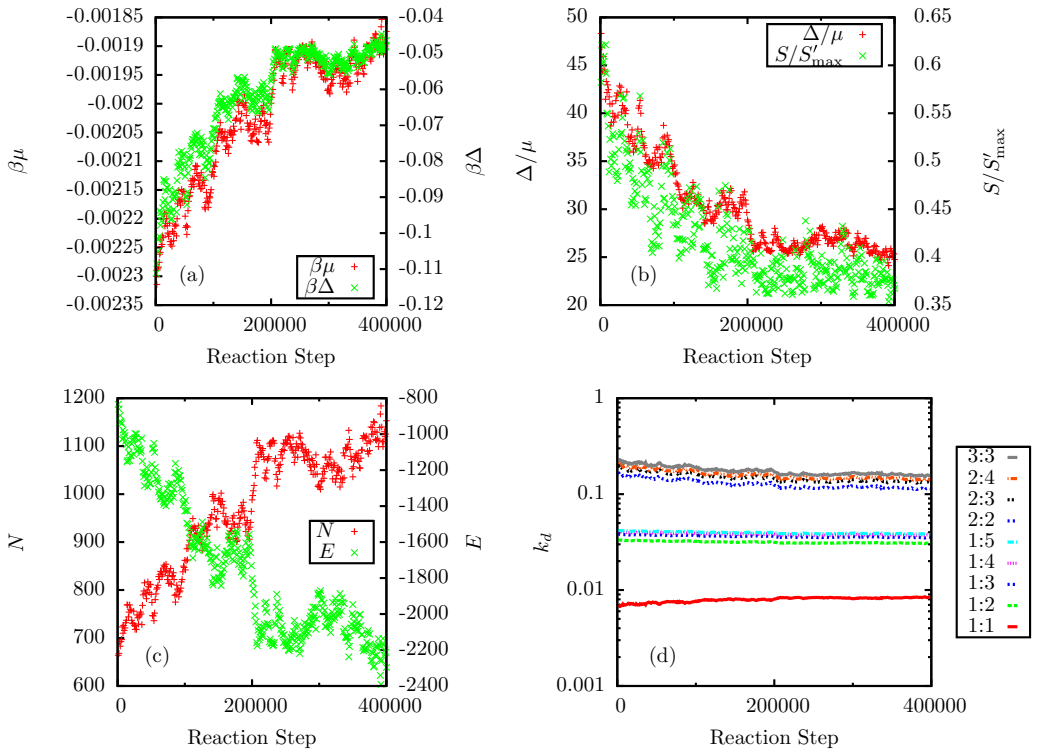

FIG. 13. Example of time variation of N, E, S, $\beta\Delta$, $\beta\mu$, and k_d during an isolated run ending in a nonequilibrium entropically steady state.

derivatives do give an estimate of the numbers of reactions required to change N and E, which is what is required for the adiabatic approximation.

Further evidence for the validity of the adiabatic approximation is provided by the actual computed values of k_d for which we supply an example in Fig. 13. Though details of the time dependence of E and N vary widely from run to run, we consistently find the type of slow variation of k_d shown there. Here $k_d(L:M)$ is the value of k_d used for ligation of polymers of lengths L and M.

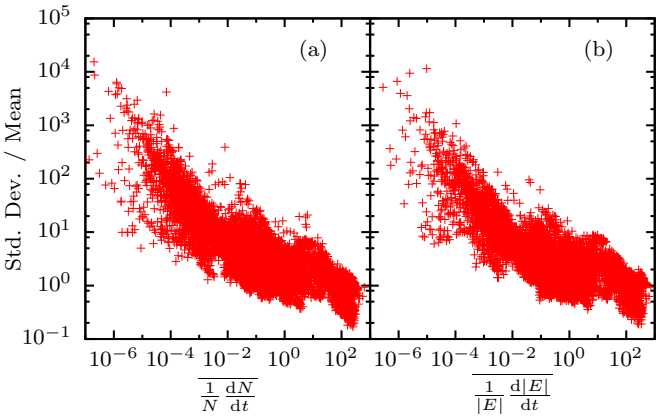

FIG. 14. (a) Average logarithmic Gillespie time derivative of the polymer number N and the ratio of the standard deviation of that quantity divided by its average for a series of runs at fixed ambient temperature given by $\beta\Delta = 0.1$, with $p = 0.002\,26$ and $l_{\max} = 8$. The data are for 9600 systems. (b) Same quantities for the energy E.

To test whether the systems generated are exponentially growing in time, the relevant measure of time is not reaction step time, as it is for evaluating the validity of the adiabatic approximation, but our approximation to real experimental time, which is given by the Gillespie algorithm as reviewed for this application in Ref. [5]. In Fig. 14 we show an example of the average values of the logarithmic derivatives of N and E calculated as before, but with the Gillespie time interval between each step in the denominator of the summand in Eqs. (A1) and (A2) for a run at ambient temperature given as $\beta\Delta = 3.16$. Characteristically we find variances which are large for small logarithmic derivatives, indicating that the log-derivative is measuring noise and not actual exponential growth, but much smaller variances at larger growth rates, indicating systematic exponential growth. In those cases in which exponential growth emerges from the noise, the difference between Gillespie time and reaction step time is usually large, with each unit of Gillespie time including many reactions steps. In the selection for exponentially growing systems (constraint four) which gives the frequencies of exponentially growing systems in Figs. 7(c) and 8(c) we counted systems with an average logarithmic N derivative with respect to Gillespie time greater than 1 as exponentially growing.

APPENDIX B: DYNAMICS ALGORITHM—COMPUTING μ AND β ON THE FLY

Before a simulation is started a coarse-grained table is made so that for given N and E values the values of μ and β can be quickly determined by linear interpolation from the values in the table. To construct the table, we use Newton's method to

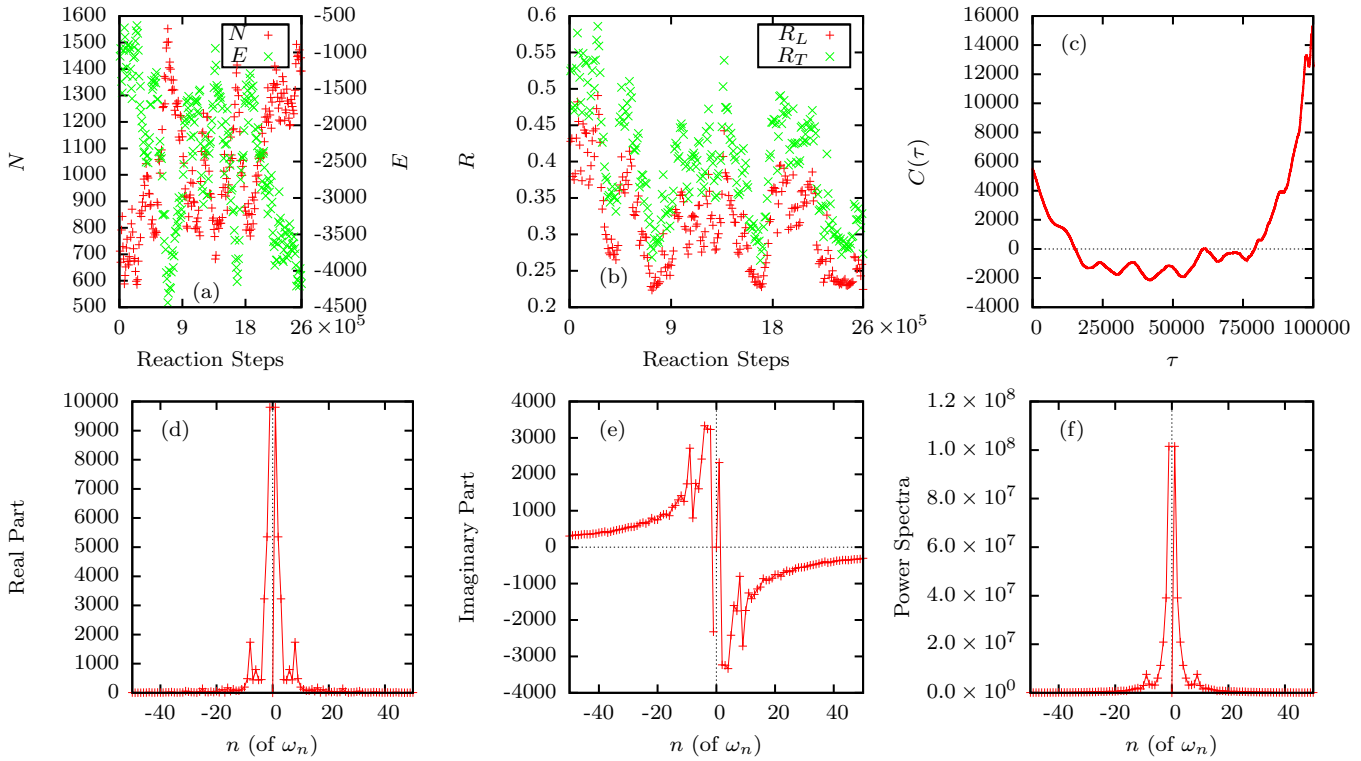

FIG. 15. Fourier transforms and the power spectrum of $C(\tau)$ of a run at fixed infinite ambient temperature ($\beta\Delta = 0$). (a) and (b) Values of N, E, R_L, and R_T over the entire history of the run, indicating that it is rather close to equilibrium. (c) Plot of $C(\tau)$ calculated for the last 10^5 steps of the run. (d)–(f) Real and imaginary parts and the power spectra of the Fourier transform of $C(\tau)$, with point weighting as described in the text, near the origin. Here $p = 0.002\,26$ and $l_{\max} = 6$.

solve Eqs. (3) and (4) as follows. Let $\vec{v}$ be the two-dimensional vector

$$\vec{v} = \begin{pmatrix} \beta\mu \\ \beta\Delta \end{pmatrix}. \tag{B1}$$

Let $N(\vec{v})$ and $E(\vec{v})$ be the values of the right-hand sides of (4) and (3) for that value of $\vec{v}$ and N and E be the values of the left-hand sides of (3) and (4). Define the two-dimensional vector $\vec{F}$ by

$$\vec{F} = \begin{pmatrix} N(\vec{v}) - N \\ E(\vec{v}) - E \end{pmatrix}. \tag{B2}$$

Consider a small variation $\delta\vec{v}$ from an initial guess for $\vec{v}$ and expand to first order about $\vec{v}$, giving

$$\vec{F}(\vec{v} + \delta\vec{v}) = \mathbf{M} \cdot \delta\vec{v} + \vec{F}(\vec{v}), \tag{B3}$$

where the matrix $\mathbf{M}$ is the Jacobian

$$\mathbf{M} = \begin{pmatrix} \frac{\partial N}{\partial \beta\mu} & \frac{\partial N}{\partial \beta\Delta} \\ \frac{\partial E}{\partial \beta\mu} & \frac{\partial E}{\partial \beta\Delta} \end{pmatrix}. \tag{B4}$$

Set $\vec{F}(\vec{v} + \delta\vec{v})$ to zero and solve for $\delta\vec{v}$, giving a first correction to the initial guess for $\vec{v}$ of

$$\delta\vec{v} = -\mathbf{M}^{-1} \cdot \vec{F}(\vec{v}). \tag{B5}$$

Iterate by evaluating $\mathbf{M}$ at $\vec{v} + \delta\vec{v}$ and getting a further correction until the corrections are sufficiently small and the components of $\vec{\mathbf{F}}$ are sufficiently close to zero. The inverse $\mathbf{M}^{-1}$ of the matrix $\mathbf{M}$ is evaluated as

$$\mathbf{M^{-1}} = \begin{pmatrix} h - 2g + f & g - f \\ g - f & -f \end{pmatrix} \times \left(\frac{\mathbf{1}}{g^2 - hf} \right) \tag{B6}$$

in which, with $x_l = -(l-1)\beta\Delta - \mu\beta$,

$$f = \sum_{l=0}^{l_{\max}} (2^l - 1)/4 \sinh^2(x_l/2), \tag{B7}$$

$$g = \sum_{l=0}^{l_{\max}} l(2^l - 1)/4 \sinh^2(x_l/2), \tag{B8}$$

$$h = \sum_{l=0}^{l_{\max}} l^2(2^l - 1)/4 \sinh^2(x_l/2). \tag{B9}$$

APPENDIX C: COLLECTIVE DYNAMICS AND DYNAMICS CUT

To characterize the collective dynamics of states in the model, we used, as in Ref. [5], the correlation function $C(\tau)$ defined as

$$C(\tau) = [1/(N_{\rm st} - \tau)] \sum_{t=1}^{N_{\rm st}-\tau} \sum_{m} [n_m(t) - \overline{n_m}][n_m(t+\tau) - \overline{n_m}]. \tag{C1}$$

Here τ and t are integers numbering successive reactions, $N_{\rm st}$ is the number of time (event) steps used to characterize the dynamics, and $\overline{n_m}$ is the time average of n_m over the $N_{\rm st}$ reaction steps.

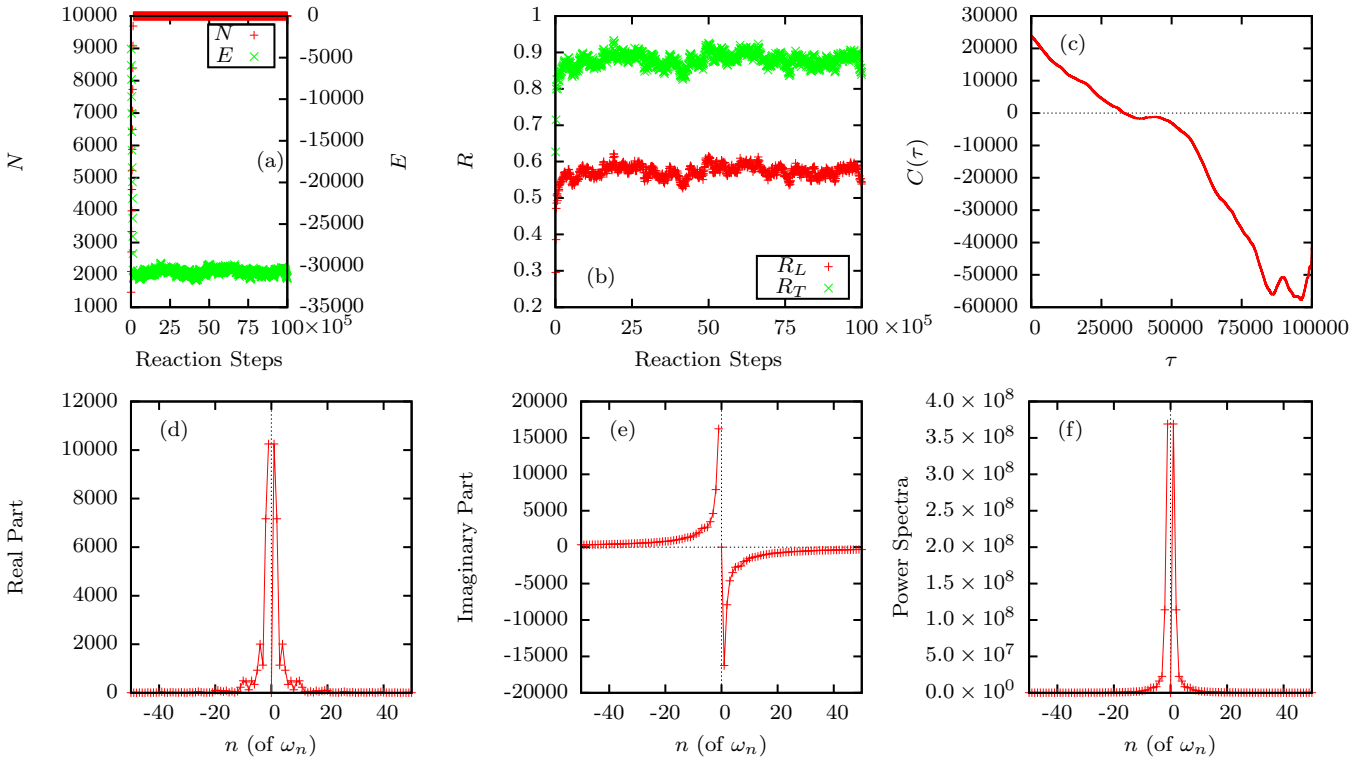

FIG. 16. Fourier transforms and the power spectrum of a run at fixed ambient temperature $\beta\Delta = 1$. Panels are defined as in Fig. 15. This system is classified as LATA based on the R values. Its population grew rapidly and its dynamic spectrum is qualitatively similar to the previous, equilibrium example. Here $p = 0.002\,26$ and $l_{\max} = 6$. The same network that was used to generate the data in Fig. 15 was used here.

The time Fourier transform, with $\omega_l = 2\pi l/N_{\rm st}$, is (with $i = \sqrt{-1}$)

$$\mathcal{S}(\omega_l) = \sum_{\tau=1}^{N_{\rm st}} \exp(-2\pi i l\tau/N_{\rm st}) C(\tau) F(\tau), \qquad \text{(C2)}$$

in which $F(\tau) = (N_{\rm st} - \tau)/N_{\rm st}$ is a weighting function which takes approximate account [11] of the fact there are fewer data points in the sum for $C(\tau)$ leading to larger uncertainties in $C(\tau)$ as τ increases. The power spectrum is

$$P(\omega_l) = |{\rm Re}[\mathcal{S}(\omega_l)]|^2 + |{\rm Im}\mathcal{S}(\omega_l)|^2. \qquad \text{(C3)}$$

We characterize the scale ω_m of the frequencies in the power spectrum by approximately solving the implicit equation

$$\frac{\sum_{l=-m}^{l=m-1} P(\omega_l)}{\sum_{l=-N_{\rm st}/2}^{N_{\rm st}/2-1} P(\omega_l)} = 1/2. \qquad \text{(C4)}$$

Over the relatively short steady-state period of 10 000 reactions steps we showed in Ref. [5] that the shape of the Fourier transform is not sensitive to choice of event time or Gillespie time. The Fourier transforms are numerically much quicker to evaluate using event time, so we have chosen to use event time in the computation of the Fourier transform and correct the values of ω_m to Gillespie time after ω_m is determined by dividing by the average Gillespie time $\delta t_{\rm Gillespie}$ between reactions during the steady-state interval. We then choose a cutoff frequency ω_c and deem states with $\omega_m > \omega_c$ to have passed the dynamics cut and to be considered dynamically alive.

We show some examples of the correlation functions and power spectra in Figs. 15–17 (in units of inverse reaction time). With our parametrization, the results fall into two, fairly distinct, dynamic classes. In one class, the power spectrum is dominated by a very strong peak around the highest possible frequency, obtained by setting $l = N_{\rm st}/2$ so that $\omega_l = \pi$ (the zone boundary). These states easily pass any dynamics test of the sort described, but it is not clear that they are intuitively very lifelike. The other states have quite symmetrical peaks of various shapes around the origin in the power spectrum. Many of these Fourier transforms have a shape reminiscent of a Drude or Debye spectrum characterized by one or a few dissipatively damped modes, while some others show evidence of underdamped modes as well. The time scale of this second class of dynamic states is much longer, usually roughly 10^4 event time steps. We do not have a detailed theory to account for this time scale, but find that it is of the order of the number of reactions in the networks (about 2500 for typical values of $l_{\max} = 8$ and $p = 0.005$).

APPENDIX D: FOOD SET INSTABILITY IN ISOLATED DYNAMICS

We hold the total food set population of dimers and monomers at a fixed value N_f (=500 in the simulations reported here). The simulations start with finite populations only in the food set. Here $N_f = N_1 + N_2$ and from (7) the rate of dimer scission to monomer monomer ligation is

$$\frac{2 \to 1+1}{1+1 \to 2} = \overline{n_1}^2/\overline{n_2} \qquad \text{(D1)}$$

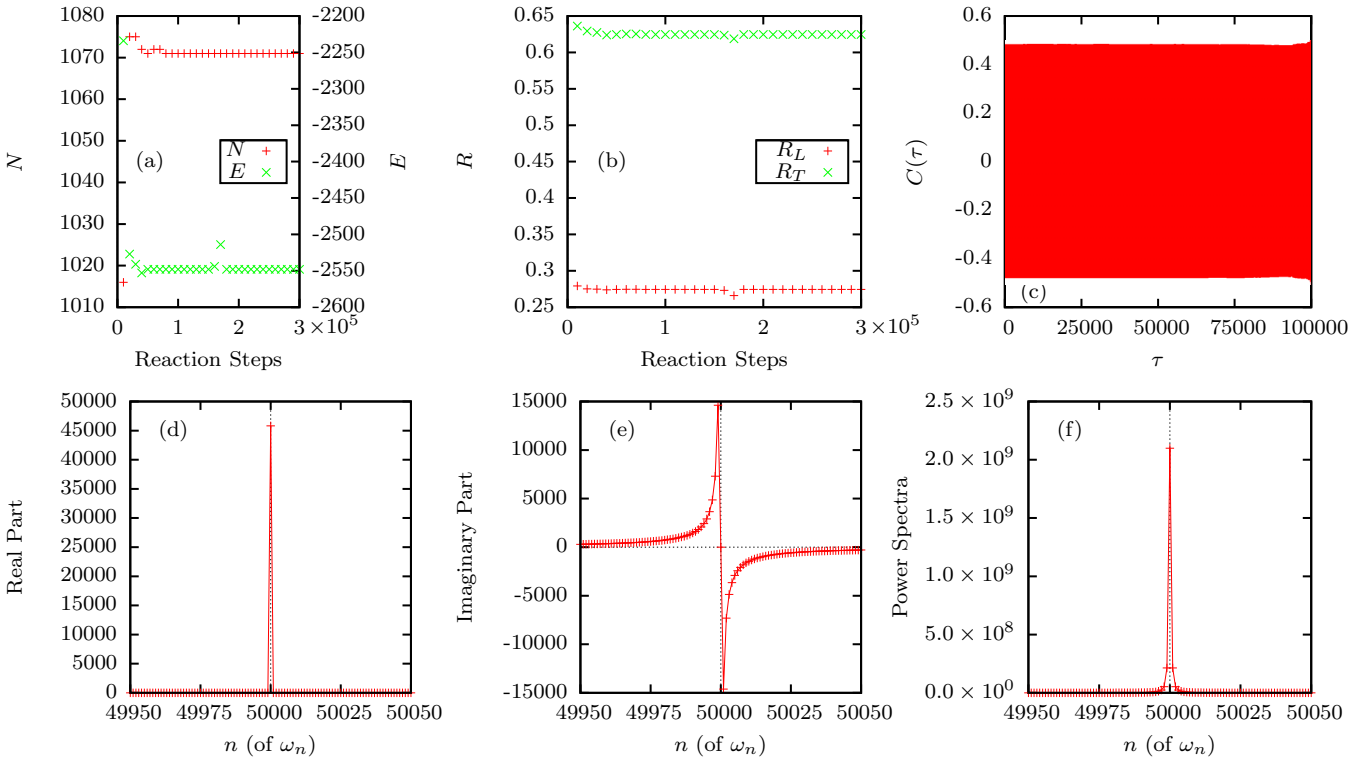

FIG. 17. Fourier transforms and the power spectrum of a run at fixed ambient $\Delta\beta = 1$ but exhibiting a sharp peak in its the power spectrum of $C(\tau)$ around the maximum frequency $\omega = \pi$. Panels as defined as in Figs. 15 and 16 except that (d)–(f) have been centered around $\omega = \pi$. This system would qualify as LDTA on the basis of the R values but its population is not growing and its dynamic spectra are very different from the previous two examples. For this run $p = 0.002\,26$, $l_{\max} = 6$, and the network is different from that used to generate Figs. 15 and 16.

for any pair of monomers. At the beginning of the simulation, there are only dimers and monomers and the instantaneous equilibrium values of the populations are, using the relations $N = 2\overline{n_1} + 4\overline{n_2}$ and $\overline{n_2} = -E/4\Delta$,

$$\overline{n_1} = (N + E/\Delta)/2, \quad \overline{n_2} = (-E/4\Delta). \tag{D2}$$

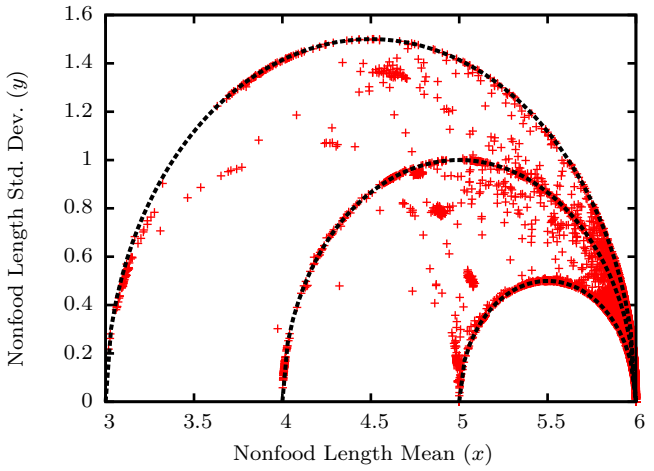

FIG. 18. Scatter plot of values of the nonfood mean polymer length and its standard deviation produced by 41 667 simulations driven toward equilibrium at fixed positive temperature with $\beta\Delta = 10$, $p = 0.012\,80$, and $l_{\max} = 6$. The dashed curves show the results of Eq. (E3) for the cases $l_1 = 6$ and $l_2 = \{3,4,5\}$ with no adjustable parameters.

Express this in terms of $f = N_1/N = (1 + E/\Delta N)$, giving

$$\frac{2 \to 1 + 1}{1 + 1 \to 2} = Nf^2/(1 - f). \tag{D3}$$

The forward and reverse rates are equal when this is 1, giving a quadratic equation for f with a positive root

$$f_c = (1/2N)(-1 + \sqrt{1 + 4N}) \approx 1/\sqrt{N} \tag{D4}$$

because $N_f = 500 \gg 1$. The critical value of N_1/N is thus about 0.05 and if we populate the food set with a 1/6 probability for each species, 95% of the time we will drive all the dimers to monomers.

This accounts for the observation that in the simulation, uncontrolled equilibration of the food set to the instantaneous values of E and N drives the dissolution of dimers into monomers. However, then the evolution cannot start, because in order to get a nonfood polymer (of length 3 or larger) from a collection of only monomers one must first make a dimer and, by the previous argument, the dimer will immediately be dissociated back to monomers.

APPENDIX E: SCARS IN MEAN LENGTH VERSUS LENGTH VARIANCE SCATTER PLOTS

We consider the case in which there are nonzero populations of nonfood polymers of just two types with lengths l_1 and l_1 and

TABLE I. Numbers of networks generated to produce the results shown in Figs. 2 and 3. For each viable network, 50 dynamics simulations were produced.

p	No. of networks
0.00226	10 000
0.00320	10 000
0.00452	10 000
0.00640	5000
0.00761	1000
0.00905	1000
0.01280	1000

TABLE II. Numbers of networks generated and numbers of dynamics runs per viable network used to produce the results shown in Figs. 7 and 8.

p	No. of networks	Dynamics runs per viable network
0.00226	10 000	50
0.00320	10 000	50
0.00452	5000	25
0.00640	2000	25
0.00761	1500	25
0.00905	1250	25
0.01076	1000	25
0.01280	1000	25

populations n_1 and n_2 and a constant total nonfood population $N = n_1 + n_2$. The mean length is then

$$x = \bar{l} = (l_1 n_1 + l_2 n_2)/N \tag{E1}$$

and the variance in the length is

$$y^2 = \overline{l^2} - \bar{l}^2 = \left(l_1^2 n_1 + l_2^2 n_2\right)/N - [(l_1 n_1 + l_2 n_2)/N]^2. \tag{E2}$$

Rearranging these relations, we obtain

$$y^2 = (l_1 - x)(x - l_2), \tag{E3}$$

which describes the loci of the scars in Figs. 4 and 6 very well. We show an example of the result of Eq. (E3) compared with scatter plot data from a thermal run in Fig. 18.

APPENDIX F: SIMULATION DETAILS

For the data in Figs. 2 and 3, the number of generated networks (including those discarded as not viable) was as shown in Table I. For each viable network, 50 dynamic simulations with different random number seeds for initiation were generated. For the data in Figs. 7 and 8, the number of networks and the number of simulations per network are shown in Table II.

For larger p, a larger number of generated networks is viable, so a smaller number of starting networks is required to generate statistically significant data. Also, at larger p it takes more reactions to reach steady state, so longer runs are required and fewer dynamics runs per viable network are computationally affordable.

[1] M. Eigen, Naturwissenschaften **58**, 465 (1971).

[2] F. Dyson, J. Mol. Evol. **18**, 344 (1982).

[3] S. A. Kauffman, *The Origins of Order* (Oxford University Press, New York, 1993), Chap. 7.

[4] J. D. Farmer, S. A. Kauffman, and N. H. Packard, Physica D **22**, 50 (1986); R. Bagley and J. D. Farmer, in *Artificial Life II*, edited by C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen (Addison Wesley, Redwood City, 1991), p. 93.

[5] A. Wynveen, I. Fedorov, and J. W. Halley, Phys. Rev. E **89**, 022725 (2014).

[6] B. F. Intoy, A. Wynveen, and J. W. Halley, Phys. Rev. E **94**, 042424 (2016).

[7] G. Caetano-Anollés, K. M. Kim, and D. Caetano-Anollés, J. Mol. Evol. **74**, 1 (2012).

[8] D. Gillespie, J. Comput. Phys. **22**, 403 (1976).

[9] L. D. Landau and E. M Lifshitz, *Statistical Physics* (Pergamon Press, London, 1958), p. 216; N. F. Ramsey, Phys. Rev. **103**, 20 (1956).

[10] N. Fischer, P. Neumann, A. L. Kovevega, L. V. Bock, R. Ficner, M. V. Rodnina, and H. Stark, Nature (London) **520**, 567 (2015); http://www.rcsb.org/pdb/explore/explore.do?structureId=5afi

[11] P. Stoica and R. Moses, *Spectral Analysis of Signals* (Prentice Hall, Upper Saddle River, 2005), Eq. (2.24).