Biodata of **Kelley Harris** and **Irene A. Chen**, authors of "***Mathematical Models of Prebiotic Replication of Informational Molecules***."

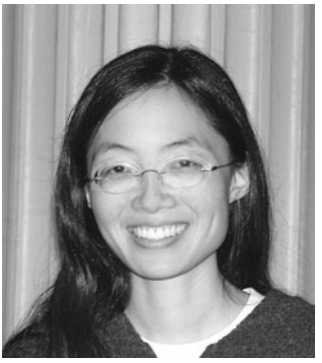
Kelley Harris is a Ph.D. student at the University of California, Berkeley. She received an M.Phil in Biological Science in 2011 from Cambridge University, working at the Wellcome Trust Sanger Genome Campus, and earned an A.B. in mathematics from Harvard in 2009. Her current interests include population genetics, identity by descent in genome data, and the inference of human demographic history.

E-mail: **harris.kelley@gmail.com**

Irene A. Chen is currently a Bauer Fellow at the FAS Center for Systems Biology at Harvard University. She received an M.D. and Ph.D. in Biophysics from Harvard in 2007. Her scientific interests include experimental and theoretical studies on the origins of life, the structure of chemical activity landscapes, and the use of bacteriophages to control the dynamics of bacterial populations.

E-mail: **ichen@post.harvard.edu**

Kelley Harris



Irene A. Chen

J. Seckbach (ed.), *Genesis - In The Beginning: Precursors of Life, Chemical Models and Early Biological Evolution*, Cellular Origin, Life in Extreme Habitats and Astrobiology 22, 67–88
DOI 10.1007/978-94-007-2941-4_4,

MATHEMATICAL MODELS OF PREBIOTIC REPLICATION OF INFORMATIONAL MOLECULES

KELLEY HARRIS[1] AND IRENE A. CHEN[2]
[1]*UC Berkeley Mathematics Department, Berkeley, CA 94720, USA*
[2]*FAS Center for Systems Biology, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA*

1. Introduction

Anyone who studies the origin of life must contend with huge gaps in the available information. The earliest stages of life have left very few traces that survived the challenges of time. It is possible for natural history to degenerate into speculation when physical and chemical fossils cannot be found, but we can still make good use of universal consequences of physics and mathematics, which have not changed over time. Surprisingly, concrete claims about early life can be derived using chemical kinetics and information theory, in particular, and this chapter surveys major claims in this area. This chapter is not meant to review the entire literature of prebiotic models but to give a few examples of major frameworks as an introduction to interdisciplinary scientists.

Attempts to mimic prebiotic reactions, from Stanley Miller's classic experiments to recent advances in RNA synthesis, have done much to explain how life's building blocks could have appeared on earth (Miller and Urey, 1959; Ricardo et al., 2004; Powner et al., 2009; Orgel, 2004). These experiments may eventually be enough to explain the terrestrial presence of chemicals like amino acids and nucleic acid monomers. However, novelties of organization like metabolism and replication remain largely unexplained.

Few biomolecules are functional outside layers of native organization. To describe a protein means to consider structure at the levels of both sequence and fold. For example, precise folding is the difference between raw and denatured egg white, and once heating destroys the information contained in the fold, the native protein cannot be recovered. In the seventeenth century, this denaturation was attributed to the expulsion of an unknown "vital force," an understandable, if inappropriate, intuition given the biological importance of organization (Weisstein, 2007). Many proteins cannot assume their bioactive, water-soluble native fold without help from cellular chaperones. However, proteins with certain energy landscapes are capable of folding spontaneously (Jackson, 1998). This ability to self-organize would have been an essential feature of the earliest living systems.

In some remarkable cases, it is possible to engineer a peptide sequence that both folds spontaneously and has a rudimentary ability to self-replicate, catalyzing the ligation of shorter peptides to produce the same sequence (Lee et al., 1996). However, in contrast to nucleic acids, such peptides appear to be special cases. Because of the generalized templating ability of nucleic acids, any sequence can, in principle, readily catalyze the formation of a complementary sequence through template-directed ligation (James and Ellington, 1997; Rohatgi et al., 1996a, b). Among nucleic acids, RNA is of particular interest to the origins of life because of its ability to serve as genetic material and to catalyze a wide variety of chemical reactions (Guerrier-Takada et al., 1983; Kruger et al., 1982). Although an RNA polymerase ribozyme capable of true self-replication has not yet been discovered in the laboratory, work in this direction has demonstrated that ribozymes are capable of both oligonucleotide ligation and processive mononucleotide polymerization chemistry (Doudna and Szostak, 1989; Green and Szostak, 1992; Bartel and Szostak, 1993; Ekland and Bartel, 1995; Hager and Szostak, 1997; Jaeger et al., 1999; Johnston et al., 2001; Wochner et al., 2011).

While synthetic biologists make progress in building specific self-replicating systems, a complementary approach to understanding the origins of life is to look for mathematical constraints that must be satisfied by any replicating system. Several such constraints arise from the quasispecies model, presented in Sect. 2 and 3 (Eigen, 1971, 1977; Eigen and Schuster, 1978a, b; Eigen et al., 1988). These models were developed to capture the dynamics of self-replicating molecules and have been shown to describe viral evolution quite accurately (Eigen, 1993; Quer et al., 1996).

Section 4 will review progress in addressing the chicken-and-egg paradox of the error threshold, which suggests that early genomes might not have been capable of encoding enough information to encode replication machinery. This section includes a discussion of hypercycles, which Eigen and Schuster developed in the context of protein translation in a world of RNA quasispecies (Eigen, 1977; Eigen and Schuster, 1978a, b). Section 5 reviews recent efforts to understand how chemical kinetics become replicator dynamics in a solution of polymers that form more or less randomly (Nowak and Ohtsuki, 2008). This model is somewhat abstract, but it could describe nontemplated nucleic acid polymerization (e.g., catalyzed by charged clay surfaces, assisted by heating and drying cycles in lipids, or carried out in the eutectic phase (Ertem and Ferris, 1996; Ferris and Ertem, 1992, 1993; Ferris et al., 1996; Rajamani et al., 2008; Monnard and Deamer, 2001, 2002)) and templated polymerization (Sawai and Orgel, 1975; Orgel, 1992; Manapat et al., 2009; Ohtsuki and Nowak, 2009).

Experimental verification of theories related to origins of life is particularly important because of the plethora of interesting, plausible, but essentially untested ideas. Quasispecies theory has proven itself to be useful in understanding real systems and particularly where viruses are concerned. One approach to combating viral infection is to design drugs that increase the viral mutation rate beyond the error threshold (Eigen, 1993, 2002; Quer et al., 1996; Domingo and Holland, 1997;

Grande-Perez et al., 2002; Crotty et al., 2004; Vignuzzi et al., 2006). It has also been shown that artificially low mutation rates actually decrease virulence by reducing the rate at which a virus can adapt and evolve (Vignuzzi et al., 2006). Recently, more attention has focused on using experimental systems to test theoretical predictions, such as a study of the error threshold in nonenzymatic nucleic acid replication (Rajamani et al., 2010). Several experimental systems are under development, including an elementary hypercycle based on fragmentation of a ligase ribozyme (Hayden et al., 2005), a system of cooperating and competing ligase ribozymes (Kim and Joyce, 2004; Lincoln and Joyce, 2009) and model protocells encapsulating replicating nucleic acids (Mansy and Szostak, 2009). Although this chapter does not survey the entire literature on interesting simulations and models of prebiotic replication, our hope is that it will facilitate the dialogue between experimentalists and theorists by reviewing some major theories regarding information in early genomes whose predictions would benefit immediately from further experimental study.

2. Introducing Kinetics of Self-Replication

Let us imagine n polymers $x_1,\ldots,x_n$ that coexist in a pool containing a limited supply of activated subunits (e.g., nucleotides or oligomers). For simplicity, we will refer to the subunits as monomers. Each polymer can use the monomers to reproduce, and we wish to predict the outcome of their struggle for existence. Enzymes and other catalysts might be present in solution, giving the polymers the ability to be copied, or each x_i might incorporate one nucleotide at a time, begetting new sequences by elongating itself. Every molecule that is formed will eventually degrade so that no species can persist without regular replenishment.

The mechanism of replenishment will be a reaction of the form

$$a_1x_1+\cdots+a_nx_n \leftrightharpoons b_1x_1+\cdots+b_nx_n, \tag{1}$$

describing the net process of reproduction and death (Eigen, 1971). Letting $\vec{k}$ and $\overleftarrow{k}$ denote the forward and reverse reaction rate constants, we recall that the system is at steady state when the concentrations $[x_1],\ldots,[x_n]$ satisfy

$$\vec{k}[x_1]^{a_1}\ldots[x_n]^{a_n} = \overleftarrow{k}[x_1]^{b_1}\ldots[x_n]^{b_n} \tag{2}$$

such that the forward and reverse reaction rates remain constant and equal. (Hereafter, we will omit concentration brackets to simplify notation.) Since every population moves toward a steady state over time, the dominant species will be the fittest species, the survivors of the evolutionary contest.

Equation (1) is not a true chemical equilibrium, as it omits the activated monomers that drive reproduction. True chemical equilibrium is incompatible with life, and some biologists see this incompatibility as more definitive of life than reproduction, postulating that metabolism, not replication, was the first

lifelike property to arise (Anet, 2004; Pross, 2004; Wachtershauser, 1988). However, we will focus on replication-first scenarios in which the precursors to life were passive replicators exploiting a reserve of digestible energy, racing against the clock to evolve metabolism before an energy crisis pulled the plug on their existence.

3. Quasispecies Theory

In this section, we present the basic theory about a model population of self-replicating molecules $x_1,\ldots,x_n$ (Eigen, 1977; Eigen et al., 1988). We first outline n differential equations that specify their replication dynamics. We assume a constant influx of activated monomers, balanced by a uniform outflux of polymers and their degradation products, so that a steady state is reached and the "fitness" of each x_i does not vary with time. In addition, we assume that x_i forms and degrades at a rate that depends linearly on $x_1,\ldots,x_n$, as is the case when one polynucleotide templates the synthesis of another.

3.1. THE REPLICATOR EQUATION

In a system close to chemical equilibrium, reactions tend to be reversible, such that the reaction $C \to A + B$ proceeds by reversing each step of $A + B \to C$. In contrast, systems far from equilibrium tend to be dominated by irreversible reactions, meaning that self-replicating molecules form and degrade by pathways with different rate dependencies. For example, x_i templates polymer synthesis at a rate $A_i x_i$ that depends more strongly on catalytic activity than does the rate $D_i x_i$ at which x_i is degraded.

Another source of x_i is the imperfect replication of competitors like x_j; when x_i and x_j differ by a simple monomer insertion, deletion, or substitution, x_j will catalyze x_i's production at a nonzero rate $w_{ij}x_j$. Conversely, there is a nonzero probability $1-Q_i$ that x_i will produce a mutant when it tries to copy itself. Finally, the uniformizing outflux will carry away x_i at rate $\Phi_i(\mathbf{X}) = \Phi_i(x_1,\ldots,x_n)$ that depends linearly on each of the polymer concentrations. If we neglect all other factors that affect the concentration of x_i, we obtain the following system of n differential equations:

$$\dot{x}_i = (A_i Q_i - D_i)x_i + \sum_{k \neq i} w_{ik} x_k - \Phi_i(\vec{x}). \tag{3}$$

We repress the time dependence of $x_1,\ldots,x_n$ for simplicity, using $\dot{x}_i$ to denote a first-order time derivative. When fitness values $W_i = A_i Q_i - D_i$ and mutation probabilities w_{ik} do not vary with time, this system of equations is solvable. An exact solution is reported in Thompson and McBride (1974), and good approximations can be derived using perturbation theory (Eigen, 1971).

Assuming that x includes the full array of mutants that can result from x_i's imperfect replication, it must hold that

$$\sum_{i=1}^{n} A_i(1-Q_i) = \sum_{j\neq i} w_{ji}. \tag{4}$$

Finally, we maintain constant reaction conditions by requiring that

$$\sum_{i=1}^{n} \Phi_i = \sum_{i=1}^{n} A_i x_i - \sum_{i=1}^{n} D_i x_i. \tag{5}$$

3.2. THE QUASISPECIES

No primitive autocatalyst could have entirely eliminated its competitors. Even when there is a "master sequence" x_i whose fitness $W_i = A_iQ_i - D_i$ is much higher than all competing W_j's, x_i continuously populates a range of nearby mutants, except in the unrealistic boundary case $Q_i = 1$. Eigen proved algebraically that the mutant mixture will stabilize with time, converging to an eigenvector of the *mutation matrix* $\mathbf{W}$ whose diagonal entries are the W_j's and whose off-diagonal entries are the w_{ij}'s. If we let $\mathbf{x}_i$ denote the n-entry vector whose ith entry is $x_i\delta_{ij}$, then by construction, the replicator population will evolve toward the stationary state $\lim_{k\to\infty} \mathbf{W}^k\mathbf{x}_i$, which is an eigenvector of $\mathbf{W}$ by the theory of finite-state Markov chains.

It is accurate to say that natural selection acts on these eigenvectors rather than on individual molecules since a dominant molecule can only persist in the context of the mutant mixture it generates. For this reason, Eigen refers to the eigenvectors of $\mathbf{W}$ as *quasispecies*. One may verify, using the Perron-Frobenius theorem (Horn and Johnson, 1991), that $\mathbf{W}$ has one positive, real eigenvalue with larger absolute value than any of its other eigenvalues; the corresponding eigenvector is the fittest quasispecies. One may also use the Perron-Frobenius theorem to prove that this dominant eigenvector has nonnegative entries, thus specifying a physically meaningful sequence distribution (Harris, 2009).

3.3. THE ERROR THRESHOLD

Theory and experiment show that the "master sequence" rarely makes up more than a few percent of a replicator population (Eigen, 1977; Quer et al., 1996). Though each individual mutant is much less abundant than the master sequence, the set of possible mutants is diverse enough that they dominate the population numerically. But despite the relatively low abundance of master copies, the fitness of a quasispecies depends very strongly upon the fitness of the master. Eigen (1977) used perturbation theory to show that the fitness of the dominant quasispecies is approximately

$$W_m + \sum_{k\neq m} \frac{w_{km}w_{mk}}{W_m - W_k}; \tag{6}$$

it follows that a dominant quasispecies replicating with high fidelity in a population with large fitness differences should have a fitness very close to W_m.

Master quasispecies proliferate because x_m's productivity exceeds the population average, meaning that

$$A_m - D_m > \frac{\sum_{k \neq m} (A_k - D_k) x_k}{\sum_{k \neq m} x_k}. \tag{7}$$

However, it is possible for the concentration of x_m to decrease at the same time that its descendents proliferate. The number of possible mutants is great enough that back mutation is a negligible source of perfect master copies, meaning that the concentration of master sequences will only increase with time if x_m makes perfect copies of itself at a rate that exceeds the average population productivity. This is contingent on the inequality

$$A_m Q_m - D_m > \frac{\sum_{k \neq m} (A_k - D_k) x_k}{\sum_{k \neq m} x_k} \tag{8}$$

and a population that satisfies (7) but not (8) will succumb to an *error catastrophe* where resources disperse evenly through sequence space. In this situation, dominance drifts randomly as every species dies out soon after it appears (Eigen et al., 1988).

Inequality (8) dictates a sharp *error threshold* that is usually represented as the minimum replication accuracy that is required to preserve a certain amount of information. It is standard to assume that replicators are assembled by adding monomers one by one to the end of a growing chain, letting q be the probability of incorporating the correct monomer during any given assembly step. If $x_1,\ldots,x_n$ are polymers of length v, then $Q_1 = \cdots = Q_n = q^v$. Thus, we can let

$$\sigma_m = \frac{A_m}{D_m + \frac{\sum_{k \neq m} (A_k - D_k) x_k}{\sum_{k \neq m} x_k}} \tag{9}$$

denote a measure of x_m's superiority and see that an error catastrophe happens when

$$v > \frac{\log \sigma_m}{1 - q}. \tag{10}$$

The error threshold is a very testable result. Experimentalists have created error catastrophes by sabotaging error correction *in vitro*, and some useful antivirals have this effect (Anderson et al., 2004; Crotty et al., 2004; Eigen, 2002).

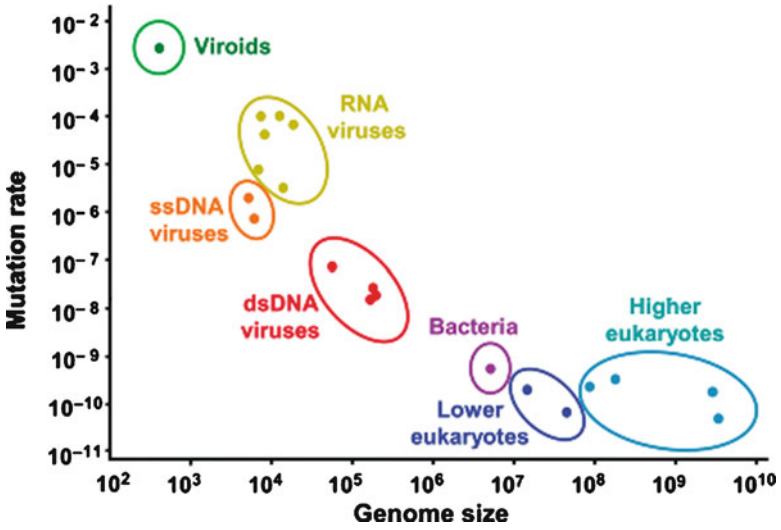


Figure 1. Inverse relationship between mutation rate and genome size, illustrating the extent to which the error threshold pressures large genomes to evolve high-fidelity replication mechanisms. (Reprinted with permission from a paper reporting that a 399-nucleotide viroid genome has the exceptionally high error rate of 1 mutation per 400 bases per generation (Gago et al., 2009).)

Researchers have also found a general inverse correlation between genome size and the effectiveness of error-correcting machinery (Gago et al., 2009; Fig. 1). Viral Qβ replicase, an RNA-dependent RNA polymerase, has a predicted error threshold around 10^4 bases, which is the observed length of the virus's single-stranded RNA genome. Double-stranded genomes allow for specialized enzymatic error checks that reduce the mutation rate as much as 1,000-fold, and bacterial genomes can reach 10^7 nucleotides (Eigen, 1977). Eukaryotes use error-correction mechanisms that are even more complex, some facilitated by diploidy, which permit even longer genomes.

Although the error threshold is a beautiful, nonobvious explanation for the inverse relationship between genome size and mutation rate, it is also true that error-correcting enzymes take up a lot of genome space and that the machinery for translating DNA into protein takes up even more space. A 10,000-base viral genome is too small to encode translation machinery (Eigen, 1977), yet it could not elongate without enzymatic stewardship. Even earlier in the emergence of life, ribozymes presumably propagated during the RNA world, but nonenzymatic template-directed polymerization appears to be too inaccurate to copy sequences of ribozyme length (error rate near 20%, while ribozymes are typically >30 bases) (Hagenbuch et al., 2005). Furthermore, a ribozyme that catalyzes template-directed RNA polymerization (Johnston et al., 2001) also appears to be too inaccurate to copy its own sequence (mutation rate of 3%, but length close to 200 bases). This apparent evolutionary barrier is known as *Eigen's paradox*.

In the next section, we describe several proposed solutions to Eigen's paradox. We begin with the solution that Eigen himself proposed when considering the evolution of the translation machinery.

4. Overcoming Eigen's Paradox

4.1. THE HYPERCYCLE

Eigen's paradox presents a serious obstacle to the invention of translation by a lone self-replicating molecule. However, we have seen that no replicator can eliminate the other members of its quasispecies. But if the members of a quasispecies could cooperate, the whole group would have the potential to defeat Eigen's paradox because it could survive at higher error rates than a single species containing the same amount of information.

We can convince ourselves of this by a thought experiment in which A, B, and C reproduce faithfully enough to preserve their own sequences, but not faithfully enough to replicate the concatenation ABC. If A, B, and C were concatenated by chance, perfect ABC master sequences would decline and disappear even as A, B, and C persisted as master subsequences. The expected frequency of perfect ABCs is proportional to the product of the frequencies of A, B, and C, and since A, B, and C are replicating so close to their own error thresholds, their abundances will be low enough that stochastic fluctuations would likely eliminate ABC entirely. But if A, B, and C are not concatenated and perfect copies have a mechanism for finding one another in solution, ABC cooperatives can persist at the concentration of their limiting component.

Eigen and Schuster used dynamical systems theory to model various catalytic replicator networks, varying the ways that participating species could catalyze one another's replication. They found that some networks dissolved as key species went extinct, while others stabilized and prospered (Eigen, 1977; Eigen and Schuster, 1978a, b). Specifically, they claim that the only stable self-replicating network is an unbranched circle they call a *hypercycle* (1978a), and they go on to describe a specific hypercycle that could account for the origin of translation (1978b).

The hypercycle is a network of n species $E_1,\ldots,E_n$, where E_i produces copies of E_{i+1} and E_n produces copies of E_1. DNA replication is a two-element hypercycle since each strand templates its complement rather than itself. In theory, if a hypercycle of ten RNA 10^4-mers could replicate without error-correcting enzymes, this system could encode a translation system.

Since a dominant quasispecies drives competitors to extinction, the members of an early hypercycle would probably have to belong to the same quasispecies. One might object that ten RNA 10^4-mers from the same quasispecies contain much less information than a typical RNA 10^5-mer since they share many common residues. Whether ten related RNA sequences would be capable of performing different functions is an interesting experimental question, but one may argue

that specialization of mutants from a quasispecies is plausible because it is analogous to gene duplication, which provides much raw material for modern evolution. Interestingly, examples of an RNA sequence that can perform two functions have been found (Schultes and Bartel, 2000; Vaidya and Lehman, 2009), supporting the idea that sequences from the same quasispecies could have different functions. However, such sequences may be relatively rare.

Eigen and Schuster proposed that early translation could proceed without ribosomes and synthetases by substituting nonspecific inorganic catalysts and modifying tRNA structure (1978b). They are left with an army of tRNAs that could, in principle, belong to one quasispecies. Phylogenetic analysis of modern tRNA sequences suggests that they could have diverged from a common ancestor within a quasispecies framework (Eigen and Winkler-Oswatitsch, 1981a, b). In addition, a computer simulation of RNA evolution that assigns a selective advantage to rows of stable internal base pairs has a very high probability of producing a cloverleaf structure that resembles tRNA (Eigen, 1971). These pieces of evidence add support to the idea that tRNA was the first part of the translation apparatus to evolve.

The scenario suggested by Eigen and Schuster is as follows. Once certain quasispecies members started to act as tRNAs, the RNA replicators $I_1,\ldots,I_n$ could template the formation of polypeptides $E_1,\ldots,E_n$. In turn, these polypeptides could catalyze nucleotide synthesis. If synthetases are not required and E_i replicates $I_{i+1-\delta_{n1}n}$ more efficiently than it replicates any other I_j, we get a hypercyclic structure (Fig. 2). The relative strengths of catalysis are important; imbalances can cause a complex network to disintegrate to a simpler hypercycle and a parasitized network to go completely extinct (Fig. 3).

Implementing hypercyclic translation in the lab would involve overcoming significant challenges (e.g., lack of synthetases and ribosomes, existence of multifunctional tRNA-like molecules). However, these challenges must be weighed against the likelihood that branched networks are fated to collapse. If Eigen and Schuster's result is broadly applicable, then perhaps efforts to build a rudimentary translation apparatus should focus on constructing cyclic networks.

Some of the hypercycle's problems can be resolved by embedding it in a spatially organized ecology. Eigen and Schuster assume that the components of the hypercycle are well mixed with each other and with competing species, but hypercycles become more robust when new species take time to disperse from their places of origin. One problem with the classical hypercycle is its susceptibility to parasites that receive catalysis without offering any in return (Eigen and Schuster, 1978b), but when spatial organization forces parasites to attack the edges of a hypercycle instead of permeating it instantly, the hypercycle is capable of overcoming most challenges (Boerlijst and Hogeweg, 1991).

In addition, spatial organization makes hypercycles less likely to destroy one another. In contrast to replicator quasispecies, where low concentrations of less fit species persist indefinitely, dominant hypercycles drive all competing networks to extinction under conditions of instantaneous mixing. These unforgiving selection

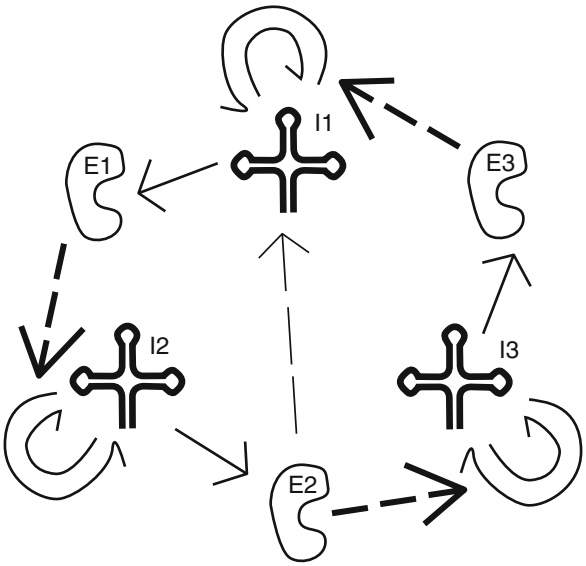


Figure 2. In this simple hypercyclic translation apparatus, $I_1,\ldots,I_3$ are tRNA-like molecules that double as templates for the peptide catalysts $E_1,\ldots,E_3$, which in turn facilitate production of the templates. A *dotted arrow* pointing to a *two-headed reaction arrow* denotes catalysis of a self-replication reaction, with *bold arrows* representing strong catalysis and the *thin arrow* representing weaker catalysis. *Solid arrows* denote the catalytic production of enzymes from monomers that are not shown. In theory, all molecular species survive indefinitely, stabilizing at finite equilibrium concentrations. If we complicate this hypercycle by introducing additional catalytic dependencies such as $E_2 \rightarrow I_1$ (*thin dotted arrow*), with E_2 catalyzing I_1's self-replication, then as long as these dependencies have weaker rate constants than the dependencies $E_i \rightarrow I_{i+1}$, they will alter the position of equilibrium only slightly without jeopardizing the stable propagation of the hypercycle.

dynamics leave hypercycles vulnerable to parasites; the slightest fitness difference is enough to let a parasite extinguish an essential network link (see Fig. 3). They would also make it hard to escape a local fitness maximum, repressing the formation of modestly fit species that could evolve into dominant species over time.

Spatial organization mitigates the parasite threat by forcing competition to occur along a threshold in space. Such a threshold borders a parasite-free zone where the hypercycle reigns undisturbed. A similar threshold can separate regions that are dominated by two different coexisting hypercycles. Such partitioning could shield pockets of hypercycles that are less fit than their neighbors, and such hypercycles might evolve further to surpass the current fitness maximum.

Depending on the tendency of real catalytic networks to remain stuck in local fitness maxima, selection of a hypercycle precursor to translation could have been a "frozen accident" that persisted because of evolutionary inertia (Crick, 1963, 1968). The modern analogy is that any organism that attempts to alter its genetic code now would likely experience a severe fitness decline in the short run, due to the number of components that work together in the existing system.

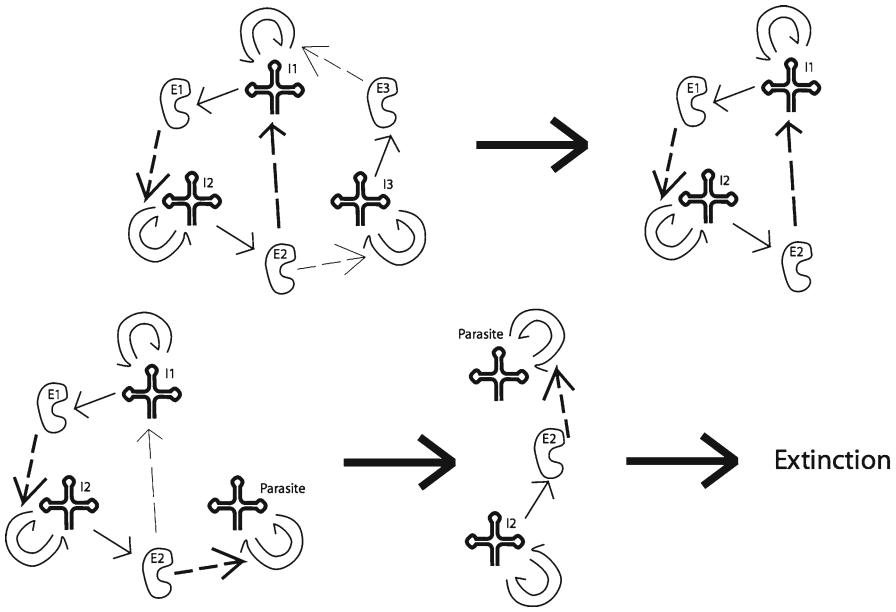


Figure 3. In Eigen and Schuster's treatment of well-mixed chemical systems (1978b), parasites invariably drive a network to extinction, while branched networks reduce to a single hypercyclic component.

However, some evidence suggests that the genetic code may actually be somewhat optimized. Plausibly, prebiotic amino acids are specified by GC-rich codons (e.g., glycine = GGC; alanine = GCC), which are the codons that would have functioned best without catalytic help (Eigen and Schuster, 1978b). Others have noted that the existing genetic code maximizes the likelihood that a point mutation will induce the substitution of an amino acid that is chemically similar to the wild-type residue (Freeland and Hurst, 1998; Freeland et al., 2000).

Historical questions such as these cannot always be satisfactorily resolved by examination of existing natural systems. Further insights may be gained through experimental efforts to implement a hypercycle. Lincoln and Joyce recently constructed a pair of cross-catalytic ribozymes, which form a kind of two-member hypercycle and can reproduce each other indefinitely in the absence of protein (2009). In addition, Hayden and Lehman were able to cleave the *Azoarcus* group I ribozyme into four parts that are inactive on their own but can find each other in solution to cooperatively cleave and recombine RNA (2006).

4.2. RELAXATION OF THE ERROR THRESHOLD

Researchers have come up with many scenarios that challenge Eigen's paradox without invoking translation or extreme hypercyclic cooperation. While spatial organization increases the robustness and believability of the hypercycle idea, it

also has a similar effect on competing theories. For example, simulations suggest that ribozymes adhering to a charged mineral surface would encounter one another often enough to evolve significant reciprocal altruism (Szabo et al., 2002), an effect that might increase their fitness enough to prevent error catastrophes. Altmeyer and McCaskill have analytically shown that the error threshold decreases monotonically with the diffusion coefficient (2001), independently of hypercycles.

In addition to spatial segregation, there are other effects that can increase the error tolerance of simple, noncooperative quasispecies. One such effect is stalling, a relaxation of the error threshold based on a realistic feature of the polymerization mechanism of nucleic acids. Replication slows during the production of mutant sequences (Huang et al., 1992; Ichida et al., 2005a, b; Mendelman et al., 1990; Perrino and Loeb, 1989), and replication fork stalling is also a known mechanism for error correction in vivo (Krasilnikova and Mirkin, 2004; Mirkin and Mirkin, 2007). If stalling after mutations dampens the production of mutants effectively enough, it may allow useful ribozymes to emerge independently of translation machinery.

Previous work has suggested that nucleic acid replication without enzymes is too error prone to replicate ribozymes or deoxyribozymes. However, a study of nonenzymatic replication suggests that stalling could have significantly raised the prebiotic error threshold (Rajamani et al., 2010). Rajamani et al. determined rate constants for the incorporation of matched and mismatched nucleotides in a nonenzymatic polymerization reaction, finding that the mutation rate of a single nucleotide extension is 7.6%, which is too high to sustain a moderately fit ribozyme. However, a mismatch stalls further extension by more than two orders of magnitude, effectively curtailing the production of mutant sequences. Eigen's equations predict that a 7.6% mutation rate will cause an error catastrophe to destroy any replicator longer than 13 bases, but stalling pushes back this theoretical barrier to 39 bases, making their system accurate enough to propagate ribozyme-length sequences. This mechanism is essentially a way to propagate the very first ribozymes (e.g., able to cooperate with one another); such sequences could then evolve greater complexity through hypercyclic organization, cooperation in spatial structure, or other means.

Once stalling or similar effects allow ribozymes to arise through nonenzymatic polymerization, several mechanisms could lead to their diversification and fixation. Simulations involving "digital organisms" suggest that diverse quasispecies with flat fitness peaks tend to outcompete quasispecies with higher but narrower fitness peaks (Wilke et al., 2001). In an RNA world setting, such "survival of the flattest" might generate relatively diverse arrays of good catalysts that might later assemble themselves into hypercycles. For example, many nucleobases in a given ribozyme can be mutated without substantially altering its function (e.g., structural residues in a hairpin). Also, these sites contain essentially no information, so they would not "count" in the calculation of the error threshold. In other words, if 25% of sites are neutral (Kun et al., 2005), then the informative

length of this ribozyme would be 75% of its physical length, allowing physically longer sequences to propagate.

We do not know how fit the earliest ribozymes could have been, and most scientists believe that early fitness was probably in the single digits (hence the approximation that $(\ln f)/\mu \approx 1/\mu$). However, invoking high fitness would increase the amount of information that could have predated the origin of translation. One study of mutagenesis data suggests that modern ribozymes have quite high chemical activity fitness, relative to nearby mutants (Kun et al., 2005).

The effects of stalling, spatial organization, and high fitness gradients take some pressure off of the hypercycle as the only way to circumvent Eigen's paradox. However, they could at most delay the need to develop higher-level organization. Hypercyclic structure is a kind of bridge between passive replicators and complex, metabolizing cells; Fontana et al. claim, based on λ-calculus models, that it is inevitable for replicators to eventually form hypercycles, particularly when a force begins to inhibit simple self-reproduction (1994).

In the final section of this chapter, we turn our attention to the origin of replication itself. Nowak and Ohtsuki have recently presented a model demonstrating how natural selection precedes replication (2008). In a sense, their work is a "prequel" to Eigen and Schuster's scenarios.

5. Prevolutionary Dynamics and the Origin of Replication

5.1. SELECTION BEFORE REPLICATION

Nowak and Ohtsuki's model of "prevolution" describes binary sequences of 0s and 1s that elongate by reacting with activated monomers. There is no black-box replication; for example, the species x_{101} can be converted to the species x_{1010} by addition of a single monomer but cannot make more copies of itself. Activated nucleotides can also be deactivated hydrolytically, producing the monomeric species x_0 and x_1. For constant population size, x_i decays at the rate Dx_i. We let $x_{i'}$ denote the precursor of x_i and let a_i denote the rate at which $x_{i'}$ is transformed into x_i. Prevolution follows a linear set of growth equations:

$$\dot{x}_i = a_i x_{i'} - (D + a_{i0} + a_{i1}) x_i \tag{11}$$

Since the empty string is the immediate precursor to any 1-element string, it follows that $x_{0'} = x_{1'}$. Furthermore, since activated nucleotides become deactivated by reacting with superabundant water molecules, we assume that $x_{0'} = x_{1'} = 1$, i.e., that the empty string is more abundant than any other string.

Since (11) is linear with constant coefficients, prevolutionary dynamics are much simpler than the dynamics of quasispecies or catalytic networks. Letting $\mathbf{x} = (x_{0'} = x_{1'}, x_0, x_1, x_{00}, x_{01}, x_{10}, x_{11}, \ldots)$ be the infinite vector whose entries are the finite binary strings, we can write system (11) as $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ for a matrix $\mathbf{A}$ with

constant entries. Its solutions are the vectors of the form $\mathbf{x} = \exp(\mathbf{A}t)\mathbf{x}_0$, where $\mathbf{x}_0$ is the constant vector of initial conditions. The endpoint of evolution is the solution of the equation system $\dot{\mathbf{x}} = \mathbf{0}$, whose vector entries satisfy

$$\frac{x_i}{x_{i'}} = \frac{a_i}{d + a_{i0} + a_{i1}} := b_i. \tag{12}$$

Given that x_i begins with the nucleotide $\sigma \in \{0,1\}$, it follows that

$$x_i = \frac{x_i}{x_{i'}} \cdot \frac{x_{i'}}{x_{i''}} \cdots \frac{x_\sigma}{x_{\sigma'}} = b_i b_{i'} \ldots b_\sigma \tag{13}$$

at equilibrium. Thus, short sequences have exponentially greater abundance than long sequences. Such length distributions are typical of polymerization reactions (for a prebiotic example, see Lawless and Yuen, 1979).

Differences in reaction rate create further asymmetries that resemble the outcome of natural selection. The next step is to include a mechanism for self-replication.

5.2. SELECTION FOR REPLICATION

Transmission of genetic information could occur via an experimentally described process (Hill and Orgel, 1993; Mansy et al., 2008), where template-directed primer extension proceeds faster than the extension of an unpaired strand, even when no enzymes are involved. This rate asymmetry would select for master sequences that are particularly good at base pairing to form extendible primer-template complexes; the dynamics of this process should follow a scenario described by Nowak and Ohtsuki.

Working with the master sequence 0^n (assumed to be taken together with its complement), Nowak and Ohtsuki suppose that 0^k becomes 0^{k+1} at the rate b whenever $k < n$, while all other extensions happen at the rate $a < b$. Fixing a and letting $a_0 = a_1 = \alpha$, we can deduce from (12) that

$$x_m = \frac{\alpha}{a}\left(\frac{b}{a+b+d}\right)^{n-1}\left(\frac{a}{2a+d}\right), \tag{14}$$

such that the abundance of x_m approaches $\alpha/(2a+d)$ as b approaches infinity. In contrast, the abundance of x_i approaches zero whenever x_i is not a subsequence of 0^n.

This computation indicates how a master sequence, once formed, could have avoided extinction and increased its abundance. However, it does not change the earlier result that sequence abundance decreases exponentially with length. Another mechanism is needed to shift the population toward long sequences, a

prerequisite for ribozymes. Manapat et al. demonstrate how another chemical process, template-directed ligation, could bias prelife toward the production of long sequences (2010).

As an aside, papers about prelife avoid the 4-base complexity of the modern genetic code, but instead consider one or two bases and ignore complementarity (i.e., consider the entire duplex together). However, Sievers and von Kiedrowski report that cross-catalytic replication and self-catalytic replication should have similar efficiencies and obey similar dynamics (1994).

If mutations occur, certain assumptions make it possible to find the accompanying error threshold. Specifically, we assume that subsequences of 0^n will form at the rate $b(1-u)$, while mutants of 0^k 1, with $0<k<n$, form at the rate $a+bu$. All other extensions proceed at the rate a. As before, the limiting master abundance is $\alpha/(2a+d)$, but x_m cannot proliferate to $1/k$ times its maximum abundance unless

$$u < \frac{\log k}{n} - \frac{a+d}{b}. \tag{15}$$

Once true replicators start proliferating, they will obey the dynamics introduced in Sect. 3.1. Hybrid replicator equations describe the coexistence of "prevolution" with true evolution:

$$\dot{x}_i = a_i x_{i'} - (d + a_{i0} + a_{i1}) x_i + r x_i (f_i - \phi). \tag{16}$$

The constant r scales the relative speeds of template-directed replication and nontemplated sequence extension, while f_i denotes the fitness of x_i. ϕ is the outflow that balances excess productivity.

Computer simulations show that the abundance of replicators shoots up immediately above the critical value of r at which there are more copies of x_i being created by replication than being converted to other sequences by random nucleotide addition (Nowak and Ohtsuki, 2008). This phase transition seems to denote the change from "prevolution" to evolution, when

$$-(d + a_{i0} + a_{i1}) + r(f_i - \phi) > 0, \tag{17}$$

such that

$$r > r_c := \frac{d + a_{i0} + a_{i1}}{f_i - \phi}. \tag{18}$$

This threshold changes with the presence of replication errors, resulting in a different kind of error threshold.

Prelife as described by Manapat et al. (2009) has begun to converge with what experimentalists can make (Mansy et al., 2008): membrane-bound polynucleotides that base pair, dissociate, and extend, all without the help of protein catalysis. Lab results confirm that the theory is more than just a fantasy, while theory suggests that protocells in the lab could evolve in very interesting ways.

Every result in theoretical chemistry is conditional on simplifications, and these simplifications limit the degree to which anything can be proved to be "impossible." This is certainly the case with Eigen's paradox, which we have seen to admit a great many loopholes. However, it would have been much more difficult to find these loopholes without understanding Eigen's theory, looking at his assumptions and deliberately constructing situations where they break down.

Quasispecies theory and its offshoots predict that whole classes of experiments should fail, challenging lab scientists to look for success in less predictable setups. But just as hypercycle theory predicts that certain early translation mechanisms would have been inviable, Nowak and Ohtsuki's work predicts that certain chemical systems should be experimentally interesting. Similarly, experimentalists could contribute much to the direction of further theoretical treatments because realistic details that appear to be minor can have major consequences (e.g., the fitness landscape is an extremely important, but currently poorly understood, detail of replication models that is often lumped into the replicative rate "r").

6. Concluding Remarks

In this chapter, we have described a handful of major frameworks for modeling prebiotic replication of informational molecules mathematically, with the intention of connecting experimentalists to this fruitful area. However, the related literature is much larger than the focus of this chapter. Outside the domain of purely mathematical models, much recent effort has been devoted to computer simulation of evolving digital organisms, such as the Avida platform (Ofria and Wilke, 2004), which has also yielded interesting insights and predictions (e.g., about complexity (Adami et al., 2000)). In addition, the consideration of cellular compartmentalization, such as by lipid membranes, leads to interesting simulations of higher organization (e.g., see Hutton, 2007, and Sole et al., 2009). We have not reviewed this extensive and rich literature.

We might never know exactly how our earliest history unfolded, but we can hope to learn the features of that history that were necessary, not contingent Fontana and Buss (1994). Even scholars of time periods that did leave fossils behind work to distinguish contingency from necessity, a goal that was popularized in Stephen Jay Gould's *Wonderful Life* (1989); after piecing together some of the history of the animals preserved in the Burgess Shale, Gould wonders how much of that history would stay the same if "the tape were played twice." The theories we have described are essentially attempts to delineate necessary constraints on early life, with the hope that whatever we learn would hold during any playback of Gould's "tape." Determining whether this is actually the case will require many joint adventures in experiment and theory.

7. References

Adami C, Ofria C, Collier TC (2000) Evolution of biological complexity. Proc Natl Acad Sci 97(9):4463–4468

Altmeyer S, McCaskill JS (2001) Error threshold for spatially resolved evolution in the quasispecies model. Phys Rev Lett 86:5819–5822

Anderson JP, Daifuku R, Loeb LA (2004) Viral error catastrophe by mutagenic nucleosides. Annu Rev Microbiol 58:183–205

Anet FAL (2004) The place of metabolism in the origin of life. Curr Op Chem Biol 8(6):654–659

Bartel DP, Szostak JW (1993) Isolation of a new ribozymes from a large pool of random sequences. Science 261:1411–1418

Boerlijst MC, Hogeweg P (1991) Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. Phys D: Nonlinear Phenomena 48(1):17–28

Crick FHC (1963) On the genetic code: deductions about the general nature of the code are drawn from results of biochemical experimentation. Science 8:461–464

Crick FHC (1968) The origin of the genetic code. J Mol Biol 38(3):367–379

Crotty S, Cameron CE, Andino R (2004) RNA virus error catastrophe: direct molecular test by using ribavirin. Proc Natl Acad Sci U S A 98(12):6895–6900

Domingo E, Holland JJ (1997) RNA virus mutations and fitness for survival. Annu Rev Microbiol 51:151–178

Doudna JD, Szostak JW (1989) RNA-catalyzed synthesis of complementary-strand RNA. Nature 339:519–522

Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 33a:465–523

Eigen M (1977) The hypercycle: a principle of natural self-organization; Part A: emergence of the hypercycle. Naturwissenschaften 64:541–565

Eigen M (1993) The origin of genetic information: viruses as models. Gene 135:37–47

Eigen M (2002) Error catastrophe and antiviral strategy. Proc Natl Acad Sci U S A 99(21):13374–13376

Eigen M, Schuster P (1978a) The hypercycle: a principle of natural self-organization; Part B: The abstract hypercycle. Naturwissenschaften 65:7–41

Eigen M, Schuster P (1978b) The hypercycle: a principle of natural self-organization; Part C: The realistic hypercycle. Naturwissenschaften 65:341–369

Eigen M, Winkler-Oswatitsch R (1981a) Transfer RNA: the early adaptor. Naturwissenschaften 68:217–228

Eigen M, Winkler-Oswatitsch R (1981b) Transfer RNA, an early gene? Naturwissenschaften 68:282–292

Eigen M, McCaskill J, Schuster P (1988) Molecular quasi-species. J Phys Chem 92:6881–6891

Ekland EH, Bartel DP (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. Nucleic Acids Res 23(16):3231–3238

Ertem G, Ferris JP (1996) Synthesis of RNA oligomers on heterogeneous templates. Nature 379:238–240

Ferris JP, Ertem G (1992) Oligomerization of ribonucleotides on montmorillonite: reaction of the 5′-phosphorimidazolide of adenosine. Science 257(5075):1387–1389

Ferris JP, Ertem G (1993) Montmorillonite catalysis of RNA oligomer formation in aqueous solution. A model for the prebiotic formation of RNA. J Am Chem Soc 115(26):12270–12275

Ferris JP, Hill AR, Liu R, Orgel LE (1996) Synthesis of long prebiotic oligomers on mineral surfaces. Nature 381:59–61

Fontana W, Buss LW (1994) What would be conserved if "the tape were played twice"? Proc Natl Acad Sci U S A 91:757–761

Fontana W, Wagner GP, Buss LW (1994) Beyond digital naturalism. Artif Life 1:211–227

Freeland SJ, Hurst LD (1998) The genetic code is one in a million. J Mol Evol 47:238–248

Freeland SJ, Knight RD, Landweber LF, Hurst LD (2000) Early fixation of an optimal genetic code. Mol Biol Evol 17:511–518

Gago S, Elena SF, Flores R, Sanjuan R (2009) Extremely high mutation rate of a hammerhead viroid. Science 323:1308

Gould SJ (1989) Wonderful life. Norton, New York

Grande-Perez A, Sierra S, Castro MG, Domingo E, Lowenstein PR (2002) Molecular indetermination in the transition to error catastrophe: systematic elimination of lymphocytic choriomeningitis virus through mutagenesis does not correlate linearly with large increases in mutant spectrum complexity. Proc Natl Acad Sci U S A 99(20):12938–12943

Green R, Szostak JW (1992) Selection of a ribozyme that functions as a superior template in a self-copying reaction. Science 258(5090):1910–1915

Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35:849–857

Hagenbuch P, Kervio E, Hochgesand A, Plutowski U, Richert C (2005) Chemical primer extension: efficiently determining single nucleotides in DNA. Angew Chem Int Ed 44:6588–6592

Hager AJ, Szostak JW (1997) Isolation of novel ribozymes that ligate AMP-activated RNA substrates. Chem Biol 4(8):607–617

Harris K (2009) On the dynamics of prebiotic evolution. A.B. honors thesis, Harvard University, http://brauer.harvard.edu/theses/senior/harris/harris.pdf

Hayden EJ, Lehman N (2006) Self-assembly of a group I intron from inactive oligonucleotide fragments. Chem Biol 13(8):909–918

Hayden EJ, Riley CA, Burton AS, Lehman N (2005) RNA-directed construction of structurally complex and active ligase ribozymes through recombination. RNA 11:1678–1687

Hill AR Jr, Orgel LE (1993) The limits of template-directed synthesis with nucleoside-5′-phosophoro (2-methyl)imidazolides. Orig Life Evol Biosph 23:285–290

Horn RA, Johnson CR (1991) Matrix analysis. Cambridge University Press, New York

Huang M-M, Arnheim N, Goodman MF (1992) Extension of base mispairs by *Taq* DNA polymerase: implications for single nucleotide discrimination in PCR. Nucleic Acids Res 20(17):4567–4573

Hutton TJ (2007) Evolvable self-reproducing cells in a two-dimensional artificial chemistry. Artif Life 13(1):11–30

Ichida JK, Horhota A, Zou K, McLaughlin LW, Szostak JW (2005a) High fidelity TNA synthesis by Therminator polymerase. Nucleic Acids Res 33(6):5219–5225

Ichida JK, Zou K, Horhota A, Yu B, McLaughlin LW, Szostak JW (2005b) An *in vitro* selection system for TNA. J Am Chem Soc 127(9):2802–2803

Jackson SE (1998) How do small single-domain proteins fold? Fold Des 3:R81–R91

Jaeger L, Wright MC, Joyce GF (1999) A complex ligase ribozyme evolved *in vitro* from a group I ribozyme domain. Proc Natl Acad Sci U S A 96(26):14712–14717

James KD, Ellington AD (1997) Surprising fidelity of template-directed chemical ligation of oligonucleotides. Chem Biol 4(8):595–605

Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. Science 292:1319–1325

Kim D-E, Joyce GF (2004) Cross-catalytic replication of an RNA ligase ribozyme. Chem Biol 11(11):1505–1512

Krasilnikova MM, Mirkin SM (2004) Replication stalling at Friedrich's ataxia $(GAA)_n$ repeats *in vivo*. Mol Cell Biol 24(6):2286–2295

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell 31:147–157

Kun A, Santos M, Szathmary E (2005) Real ribozymes suggest a relaxed error threshold. Nat Genet 37(9):1008–1011

Lawless JG, Yuen GU (1979) Quantification of monocarboxylic acids in the Murchison carbonaceous meteorite. Nature 282:396–398

Lee DH, Granja JR, Martinez JA, Severin K, Ghadiri MR (1996) A self-replicating peptide. Nature 382:525–528

Lincoln TA, Joyce GF (2009) Self-sustained replication of an RNA enzyme. Science 323(5918): 1229–1232

Manapat ML, Ohtsuki H, Burger R, Nowak MA (2009) Originator dynamics. J Theor Biol 256: 586–595

Manapat ML, Chen IA, Nowak MA (2010) The basic reproductive ratio of life. J Theor Biol 263(3):317–327

Mansy SS, Szostak JW (2009) Reconstructing the emergence of cellular life through the synthesis of model protocells. Cold Spring Harb Symp Quant Biol 74:47–54

Mansy SS, Schrum JP, Krishnamurthy M, Tobe S, Treco DA, Szostak JW (2008) Template-directed synthesis of a genetic polymer in a model protocell. Nature 454:122–125

Mendelman LV, Petruska J, Goodman MC (1990) Base mispair extension kinetics: comparison of DNA polymerase alpha and reverse transcriptase. J Biol Chem 265:2338–2346

Miller SL, Urey HC (1959) Organic compound synthesis on the primitive Earth: several questions about the origin of life have been answered, but much remains to be studied. Science 130(3370): 245–251

Mirkin EV, Mirkin AM (2007) Replication fork stalling at natural impediments. Microbiol Mol Biol Rev 71(1):13–35

Monnard PA, Deamer DW (2001) Nutrient uptake by protocells: a liposome model system. Orig Life Evol Biosph 31(1):147–155

Monnard PA, Deamer DW (2002) Membrane self-assembly processes: steps toward the first cellular life. Anat Rec 268:196–207

Nowak MA, Ohtsuki H (2008) Prevolutionary dynamics and the origin of evolution. Proc Natl Acad Sci U S A 105(39):14924–14927

Ofria C, Wilke CO (2004) Avida: a software platform for research in computational evolutionary biology. Artif Life 10(2):191–229

Ohtsuki H, Nowak MA (2009) Prelife catalysts and replicators. Proc R Soc B 279:3783–3790

Orgel L (1992) Molecular replication. Nature 358:203–209

Orgel L (2004) Prebiotic chemistry and the origin of the RNA world. Crit Rev Biochem Mol Biol 39:99–123

Perrino FW, Loeb LA (1989) Differential extension of 3′ mispairs is a major contribution to the high fidelity of calf thymus DNA polymerase-alpha. J Biol Chem 264:2898–2905

Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature 459:239–242

Pross A (2004) Causation and the origin of life: metabolism or replication first? Orig Life Evol Biosph 34:3307–3321

Quer J, Huerta R, Novella IS, Tsimring L, Domingo E, Holland JJ (1996) Reproducible nonlinear population dynamics and critical points during replicative competitions of RNA virus quasispecies. J Mol Biol 264:465–471

Rajamani S, Vlassov A, Benner S, Coombs A, Olasagasti F, Deamer D (2008) Lipid-assisted synthesis of RNA-like polymers from mononucleotides. Orig Life Evol Biosph 38(1):57–74

Rajamani S, Ichida JK, Antal T, Treco DA, Leu K, Nowak MA, Szostak JW, Chen IA (2010) Effect of stalling after mismatches on the error catastrophe in non-enzymatic nucleic acid replication. J Am Chem Soc 132(16):5880–5885

Ricardo A, Carrigan MA, Olcott AN, Benner SA (2004) Borate minerals stabilize ribose. Science 303(5655):196

Rohatgi R, Bartel DP, Szostak JW (1996a) Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. J Am Chem Soc 118(14):3332–3339

Rohatgi R, Bartel DP, Szostak JW (1996b) Nonenzymatic, template-directed ligation of oligoribonucleotides is highly regioselective for the formation of 3′-5′ phosphodiester bonds. J Am Chem Soc 118(14):3340–3344

Sawai H, Orgel LE (1975) Oligonucleotide synthesis catalyzed by the Zn^{2+} ion. J Am Chem Soc 97:3532–3533

Schultes EA, Bartel DP (2000) One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science 289(5478):448–452

Sievers D, von Kiedrowski G (1994) Self-replication of complementary nucleotide-based oligomers. Nature 369:221–224

Sole RV, Macia J, Fellermann H, Munteanu A, Sardanyes J, Valverde S (2009) Models of protocell replication. In: Rasmussen S, Bedau MA, Chen L, Deamer D, Krakauer DC, Packard NH, Stadler PF (eds) Protocells: bridging nonliving and living matter. MIT Press, Cambridge

Szabo P, Scheuring I, Czaran T, Szathmary E (2002) *In silico* simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. Nature 420:340–343

Thompson CJ, McBride JL (1974) On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. Math Biosci 21:127

Vaidya N, Lehman N (2009) One RNA plays three roles to provide catalytic activity to a group I intron lacking an endogenous internal guide sequence. Nucleic Acids Res 37(12):3981–3989

Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R (2006) Quasispecies diversity determines pathogenesis through cooperative interactions within a viral population. Nature 439(7074): 344–348

Wachtershauser G (1988) Before enzymes and templates: theory of surface metabolism. Microbiol Rev 52(4):452–484

Weisstein EW (2007) Vitalism theory—from Eric Weisstein's world of chemistry. Accessed at http://scienceworld.wolfram.com/chemistry/VitalismTheory.html on 14 Jan 2009

Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C (2001) Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412:331–333

Wochner A, Attwater J, Coulson A, Holliger P (2011) Ribozyme-catalyzed transcription of an active ribozyme. Science 332(6026):209–212