

Richard J. Bagley and J. Doyne Farmer
Complex Systems Group, Theoretical Division, and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545 and Santa Fe Institute, 1120 Canyon Road, Santa Fe, NM 87501

Spontaneous Emergence of a Metabolism

Networks of catalyzed reactions with nonlinear feedback have been proposed to play an important role in the origin of life. We investigate this possibility in a polymer chemistry with catalyzed cleavage and condensation reactions, studying the properties of a well-stirred reactor driven away from equilibrium by the flow of mass. Near equilibrium the distribution of material is uninteresting; it favors short polymers but is otherwise homogeneous. However, under appropriate non-equilibrium conditions, the situation changes radically: The nonlinear feedback of the reaction network focuses the material of the system into a few specific polymer species, whose concentrations can be orders of magnitude above the background. Like a metabolism, the network of catalytic reactions "digests" the material of its environment, incorporating it into its own form. For this reason we call it an *autocatalytic metabolism*. We vary the diet of an autocatalytic metabolism, and demonstrate that under some variations it persists almost unchanged, while in other cases it dies. We argue that the dynamical stability of autocatalytic metabolisms gives them regenerative properties that allow them to repair themselves and to propagate through time.

CONTENTS

1 Motivation	95
1.1 Setting the Stage for an Origin of Life	95
1.2 Nontrivial Dissipative Structures	97
1.3 A Simple Model for Studying Evolution with an Emergent Notion of Fitness	98
2 Background	98
2.1 Spontaneous Reactions	99
2.2 Equilibrium Distribution of Polymers	99
2.3 Catalyzed Reactions	100
2.4 Driving from Equilibrium	101
2.4.1 Mass Flow	101
2.4.2 Pyrophosphates	102
2.5 Reaction Graph	103
2.6 Catalytic Focusing	104
2.7 Autocatalytic Sets and Metabolisms	105
3 Simulation	107
3.1 Artificial Chemistry	107
3.1.1 Random Assignment of Reactions	108
3.1.2 Assignment of Reactions by String Matching	109
3.1.3 Other Kinetic Parameters	109
3.2 Metadynamics	110
3.3 The Background of Uncatalyzed Reactions	112
4 Emergence	115
4.1 Departure from Equilibrium	115
4.2 Dependence on Parameters	118
5 Sensitivity	123
5.1 Sensitivity to Topology	123
5.2 Sensitivity to Parameters	124
5.3 Role of Diversity	126
6 Metabolic Robustness	126
6.1 What is a Metabolism?	127
6.2 Robustness	127
7 Discussion	129
7.1 Relevance to Experiments	129
7.2 Possible Scenario for the Origin of Life	130
7.3 Critique	131
7.4 Are They Alive?	132
7.5 Summary	133
Appendix	136
References	138

1. MOTIVATION

1.1 SETTING THE STAGE FOR AN ORIGIN OF LIFE

When Miller and Urey discovered that amino acids could be formed under conditions that might be similar to those of the prebiotic earth,²⁸ the spontaneous synthesis of proteins seemed just around the corner. However, this turns out to be much more difficult than the spontaneous synthesis of individual amino acids. Equilibrium conditions tend to favor dissociation, and generate a concentration profile that is fairly uniform. Except for occasional fluctuations, for long polymers the population of any given molecular species is typically zero. The population distribution of polymers is homogeneous, nonspecific, and uninteresting. This is in contrast to living organisms, which have high concentrations of a few *specific* polymer species.^[1]

Contemporary organisms achieve specificity through a codependent relationship between templates and enzymes. Proteins and nucleic acids synthesize each other through a replication mechanism in which none of the components synthesizes itself. Even for the simplest organisms, this process is highly complex. There seems to be a minimum level of complexity below which a replicating machine based on proteins and nucleic acids simply cannot function. While it is easy to understand how such a replicating machine perpetuates itself, it is difficult to understand how the necessary initial conditions ever arose on their own. The probability that both enzymes and templates could be created through a statistical fluctuation is effectively nil. This suggests that other processes preceded contemporary life.

The idea that enzymatic activity might have set the stage for the origin of life was developed by Oparin, who suggested that coacervates may have played a major role.³³ Early experiments unsuccessfully attempted to use clays and other materials as nonspecific catalysts for polymerization.³⁴ Calvin studied several different scenarios through which catalytic activity could provide a selection mechanism, even without self-replication.^{5,6} In 1971 Rössler,^{37,38} Eigen,¹⁰ and Kauffman²³ developed this idea further. In particular, Rössler envisioned a form of chemical evolution similar to that studied here. He emphasized the importance of specific catalysts which catalyze only a small fraction of all possible reactions. Along these same lines, Kauffman²⁴ later modeled the problem in terms of random graphs, and showed that under reasonable assumptions the probability of catalytic closure is quite high.^[2] The random graph model was developed into a kinetic model that could be simulated on a computer by Farmer et al.¹¹ This line of investigation, which attempts to find possible precursors facilitating the emergence of life should be contrasted with

[1] An exception is provided by the experiments of Sidney Fox, who by heating a mixture of amino acids demonstrated the formation of polypeptides, called proteinoids.^{20,21} The structure of proteinoids is not random; some subsequences, such as certain hexapeptides, occur much more frequently than others. In contrast, our goal is to increase the concentration of entire molecules, so that it is several orders of magnitude above equilibrium.

[2] Another toy model investigating the possibility that a metabolism might have spontaneously emerged without a replicator is due to Dyson.⁸

other work that addresses the (also very interesting) question of the early evolution of life once replication has already begun.^{1,9,10,39}

In this paper we study the behavior of a network of catalyzed chemical reactions, along the lines laid out by Farmer et al.¹¹ We make several enhancements of the model and analytically study a few simple cases to gain better intuition about the dynamics. We also improve the simulation so that it is several orders of magnitude faster. This allows us to simulate the kinetics of a complicated reaction network in a matter of seconds. As a result, we are able to widely explore the parameter space and answer many of the questions originally raised in earlier papers.

Our main result is that, under appropriate conditions, a catalytic reaction network can focus most the material of its environment into a few chemical species. For this to happen the system must be driven the appropriate distance from equilibrium, polymerization must be favored, and it must have diverse kinetic parameters. Favoring polymerization may require the addition of energy, for example, through pyrophosphates energized by light. In spite of these restrictions, there is a wide range of parameters in which the material of the system is focused into only a few species, which dominate over the background.

Focusing radically alters the material composition of the environment. The species that emerge reinforce each others' production and largely take over the reaction vessel, excluding other possibilities. Since this behavior is analogous to that of a metabolism, we call the resulting set of species and reactions an *autocatalytic metabolism*. Under appropriate conditions autocatalytic metabolisms can evolve out of a simple, undifferentiated initial state, generating a sequence of complex, highly differentiated, final states. Like contemporary organisms, these final states are composed of a highly focused, specific set of long polymers. While the autocatalytic metabolism does not replicate itself in the usual sense, it propagates itself by taking over any medium with suitable properties, sustaining itself as long as the appropriate conditions are met. Furthermore, it may generate a lineage of related autocatalytic metabolisms.

The model that we study here applies to any system in which polymers can catalyze the formation of other polymers through cleavage and condensation reactions. If the basic building blocks are amino acids, then the polymers are called polypeptides or proteins. Such reactions are common among proteins, forming the basis for many of the functions of living organisms. If the basic building blocks are nucleotides, then the polymers are called nucleic acids (RNA or DNA). It is well known that polypeptides possess a large repertoire of catalytic activities of this type; the recent discovery of specific catalysis reactions in RNA suggests that nucleic acids may also possess the necessary properties.⁷ Whether the polymers are polypeptides or nucleic acids changes the parameters but not the basic form of the model.

Even if the model we discuss here has nothing to do with the actual origin of life on earth, it might provide a *possible* origin of life in the laboratory. Although at present we cannot predict the outcome of experiments in detail, we can make qualitative predictions that provide broader experimental guidelines. The accumulation of more experimental knowledge can be used to determine the unknown parameters

of our model, which in turn should sharpen its predictive value for experiments. Although many important experimental details are still unknown, and some important questions await further study, our numerical simulations suggest that it may be possible to synthesize autocatalytic metabolisms in the laboratory.

1.2 NONTRIVIAL DISSIPATIVE STRUCTURES

The problem of the emergence of life is embedded in the broader problem of understanding self-organizing phenomena, which from the point of view of a physicist may be more interesting anyway. Many non-living systems exhibit self-organizing properties, albeit much weaker than those of living systems. Is there a sharp distinction between living and nonliving systems? Or can there exist levels of organization that are between those of present living and non-living systems? Can evolution and other self-organizing properties of living systems be viewed as manifestations of a general law that describes the tendencies of matter to organize itself?

There are many simple examples that have been cited as instances of self-organizing phenomena in nature. For example, when a fluid is heated from below, under appropriate conditions, patterns of convection cells form. The macroscopic structure of these cells is internally generated by the system itself, and is not apparent in its initial conditions. Such patterns are often called *dissipative structures*, because they occur when energy flows into a system and then is dissipated.³²

Several researchers have asserted that life is a non-equilibrium phenomenon, associated with dissipative structures.³² This is certainly true, but it is a very weak statement. While deviation from equilibrium is a necessary condition for life, it is far from sufficient. Driving a system away from equilibrium does not necessarily cause the emergence of order—in fact, it often has precisely the opposite effect. A central question that must be addressed in a theory of self-organizing phenomena is: Why do some non-equilibrium situations foster the spontaneous emergence of organization, while others do not?

There is a big gap between the dissipative structures of simple non-living systems, such as patterns in fluid convection, and the much richer dissipative structures associated with living systems. The model discussed here is intended to bridge this gap, at least to some extent, by showing the possibility for dissipative structures that are intermediate in complexity between living and non-living systems. Autocatalytic metabolisms are more complex than convection patterns, in that they propagate specific information through time. One autocatalytic metabolism can seed the formation of another, similar metabolism. The autocatalytic metabolisms of this model can be viewed as proto-life forms, since they have a metabolism, they evolve and store information, and they reproduce (although more continuously and with less fidelity than contemporary organisms). They are also dynamically stable, and so capable of self-repair. They, thus, have many of the essential properties of living systems, albeit in a much less sophisticated form.

Besides demonstrating the possibility for the spontaneous generation of autocatalytic metabolisms, one of our main purposes in this paper is to discover under

what conditions they can be expected to form. How does their formation depend on the parameters of the system, such as the flow of energy, or the inherent diversity of the underlying dynamics? Although our results are specific to this model, they nonetheless suggest several rules that may pertain to the more general problem of self-organization.

1.3 A SIMPLE MODEL FOR STUDYING EVOLUTION WITH AN EMERGENT NOTION OF FITNESS

In principle, it is possible to describe biological systems at a fundamental level in terms of their dynamics. At this level of description, "selection" is an emergent property of the dynamics. In practice, however, for most systems this is hopelessly intractable. As a result, studies of evolution are typically couched in terms of the fitness function, which is an empirical construct, disconnected from the laws of physics. Even so, in most systems the fitness function is known only in very special circumstances where all but a few relevant factors are neglected. In general, fitness is a complicated function of the external environment, which includes other organisms. As a result, most theoretical models for evolution make many *ad hoc* assumptions, postulating fitness functions that may be qualitatively different from those in the real world.

As pointed out by Eigen,¹⁰ Rössler,³⁷ and others, chemistry provides an excellent forum for studying evolution. The laws of chemical kinetics are well understood, and make it possible to model the behavior of the system at a fundamental level. These laws determine population levels and therefore determine fitness. As in biological systems, fluctuations are always present, generating random variation. Thus, for chemical networks we can describe the fitness at the fundamental level of dynamics.

Even though autocatalytic metabolisms do not have templates or a genetic code, because of their specificity they are nonetheless capable of evolution. This is discussed in a companion paper.⁴ Autocatalytic metabolisms, therefore, provide an interesting alternative for studying evolution in a chemical setting. It is also interesting to note that autocatalytic structures analogous to those we study here occur spontaneously in more abstract environments, as observed by Fontana,^{16,17} and Rasmussen et al.³⁶

2. BACKGROUND

In this section we discuss some of the properties of catalyzed reaction networks, providing a background for the development of the simulation in Section 3. We discuss the reactions we are going to consider, and show why they are uninteresting at equilibrium. We then explain how the situation is altered as we move away from equilibrium, and how catalysis can play an important role in focusing the material

of the system into just a few chemical species. We define autocatalytic sets and the related notion of autocatalytic metabolisms.

2.1 SPONTANEOUS REACTIONS

We are interested in reversible polymerization reactions, in which two polymers either *condense* to form a single longer polymer, or a single polymer *cleaves* into two shorter polymers. Cleavage and condensation can be considered together as a single reversible reaction. The reaction in which polymers *A* and *B* join together to form *C*, giving off water, or equivalently, in which *C* hydrolyzes into *A* and *B*, can be written



where *H* represents water.

Providing the concentrations are sufficiently high and the solution is well stirred, the law of mass action provides a good approximation of the kinetics. Let k_f be the rate constant for the forward reaction, $A + B \rightarrow C + H$, and k_r be the rate constant for the backward reaction $C + H \rightarrow A + B$. The rate equation for *C* is then

$$\dot{C} = \frac{dC}{dt} = k_f AB - k_r HC. \quad (2)$$

For convenience, whenever the meaning is unambiguous, we use the same symbol to represent both a polymer and its concentration. Similar equations apply for *A* and *B*.

2.2 EQUILIBRIUM DISTRIBUTION OF POLYMERS

At equilibrium the concentrations of the polymers of a given length can be computed analytically using the classical theory of polycondensation reactions developed by Flory and Stockmayer.^{15,40,44} For simplicity we assume that all the reactions have the same forward and backward rate constants, and that the reaction vessel is well stirred. Furthermore, we assume that the monomers are oriented so that each monomer has two sites, which we arbitrarily designate as the "+" site and "-" site.

We follow the treatment of Macken and Perelson.²⁷ Rather than solving for the concentrations of each polymer, it is more convenient to use an aggregate variable *y*, which is the concentration of free sites of a given kind (either "+" or "-"). We assume that the polymers are unbranched, and that they cannot form rings. For reactions of the form of Eq. (2),

$$\dot{y} = -k_f y^2 + k_r (m_0 - y) H, \quad (3)$$

$$C = K_f \frac{A B}{H} = \frac{m_0 - y}{2} H$$

where m_0 is the total concentration of monomers, which is equal to the concentration of free sites if nothing is bound. At steady state $y = 0$ and the concentration of free sites is

$$y = \frac{(1 + 4\kappa y_0)^{1/2} - 1}{2\kappa}, \quad (4)$$

where $\kappa = k_f/Hk_r$ is the equilibrium constant.

We now compute the concentration of polymers of length n . At equilibrium, let ρ be the probability for the formation of a bond. This is the ratio of bound sites to the total number of sites, i.e.,

$$\rho = \frac{m_0 - y}{y_0}. \quad (5)$$

Assume that each binding event is independent. For a given free site, the probability that it is attached to $n - 1$ bonds followed by another free site is $\rho^{n-1}(1 - \rho)$. Solving Eq. (5) for y shows that the concentration of free sites for a given value of ρ is $y = m_0(1 - \rho)$. Thus, the concentration of polymers of length n is

$$x_n = m_0(1 - \rho)^2 \rho^{n-1}. \quad (6)$$

At equilibrium, inserting Eq. (4) into Eq. (5) gives

$$\rho = 1 - \frac{(1 + 4\kappa m_0)^{1/2} - 1}{2\kappa m_0}. \quad (7)$$

Note that $\rho < 1$. Thus, Eq. (6) implies that the concentration of polymers of length n decreases exponentially with n , at a rate that depends only on the product of the equilibrium constant and the concentration of monomers. In a system with m distinct monomers, present initially at equal concentrations, the concentration of any particular polymer species of length n is further decreased by a factor of m^{-n} . For $m > 1$, even if $\rho \approx 1$, so that polymerization is favored, for large n the concentration of any particular species is quite small. For example, for polypeptides $m = 20$; the concentration of a polypeptide of length $n = 30$ is roughly 20^{-30} less than that of a monomer. For a container of finite size, this implies that, except for occasional fluctuations, most longer species are not present.

2.3 CATALYZED REACTIONS

The presence of a catalyst (enzyme) E can accelerate a reaction.



At equilibrium the rate of the forward reaction equals that of the backward reaction, so that $\dot{A} = \dot{B} = \dot{C} = 0$. Catalysis speeds up the rate at which the system approaches equilibrium, but does not change the concentrations at equilibrium.

However, when the reaction is driven away from equilibrium, for example by externally supplying one of the participants in the reaction, catalysis can shift the steady state. This is the basis for the effect we study here.

Catalysis increases both the forward and backward rate constants by the same amount. This can be taken into account by defining a quantity ν that we call the *catalytic efficiency*. For fixed concentration of the reactants, the increase in the velocity of the reaction is proportional to the product of ν and the concentration E of the catalyst. The kinetic equation for C can be crudely approximated as

$$\dot{C} = (1 + \nu E)(k_f AB - k_r HC). \quad (9)$$

Similar equations apply to \dot{A} and \dot{B} . When the catalytic efficiency $\nu = 0$, this reduces to the kinetic equation for a spontaneous reaction.

Note that, for a population of polymers, this reaction is just one reaction in a network of many. A given polymer may play the role of A in some reactions, and the role of C in others. To compute the rate of production of any given species, it is necessary to sum all the relevant reaction terms.

The approximation made in Eq. (9) neglects the effect of saturation, which comes about because the enzyme and the reactants are bound together for a finite time. During this time they cannot participate in new reactions, which lowers the effective reaction rates. If this is a dominant effect, so that most of the enzyme or product is bound at any given time, the reaction is *saturated*. To take this into account we do not use Eq. (9), but rather use a more accurate approximation. We keep track of the concentration of any given species x_i that is bound into complexes through an auxiliary variable \bar{x}_i , which is equal to the sum of the concentrations of all the complexes in which x_i is bound. To keep the simulation tractable, we assume that all complexes unbind at the same rate k_u . This approximation is described in more detail in the Appendix.

2.4 DRIVING FROM EQUILIBRIUM

To make anything interesting happen in a reaction network it must be driven away from equilibrium. In this model we investigate two different mechanisms. The first involves a flux of mass, and the second involves the formation of energetic pyrophosphate molecules, driven by light.

2.4.1 MASS FLOW We model a reaction vessel with a steady input flux of monomers or short polymers, and an output flux due either to diffusion or overflow of the reaction vessel. This might correspond to a prebiotic environment, or it might correspond to a chemostat in a laboratory experiment. The chemical species that are input are collectively called the *food set*. For simplicity, we assume an inflow rate δ of concentration per unit time, and an outflow that is proportional to concentration, with rate constant K .

For an element of the food set, the kinetic equations are of the form

$$\frac{dx_i}{dt} = k_a \bar{x}_i + \sum (\text{reaction terms}) + \delta - Kx_i, \quad (10)$$

$$\frac{d\bar{x}_i}{dt} = -k_a \bar{x}_i + \sum (\text{reaction terms}) - K\bar{x}_i. \quad (11)$$

The (*reaction terms*) are defined in the appendix in the discussion following Eq. (25). For a species outside the food set, the kinetic equations are of the same form, except that $\delta = 0$. There is a net flow of mass from the food set to the other elements of the system, which drives it away from equilibrium.

Because the reaction terms conserve mass, the total mass in the reaction vessel always goes to a fixed point, independent of initial concentrations. To see this, note that only the last two terms in Eq. (10) and the last term in Eq. (11) change the total mass. The total mass concentration is proportional to^[3] $m = \sum n(i)x_i$, where $n(i)$ is the length of the i th species. Letting N_f be the number of elements in the food set, the rate of change of the total mass concentration is given by a simple differential equation,

$$\frac{dm}{dt} = N_f \delta - Km, \quad (12)$$

which has a global fixed point $m_0 = N_f/K\delta$. This means the initial mass is irrelevant anyway, and so for convenience in our simulations, we choose $m(0) = m_0$. Thus, there are effectively only two parameters relating to the flow of mass through the system, which can be δ and K , or equivalently δ and m_0 .

Since δ and K are not intuitively easy to interpret, it is sometimes useful to quote results in terms of the *mean reaction number* r . This is defined as the mean number of times a given monomer participates in a reaction, on average, from the time it enters the vessel until the time it is flushed out. At equilibrium r is infinite, and when the other parameters are fixed, it decreases monotonically as δ increases.

2.4.2 PYROPHOSPHATES As we demonstrate in subsection 2.6, catalytic focusing requires conditions that favor polymerization. The tendency to polymerize can be enhanced by an appropriate input of energy. The mechanism that we investigate here involves pyrophosphate molecules, which play a role analogous to that of *ATP* in contemporary organisms. This mechanism is supported by early experiments.³⁵ The detailed sequence of reactions was suggested by Ron Fox,^[4] and is illustrated in Table 1.

It proceeds as follows: Light causes the formation of pyrophosphate molecules (p_2), which is balanced by hydrolysis. When a pyrophosphate molecule binds to polymer A , it creates the energized form A^* and releases a phosphate atom in the

[3] To convert this to units of mass/volume there is a constant of proportionality, that depends on the mass of a monomer.

[4] Private communication.

TABLE 1 Pyrophosphate energizes and enhances polymerization. The first column lists the reactions and the second column the reaction rate. γ represents a photon (and in the column on the right represents the intensity of light), p a phosphate atom, and H water. A and B represent the polymers that condense to form C . E is the catalyst. A bar indicates a complex that is bound to the enzyme E . k_f is the rate constant for polymerization, k_r for hydrolysis, k_a for the dissociation of the bound complex, k_e for the polymerization of phosphate, and k_g for the activation of a polymer. Activated polymers are indicated with a "*" superscript. k_f^* is the rate constant for the condensation of an activated polymer with another polymer.

Reaction	Rate
$2p$	$k_e \gamma p^2$
$p_2 + HC$	$k_r p_2 H$
$A + p_2$	$k_a A p_2$
$A^* + H$	$k_r A^* H$
$A^* + B + E$	$k_f^* v E A^* B$
$2p$	$k_f \gamma p^2$
$p_2 + H$	$k_r p_2 H$
$A + p_2$	$k_a A p_2$
$A^* + H$	$k_r A^* H$
$A^* + B + E$	$k_f^* v E A^* B$

process. A^* may hydrolyze, releasing the other phosphate atom, or it can bind to another polymer B (in the presence of the catalyst E). This occurs with a rate constant k_f^* , which is greater than the unenergized rate constant k_f . Thus, the addition of energy favors polymerization.

Based on simulations involving the full reaction scheme shown in Table 1, we found that when the concentration of pyrophosphate and the input of light are sufficiently high, the behavior is roughly equivalent to that obtained by simply using the equations given in the Appendix, with the effective forward rate constant equal to k_f^* . For convenience, in the numerical experiments described here we simply assume that the parameters quoted correspond to k_f^* , and use the simpler equations which do not involve pyrophosphate.

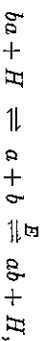
2.5 REACTION GRAPH

Each distinct monomer can be assigned a character from a fixed alphabet, a, b, c, \dots . A polymer can then be represented as a character string, for example (*acabcbac*, \dots). We assume that the polymers are oriented, so that *abc* and *cba* are different strings. The topological structure of a network of reactions, each of the form of Eq. (3), can be represented as a polygraph with two types of nodes and two types of connections,¹¹ as illustrated in Figures 1 and 2. One type of node represents the

polymer species and is labeled by the corresponding string. The other type of node represents the catalyzed reaction and is labeled by a black dot. The polymers that participate in a reaction are connected to the corresponding reaction node by reaction links (black arrows), which point in the direction of condensation. Each polymer is connected to the reactions it catalyzes by a catalytic link (dotted line). Each reaction has at least four links: three reaction links, and one or more catalytic links.

2.6 CATALYTIC FOCUSING

Under appropriate conditions catalysis can focus most of the material of a reaction network into only a few species. The basic idea can be grasped by considering the simple reaction network



as shown in Figure 1. Assume a and b are supplied at rate δ , and diffuse out of the container with rate constant K , as described in subsection 2.4. For simplicity, assume the concentrations of E and H are maintained at fixed values. Neglecting saturation, according to the approximation of Eq. (9), the rates of change of ab and ba are

$$[ab] = \gamma(k_f[a][b] - k_r H[ab]) - K[ab], \quad (13)$$

$$[ba] = k_f[a][b] - k_r H[ba] - K[ba], \quad (14)$$

where $[ab]$ is the concentration of polymer ab . Setting the derivatives to zero and using the mass conservation condition of Eq. (12) gives

$$\frac{[ab]}{[ba]} = \frac{1 + \beta}{1 + \frac{\beta}{\gamma}}. \quad (15)$$

$\beta = \delta/m_0 k_r H$ is a dimensionless parameter related to the deviation from equilibrium, where $m_0 = a(0) + b(0)$ is the total concentration of monomers. Note that $\beta \geq 0$. $\gamma = 1 + \nu E$ is a dimensionless parameter that characterizes the strength of catalysis. $\gamma \geq 1$, and $\gamma = 1$ corresponds to an uncatalyzed reaction.

Under what circumstances is the concentration of ab much greater than that of ba ? At equilibrium $\beta = 0$ and the ratio of $[ab]$ to $[ba]$ is one. This ratio can become large only when $\beta \gg 0$, i.e., only when the system is driven well away from equilibrium. When $\beta \gg \gamma$ this ratio approaches γ ; when $\beta \ll \gamma$ it approaches β . Thus, by varying γ and β the concentration of ab relative to ba can be made arbitrarily large.

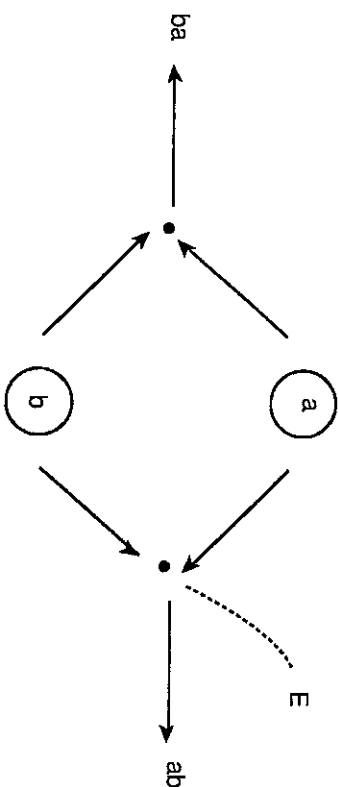


FIGURE 1 A simple network illustrating how steady-state concentration can be boosted by catalytic activity. a and b are driven at a fixed rate δ , and the enzyme E is maintained at a fixed concentration.

Note that the ability to focus comes about because the formation of one species is catalyzed, while that of the other is not. If all reactions were catalyzed equally, with equal kinetic parameters, there would be no focusing; the concentration of ab would equal that of ba . Focusing thus requires *specific catalysis*, in which some reactions are catalyzed more strongly than others.

2.7 AUTOCATALYTIC SETS AND METABOLISMS

To achieve catalytic focusing the enzyme E must be maintained at high concentration. One way for the system to accomplish this by itself is through an autocatalytic reaction, in which one of the products catalyzes its own formation. A simple example is



If we set $C = ab = E$ in reaction (13), then the enzyme is produced automatically, and the focusing maintains itself.

Simple autocatalytic reactions such as reaction (16) are obviously very special. A more common situation occurs when autocatalysis involves a cooperation between reactions, in which one species catalyzes the formation of another. An *autocatalytic set* is defined as a set of chemical species such that each member of the set is produced by at least one catalyzed reaction involving only members of the set. This notion was introduced by Calvin,⁶ Eigen,¹⁰ Kauffman,²³ and Rössler.³⁷ Since the reactions we are considering are reversible, a species can be produced either by cleavage or condensation. Thus reaction (16) is an autocatalytic set, and so is



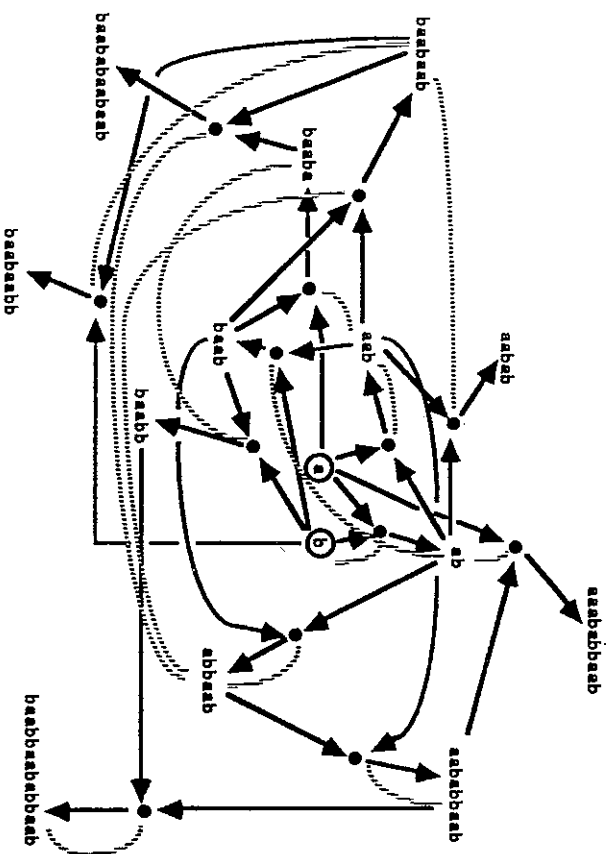


FIGURE 2 An autocatalytic network consisting of 15 species. The monomers a and b , which are circled, comprise the food set. Character strings represent the polymers. The fat dots represent reaction nodes. The arrows connecting the reactants to the nodes are reaction links and point in the direction of condensation. The broken lines connecting polymer species to reactions are catalytic links, indicating which species catalyze which reactions.

A more complicated (and more typical) autocatalytic set is shown in Figure 2. This network happens to have one catalytic link per reaction, a rather special property. We will use this reaction network, or variants with more catalytic links, for many of our numerical experiments. Note that autocatalytic sets may contain other autocatalytic sets as subsets.

Using the random graph model described in the next section, Kauffman²⁴ showed that for any given probability of catalysis, if the food set is sufficiently large, the resulting graph will almost always have an autocatalytic set. This result is encouraging, since it suggests that autocatalytic sets exist under fairly reasonable conditions. We must emphasize, however, that the presence of an autocatalytic set in a reaction network, in and of itself, does not imply that there will be any interesting departures from equilibrium behavior. To achieve catalytic focusing it is critical that the kinetic parameters are favorable. Thus the graph-theoretic notion of an autocatalytic set is a necessary but not a sufficient condition.

To make this distinction, we define an *autocatalytic metabolism* as an autocatalytic set whose concentrations make significant departures from the values they would have if none of the reactions were catalyzed. The phrase "significant departures" is subjective, and is admittedly rather vague. However, from an operational point of view, in our simulations we often see a clear distinction between autocatalytic sets that can function as metabolisms and those that cannot, as shown in Section 6.

3. SIMULATION

In principle the kinetic equations are all we need to know in order to simulate the behavior of a reaction network. In practice, however, there are two major problems: The first is that the kinetic parameters cannot be determined from first principles. To deal with this we construct an artificial chemistry, as discussed in subsection 3.1. The second problem is that there are an infinite number of possible reactions, and it is intractable to solve all of them; we must focus our computational resources on only the most relevant ones. Our method for doing this is discussed in subsection 3.2.

3.1 ARTIFICIAL CHEMISTRY

In a real chemical system the efficiencies and rate constants of the reactions depend on detailed properties of chemical composition, as well as on thermodynamic parameters such as temperature and pressure. While a computation of these constants from quantum mechanics and statistical mechanics is possible in principle, from a practical point of view, at this point in time it is hopelessly intractable.

To circumvent this problem, the approach introduced by Kauffman,²⁴ Farmer et al.,^{11,12} and Bagley et al.² is to invent an artificial chemistry, a set of rules stating which catalyzed reactions occur, and with what strength. An artificial chemistry cannot reproduce the behavior of a real chemistry in detail, but it may reproduce many of the correct qualitative properties. An artificial chemistry can produce complex behavior, even though it is simple from a calculational point of view. By exploring different artificial chemistries, we can discover which properties cause significant changes in behavior, and which do not. We can begin with simple chemistries and move toward more complex chemistries, adding layers of realism as needed. The knowledge gained in this way can be useful in guiding experimental investigations of real systems, by pinpointing the essential quantities that need to be measured in experiments in order to make the model more realistic.

Since our primary interest is in understanding the effect of catalysis, we first address the problem of assigning a catalytic efficiency to each reaction. We do this using two different methods. In the first, we construct a completely disordered artificial chemistry, by assigning catalytic efficiencies at random, and in the second,