ACCOUNTS
of chemical research

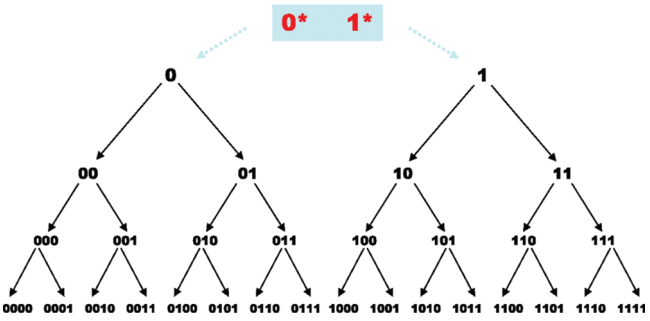
From Prelife to Life: How Chemical Kinetics Become Evolutionary Dynamics

IRENE A. CHEN*,† AND MARTIN A. NOWAK*,‡

†FAS Center for Systems Biology, ‡Program for Evolutionary Dynamics, Department of Mathematics, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States

RECEIVED ON OCTOBER 20, 2011

CONSPECTUS



Life is that which evolves. Living systems are the products of evolutionary processes and can undergo further evolution. A crucial question for the origin of life is the following: when do chemical kinetics become evolutionary dynamics? In this Account, we review properties of "prelife" and discuss the transition from prelife to life. We describe prelife as a chemical system where activated monomers can copolymerize into macromolecules such as RNA. These macromolecules carry information, and their physical and chemical properties depend to a certain extent on their particular sequence of monomers. We consider prelife as a logical precursor of life, where macromolecules are formed by copolymerization, but they cannot replicate. Prelife can undergo "prevolutionary dynamics", including processes such as mutation, selection, and cooperation. Prelife selection, however, is blunt: small differences in rate constants lead to small differences in abundance. Life emerges with the ability of replication. In the resulting evolutionary dynamics, selection is sharp: small differences in rate constants can lead to large differences in abundance.

We also study the competition of different "prelives" and find that there can be selection for those systems that ultimately give rise to replication. The transition from prelife to life can occur over an extended period of time. Instead of a single moment that marks the origin of life, prelife may have seeded many attempts for the origin of life. Eventually life takes over and destroys prelife.

Introduction

Imagine an aqueous solution of small molecules on the early earth. Now try to picture how that prebiotic soup might assemble itself into even the simplest, tiniest living organism, perhaps a few hundred nanometers across. At first glance, this process may seem like an impossible leap because so many transitions must occur to transform the jittery molecules into a living structure. To understand the origin of life, one must break it down into a series of smaller transitions and look for simple ways that physical and chemical effects could accomplish each transition. One successful synthetic approach is to focus on the emergence of structures: the synthesis of monomers, polymerization of monomers into sequences, the formation of protocells by membrane encapsulation of sequences, and so forth. Significant experimental work has been directed at producing ribozymes and protocells and is reviewed elsewhere.[1,2] Substantial theoretical work has also been particularly directed toward understanding the emergence of well-folded RNA.[3,4] But a complementary viewpoint that comes naturally from a mathematical perspective is to study the emergence of dynamics that accompany the structural transitions.

Perhaps the most well-known study of chemical evolution was introduced in the 1970s by Manfred Eigen and Peter Schuster,[5−8] who described a population of sequences

Published on the Web 02/15/2012 www.pubs.acs.org/accounts
10.1021/ar2002683

undergoing mutation and selection. Because of mutation, the population typically exists as a "quasispecies", a collection of similar sequences sometimes centered around a fittest sequence.[9−13] The standard approach to quasispecies theory studies evolution in a population of sequences all of which have the same length. The theory does not explore how longer sequences arise from shorter ones or how the ability of replication emerges. In contrast, this is the topic of our current investigation. Therefore we examine an earlier phase in the origin of life where simple polymerization chemistry generates sequences of variable length and some of those sequences might have the ability to replicate. Our approach is a logical precursor to the celebrated quasispecies theory.

In this Account, we highlight our theoretical work on understanding how the chemical kinetics of RNA polymerization become the competitive, evolutionary dynamics of replicators.[14−17] Although the origin of life included several intermediate forms of increasing complexity from a prebiotic soup to a definite living organism, we refer to the system of polymerization as "prelife" and the system of replicators as "life" for a descriptive shorthand. Many possible definitions of life have been proposed (reviewed in refs 18−20, some are discussed in refs 21−26). Most definitions, such as the widely used working definition put forward by the Exobiology Program at NASA,[27,28] include self-replication and evolution as important features. In this paper, "life" refers to a chemical system that includes template-based replication, and the ability to evolve is a consequence of replication errors and the differential fitness of mutants. The underlying goal of these studies is to build the simplest possible model that captures the emergence of evolutionary dynamics. In a sense, this is analogous to the bottom-up approach for synthesizing life-like entities. Synthetic biologists hope to learn about the origin of life by combining a minimal set of molecules to construct an entity that satisfies an operational definition of life (e.g., a self-replicating chemical system capable of Darwinian evolution[28]). Similarly, the goal of this line of mathematical biology is to find a minimal set of chemical equations that describes a system with life-like evolutionary dynamics. There are at least two motivations for this minimalist style of approach: (1) the first living entities must have been quite simple and obeyed the laws of chemical kinetics, and (2) if there are relatively few components to the model, we can best understand the principles behind its emergent properties. The models, whether experimental or theoretical, may not be perfect mimics of a natural system, but we hope that they capture essential features and thereby improve our fundamental understanding of the origin of life.

The following sections describe (1) the chemical kinetics and equilibrium distribution of a polymerizing system, or prelife, (2) the dynamical changes that occur if the system has the ability to self-replicate (e.g., through nonenzymatic, templated synthesis), and (3) the competition between systems with different properties (e.g., different polymer types). Prelife is characterized by gentle changes in the abundance of different sequences in response to differences in reactivity. Such a response of the system would be familiar to those who study chemical systems. On the other hand, if the polymers are able to template and thereby self-replicate, the dynamics change abruptly, and the fittest sequences dominate the pool in large excess even if they are only slightly better replicators than the rest. And if two systems compete for resources, one can exclude the other. Such features would be familiar to those who study biological systems.

Prelife

The prebiotic synthesis of small organic building blocks has been investigated since the Miller−Urey experiment nearly 60 years ago, which produced amino acids and other compounds from a gaseous mixture meant to simulate the atmosphere of the early earth.[29] Since then, the field has advanced considerably and now includes the aqueous synthesis of RNA nucleotides.[30] Condensation of these nucleotides can also be achieved under plausibly prebiotic conditions promoted by various means, including divalent cations, clay or lipid surfaces, and solute concentration by freezing conditions.[31−34] Our model of prelife presumes the availability of activated monomers and the presence of conditions conducive to polymerization.

In the spirit of studying a model system that is as simple as possible while retaining key features, we consider binary sequences that grow by the addition of monomers one at a time.

Let us consider a prebiotic chemistry that produces activated monomers, 0* and 1*, that can copolymerize into binary sequences according to the following chemical reactions:

$$i + 0^* \rightarrow i0$$
$$i + 1^* \rightarrow i1 \qquad (1)$$

Here i denotes any binary sequence of any length, while $i0$ and $i1$ denote the extension product of i by addition of a

monomer of type 0 and 1, respectively. We assume that all sequences grow only in one direction, corresponding to the orientation of a nucleic acid or polypeptide sequence (e.g., polymerization in the 5′ to 3′ direction). Therefore each sequence, i, has one precursor, which we denote by i', and two followers, $i0$ and $i1$. Each sequence is produced by a particular unique "lineage" in prelife. For example, the lineage that leads to sequence 0100 is $0 \rightarrow 01 \rightarrow 010 \rightarrow 0100$. The reaction scheme 1 can generate any binary sequence.

The rate at which sequence i is generated from its precursor is given by a_i. If the polymerization rate were independent of the sequence, all a_i would be the same. However, in reality the polymerization rate does depend on the reactants, and therefore some sequences may extend faster than others.

For example, polymerization of RNA monomers on montmorillonite clay exhibits definite preferences. The rate of extension seems to be highest if the primer contains a purine at the 3′ end and the activated monomer is a pyrimidine.[35] The yields from different combinations of reactants in the montmorillonite-promoted system cover nearly 2 orders of magnitude. Similarly, in RNA polymerization catalyzed by Zn^{2+}, the monomer A reacts about 10 times faster than U.[31] One prebiotic polymerization reaction gave a relatively even ratio of incorporation of different nucleotides (within 2-fold), but this was likely because the reaction was carried out for a long time, such that most monomers would be incorporated even if the reactivities were biased.[34] Variation in a_i is apparent in more widely used synthetic reactions as well. For example, in the synthesis of degenerate oligonucleotides using a mixed pool of nucleoside phosphoramidites, the reaction efficiency depends on the identities of the phosphoramidite and the nucleotide on the 5′ terminus of the elongating oligonucleotide.[36]

The list (a_0, a_1, a_{00}, a_{01}, ...) specifies the "prelife landscape" of sequence formation. Some route for removal of sequences is required to avoid having the population grow infinitely large. In a real situation, this would correspond to RNA degradation or a fluid flow that removes sequences from the reacting pool. For simplicity, let us assume that all sequences are removed from the population at the same rate, d.

The following system of infinitely many differential equations describes the chemical kinetics of prelife:[14]

$$\dot{x}_i = a_i x_{i'} - (d + a_{i0} + a_{i1}) x_i \qquad (2)$$

The abundance of sequence i is given by x_i. The notation $\dot{x}_i$ denotes the time derivative. The index i runs through all binary strings, i = 0, 1, 00, 01, ..., of finite length. For the "precursors" of 0 and 1, we set $x_{0'} = x_{1'} = 1$. We also assume that the concentrations of the active monomers, 0* and 1*, are constant (although not necessarily the same) and factored into the rate constants a_i.

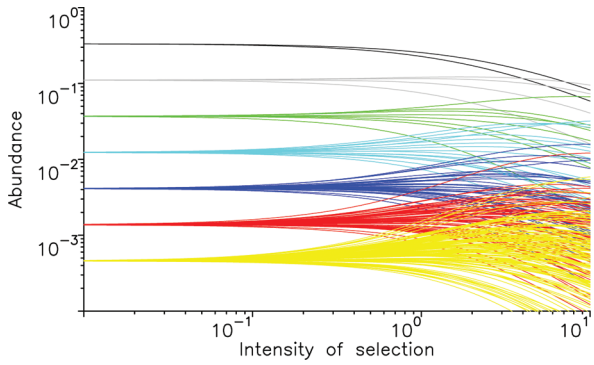


FIGURE 1. Prelife with a random fitness landscape. The equilibrium configuration of eq 2 is shown. There are 2^k different sequences of length k. The color code is $k = 1$ (black), 2 (gray), 3 (green), 4(cyan), 5 (blue), 6 (red), and 7 (yellow). Longer sequences are not shown. Parameter values: $d = 1$; $a_0 = a_1 = 1$; for all other i, we have $a_i = 1 + s\xi_i$ where s is the intensity of selection, shown on the x-axis and ξ_i is a random number taken from a uniform distribution on the interval [0,1]. We observe that longer sequences are exponentially less common. As the intensity of selection, s, increases some sequences become more abundant than others.

If all rate constants are greater than zero, then system 2 has a unique, globally stable equilibrium, where all sequences are present. Using the notation $b_i = a_i/(d + a_{i0} + a_{i1})$, we can write the equilibrium abundance of sequence i as $x_i = b_i b_{i'} b_{i''} ... b_\sigma$. The product is over the entire lineage leading from sequence i back to the monomer, σ (= 0 or 1). At equilibrium, the total population size, $\Sigma_i x_i$, is given by $(a_0 + a_1)/d$.

Although we do not know the exact fitness landscape for prebiotic sequences, we can still get an idea of how fitness affects sequence abundance by assigning particular rate constants to different reactions and observing the resulting abundances. Figure 1 shows the equilibrium configuration of prelife for a random prelife landscape. In this case, we set $a_0 = a_1 = d = 1$. All other rate constants are given by $a_i = 1 + s\xi_i$ where s is a parameter signifying the "intensity of selection" and ξ_i is a random number taken from a uniform distribution on the interval [0,1]. When there is little variation among the rate constants (small s), all sequences of equal length have approximately the same abundance. As s increases, some sequences become more abundant than others of the same length. The variation in sequence abundance is relatively gentle in prelife, although this variation increases for longer

sequences as the differences in rate constants compound over the length. The value of s based on reactivity differences in RNA polymerization catalyzed by clay or Zn^{2+} would be roughly 10−100. While removing and replacing spent monomers can substantially improve the yield of nonenzymatic templated RNA polymerization reactions, reactivity differences of at least 20-fold still remain.[37] In addition, s would be influenced by variation in the concentration of monomers. For example, one prebiotic synthesis of nucleobases produced about 10 times more A than other monomers.[38] The ratio of nucleobases found in carbonaceous chondrite meteorites also appears to span about 1 order of magnitude (reviewed in ref 39). Therefore, a realistic range of s would be fairly large, probably around 1 or 2 orders of magnitude. Thus, selection can operate in prelife in the sense that some sequences would attain high equilibrium abundance relative to others. This distribution is illustrated in Figure 1, where abundances of different sequences of the same length range over nearly 2 orders of magnitude at a realistic value of $s = 10$. Larger disparities would be expected for higher s and longer sequences.

It is conceivable that certain prelife sequences have catalytic activities. Some sequences could enhance the rate of particular prelife reactions.[15] If one sequence augments the rate at which another sequence is produced, then already in prelife we can encounter aspects of cooperation and defection.[40]

The prelife distribution of sequences resembles an equilibrium distribution with different chemical species coexisting. Differences in abundance depend on environmental conditions (e.g., the monomer concentration and polymerization chemistry). Even at a relatively conservative selection strength ($s = 10$), the abundance of polymers as short as 5 or 6 can vary by 100-fold (Figure 1). In prelife, variation is present, but it is not heritable. For example, prelife would describe a polymer in environmental conditions that do not yet support templating (e.g., a salt concentration that is too low to screen backbone charges, such that electrostatic repulsion prevents the annealing of complementary nucleic acid strands), or a polymer that lacks the ability to template (e.g., the condensation of polypeptides or a primitive nucleic acid that lacked the correct backbone or combination of nucleobases necessary for templating).

Life and Prelife

Let us now suppose that some prelife sequences have the ability of replication. For example, at an early stage of the development of the RNA world, many possible nucleobases would have coexisted. But not all of these nucleobases could support templating, which is required for replication. Some sequences would happen to contain nucleobases that could template a complementary strand, and these sequences would be able to replicate nonenzymatically by assembling complementary nucleotides or oligonucleotides. In addition, some template sequences are better than others.[41] For example, nonenzymatic polymerization on pyrimidine templates proceeds much more efficiently than on purine templates.[42−44] To include the ability of some sequences to replicate in our model, the overall replication rate, which may depend on chemical concentrations, temperature oscillations, and other environmental factors, is given by the parameter r. In addition, every specific sequence i has a relative replication rate (fitness), f_i. The list (f_0, f_1, f_{00}, ...) defines the fitness landscape. The resulting evolutionary dynamics can be described by the following system of differential equations:

$$\dot{x}_i = a_i x_{i'} - (d + a_{i0} + a_{i1})x_i + x_i(rf_i - \phi) \qquad (3)$$

As before, the index i runs over all binary sequences of finite length. There is a density-dependent death rate, ϕ, which prevents the total population size from growing to infinity. If we chose $\phi = a_0 + a_1 - d + r\Sigma_i x_i f_i$, then the total population size remains at a constant value, $\Sigma_i x_i = 1$.

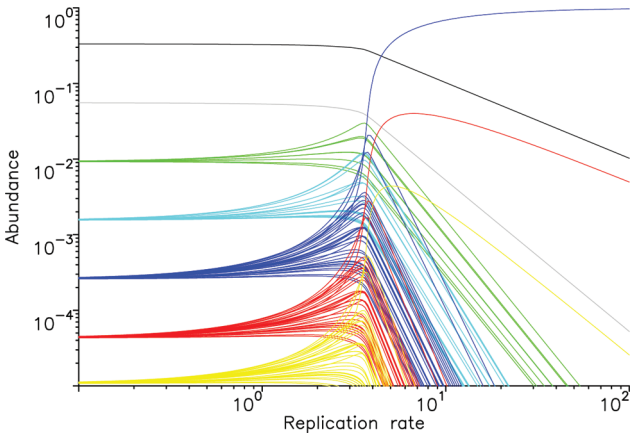


FIGURE 2. Prelife and life. The equilibrium configuration of eq 3 is shown. Again there are 2^k different sequences of length k. The color code is the same as for Figure 1. Parameter values: $d = 1$; $a_0 = a_1 = 1$; all other $a_i = 0.5$. Some cutoff in the numerical simulation is needed, and therefore we assume that sequences of length $k = 7$ (or greater) do not extend. Sequences of length $k = 3$, 4, and 5 can replicate; their relative replication rates, f_i, are random numbers taken from a uniform distribution on [0,1]. As the overall replication rate r (shown on the x-axis) increases, life replaces prelife. For high replication rate, the population is dominated by a particular sequence of length 5, which has the maximum relative fitness f_i. Prelife allows coexistence. Life leads to competitive exclusion. (Note that $\Sigma x_i = 1$.)

Figure 2 shows the equilibrium configuration of system 3 as a function of the parameter r. In this particular simulation, we assume that sequences of length 3, 4, and 5 can replicate to illustrate the difference between the template and nontemplate sequences. For template sequences, the relative replication rates, f_i, are drawn at random from a uniform distribution on the interval [0,1]. For nontemplate sequences, we have $f_i = 0$. For small r, we obtain the equilibrium configuration of prelife, where longer sequences are exponentially less frequent than shorter ones. As r increases, we note a threshold phenomenon: the fittest sequence (with the largest f_i value) is selected. For large r, this sequence dominates the population. In other words, evolutionary dynamics appear when r is large enough, that is, when templating is good enough for substantial replication. Prelife has the property of coexistence, while life with replication leads to competitive exclusion.

The critical r can be predicted from the relative magnitude of the highest fitness. For the conservative example shown in Figure 2, the highest fitness is roughly 1 and the average fitness of replicators is about 0.5, corresponding to a relatively modest variation among templating efficiencies of different sequences. Templating efficiencies in DNA- and RNA-based experimental systems vary by an order of magnitude or more for different template bases.[45,46] In Figure 2, the critical r is relatively low and templating only needs to lead to a roughly >4-fold increase in the production rate compared with nontemplated polymerization. This magnitude of increase is readily met by DNA and RNA at reasonable monomer concentrations. Although a templated reaction involves three molecules instead of two, the replication reaction is faster than the prelife reaction if the molecular concentrations are high enough. For example, self-condensation of activated G monomers (guanosine 5′-phospho-2-methylimidazolide) occurs spontaneously with a bimolecular rate constant of 0.09 M^{-1} h^{-1}. When a poly(C) template is present, the effective rate constant is 430 M^{-2} h^{-1}.[47] Therefore, the critical r is reached when the monomer concentration is less than 1 mM. If the reactive subunits are longer than monomers, the templating effect can be greater since annealing is more efficient; with rate constants for condensation of trimer oligonucleotides onto a hexamer template, the critical r is reached at a monomer concentration of 50 μM.[48]

A counterintuitive feature of this simple life is that long sequences can actually be much more abundant than short sequences. In prelife, long sequences can become relatively abundant if selection is very strong and there are large differences in reaction rate, such that the high reaction rates for a particular sequence effectively overcome the number of chemical steps required to produce the longer sequence. However, in life, the longest, best replicator can dominate dramatically. If several replicators of different length have the same fitness, the longest can actually win because its precursors help channel biomass into the lineage.[16,17] This results in selection for greater complexity.

Sharp transitions have also been observed in simulations of the appearance of ribozymes. In models by Wu and Higgs,[49,50] template-independent polymerization produces sequences of increasing length,[14] and a small fraction of the long sequences is assumed to possess catalytic activity and increase the general rate of polymerization. Stochastic fluctuations can cause an abrupt switch to a state that is characterized by a high concentration of long sequences. These models illustrate the principle that the transition to a ribozyme-rich world can occur suddenly, when the catalytic activity reaches a critical threshold. In our terminology, these models study "prelife catalysis". The critical transition and bistability between states with high and low concentrations of catalysts (long sequences) was also described by Ohtsuki and Nowak.[15] Wu and Higgs do not study the onset of replication, while our work shows that it is much easier to select for long replicators than for long prelife catalysts. Detailed comparisons of the two approaches will be of great value.

Interestingly, life with replication wins over prelife even if prelife produces catalysts that accelerate all of the reactions that lead to their own sequences. This is essentially because such catalysts would undergo many associations and dissociations to produce a single product, while replication would ideally only need one annealing event.[15] Another feature that causes a transition from blunt to sharp selection is product inhibition, in which the newly synthesized strand remains bound to the original template, resulting in parabolic rather than exponential growth. Under this condition, there is "survival of everybody", analogous to prelife.[51,52] But when the fact that single-stranded RNA is more sensitive to degradation than double-stranded RNA is taken into account, this can, in some parameter regimes, lead to strong competition.[53] The general theme to emerge from these theoretical studies is that a new structural property can lead to new system dynamics.

Competition of Two Prelives

Suppose we have different biopolymer systems, some with better replication properties than others. What would

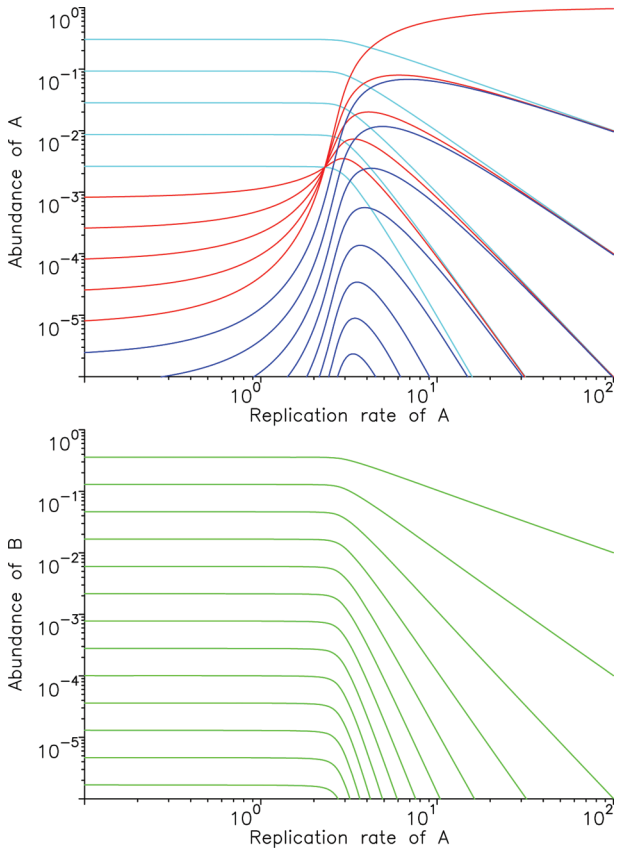


FIGURE 3. Competition of two prelives as described by eq 4. Sequences of type A are somewhat less stable than sequences of type B. Type A sequences of length less than 6 (cyan) or greater than 10 (blue) cannot replicate. But A sequences of length 6 to 10 (red) can replicate. As the overall replication rate increases, A outcompetes B (green). Parameter values: $a = b = 1$, $d_A = 1$, $d_B = 0.5$, $f_i = 1$ for $i = 6, ..., 10$, $f_i = 0$ otherwise; maximum sequence length is 50. Hence, different prelives can compete with each other, and those "fertile" prelives can be selected that give rise to life.

happen if these systems compete in the prebiotic soup? We can study the competition of two different sets of monomers that can polymerize into distinct sequences. Let us label these two different prelives as A and B. The two prelives could compete for common resources, and one prelife might have the ability to produce some sequences that can replicate. For example, this situation might correspond to the competition between two distinct nucleic acid systems (e.g., utilizing different nucleobases or a different backbone). One type of nucleic acid might have superior templating properties; for example, this could represent the competition between RNA and a different nucleic acid polymer.[54] To illustrate the major competitive dynamics at work, let us study this system with a simplified approach, where we keep track of sequence length but not of sequence diversity.

Let x_i denote the abundance of A sequences of length i, while y_i denotes the abundance of B sequences of length i. Thus we lump all the different sequences of a given length into a single variable; we only distinguish between type A and type B. Denote by a and b, respectively, the rate at which A and B sequences grow in length. The death rates of A and B sequences is given by d_A and d_B. It is conceivable that one type of sequence has faster polymerization kinetics than the other, or that one type has a faster death rate than the other. For example, RNA is particularly prone to hydrolysis from nucleophilic attack of the 2′-hydroxyl on the phosphodiester bond, while other backbones are more stable in this respect.[55] Let us further assume that type A sequences of certain length have the ability to replicate, but not type B sequences. For example, substituting the phosphates of an RNA backbone with sulfones results in a nucleic acid that is prone to aggregation and is therefore a poor replicator.[56] We obtain the following system of differential equations:

$$\dot{x}_i = a x_{i'} - (a + d_A) x_i + x_i (r f_i - \phi)$$
$$\dot{y}_i = b y_{i'} - (b + d_B) y_i - \phi y_i \qquad (4)$$

The index i runs through all positive integers, $i = 1, 2, 3, ...$ denoting sequence length. Again the density-dependent death rate, ϕ, is chosen to keep the total population size constant, $\Sigma_i (x_i + y_i) = 1$. Therefore, we have $\phi = a + b + \Sigma_i (r f_i x_i - d_A x_i - d_B y_i)$.

Figure 3 shows the equilibrium configuration of the population. In this case, A has the ability to replicate, but B has a longer lifetime, as we might expect from a competition between RNA and a pre-RNA that is more chemically stable but a poorer template. For low replication rate, r, we observe the equilibrium structure of prelife. Longer sequences are exponentially less common. Moreover, B sequences have a lower death rate than A sequences, $d_B < d_A$, and hence they are more abundant for low r. In this regime, greater chemical stability is the dominant factor. Nevertheless, both A and B coexist at significant frequencies.

As r increases, however, a critical point is reached when the replication advantage of A is realized, and consequently A outcompetes B. When replication dominates the system, the outcome is highly skewed with A essentially taking over the pool. This simple example shows there can be competition between different prelife systems and selection for those prelives that give rise to replication.

Discussion

These minimalist models illustrate how evolutionary dynamics emerge from chemical kinetics once the ability to template (and therefore replicate) appears. The sequences in

prelife experience mild selection, even without replication, based on the variation in reaction rates. Once templating arises, the ability to replicate itself is selected. The dynamics change abruptly at a certain replication rate that demarcates qualitatively different regimes. While the relatively egalitarian sequences in prelife proliferate relatively independently from one another, the life-like replicators compete mercilessly among themselves and exaggerate intrinsic reactivity differences into very large inequities of abundance. This dynamic is also common in Darwinian evolution, where an incremental increase in fitness can translate into an explosive takeover of the population due to exponential amplification of the difference.

Does our modeling have any bearing on the metabolism-first vs gene-first debate about the origin of life? In the metabolism-first view, reaction cycles of small molecules arose spontaneously, eventually producing the monomers needed for informational molecules. On the other hand, in the gene-first view, the informational polymers were made directly through a relatively short series of chemical steps and later evolved to catalyze those synthetic reactions. The life-like dynamics of our model result naturally from a minimal set of reasonable assumptions about a life form based on linear polymers, such as, but not necessarily limited to, nucleic acids. We focus on the genetic properties of early life, and we take for granted the chemical reactions that could produce monomers capable of condensation into long chains. While there is substantial debate about the relative importance and temporal order of the emergence of the genetic and metabolic components of life, we focus on the genetic aspect because it has proven to be a fruitful avenue for investigating the emergence of evolutionary dynamics. In addition, the chemical reactions (polymerization, both template-directed and template-independent) are relatively well-understood, so our modeling could be based on a fairly solid experimental foundation. This model does fall naturally into the realm of the gene-first view and the idea of the RNA world, though it could also apply to self-replicating peptides.[57] In the context of the metabolism-first hypothesis, although this modeling does not address the very initial metabolic cycles, it would still be relevant for the stage at which the genetic polymers arose.

An interesting aspect of the prelife modeling is to clarify the process by which a templating genetic polymer took hold given numerous other possible backbones and nucleobases that one might imagine could be synthesized by the same prebiotic process. The ability to replicate nonenzymatically would give a tremendous advantage to a templating polymer, even given a disadvantage of decreased chemical stability. Nucleobases that were poor for templating would be eliminated as sequences containing them lost the competition, effectively selecting for a nucleobase system that best supported base pairing and replication. Later takeovers could happen as even better replicators arose, ultimately resulting in RNA.

Outlook

Testing the predictions of prelife modeling would require an experimental system in which monomers can polymerize with or without a template in the absence of enzymes. The activated nucleotides first pioneered by Leslie Orgel's group (reviewed in ref 58) have probably been best studied in this context. These monomers polymerize much faster in the presence of a template, but the nontemplated reaction also occurs under the same reaction conditions.[31,47] Nontemplated polymerization of activated nucleotides can be sped up by the clay montmorillonite, a reaction that produces sequences capable of templating after purification.[59] Polymerization with and without templates can also be promoted in the "eutectic" phase of an aqueous solution, in which water and ice coexist such that the effective concentration of the solutes is greatly increased.[34] Alternatively, nontemplated polymerization of ribonucleoside monophosphates can be driven by wet–dry cycling in the presence of lipids,[33] and similar conditions promote the template-directed synthesis of deoxyribonucleotide monophosphates.[60] In general, one-pot experiments enabling both templated and nontemplated polymerization have been avoided in order to better analyze the products and interpret the results, but in principle a one-pot reaction could be attempted to mimic the combination of prelife with life. Perhaps the most serious challenge that experimental prelife-to-life reactions would pose is analytical. While liquid chromatography and mass spectrometry could be used to distinguish short sequences (up to several nucleotides long) given enough material, sequencing techniques for longer polymers can only analyze consistently 5′–3′ linked chains, but RNA polymerization tends to give a mixture of 5′–2′ and 5′–3′ linkages. The analytical limitation is due to sequencing technology, which relies on enzymes that evolved to recognize nucleic acids containing the biological linkages. Biases of the enzymes favoring certain sequences would also need to be characterized in order to infer true frequencies. We are optimistic, however, that an emerging single-molecule, enzyme-free sequencing technology could soon sidestep this problem,

opening the door to thorough sequence analysis of a complex prelife-to-life "prebiotic soup".

A clear future direction for this line of modeling is the addition of more realistic features. Experimental implementations of polymerization inevitably suffer from inefficiencies and drawbacks that are not included in our model so far, such as the problem of heterogeneous backbone linkages. An important goal is to determine whether there exist conditions under which experimental problems, such as the incorporation of nucleotides that undergo further reaction only poorly and therefore effectively terminate the chain,[58] could be overcome through prelife evolution. One might model the copolymerization of nucleotides that are capable or incapable of templating to understand how a subset of bases is finally selected. A particularly interesting variant of this problem is the copolymerization of nucleotides containing D- or L-sugars, which might polymerize passably well together in prelife, but a mixed backbone would effectively poison the templating ability and replication of its sequence.[61] Other realistic additions would be to include more bases in the alphabet and to model scenarios for separating the product strands after replication (e.g., by thermal melting).

The models we have described so far are deterministic and assume that the population of molecules is very large and well-mixed. Since model protocells might be as small as 100 nm in diameter and RNA might be relatively dilute, it is likely that the number of polymers could be quite low, so that stochastic effects might become important. Small-number effects can be surprisingly important for early life, from favoring a ribozyme population explosion[49] to limiting genomic information.[13] An important goal is to put prelife modeling in the context of a spontaneously dividing protocell.[17]

A particularly exciting future direction for prelife modeling is to include functional importance of the sequence information. Although we do not have a detailed knowledge of the fitness landscape for RNA, one proxy for function is folding into a defined structure. For example, the possibility that stable folding could be protective against degradation, and therefore help select for structured molecules, has been studied theoretically in the context of nontemplated ligation.[62] In that study, the simulated RNA pool showed a surprising degree of heritability even though ligation was not templated. One wonders how this effect would alter the evolutionary dynamics of prelife, perhaps by enhancing replication. Ultimately, the function of the folded structures should be taken into account. We have considered the possibility that the RNAs could catalyze polymerization,[15] but other activities should be studied as well.

A fundamental challenge of theoretical modeling is to ensure its relevance to experimental reality. We have attempted to connect our model to experimental systems whenever possible and to avoid unnecessary assumptions by making the model as simple as possible. What is the importance of theoretical models for the origin of life? Modeling can help crystallize intuition, outlining the dynamical transition caused by the advent of replication, as well as point toward counterintuitive ideas like the dominance of long sequences. Perhaps the most important contributions of theory to the origin of life field was the prediction by Woese, Crick, and Orgel that RNA sequences could be catalysts, inferred from the ability of RNA to fold into complex structures reminiscent of protein folds. This prediction formed the foundation for the RNA world theory, which has since gained credibility through considerable accumulated circumstantial evidence.[63] In this spirit, we hope to contribute in some way to the basic understanding of the dynamics that accompany the emergence of the RNA world.

I.A.C. is a Bauer Fellow at the FAS Center for Systems Biology supported by NIH Grant P50GM068763. Support from the John Templeton Foundation and the NSF/NIH joint program in mathematical biology (NIH Grant R01GM078986) is gratefully acknowledged.

BIOGRAPHICAL INFORMATION

Irene A. Chen is a Bauer Fellow at the FAS Center for Systems Biology at Harvard University. She received a B.A. in chemistry and a M.D.—Ph.D. in biophysics from Harvard, advised by Jack Szostak. She received the G.E. and *Science* prize for young life scientists, the Harold Weintraub graduate student award, and the David White award for research on the origin of life.

Martin A. Nowak is a Professor of Mathematics and Biology at Harvard University and Director of the Program for Evolutionary Dynamics. He received a M.Sc. and Ph.D. from the University of Vienna in biochemistry and mathematics, headed the mathematical biology group at Oxford University, and began the theoretical biology group at Princeton's Institute for Advanced Study. A corresponding member of the Austrian Academy of Sciences, Dr. Nowak is the recipient of numerous awards including the Weldon Memorial and the David Starr Jordan Prize. His most recent book, "SuperCooperators" (with Roger Highfield) was published in 2011.

FOOTNOTES

*E-mail addresses: martin_nowak@harvard.edu; ichen@post.harvard.edu.

REFERENCES

1 *Protocells: Bridging Nonliving and Living Matter*; Rasmussen, S., Bedau, M. A., Chen, L., Deamer, D., Krakauer, D. C., Packard, N. H., Stadler, P. F., Eds.; MIT Press: Cambridge, MA, 2008.

2 *The Origins of Life*; Deamer, D., Szostak, J. W., Eds.; Cold Spring Harbor Laboratory Press: New York, 2010.

3 Fontana, W.; Schuster, P. Continuity in evolution: On the nature of transitions. *Science* **1998**, *280*, 1451–1455.

4 Briones, C.; Stich, M.; Manrubia, S. C. The dawn of the RNA World: Toward functional complexity through ligation of random RNA oligomers. *RNA* **2009**, *15*, 743–749.

5 Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, *58*, 465–523.

6 Eigen, M.; Schuster, P. Hypercycle - Principle of natural self-organization. A. Emergence of hypercycle. *Naturwissenschaften* **1977**, *64*, 541–565.

7 Eigen, M.; Schuster, P. Hypercycle - Principle of natural self-organization. B. Abstract hypercycle. *Naturwissenschaften* **1978**, *65*, 7–41.

8 Eigen, M.; Schuster, P. Hypercycle - Principle of natural self-organization. C. Realistic hypercycle. *Naturwissenschaften* **1978**, *65*, 341–369.

9 Nowak, M. A. What is a quasispecies? *Trends Ecol. Evol.* **1992**, *7*, 118–121.

10 Schuster, P.; Swetina, J. Stationary mutant distributions and evolutionary optimization. *Bull. Math. Biol.* **1988**, *50*, 635–660.

11 Eigen, M.; McCaskill, J.; Schuster, P. Molecular quasi-species. *J. Phys. Chem.* **1988**, *92*, 6881–6891.

12 Fontana, W.; Schuster, P. A computer model of evolutionary optimization. *Biophys. Chem.* **1987**, *26*, 123–147.

13 Nowak, M. A.; Schuster, P. Error thresholds of replication in finite populations. Mutation frequencies and the onset of Muller's ratchet. *J. Theor. Biol.* **1989**, *137*, 375–395.

14 Nowak, M. A.; Ohtsuki, H. Prevolutionary dynamics and the origin of evolution. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14924–14927.

15 Ohtsuki, H.; Nowak, M. A. Prelife catalysts and replicators. *Proc. R. Acad. Sci. B-Biol.Sci.* **2009**, *276*, 3783–3790.

16 Manapat, M.; Ohtsuki, H.; Burger, R.; Nowak, M. A. Originator dynamics. *J. Theor. Biol.* **2009**, *256*, 586–595.

17 Manapat, M. L.; Chen, I. A.; Nowak, M. A. The basic reproductive ratio of life. *J. Theor. Biol.* **2010**, *263*, 317–327.

18 Tirard, S.; Morange, M.; Lazcano, A. The definition of life: A brief history of an elusive scientific endeavor. *Astrobiology* **2010**, *10*, 1003–1009.

19 Luisi, P. L. About various definitions of life. *Origins Life Evol. Biospheres* **1998**, *28*, 613–622.

20 Orgel, L. E. The origin of life—a review of facts and speculations. *Trends Biochem. Sci.* **1998**, *23*, 491–495.

21 Shapiro, R. A simpler origin for life. *Sci. Am.* **2007**, *296*, 46–53.

22 Rosen, R. *Life Itself*; Columbia University Press: New York, 1991.

23 Kauffman, S. Question 1: Origin of life and the living state. *Origins Life Evol. Biospheres* **2007**, *37*, 315–322.

24 Copley, S. D.; Smith, E.; Morowitz, H. J. The origin of the RNA world: Co-evolution of genes and metabolism. *Bioorg. Chem.* **2007**, *35*, 430–443.

25 Morowitz, H. *Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis*; Yale University Press: New Haven, CT, 2004.

26 Orgel, L. E. The origin of life on the earth. *Sci. Am.* **1994**, *271*, 76–83.

27 Horowitz, N.; Miller, S. Current theories on the origin of life. *Fortschr. Chem. Org. Naturst.* **1962**, *20*, 423–459.

28 Joyce, G. Foreword. In *Origins of Life: The Central Concepts*; Deamer, D., Fleischaker, G., Eds.; Jones and Bartlett: Boston, MA, 1994; pp xi—xii.

29 Miller, S. L.; Urey, H. C. Organic compound synthesis on the primitive earth. *Science* **1959**, *130*, 245–251.

30 Powner, M. W.; Sutherland, J. D.; Szostak, J. W. Chemoselective multicomponent one-pot assembly of purine precursors in water. *J. Am. Chem. Soc.* **2010**, *132*, 16677–16688.

31 Sawai, H.; Orgel, L. E. Oligonucleotide synthesis catalyzed by the Zn-2+ ion. *J. Am. Chem. Soc.* **1975**, *97*, 3532–3533.

32 Ferris, J. P.; Hill, J., A. R.; Liu, R; Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **1996**, *381*, 59–61.

33 Rajamani, S.; Vlassov, A.; Benner, S.; Coombs, A.; Olasagasti, F.; Deamer, D. Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Origins Life Evol. Biospheres* **2008**, *38*, 57–74.

34 Monnard, P. A.; Kanavarioti, A.; Deamer, D. W. Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. *J. Am. Chem. Soc.* **2003**, *125*, 13734–13740.

35 Miyakawa, S.; Ferris, J. P. Sequence- and regioselectivity in the montmorillonite-catalyzed synthesis of RNA. *J. Am. Chem. Soc.* **2003**, *125*, 8202–8208.

36 Ruff, K. M.; Snyder, T. M.; Liu, D. R. Enhanced functional potential of nucleic acid aptamer libraries patterned to increase secondary structure. *J. Am. Chem. Soc.* **2010**, *132*, 9453–9464.

37 Deck, C.; Jauker, M.; Richert, C. Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat. Chem.* **2011**, *3*, 603–608.

38 Miyakawa, S.; Cleaves, H. J.; Miller, S. L. The cold origin of life: B. Implications based on pyrimidines and purines produced from frozen ammonium cyanide solutions. *Origins Life Evol. Biospheres* **2002**, *32*, 209–218.

39 Martins, Z.; Botta, O.; Fogel, M. L.; Sephton, M. A.; Glavin, D. P.; Watson, J. S.; Dworkin, J. P.; Schwartz, A. W.; Ehrenfreund, P. Extraterrestrial nucleobases in the Murchison meteorite. *Earth Planet. Sci. Lett.* **2008**, *270*, 130–136.

40 Nowak, M. A. Five rules for the evolution of cooperation. *Science* **2006**, *314*, 1560–1563.

41 Joyce, G. F.; Orgel, L. E. Non-enzymatic template-directed synthesis on RNA random copolymers - poly(C,A) templates. *J. Mol. Biol.* **1988**, *202*, 677–681.

42 Ninio, J.; Orgel, L. E. Hetero-polynucleotides as templates for non-enzymatic polymerizations. *J. Mol. Evol.* **1978**, *12*, 91–99.

43 Joyce, G. F. Nonenzymatic template-directed synthesis of informational macromolecules. *Cold Spring Harb. Symp. Quant. Biol.* **1987**, *52*, 41–51.

44 Stribling, R.; Miller, S. L. Template-directed synthesis of oligonucleotides under eutectic conditions. *J. Mol. Evol.* **1991**, *32*, 289–295.

45 Rajamani, S.; Ichida, J. K.; Antal, T.; Treco, D. A.; Leu, K.; Nowak, M. A.; Szostak, J. W.; Chen, I. A. Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *J. Am. Chem. Soc.* **2010**, *132*, 5880–5885.

46 Leu, K.; Obermayer, B.; Rajamani, S.; Gerland, U.; Chen, I. A. The prebiotic evolutionary advantage of transferring genetic information from RNA to DNA. *Nuc. Acids. Res.* **2011**, *39*, 8135–8147.

47 Kanavarioti, A.; White, D. H. Kinetic-analysis of the template effect in ribooligoguanylate elongation. *Origins Life Evol. Biospheres* **1987**, *17*, 333–349.

48 Sievers, D.; Vonkiedrowski, G. Self replication of complementary nucleotide-based oligomers. *Nature* **1994**, *369*, 221–224.

49 Wu, M.; Higgs, P. G. Origin of self-replicating biopolymers: Autocatalytic feedback can jump-start the RNA world. *J. Mol. Evol.* **2009**, *69*, 541–554.

50 Wu, M.; Higgs, P. Comparison of the roles of nucleotide synthesis, polymerization, and recombination in the origin of autocatalytic sets of RNAs. *Astrobiology* **2011**, *11*, 895–906.

51 von Kiedrowski, G. Minimal replicator theory I: Parabolic versus exponential growth. *Bioorg. Chem. Front.* **1993**, *3*, 113–146.

52 Szathmary, E. Simple growth laws and selection consequences. *Trends Ecol. Evol.* **1991**, *6*, 366–370.

53 Scheuring, I.; Szathmary, E. Survival of replicators with parabolic growth tendency and exponential decay. *J. Theor. Biol.* **2001**, *212*, 99–105.

54 Nielsen, P. E. DNA analogs with nonphosphodiester backbones. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 167–183.

55 Schoning, K.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. Chemical etiology of nucleic acid structure: The alpha-threofuranosyl-(3′→2′) oligonucleotide system. *Science* **2000**, *290*, 1347–1351.

56 Benner, S. A. Understanding nucleic acids using synthetic chemistry. *Acc. Chem. Res.* **2004**, *37*, 784–797.

57 Lee, D. H.; Granja, J. R.; Martinez, J. A.; Severin, K.; Ghadiri, M. R. A self-replicating peptide. *Nature* **1996**, *382*, 525–528.

58 Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.

59 Ertem, G.; Ferris, J. P. Template-directed synthesis using the heterogeneous templates produced by montmorillonite catalysis. A possible bridge between the prebiotic and RNA worlds. *J. Am. Chem. Soc.* **1997**, *119*, 7197–7201.

60 Olasagasti, F.; Kim, H. J.; Pourmand, N.; Deamer, D. W. Non-enzymatic transfer of sequence information under plausible prebiotic conditions. *Biochimie* **2011**, *93*, 556–561.

61 Joyce, G. F.; Inoue, T.; Orgel, L. E. Non-enzymatic template-directed synthesis on RNA random copolymers - poly(C,U) templates. *J. Mol. Biol.* **1984**, *176*, 279–306.

62 Obermayer, B.; Krammer, H.; Braun, D.; Gerland, U. Emergence of information transmission in a prebiotic RNA reactor. *Phys. Rev. Lett.* **2011**, *107*, No. 018101.

63 *RNAWorlds*; Atkins, J. F., Gesteland, R. F., Cech, T. R., Eds.; Cold Spring Harbor Laboratory Press: New York, 2011.