Complex Autocatalysis in Simple Chemistries

Nathaniel Virgo*,**
University of Tokyo
and
Tokyo Institute of Technology

Takashi Ikegami†
University of Tokyo

Simon McGregor‡
University of Sussex

Abstract Life on Earth must originally have arisen from abiotic chemistry. Since the details of this chemistry are unknown, we wish to understand, in general, which types of chemistry can lead to complex, lifelike behavior. Here we show that even very simple chemistries in the thermodynamically reversible regime can self-organize to form complex autocatalytic cycles, with the catalytic effects emerging from the network structure. We demonstrate this with a very simple but thermodynamically reasonable artificial chemistry model. By suppressing the direct reaction from reactants to products, we obtain the simplest kind of autocatalytic cycle, resulting in exponential growth. When these simple first-order cycles are prevented from forming, the system achieves superexponential growth through more complex, higher-order autocatalytic cycles. This leads to nonlinear phenomena such as oscillations and bistability, the latter of which is of particular interest regarding the origins of life.

1 Introduction

How can abiotic chemistry give rise to phenomena such as replication, metabolism, and evolution? This question is at the core of the origins of life as a field of inquiry, and there are many approaches to it. One such methodology is *artificial chemistry*, the study of more-or-less abstract artificial chemical reaction networks and their emergent dynamical properties. This field can be traced back to the work of Fontana and Buss [5] and Kauffman [6]. More recently it has been used to show that evolution can occur without the need for template replication [12], and to investigate the mechanisms behind real reactions, particularly those that are too complex to model with traditional approaches, such as as HCN polymerization [1].

In these models, thermodynamic principles play at best a minor role. However, the nature of living organisms as far-from-equilibrium structures is fundamental to our understanding of biology at the chemical level. Our work aims to show that new insights can be gained by putting thermodynamics at the heart of artificial chemistry approaches.

Recently we presented a simple yet thermodynamically reasonable artificial chemistry [15]. When held out of equilibrium, this chemistry formed moderately large and complex autocatalytic cycles

* Contact author.
** Ikegami Laboratory, University of Tokyo, Japan, and Earth-Life Science Institute, Tokyo Institute of Technology, Japan. E-mail: nathanielvirgo@gmail.com
† Ikegami Laboratory, University of Tokyo, Japan. E-mail: ikeg@sacral.c.u-tokyo.ac.jp
‡ University of Sussex, U.K. E-mail: s.mcgregor@sussex.ac.uk

(that is, moderately large sets of species that are collectively capable of exponential growth). Unlike many previous studies, this chemistry did not include any a priori notion of catalysis. Instead, catalytic effects emerged as dynamical properties of the reaction network. The system was allowed to dynamically "choose" the direction of its reactions according to thermodynamic principles, and it was this property that allowed it to self-organize into autocatalytic cycles.

This phenomenon of self-organizing autocatalysis happens under some circumstances but not others. In order to fully understand the result, we must investigate which classes of reaction networks will lead to autocatalysis and which will not. Moreover, the resulting cycles can be more or less complex, and we wish to understand what factors affect this. A comprehensive answer to this question would be a great help in understanding the origins of life. If we know which types of chemical system are likely to lead to complex, metabolism-like reactions, then it will give us clues about what to look for in identifying the types of abiotic chemistry from which life eventually arose.

In this article we continue this project. By constructing a series of simple models, we first demonstrate that autocatalysis arises quite easily in a fairly simple class of chemistries. More importantly, we show that it must *necessarily* arise in any chemistry that meets a fairly mild set of preconditions. We demonstrate this by blocking the simplest form of autocatalysis in our models, and observing that autocatalysis still occurs, but this time through a more complicated, second-order mechanism.

As with all nonequilibrium phenomena, the autocatalysis in our systems is caused by the tendency of energy (and other conserved currents) to "flow downhill" from more to less constrained states, that is, from low to high entropy. Informally, it seems that nature likes to take the "path of least resistance" in order to achieve this. If the easiest path is a simple, direct chemical reaction, then no complex behavior will be observed. Without a direct pathway from initial conditions to equilibrium, autocatalytic pathways tend to be observed. Blocking the simplest type of autocatalytic cycle results in the formation of more complex cycles, giving rise to nonlinear phenomena such as bistability and oscillations.

Whether the easiest path toward equilibrium is given by a direct reaction, or a simple or more complex autocatalytic cycle, is determined by the topology of the reaction network. In this article we start with a simple polymerization chemistry and block reaction pathways simply by saying that certain molecules cannot exist. In real chemical systems the network topology depends on the details of molecular physics, which determines the set of molecules that can exist and the reactions that can occur between them.

Autocatalytic cycles can only emerge if they are possible within the reaction network. In this article we deal with reversible chemistries, in which the direction in which reactions occur is not fixed a priori. In this case we will see that even the simplest reaction schemes contain many autocatalytic subnetworks.

Briefly, our argument is that for a non-driven thermodynamic chemical system in the reversible regime, there exists a unique stable fixed point, namely the thermodynamic equilibrium. However (as we demonstrate), it is quite possible for there to be other fixed points, which must necessarily be unstable. At these nonequilibrium fixed points, some of the species' concentrations are zero. For trajectories in a region surrounding such a fixed point these concentrations are small but nonzero, and must increase at an increasing rate as the trajectory accelerates away from the repelling fixed point.

The results we present are very robust. With one minor exception (the oscillatory behavior shown in Figure 3), they require no parameter tuning at all. Moreover, our mathematical arguments are applicable to a much broader class of chemical systems than the specific models we present. Thus, although our model is relatively simple and abstract, we expect that in future work similar results will be found in much more realistic scenarios.

2   Thermodynamics in Artificial Chemistry

A glance at a chemistry textbook will reveal a large number of formulae that appear to be largely empirical in nature, and this can give the impression that chemical thermodynamics is something very contingent upon the details of molecular interactions. Perhaps because of this, there is a tendency

in artificial chemistry to "abstract away" thermodynamics, choosing a set of reactions arbitrarily, without regard to energetic considerations. This may be justified by saying that some of the reactions are driven by an energy source that is external to the system and is not explicitly modeled. This methodology dates back to the origins of artificial chemistry in the work of Fontana and Buss [5].

Such an abstraction is of course perfectly valid, but we hope to show that important insights can be gained if we choose not to make it. The dynamics of a system where there are only a few energy sources and the majority of reactions must "flow downhill" are quite different from those in which the reactions proceed in arbitrarily chosen directions. Moreover, the flow of free energy through the network can be tracked, giving us insights into its dynamics that would be missing if the energetics were not modeled.

Although chemical thermodynamics might appear to be a rather empirical set of results, its core principle (the second law) is an extremely fundamental property of physical systems. Essentially, it boils down to a requirement that for a system with no external supply of energy, there must exist a global Lyapunov function, which may be called the entropy or the free energy, depending on whether the system is connected to a heat bath. The existence of such a function is derived from statistical mechanics, and is a consequence of the time reversibility of physics on the microscopic level. The details of this formalism are beyond the scope of this article and may be found, for example, in [9]. We mention it in order to emphasize that our results are less contingent upon the details of molecular interactions than they might at first appear.

The key feature of a thermodynamically reasonable artificial chemistry is that the reaction network doesn't specify the *direction* of the reactions. If the network includes a reaction, such as A + B $\longrightarrow$ C, it must also include the reverse reaction, C $\longrightarrow$ A + B. In general, these two reactions may proceed at different rates. We refer to the difference in their rates as the *net rate* of the bidirectional reaction, A + B $\rightleftharpoons$ C.

The dynamics must be such that, in the absence of any energy sources, the system will eventually reach a state of *detailed balance* in which the net rate of every reaction is zero. That is, every forward reaction must be balanced, pairwise, by its corresponding reverse reaction. (This point is of fundamental importance. In a general artificial chemistry a dynamical equilibrium can be formed by a cycle, such as A $\longrightarrow$ B; B $\longrightarrow$ C; C $\longrightarrow$ A. In a thermodynamic system with no driven reactions this is a physical impossibility.) Detailed balance is a fundamental physical property of thermodynamically closed or isolated chemical systems at equilibrium. It need not apply to a non-isolated system that has some externally imposed flow of matter or energy, and in particular it doesn't hold in biological systems. Nevertheless, its importance cannot be overstated, because the constraints it imposes on the dynamics are felt even when the system is far from equilibrium.

The condition of detailed balance may be seen as a property of time symmetry held by the equilibrium state. Away from equilibrium this symmetry may be broken. Every net reaction must always proceed in the direction that takes the system closer to this equilibrium state. Thus, no reaction can occur at a nonzero net rate unless the system either is in a transient state or is being driven by an external power source.

In many real reactions the equilibrium ratio between products and reactants is so high that the reverse reaction is never observed in practice; in this case we think of reactions as occurring strictly in the "downhill" direction. In this article we are concerned with reversible chemistries, in which this is not the case for any reaction. The chemistry of the formose cycle is a real-world example of such a reversible network, in which autocatalysis does indeed emerge.

Many of the reactions in a living cell are reversible, but there are also many key irreversible steps. We strongly suspect that the main results of this article can be extended to chemistries with irreversible reactions under some appropriate set of additional assumptions; however, this remains a task for future work.

In our model we use a simple implementation of chemical dynamics, known as *mass-action kinetics*. Essentially this corresponds to a model of a well-mixed reactor in which molecules collide with each other at random. When two molecules collide, they react with a probability determined by the *rate constant* of the reaction. Thermodynamics puts constraints on the values of the rate constants; these constraints are satisfied by the dynamics described below.

3   The A-Polymerization Model

When implemented in a thermodynamically plausible way, even very simple reaction schemes can give rise to interesting behavior. Our model consists of *unary strings* A, AA, AAA, . . ., which we write as A_1, A_2, A_3, ... In principle these are countably infinite, but we impose a limit for computational reasons. These species can be thought of as (non-oriented) polymers, with A_1 (or simply A) representing a monomer, and A_i a polymer (or oligomer) of length i. All reactions are of the form

$$A_i + A_j \rightleftharpoons A_{i+j}. \tag{1}$$

That is, two oligomers may join together into a longer one, or an oligomer may split into two smaller ones. Reactions of this form conserve the total number of A monomers.

If every reaction of the form in Equation 1 is included in the network, then the results are not especially interesting. However, we can construct A-polymerization chemistries with more complex dynamics by removing particular reactions from the network. In real chemistry, reactions may be blocked (or happen at very slow rates) for a variety of reasons, such as conservation laws or steric effects. Our goal here is not to model such effects specifically, but rather to understand, in general, the effects of imposing constraints on the reaction network. We write $\mathcal{R}$ for the set of reactions included in the network, and we call the members of $\mathcal{R}$ *permitted* reactions.

A reaction $A_i + A_j \rightleftharpoons A_{i+j}$ may be thought of as two reactions occurring simultaneously: the *synthesis* reaction $A_i + A_j \longrightarrow A_{i+j}$ and the *decomposition* reaction $A_{i+j} \longrightarrow A_i + A_j$. Writing a_k for the concentration of species A_k, mass-action kinetics tells us that the synthesis reaction occurs at a rate $k_s a_i a_j$ and the decomposition reaction occurs at a rate $k_d a_{ij}$, where the constants k_s and k_d are known as the rate constants for the two reactions. This means that the reaction uses up A_i and A_j and generates A_{ij} at a (possibly negative) *net rate* given by $k_s a_i a_j - k_d a_{ij}$, which we denote $R_{A_i + A_j \rightleftharpoons A_{i+j}}$.

The constants k_s and k_d cannot be chosen arbitrarily, but instead are constrained by thermodynamics. In general they can be different for different reactions, but in our models we take k_s and k_d to be the same for every reaction, with one small exception, as explained below. (In reality, reaction rates can vary between reactions by many orders of magnitude, but we have made this assumption for the sake of simplicity. The key results of this article do not depend strongly upon this assumption.)

In general, a set of mass-action equations may be thought of as a meanfield approximation to a stochastic model. In our case we may think of the underlying stochastic picture as follows: There is a well-mixed container containing a solution of oligomers of various lengths. In order for two oligomers to join together, their ends must meet, and since every oligomer has two ends, we can assume the probability of this happening is proportional to the product of the two species' concentrations, regardless of the molecules' length. (This assumption is standard in the study of polymerization kinetics.) If the ends of two molecules of species A_i and A_j do meet, there are two possibilities. If the reaction $A_i + A_j \rightleftharpoons A_{i+j} \in \mathcal{R}$, then a bond is formed at the point where the two ends meet, and a new molecule of species A_{i+j} is created. Otherwise the reaction is not permitted, and nothing happens; the two molecules simply bounce off one another unchanged.

In addition to this process, a molecule may spontaneously split into two smaller molecules. An oligomer of species A_k consists of k A-monomers, linked together by $k - 1$ bonds. Each of these bonds has a constant probability per unit time of becoming unstable. If the ith bond of an A_k oligomer becomes unstable (with $i < k$), it will break if the reaction $A_i + A_{k-i} \rightleftharpoons A_k \in \mathcal{R}$, forming a molecule each of A_i and A_{k-i}. Otherwise, nothing happens, and the molecule remains unchanged.

In converting this to a set of differential equations, some care must be taken over reactions $A_i + A_j \rightleftharpoons A_{i+j}$ for which $i = j$. In the stochastic picture, a molecule of A_4 may split into A_1 and A_3 in two different ways: by breaking at its first bond or its third one (assuming such reactions are permitted). In contrast, there is only one way in which it can split into two molecules of A_2: by breaking

at its midpoint. Thus, reactions of the form $2A_i \rightleftharpoons A_{2i}$ should happen at half the rate of the other reactions.

However, as a notational convenience we can get around this by considering $A_i + A_j \rightleftharpoons A_{i+j}$ and $A_j + A_i \rightleftharpoons A_{i+j}$ to be two distinct reactions, which both happen to have the same set of products and reactants. Such pairs of reactions are always either both included in $\mathcal{R}$ or both excluded. With these conventions we may write $r_{i,j} = 1$ if the reaction $A_i + A_j \rightleftharpoons A_{i+j} \in \mathcal{R}$, and 0 otherwise. Mass action kinetics then leads to the ordinary differential equations

$$\frac{da_i}{dt} = \sum_{j=1}^{i-1} r_{i,j-i}\left(k_s a_j a_{i-j} - k_d a_i\right) \tag{2}$$
$$+ \sum_{j=1}^{\infty} \left(r_{i,j} + r_{j,i}\right)\left(k_d a_{i+j} - k_s a_i a_j\right) + \phi_i.$$

The term $k_s a_j a_{i-j}$ represents the formation of A_i through the joining of A_j to A_{j-i}; the term $k_d a_i$ represents the loss of A_i through splitting into A_j and A_{j-i}; the term $k_d a_{i+j}$ represents the gain in A_i due to A_{i+j} splitting into A_i and A_j; and $k_s a_i a_j$ represents the loss of A_i by joining to A_j to make A_{i+j}.

The ϕ_i terms represent the flux of each species A_i in or out of the system. These terms may be set to nonzero values in order to model an open system with external energy sources. In most of this article these terms will be set to zero, representing a closed system whose free energy comes purely from a nonequilibrium initial state.

Finally, we must show that these equations are thermodynamically reasonable. For this to be the case, we must be able to associate with every species A_k a constant number ψ_k, known as the Planck potential, such that, for every reaction $A_i + A_j \rightleftharpoons A_{i+j}$,

$$\frac{k_s}{k_d} = e^{\Delta\psi}, \tag{3}$$

where $\Delta\psi$ is the difference between the total Planck potential of the reactants and the products, given in our case by $\Delta\psi = \psi_i + \psi_j - \psi_{i+j}$. It can be shown that if Equation 3 holds, then the system has a Lyapunov function given by $\sum_k x_k(\psi_k + \log x_k)$, which corresponds to the Gibbs free energy in chemistry. To show that this is the case for our model, we note that Equation 3 tells us that $\psi_i + \psi_j - \psi_{i+j}$ is independent of i and j. This is satisfied by $\psi_k = B + Ck$ for any constants B and C, with $\Delta\psi$ being given by B.

The notation in terms of ψ_k is a nonstandard one that we have introduced in order to emphasize the simplicity of the formalism. In chemistry we would more usually talk about Gibbs energies of formation rather than Planck potentials, and in the standard notation our ψ_k would be written $\Delta_f G_k^\circ / k_B T$, where k_B is Boltzmann's constant and T the temperature. Equation 3 usually appears in chemistry textbooks in the form $k_s/k_d = \exp(\Delta_r G^\circ / k_B T)$. A comprehensive introduction to chemical thermodynamics, and the language used by chemists to describe it, may be found in [9].

In terms of Gibbs energies, we may further justify our use of constant k_s and k_d by assuming that each monomer has a constant Gibbs energy of formation $\Delta_f G_A^\circ$ and that forming each A–A bond involves a constant amount of energy E_{AA}. We then have $\Delta_f G_i^\circ = i\Delta_f G_A^\circ + (i-1)E_{AA}$. This gives $\Delta_r G^\circ = E_{AA}$ for every reaction, with the $\Delta_f G_A^\circ$ term canceling out because reactions neither create nor destroy monomers. We therefore have that

$$\frac{k_s}{k_d} = \exp\frac{E_{AA}}{k_B T},$$

which has the same value for every reaction, as desired.

We now have a recipe to construct chemical reaction networks out of simple synthesis and decomposition reactions. Despite the simplicity of the scheme, we will now show that the resulting networks can exhibit surprisingly sophisticated autocatalytic mechanisms.

4   The Linear Case: Self-organization of First-Order Autocatalysis

In this section we present a simple nonequilibrium A-polymerization chemistry that can be solved semi-analytically. This gives some insight into some sufficient conditions for exponential growth, and the reasons why it emerges under these conditions.

In this chemistry we set $r_{1,1} = 0$ and all other $r_{i,j} = 1$. That is, two monomers (A_1) are not allowed to directly join to form a dimer (A_2), and, vice versa, a dimer cannot directly split into monomers. All other reactions are permitted. There are multi-step reactions that can convert monomers into dimers in this system, such as the autocatalytic cycle

$$
\begin{aligned}
A_1 + A_2 &\longrightarrow A_3, \\
A_1 + A_3 &\longrightarrow A_4, \\
A_4 &\longrightarrow 2 A_2,
\end{aligned}
\tag{4}
$$

which has the net reaction $2 A_1 + A_2 \longrightarrow 2 A_2$. (Here, as below, we use unidirectional arrows to represent the net direction in which these reversible reactions proceed.) This is called a first-order autocatalytic cycle, because there is only one A_2 on the left-hand side. First-order autocatalysis leads to exponential growth, as we will see below. (See [7, 8] for background on the theory of autocatalytic cycles, and [2] for a formal definition.)

Let us suppose that the initial conditions contain a very high concentration of A_1 and trace amounts of every other species. That is, a_1 is large, and a_i very small for $i > 1$. Under these assumptions, the rate of change of a_1 will be small in comparison to its value, so that we can treat it as constant. Using the approximation $a_i a_j = 0$ for $i, j > 0$, Equation 2 becomes

$$
\frac{da_2}{dt} = 2\sum_{j=1}^{\infty} \left(k_d a_{2+j}\right) - 2k_s a_1 a_2,
$$

$$
\frac{da_i}{dt} = 2k_s a_1 a_{i-1} - (i-1)k_d a_i
\tag{5}
$$

$$
+ 2\sum_{j=1}^{\infty} \left(k_d a_{i+j}\right) - 2k_s a_1 a_i
$$

for $i > 2$. This may be written $d\mathbf{a}/dt = R\mathbf{a}$, where $\mathbf{a}$ is a column vector with elements $a_2, a_3, \ldots$, and R is the matrix

$$
k_d \begin{pmatrix}
-2K & 2 & 2 & 2 & \cdots \\
2K & -2K-2 & 2 & 2 & \cdots \\
0 & 2K & -2K-3 & 2 & \cdots \\
0 & 0 & 2K & -2K-4 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix},
\tag{6}
$$

with $K = a_1 k_s / k_d$.

It should be noted that although this approximation leads to a linear system of equations, it is not a "near-equilibrium" approximation in the thermodynamic sense. The thermodynamic equilibrium of this type of system will in general contain a positive concentration of every species. We are
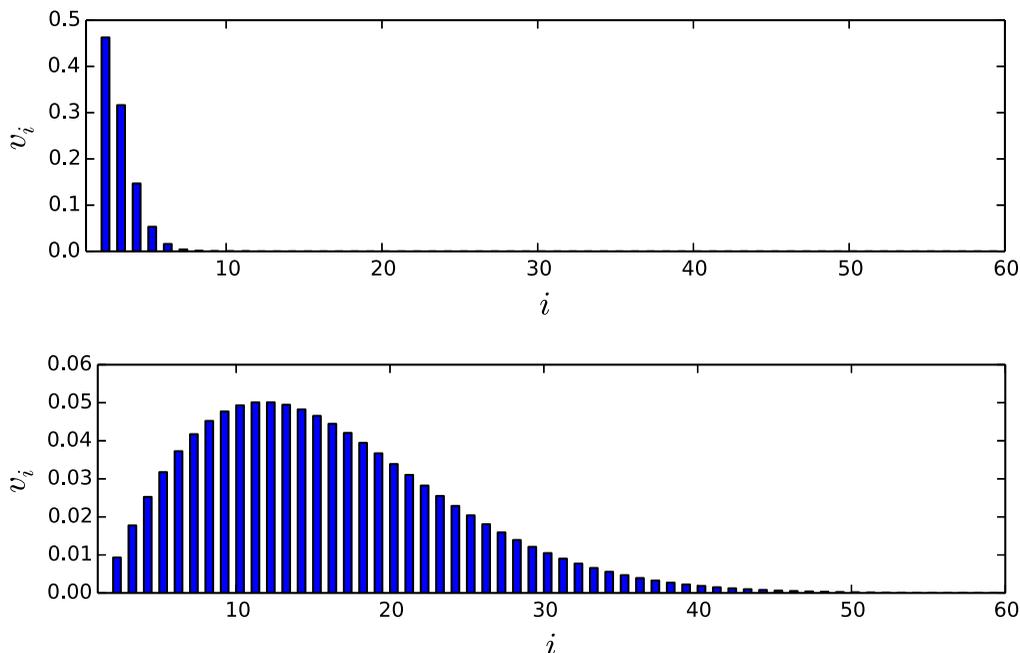
Figure 1. The elements of the leading eigenvector of the matrix R, for $K = 1.0$ (top) and $K = 100.0$ (bottom). The corresponding eigenvalues depend on k_d, but are positive in both cases, meaning that the concentrations grow exponentially while remaining proportional to the elements of the leading eigenvector. Both represent exponential growth, but through different mechanisms.

linearizing the system around the state where most of the species are at zero concentrations; this state is a dynamical fixed point, but it is very far from thermodynamic equilibrium.

Since this is a linear dynamical system, its asymptotic behavior is determined by the eigenvectors and eigenvalues of R. In particular, since the only negative elements of R are on the diagonal, R has only one eigenvector **v** with all positive entries (by application of the Perron–Frobenius theorem[1] to the irreducible matrix $e^{\varepsilon R} \approx I + \varepsilon R$ for sufficiently small ε), whose corresponding eigenvalue λ is real. For sufficiently large t we have $a_i(t) \approx C v_i e^{\lambda t}$, with C a real constant that depends on the initial conditions. In the case where every $a_i = 0$ (for $i > 1$), we have $C = 0$. For all other initial conditions, C is positive.

The eigenvector and corresponding eigenvalue can readily be found by power iteration of $e^{R^{(N)}}$, where $R^{(N)}$ is an $N \times N$ matrix whose entries are the same as those of R (with the exception that $R^{(N)}_{NN} = -N$, due to the absence of the reaction $A_N + A_1 \longrightarrow A_{N+1}$; for sufficiently large N this makes very little difference).

Figure 1 shows two examples. In both, the leading eigenvalue of R is positive, meaning that the species other than A_1 will grow exponentially. In the case where $K = 1$, this growth is mostly due to the minimal autocatalytic cycle shown in Equation 4, with a few side reactions producing longer polymers.

---

1 The Perron–Frobenius theorem states that if a matrix (a) has all non-negative entries and (b) is *irreducible*, meaning that the matrix can be seen as the adjacency matrix of a strongly connected graph, then the following conditions hold (among others): (i) the eigenvalue with the greatest magnitude is real, positive, and unique, and (ii) the eigenvector corresponding to this eigenvalue has all non-negative entries, and is the only eigenvector with this property. It is a general property of the matrix exponential that the eigenvectors of $e^{\varepsilon R}$ will be the same as those of εR, so we can conclude that R has only one eigenvector whose entries are all non-negative, which must then represent the asymptotic distribution of the species' concentrations.

However, the case with a larger K (meaning proportionally faster reaction rates in the synthesis direction, or just a higher concentration of A_1 initially) is dominated instead by larger molecules, and its growth is therefore due to much larger autocatalytic cycles. There is no one cycle that dominates; instead the microscopic picture is of a collection of polymers of many different lengths, which grow at their ends due to monomer addition, but which also occasionally split somewhere in the middle, giving rise to two new polymers.

It is a theorem in mass-action kinetics that there is a global Lyapunov function given by the Gibbs energy, whose minimum is at the thermodynamic equilibrium point. Since we are linearizing around a different fixed point, this guarantees that it will not be stable; at least one of its eigenvalues must be non-negative. Because of this, there are no parameters that have to be tuned in order to observe autocatalysis in this model. For any value of K there will always be a region around the point $a_2 = a_3 = \cdots = 0$ in which autocatalytic growth occurs.

To derive these results we assumed that a_i was small for $i > 1$, but since the concentrations of these are increasing exponentially, this assumption will be violated eventually. This means that either a_1 will start to decrease at a substantial rate, changing the value of K, or terms of the form $a_i a_j$ will start to be important, changing the dynamics. In a real system, this may or may not happen before the exponential growth becomes manifest enough to be clearly observable.

It is worthwhile examining the assumptions we used to obtain this result. One important part of the model is that the reaction $2A_1 \rightleftharpoons A_2$ is removed from the reaction network. The reason for this is that this reaction would rapidly produce enough A_2 that the assumption of small a_2 would be violated. We propose that in general any such direct route to product formation must be absent, or at least occur only slowly, in order for autocatalysis to be observed.

The other important assumption was that the polymer species have (initially) very low concentrations, so that the $a_i a_j$ terms vanish. Physically, this corresponds to a solution so dilute that two polymer molecules are vanishingly unlikely to meet and react. (A_1 monomers, however, are highly concentrated, so the probability of reacting with a monomer is high.) Under these conditions, exponential growth can occur through the molecules (a) growing into larger molecules by reacting with monomers and (b) spontaneously splitting into two smaller molecules.

Such dynamics can be expected in much more complex thermodynamically reasonable chemistries, as long as these conditions are obeyed; our reasoning is applicable to a much broader class of models than the one presented here. The chemistry need not take the form of A-polymerization. This reasoning will hold as long as there is some set of *food* species that cannot directly react to form products, and so long as there are no energetically irreversible sinks in the network. No autocatalytic cycles can arise unless they already exist in the network of potential reactions, but our example shows that for even such a simple chemistry as A-polymerization there are many such potential loops. We suspect these conditions will be satisfied by many chemistries, as long as they are sufficiently densely connected and all the reactions have some degree of reversibility.

The mechanism as described so far results in *exponential* growth. Equations 5 is a linear dynamical system, and therefore the only possibilities are exponential growth, exponential decay, and neutral stability. (Oscillation around the initial point is impossible, because it would require concentrations to become negative.) The more general assumptions outlined in the previous paragraph will also lead to linear dynamical systems, resulting either in exponential growth or decay. In the next section we explore the importance of nonlinearity in chemical networks, and we demonstrate that with some changes to the reaction network, self-organization can produce superexponential (or *hyperbolic*) growth in a simple A-polymerization chemistry.

5   Nonlinearity and Superexponential Growth

The constituents of a living cell are not capable of exponential growth at low concentrations; one of the many functions of the cell membrane is to keep the concentrations high enough for growth to occur. If the components of a living cell were to be placed in a dilute solution, it would be unlikely

for an enzyme to collide with its substrate in order to bind to it, and therefore enzyme-catalyzed reactions would occur at vanishingly low rates.

Exponential autocatalysis is interesting from an origins-of-life point of view [15], precisely because it doesn't require high concentrations, and therefore could exist before the evolution of the cell membrane or other forms of compartmentalization. However, the linear analysis in the previous section makes it clear that the concentrations will always converge towards the *same* profile, regardless of the initial conditions (as long as they obey the low-concentration assumption). This seems to leave no room for evolution: There is no way one exponential replicator can outcompete another.

Many models of prebiotic chemistry are nonlinear, because they involve reactions where two non-food molecules must collide with each other. These include all models based on explicit catalysis, such as the work of Kauffman [6] or the hypercycle model of Eigen and Schuster [4].

For these reasons it is worth exploring under which circumstances nonlinear, superexponential growth can emerge spontaneously. In this section we demonstrate circumstances in which this can happen in A-polymerization chemistries. The key is to constrain the reaction network so that first-order, exponential autocatalytic cycles cannot occur. Under these circumstances, growth occurs if we assume that reactions of the form $a_i a_j$ (for $i, j > 0$) cannot occur. However, when integrating the model numerically, we find autocatalytic cycles that rely on these reactions. Their kinetics are nonlinear (in fact superexponential), because they involve these quadratic terms. We show that if the system is modeled as a flow reactor, this nonlinearity can give rise to behaviors such as bistability and oscillations.

The exponential growth in the previous section occurs because a molecule can split into two shorter molecules, each of which can then grow through monomer addition until it reaches the length of the original molecule, giving a net reaction $A_n + nA_1 \longrightarrow 2A_n$. Now we ask what happens if we prevent this simple first-order autocatalysis mechanism while retaining many of the possible reactions from the original network. We show that autocatalysis still self-organizes, but it does so through a more complex, second-order mechanism that results in superexponential growth.

In order to prevent this, we constrain the reaction network in the following way: We define a set of *banned* species B. A reaction $A_i + A_j \rightleftharpoons A_{i+j}$ is included in the reaction network *unless* either $A_i \in B$, $A_j \in B$, or $A_{i+j} \in B$.

This technique of banning species is merely a convenient way to prevent the formation of first-order autocatalytic cycles while retaining much of the original network; it is not supposed to directly represent something one would expect to happen in polymerization chemistry. However, in more complex chemistries a species may be stiochiometrically possible but have such a high free energy that it is never formed, and so networks with broadly this type of structure might not be entirely absent from real chemistry. We suspect that there are also many other reasons why a complex network would fail to contain first-order loops.

To see how banning species can prevent first-order autocatalysis, let us imagine that $B = \{A_{2^{i+1}} : i \in \mathbb{Z}\}$. That is, molecules whose lengths are powers of two are banned, apart from A_1. A molecule of A_n (with n not a power of two) may split into two smaller molecules, A_m and A_{n-m}, as long as neither m nor $n - m$ is a power of two. However, at least one of these molecules is no longer able to grow by monomer addition until it reaches length n. For let $r = 2^{[\log_2 n]}$ be the largest power of two less than n. Then either $m < r$ or $n - m < r$, since $r \geq n/2$. A molecule of size less than r cannot grow to size n by monomer addition alone, due to the absence of the reaction $A_{r-1} + A_1 \longrightarrow A_r$. Therefore net reactions of the form $A_n + nA_1 \longrightarrow 2A_n$ are impossible in this chemistry.

By excluding even more species we can obtain quite complex behavior. Here we present results from a system that excludes the species $B = \{A_{3i} : i \in \mathbb{Z}\}$, and also excludes the reaction $2A_1 \rightleftharpoons A_2$. For computational reasons we also exclude species longer than 60 monomer units.

If, as in the previous section, we linearize this system around the point where all non-food species have zero concentration, we find that the leading eigenvalue is zero. In such cases the linear stability analysis is inconclusive, and the stability of the fixed point is determined by the
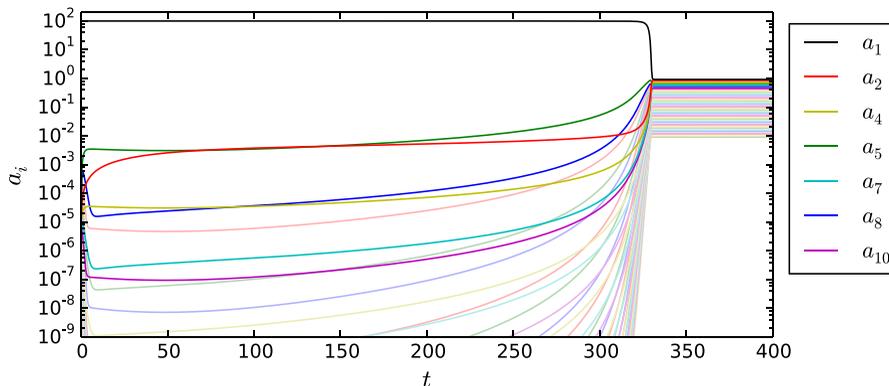
Figure 2. The results of numerically integrating an A-polymerization chemistry over time. This chemistry excludes the species $B = \{A_{3i} : i \in \mathbb{Z}\} \cup \{A_i : i > 60\}$, and in addition excludes the reaction $2A_1 \rightleftharpoons A_2$. The parameters used are $k_s = k_d = 1$. The initial concentration of A_1 is 100, and the initial concentrations of all other species are 10^{-5}. (Many species initially decay to much lower concentrations, only to recover later.) The curves appear roughly exponential on a logarithmic scale, indicating superexponential growth occurring through higher-order autocatalysis. When plotted on a linear scale, most species' concentrations appear negligible until the sudden transition to the equilibrium state at around $t = 330$.

nonlinear terms. The theoretical apparatus for dealing with such cases is called center manifold theory. However, for our purposes we can simply note that we know there is a global Lyapunov function (the Gibbs energy) whose minimum is at a different point, the thermodynamic equilibrium. We should therefore expect the nonlinear terms to result in a second-order instability around this fixed point, regardless of the details of how this comes about. Our numerical results in this section show that this is indeed the case.

Figure 2 shows the results of numerically integrating this system, using Equation 2 with no approximations, from an initial condition consisting of a high concentration of A_1 and small amounts of everything else. The result is that at around $t = 450$ there is a sudden transition to the equilibrium state, in which all species are present, exponentially distributed according to their lengths. These dynamics indicate that the species other than A_1 grow (super)exponentially at low rates until their concentrations become high enough for the growth's effects to be felt. This is similar to the results from first-order autocatalytic systems presented in [15]. However, the mechanism is rather more complicated.

At time $t = 100$, the growth in the system is largely due to the following set of reactions:

$$
\begin{aligned}
A_8 &\longrightarrow 2A_4, \\
A_4 + A_1 &\longrightarrow A_5, \\
A_4 &\longrightarrow 2A_2, \\
A_5 + A_2 &\longrightarrow A_7, \\
A_7 + A_1 &\longrightarrow A_8.
\end{aligned}
\tag{7}
$$

These reactions may be stiochiometrically balanced to form the autocatalytic net reaction $3A_8 + 8A_1 \longrightarrow 4A_8$. (That is true even though these are not the only reactions in the system, and they are not necessarily happening at stoichiometrically balanced rates, because the concentrations are changing over time. See Table 1 for the true reaction rates.) The nonlinearity comes from the reaction $A_5 + A_2 \longrightarrow A_7$, which gives rise to a second-order nonlinear term because it requires an A_5 to collide with an A_2 dimer in order to proceed; it will therefore increase in rate as the concentrations of these species increase.

Table 1. The top ten reactions in the system shown in Figure 2, at time 100.

Reaction	Rate
$A_4 + A_1 \longrightarrow A_5$	3.194×10^{-5}
$A_7 + A_1 \longrightarrow A_8$	1.888×10^{-5}
$A_8 \longrightarrow 2A_4$	1.882×10^{-5}
$A_4 \longrightarrow 2A_2$	1.366×10^{-5}
$A_5 + A_2 \longrightarrow A_7$	1.084×10^{-5}
$2A_5 \longrightarrow A_{10}$	8.050×10^{-6}
$A_{10} + A_1 \longrightarrow A_{11}$	8.021×10^{-6}
$A_{11} \longrightarrow A_7 + A_4$	7.942×10^{-6}
$A_{13} + A_1 \longrightarrow A_{14}$	9.351×10^{-8}
$A_{10} \longrightarrow A_8 + A_2$	8.598×10^{-8}

Notes. The rates shown are absolute net rates, in moles per unit time.

By $t = 400$ the system is dominated by the following overlapping but different set of autocatalytic reactions, as can be seen in Table 2:

$$
\begin{aligned}
A_{11} &\longrightarrow A_7 + A_4, \\
A_7 + A_1 &\longrightarrow A_8, \\
A_8 &\longrightarrow 2A_4, \\
A_4 + A_1 &\longrightarrow A_5, \\
2A_5 &\longrightarrow A_{10}, \\
A_{10} + A_1 &\longrightarrow A_{11}.
\end{aligned}
\tag{8}
$$

This is a second-order autocatalytic cycle with the net reaction $2A_{11} + 11A_1 \longrightarrow 3A_{11}$. Where the cycle in Equation 7 produces A_7 by splitting A_4 into A_2 and then joining it with A_5, this network instead produces A_7 by forming and then splitting A_{11}. The transition from Equation 7 to 8 is probably the result of the general increase in concentrations, making the reaction $2A_5 \longrightarrow A_{10}$ more likely to proceed. In this network, $2A_5 \longrightarrow A_{10}$ is the second-order reaction that gives rise to a nonlinear term.

We reiterate that this nonlinear autocatalytic increase in concentrations is a logical consequence of the reaction network's structure. The values of the numerical parameters are more or less arbitrary; the only tuning required is to set the initial concentrations small enough for the effect to be observed.

This nonlinearity can lead to bistability, which is important because it provides an analogue of biological death, which in turn is necessary in order for one system to be able to outcompete another. Suppose that additional reactions are added such that the polymer species decay exponentially (and irreversibly) into inert products. Then at low concentrations the decay reactions will be faster than the autocatalytic cycle and the polymer species will not be able to persist in the system. However, at higher concentrations the autocatalytic species will grow at a faster rate than their decay. Thus the polymer species can persist in the system only if they are initially present

Table 2. The top ten reactions in the system shown in Figure 2, at time 400.

Reaction	Rate
$A_4 + A_1 \longrightarrow A_5$	3.745×10^{-3}
$A_7 + A_1 \longrightarrow A_8$	1.549×10^{-3}
$A_8 \longrightarrow 2A_4$	1.429×10^{-3}
$A_{10} + A_1 \longrightarrow A_{11}$	1.202×10^{-3}
$2A_5 \longrightarrow A_{10}$	1.150×10^{-3}
$A_{11} \longrightarrow A_7 + A_4$	1.088×10^{-3}
$A_5 + A_2 \longrightarrow A_7$	3.669×10^{-4}
$A_4 \longrightarrow 2A_2$	2.466×10^{-4}
$A_{13} + A_1 \longrightarrow A_{14}$	7.950×10^{-5}
$A_8 + A_5 \longrightarrow A_{13}$	3.669×10^{-5}

Notes. The overall difference in magnitudes compared to Table 1 is due to increases in concentrations over time.

in concentrations above a certain threshold. This is in contrast to the case of exponential growth, where the zero concentrations state is either a stable or an unstable fixed point, depending on whether the growth rate of the autocatalytic cycle is greater or less than its decay rate. If the zero-concentrations point is unstable, a single molecule can start the growth process and will eventually dominate the system.

The existence of a concentration threshold leads to a kind of *precariousness* that is an important property of lifelike systems, since it is analogous to biological death [13, 3]. Precariousness means
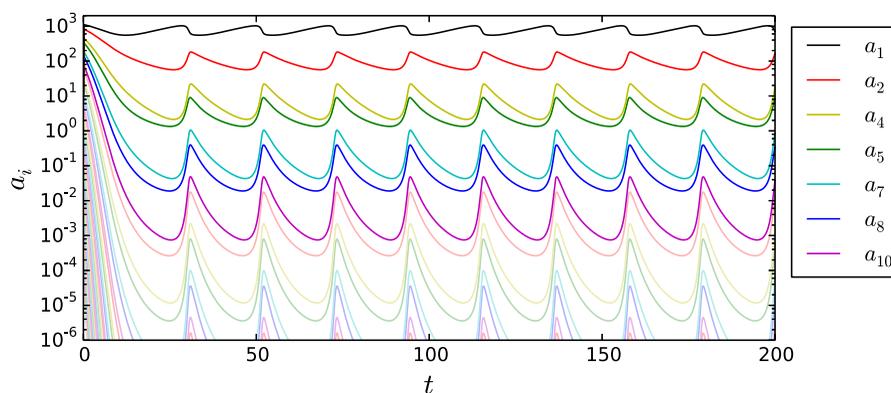

Figure 3. Convergence to a limit cycle due to a higher-order autocatalytic cycle in a flow reactor. The reaction network is the same as for Figure 2. The flow reactor is modeled with the flux terms $\phi_1 = 0.01 (10000 - a_1)$ and $\phi_i = -0.4a_i$ for $i > 1$. The parameters used are $k_s = 1.0$ and $k_d = 1500$. The autocatalysis is primarily due to the reactions shown in Equation 7.

there are some perturbations from which a system cannot recover; in this case some perturbations push the concentrations below the threshold, and the concentrations must then eventually decay to nothing.

This same kind of nonlinearity can also lead to a more easily visualized phenomenon, namely temporal oscillations, as shown in Figure 3. To obtain this result, the reaction is modeled as if in a flow reactor. There is a constant input of A_1 and exponential decay of every other species, which can be thought of as leaving the system in the reactor's outflow. The oscillations occur through a similar mechanism to the famous lynx–hare cycle in ecology: The autocatalytic cycle consumes too much of its food, and as a result its constituent species decay somewhat. This gives the food concentration a chance to recover (since it is being continually fed into the system), and the autocatalytic species once again increase in concentration, completing the cycle.

This oscillatory behavior does not occur for every choice of parameter values. However, it demonstrates nicely that the nonlinearity caused by higher-order autocatalysis can lead to phenomena that are not possible in the linear, first-order case. We reiterate that all of the other results in this article are very robust and require no parameter tuning at all.

6   Analysis and Conclusions

The second law of thermodynamics implies that physical systems must stochastically tend to "flow downhill," where "downhill" means in the direction of increasing entropy, or decreasing free energy. We have implemented this feature in a very simple class of chemical reaction networks and found that it can lead to self-organization of complex autocatalytic cycles.

This is in some way analogous to the self-organization of dissipative structures in physical systems, such as convection cells. These also involve the formation of cycles, and also increase the rate at which energy is dissipated.

Our informal, intuitive understanding of these results is as follows: Under some (but by no means all) circumstances, physical systems far from equilibrium not only "flow downhill," but also tend to seek out faster and faster ways to do this. The formation of convection cells is again an example of this; the breaking of a dam is perhaps a more vivid one. The *full* A-polymerization reaction network (with all reactions included) will flow to its equilibrium rather rapidly. However, if we block the most direct route (the $2A_1 \rightleftharpoons A_2$ reaction), then an indirect route must be found, by taking advantage of catalytic cycles. Blocking the zeroth-order route leads to the self-organization of first-order autocatalytic cycles, because (in a very informal sense) these lead to faster rates of dissipation.

Blocking these first-order solutions then leads to higher-order solutions. It is as if the system tries to find the easiest route towards its equilibrium, and only once the simplest paths are blocked does it find the more complex ones. We were surprised by the elegance and novelty of the networks that emerged in the system, especially in comparison with second-order networks that we had tried to design by hand.

From a more formal point of view, these results occur because the initial conditions are near a fixed point that is not the thermodynamic equilibrium point, and therefore cannot be stable. Creating such a far-from-equilibrium fixed point was simply a case of removing the direct reaction from the network, so that monomers could not join together to form polymers in the absence of other polymers. This suggests that many reaction networks should share this property of exhibiting a non-equilibrium fixed point close to an experimentally realizable initial condition. We expect autocatalysis to be empirically observable in such situations.

By preventing first-order instability, we forced the initial fixed point to exhibit a second-order instability. The nonlinearity of the resulting systems leads to phenomena such as precariousness and oscillations. A nonlinear system with many equilibria is more interesting from an origins-of-life point of view, because it is closer to a system in which multiple different autocatalytic systems can compete with one another, leading to evolution.

It is interesting to note that we found complex behavior in this system by removing reactions rather than by adding them. If we consider the full A-polymerization chemistry, including both forward and reverse reactions, all of these autocatalytic cycles—both first- and higher-order—are already present as its subnetworks. Whether the more complex networks become manifested or not is a matter of whether the right constraints exist to prevent simpler networks from being manifested instead.

Many authors in the origins of life have worried about the problem of *side reactions* that can prevent autocatalysis from occurring (e.g., [11, 8, 14]). Our results suggest that under the right energetic conditions, such issues might not arise. Our reaction networks contain many possible reactions besides the ones that eventually form the autocatalytic cycles, yet these do not seem to disrupt the autocatalysis to any meaningful extent. The reason is that any molecules produced by side reactions eventually end up participating in autocatalytic cycles of their own. We therefore make the prediction that when one complex autocatalytic cycle is observed experimentally, it will tend to be accompanied by many others.

A closely related issue in the origins of life is the combinatorial explosion that can arise from several small molecular components combining in different ways. Our model is incomplete in regard to this issue—we have monomers of only one type, and they combine only linearly—but nevertheless our results suggest that the combinatorial explosion might not be as serious as it might at first seem. One can imagine a class of molecules that can grow in multiple ways, leading to a combinatorial explosion, but that can also undergo reactions that decompose them into smaller molecules. With such a chemistry, at low concentrations one should still expect exponential growth through the mechanism that arises in the first of our models. We therefore suggest that the future of origins-of-life models might be not in taming the combinatorial explosion [10] but in embracing it. If every molecule forms part of an autocatalytic cycle, then the combinatorial explosion represents an efficient way to explore the space of possible replicators, and it may be that it was through such a process that the first high-fidelity replicators arose.

This research represents a step towards understanding what properties a chemical reaction network must have in order for complex, lifelike dynamics to emerge. There is still a lot of work to do to understand the effects of network topology and energetics on the emergence of complex dynamics, but the results in this article represent a substantial step in the right direction.

Acknowledgments

References

1. Andersen, J. L., Andersen, T., Flamm, C., Hanczyc, M., Merkle, D., & Stadler, P. F. (2013). Navigating the chemical space of HCN polymerization and hydrolysis: Guiding graph grammars by mass spectrometry data. *Entropy*, *15*, 4066–4083.

2. Andersen, J. L., Flamm, C., Merkle, D., & Stadler, P. (2012). Maximizing output and recognizing autocatalysis in chemical reaction networks is NP-complete. *Journal of Systems Chemistry*, *3*(1), 1–9.

3. Barandiaran, X., & Egbert, M. (2013). Norm-establishing and norm-following in autonomous agency. *Artificial Life*, *20*(1), 5–28.

4. Eigen, M., & Schuster, P. (1979). *The hypercycle: A principle of natural self-organisation*. Berlin/Heidelberg: Springer.

5. Fontana, W., & Buss, L. W. (1994). What would be conserved if "the tape were played twice"? *Proceedings of the National Academy of Sciences*, *91*, 757–761.

6. Kauffman, S. (1986). Autocatalytic sets of proteins. *Journal of Theoretical Biology*, *119*, 1–24.

7. King, G. A. M. (1978). Autocatalysis. *Chemical Society Reviews*, *7*(2), 297–316.

8. King, G. A. M. (1982). Recycling, reproduction and life's origins. *BioSystems*, *15*, 89–97.

9. Kondepudi, D., & Prigogine, I. (1998). *Modern thermodynamics: From heat engines to dissipative structures*. Hoboken, NJ: Wiley.

10. Schuster, P. (2000). Taming combinatorial explosion. *Proceedings of the National Academy of Sciences*, *97*(14), 7678–7680.

11. Szathmáry, E. (2000). The evolution of replicators. *Philosophical Transactions of the Royal Society, Series B*, *355*, 1669–1676.

12. Vasas, V., Fernando, C., Santos, M., Kauffman, S., & Szathmáry, E. (2012). Evolution before genes. *Biology Direct*, *7*(1).

13. Virgo, N. (2011). *Thermodynamics and the structure of living systems*. Unpublished doctoral dissertation, University of Sussex, UK.

14. Virgo, N., Fernando, C., Bigge, B., & Husbands, P. (2012). Evolvable physical self-replicators. *Artificial Life*, *18*, 129–142.

15. Virgo, N., & Ikegami, T. (2013). Autocatalysis before enzymes: The emergence of prebiotic chain reactions. In P. Liò, O. Miglino, G. Nicosia, S. Nolfi, & M. Pavone (Eds.), *Advances in artificial life, ECAL 2013: Proceedings of the Twelfth European Conference on the Synthesis and Simulation of Living Systems* (pp. 240–247). Cambridge, MA: MIT Press.