ACCOUNTS
of chemical research

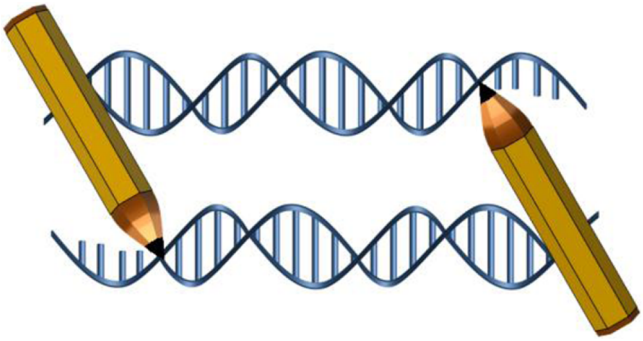
Abiotic Self-Replication

ADAM J. MEYER, JARED W. ELLEFSON, AND
ANDREW D. ELLINGTON*

Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712, United States

RECEIVED ON DECEMBER 12, 2011

CONSPECTUS



The key to the origins of life is the replication of information. Linear polymers such as nucleic acids that both carry information and can be replicated are currently what we consider to be the basis of living systems. However, these two properties are not necessarily coupled. The ability to mutate in a discrete or quantized way, without frequent reversion, may be an additional requirement for Darwinian evolution, in which case the notion that Darwinian evolution defines life may be less of a tautology than previously thought.

In this Account, we examine a variety of *in vitro* systems of increasing complexity, from simple chemical replicators up to complex systems based on *in vitro* transcription and translation. Comparing and contrasting these systems provides an interesting window onto the molecular origins of life.

For nucleic acids, the story likely begins with simple chemical replication, perhaps of the form A + B → T, in which T serves as a template for the joining of A and B. Molecular variants capable of faster replication would come to dominate a population, and the development of cycles in which templates could foster one another's replication would have led to increasingly complex replicators and from thence to the initial genomes. The initial genomes may have been propagated by RNA replicases, ribozymes capable of joining oligonucleotides and eventually polymerizing mononucleotide substrates. As ribozymes were added to the genome to fill gaps in the chemistry necessary for replication, the backbone of a putative RNA world would have emerged.

It is likely that such replicators would have been plagued by molecular parasites, which would have been passively replicated by the RNA world machinery without contributing to it. These molecular parasites would have been a major driver for the development of compartmentalization/cellularization, as more robust compartments could have outcompeted parasite-ridden compartments. The eventual outsourcing of metabolic functions (including the replication of nucleic acids) to more competent protein enzymes would complete the journey from an abiotic world to the molecular biology we see today.

While the hallmark of biology is the ability to self-replicate, this ability is found only rarely outside of cells. Cell-based organisms have an almost unique claim on the ability to continuously reproduce their information content (sequence) and functionality. That said, the molecular basis of replication, structural complementarity, has been demonstrated in a number of chemical and abiotic systems. Comparison of different abiotic, self-replicating systems leads to speculations regarding whether different life forms based on entirely different chemical species might exist elsewhere. In order to better illustrate (if only by comparison) the uniqueness of cellular biology and also the possibilities for going beyond terrestrial biology, it would be worthwhile to understand the extent and degree of replication that is possible in abiotic systems.[1]

From a different vantage, that of the origins of life, the journey from prelife chemistry to modern molecular biology likely began with (or at least included) chemical replication of linear polymers. As the complexity of replicators increased, they would have given rise to a nascent RNA world in which ribozymes were the primary catalysts of metabolism. The engines of any putative RNA world would have been ribozyme polymerases, possibly augmented by ribozyme ligases or recombinases. As with any

Published on the Web 08/15/2012 www.pubs.acs.org/accounts
10.1021/ar200325v

self-replicating species, these catalysts could have been parasitized, which would have driven the evolution of compartmentalization.[2]

In this Account, we examine a variety of *in vitro* systems of increasing complexity, going from simple chemical replicators up to complex systems based on *in vitro* transcription and translation. Comparing and contrasting these systems should provide interesting windows onto the molecular origins of life.

Chemical Templating of Replication, a General View

In Watson and Crick's seminal paper on the structure of DNA,[3] they suggest that nucleobase complementarity was the means by which biological information could be copied. Not only is this true of biological species, but it can be demonstrated *in vitro*, as well. The von Kiedrowski, Orgel, and Lynn laboratories explored how nonenzymatic ligation could lead to self-replication.[4−6] Von Kiedrowski showed that a 5′-CCG-3′ "A" trimer and a 5′-CGG-3′ "B" trimer annealed to a 5′-CCGCGG-3′ "AB" template, and upon nonenzymatic ligation a palindromic duplex was formed. Thermal fluctuations were sufficient to split the DNA duplex into the component 5′-CCGCGG-3′ hexamers, each of which could serve as the template for further ligation reactions. Beginning with an excess of substrate, autocatalytic conversion of substrate into template was observed. However, product inhibition resulted in weaker parabolic replication, rather than the ideal exponential replication, a phenomenon that von Kiedrowski has called the "survival of everyone".[7]

In order to create a more realistic model for nucleic acid replication, the von Kiedrowski laboratory allowed A and B to be joined in any combination: AA (CCGCCG), AB (CCGCGG), BA (CGGCCG), and BB (CGGCGG).[8] The sequence of the template determined what substrates could anneal to it. AA was complementary to two B trimers, thus it catalyzed formation of BB. Likewise, BB catalyzed AA formation. AB and BA were each palindromic, and thus self-templating. Given a starting population of A and B trimers, the dominant hexamer was determined by what hexamer was used to "kickstart" autocatalysis. For example, starting with free A, free B, and BB template would allow AA to dominate (it should be noted that the AA that formed would then serve as template for BB formation). This cross-catalytic (AA, BB) system has no requirement for palindromic sequence. This is advantageous because larger palindromic sequences form stable hairpin structures, thus hampering their ability to interact with other molecules. Cross-catalytic molecular replication may more closely resemble early self-replicators.

Nucleic acids are somewhat special in forming a set of highly specific, isosteric, and therefore encodable interactions. It has even been postulated that the charged backbone of nucleic acids is a requisite feature of any Darwinian replicator.[9] However, this uniqueness is best appreciated by looking at other molecules' attempt to mimic complementarity. The Ghadiri laboratory developed a peptide-based molecular replicator based on a coiled-coil leucine zipper.[10] Two 32-residue peptides could align in a parallel fashion, with interdigitating hydrophobic residues. After splitting one 32-mer into 17- and 15-mer substrates, the other 32-mer could act as a template for peptide condensation. Further work was done in which two different 15mers (N1 and N2) that could each react with a single 17mer (E), yielding two distinct 32mers (R1 and R2). R2 contained isoleucince and was more efficient at autocatalysis than R1, which contained valine. In consequence, a mixture of N1, N2 and E resulted in the formation of R1 and R2, with R2 having a slight advantage. Interestingly, seeding the reaction with R1 gave R2 a larger advantage, indicating that R1 cross-catalyzed the formation of R2 more efficiently than it autocatalyzed its own formation. Seeding the reaction with R2 yielded roughly equivalent rates of R1 and R2 formation, also indicating a preference for cross-catalysis. This strong tendency to cross-catalyze is ideal for a mutualistic relationship.

Such mutualistic interactions are very different than what is seen with nucleic acid replicators (and with classic Darwinian evolution; Figure 1). Nucleic acid identity is set by the rules of base-pairing, while information transfer in a peptide replicator is less exact. When a nucleic acid replicator sustains a mutation, there is a quantized, compensatory change in the preferred substrate. The informationally segregated mutant is forced to compete with its parent for resources. Thus, a nucleic acid replicator can evolve to be a better template without benefiting competing substrates (Figure 1A). However, mutation in a peptide replicator does not result in a quantized change in substrate preference. This means that cross-catalysis is commonplace, and it is more difficult to imagine a peptide replicator selfishly[11] evolving to be a better catalyst (Figure 1B). That said, for both types of replicators, some form of physical segregation (cellularization) will greatly assist selfishness, and help avoid parasitism, as we discuss below.

Ribozyme-Mediated Replication of Nucleic Acids

It seems plausible that at some point in life's evolutionary history, RNA served as the primary genetic as well as the catalytic molecule. However since the initial molecule that started current life has long since been lost to evolutionary history, the best scientists can do is try to mimic what the first replicators might have been like. Following a period of template-mediated replication, similar to what was described above for short oligonucleotides, catalysts capable of synthesizing themselves may have emerged, either from oligonucleotides or from mononucleotides (Figure 2A). Such a self-replicase (or "Xeroxase") would later have duplicated or elaborated function, creating substrates for itself and a nascent ribozyme-based metabolism.

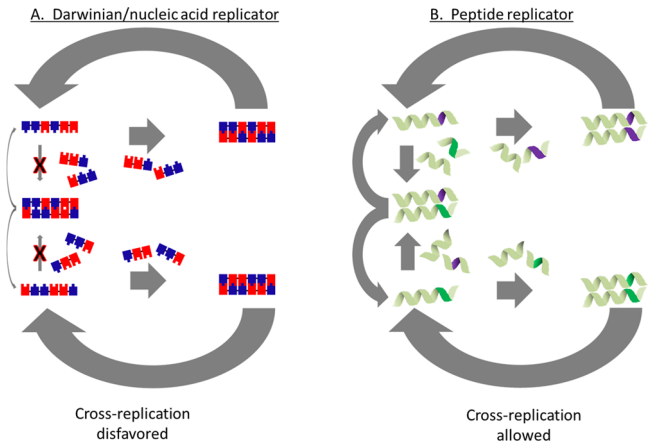


FIGURE 1. Base-pairing facilitates Darwinian competition. Molecular replicators in which the template catalyzes the creation of additional template from smaller substrates. (A) For nucleic acids replicators, templating is based on base pairing, so the formation of a mutant template is rare. Once formed, the mutant replicator forms a competing replication cycle. (B) For a peptide replicator, templating is less exact, so the formation of a mutant template is common. The mutant template can catalyze formation of mutant progeny or parental progeny, and the two species form a mutualistic network.

The quintessential example of how complex ribozymes, such as a replicase, could have emerged from relatively simple prebiotic systems came from an experiment carried out by David Bartel in the Szostak laboratory.[12] The fact that ribozymes existed and that functional nucleic acids could be selected from random sequence pools[13] begged the question of whether and how frequently ribozymes might also be selected *in vitro*. A ribozyme of particular interest was one that could catalyze 3′−5′ phosphodiester linkages, akin to how modern life's proteinaceous enzymes polymerize nucleic acids. A large randomized pool (>200 nucleotides) was generated and ribozyme ligases that could append a specific sequence tag to themselves were selectively amplified by reverse transcription and PCR. After multiple cycles of selection and amplification, ligase activity was indeed enriched in the pool. Further characterization revealed seven different families of ligases whose core catalytic sequences spanned from 56 to 191 nts. Surprisingly, all the ribozymes found catalyzed a 2′−5′ ligation reaction with the exception of one, the class I Bartel ligase.[14] This ribozyme was especially interesting in that it was large and relatively complex; indeed additional experiments that determined its

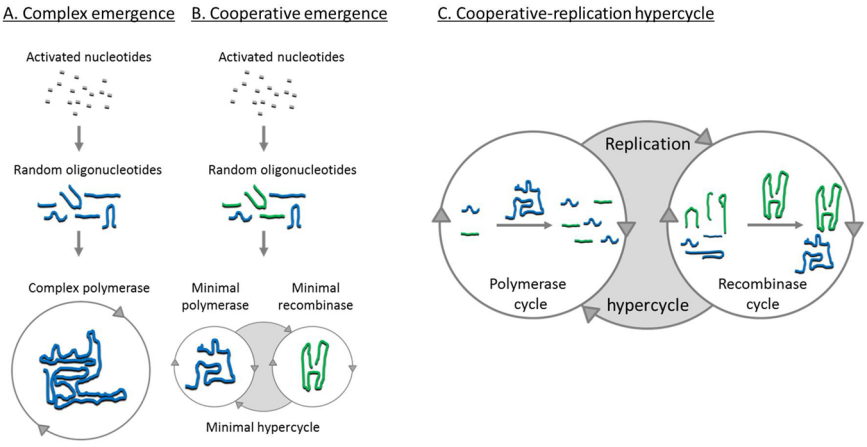


FIGURE 2. Emergence of a self-replication system through hypercycles. (A) In the first scenario, a complex RNA-dependent RNA polymerase capable of full self-replication emerges from random oligonucleotides. The ribozyme will likely be more complex than the tC19Z, because it will have to replicate structured templates. (B) In the second scenario, a minimal polymerase and minimal recombinase emerge from random oligonucleotides. These ribozymes cooperate to perform replication. (C) The replication hypercycle consists of two intertwined polymerization and recombination cycles. In one cycle, polymerization of the short RNA fragments comprising the polymerase and recombinase occurs through primer extension and dissociation of sense/antisense strands. In the other cycle, the reconstituted recombinase stitches the RNA fragments. Recombination is directed by internal guide sequences, forming longer, more complex ribozymes.

informational complexity suggested that it should only have been selected about once in every 10000 times the experiment was carried out.[14] This has been taken to mean that there may be many different ligases of roughly equal complexity in the vast sequence space that was explored and thus that complex structures and catalytic functionalities could in fact have been discovered in early evolution. While the initial catalytic rate of the class I ligase was modest, it was improved after rational engineering and further selection to the point that it was comparable to at least some protein enzymes that catalyze similar reactions.

If a ribozyme could form one 3′−5′ phosphodiester bond, perhaps the same ribozyme (or a derivative thereof) could form multiple such bonds.[15] Initially, the class I ligase proved capable of adding six mononucleotides to a primer with over 90% fidelity (which is too error-prone for sustained self-replication). The class I ligase became less efficient with each nucleotide extended, and it fell far short of being able to replicate itself. One issue was that the primer was attached to the ribozyme by base-pairing. This made it difficult for new nucleotides to move into the active site and be added. In order to reduce this steric constraint, it was hypothesized that an additional domain might bind a primer without base-pairing interactions, and thereby guide more general polymerization reactions.[16] To identify a new primer-binding domain, a 76 nt randomized region was added to the 3′ end of the ligase and a new selection scheme was devised to recover ribozymes with the best polymerization efficiency. One isolate of this selection, the R18 ribozyme, could add up to 14 nucleotides to the primer under optimal conditions. The fidelity had improved to 96.7% but was still too low to avoid an Eigen's "error catastrophe".[17] Since only about 1 in 1000 progeny would be identical to the parent, the ribozyme would need a several-hundred fold advantage over mutant progeny in order to survive.

Additional progress toward a replicase ribozyme was made when a novel selection scheme was developed by the Holliger laboratory.[18] DNA templates encoding ribozymes were attached to beads at a 1:1 ratio, and the beads were then physically segregated from one another in a water-in-oil emulsion (described below). After transcription, the ribozymes were also covalently attached to the beads, which allowed the emulsion to be broken and the beads to be transferred to a second emulsion where active ribozymes could extend a primer. After the second emulsion was broken, beads with extended primers (and therefore active ribozymes) were fluorescently labeled. Fluorescence-activated cell sorting was used to isolate the most fluorescent beads harboring the most active ribozymes, whose templates could be amplified to carry out additional cycles of selection and amplification.

This selection sought not only to improve the core of the ribozyme itself but also to add domains to the 5′ end of the ribozyme, which had been largely unaltered since the original class I ligase. A highly active version of the polymerase ribozyme, dubbed tC19Z, was generated.[18] The combined improvements in the ribozyme core and the evolution of a 5′ template hybridization domain greatly increased the ribozyme's polymerization activity. Ribozyme tC19Z can extend a primer nearly 100 nucleotides on a specially designed template. Again, improvements in processivity were mirrored by improvements in the fidelity of the ribozyme, which increased nearly 5-fold compared with the R18 ribozyme. Unfortunately, structured templates still transcribed poorly. In a monumental demonstration, the improved ribozyme could actually transcribe an active version of the self-cleaving hammerhead ribozyme.[18]

While the demonstration buoys the RNA world model, in which there was one polymerase and many templates would eventually have to evolve, the Xeroxase remains elusive. This is interesting, given the amount of effort that has been expended. It may just be that the problem is suitably hard (a few laboratory years versus hundreds of millions of years of evolution). A better explanation is that it is unlikely that a single, nascent ribozyme could both have been complex enough to achieve efficient polymerization (Figure 2A) and in turn have readily acted on such a complex, structured template. There must have been a different, better path to origins, perhaps via the evolution of a hypercyclic network[17] rather than a single ribozyme (Figure 2B,C).

Cross-Catalytic Ribozyme Replicators

An initial cross-catalytic replicator was derived from the ribozyme ligase named R3C, which has a surprisingly simple, three-way junction structure and plasticity with respect to ligation substrates. With just a little engineering, the R3C could be modified to ligate a number of different substrates in *trans*.[19] This in turn led to the division the ribozyme itself, so that there were two half-ribozymes (A and B) that could serve as substrates for the whole ribozyme.[20] R3C had to be redesigned to be palindromic across the ligation junction so that the half-ribozyme substrates would anneal to the whole-ribozyme template and be ligated to form another whole-ribozyme. The reaction was short-lived and highly dependent on the order of substrate addition. As had been

the case for parabolic replicators going back to von Kiedrowski, the two half-ribozymes could base pair to form long-lived, inactive complexes.

To alleviate the problem of self-binding, two R3C-derived species were created that could act as templates for one another without being self-complementary.[21] This system format proved to be more robust compared with the initial self-replicator. The system could undergo a full round of replication, whereby T catalyzed the ligation of A′:B′ creating T′, which then catalyzed the ligation of A:B to T. However, the redesign led to a reduction in the catalytic rate and the production of unwanted side species (AB′), which stymied continual self-replication. These difficulties were partially overcome by thermocycling,[22] but true self-replication that led to the production of more copies than parents was still not achieved.

To increase the rate of catalysis for both R3C partners, *in vitro* selection was utilized.[23] A mutation was found at the ligation junction that changed the A:U Watson−Crick pair to a G:U wobble pair and in turn increased the efficiency of ligation by 10−40-fold. To determine whether the system could undergo continuous evolution, a small amount of starting ribozyme was added to a reaction containing substrates. A small portion of this reaction was serially transferred into a fresh reaction only containing substrates; for the ribozyme to persist in the face of continuing dilution, cross-replication must occur. As expected for an autocatalytic system, the ribozyme persisted after 30 h of serial transfer, and each parental ribozyme was amplified 100-million fold. The robustness of the system was also put to the test by competing different versions of the ribozyme in the same reaction mixture, and certain variants were found to predominate, proving the system could support evolution.

Cross-catalytic replication provides a potential means around replicating through complex secondary structures; it also suggests how multiple small amplicons can be combined into a larger whole. For example, the group I self-splicing ribozyme had been shown to be capable of recombining short RNA oligonucleotides in a template-directed manner. These reactions are akin to polymerization, and attempts were made to replicate portions of the group I ribozyme, but these were stymied by the low efficiency of the multiple recombination reactions necessary.[24] The Lehman laboratory then took a new approach by eliminating the need for template-directed recombination.[25] In their system, recombination is directed by the internal guide sequence of the *Azoarcus* ribozyme, rather than an external template. The utility of this method was demonstrated by construction of an active hammerhead ribozyme and the Bartel class I ligase from otherwise inactive precursors. With the knowledge that group I ribozymes retain function as split complexes, the *Azoarcus* ribozyme was split into four fragments that contained guide sites at the peripheral loops (to limit loss of function associated with mutations and secondary structure restrictions). Mixing these four pieces indeed resulted in the construction of full-length ribozymes. The pieces first self-assemble into a noncovalent, active complex, which joins other pieces into a covalently contiguous ribozyme. A given fully assembled *Azoarcus* ribozyme then acts on noncovalently joined complexes, leading to the autocatalytic assembly of up to ∼20% of the initial substrate into active ribozymes.

While each of the aforementioned ribozyme replicators seems an unlikely candidate for origins, they nonetheless may suggest a plausible mechanistic pathway to origins. For example, it may be possible to construct a system in which the tC19Z polymerase and *Azoarcus* ribozymes work in a two part hypercycle (Figure 2B,C). Hypercycles are unique in that individual reproductive cycles are connected by functional linkages. Within the context of the proposed hypercycle, short RNA fragments are replicated by the polymerase and then stitched together by the *Azoarcus* ribozyme. This obviates the requirement that a complex replicase spontaneously emerge from simple precursors as well as the need for the polymerase to act on structured templates. Thus, while neither species can efficiently self-replicate, each can survive in the context of the hypercycle. Envisioning this mutualistic network begs the problem of how to confine benefits to the two members of the network. Informational segregation by molecular recognition (i.e., base-pairing) could allow members to discriminate, but ultimately covalent bond formation or compartmentalization may be needed to avoid a "molecular tragedy of the commons".[2]

Replication of Nucleic Acids by Protein Enzymes

It is likely that once ribozymes had lifted metabolism to prominence that translation and protein enzymes took over, leading to the organisms we see today. The intermediate state in which protein polymerases began to impact nucleic acid replication was first examined by the legendary Sol Spiegelman.[26] Since bacteria infected with RNA viruses do not contain the cDNA, Spiegelman reasoned there would likely be a catalyst capable of replicating the RNA genome itself (proceeding through a complementary RNA

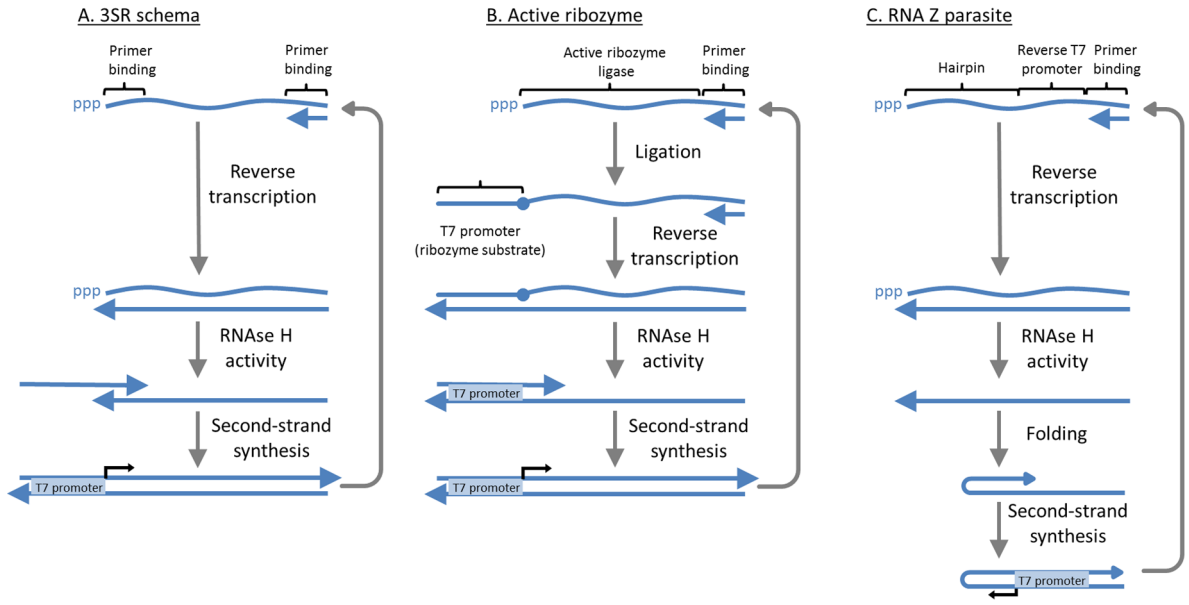


FIGURE 3. Continuous sequence amplification and the emergence of an RNA parasite. (A) In 3SR, an RNA molecule is reverse transcribed by MMLV reverse transcriptase. The primers append the T7 promoter, allowing for transcription. (B,C) A partially randomized RNA sequence is transcribed by T7 RNA polymerase. (B) An active ribozyme ligase is able to append a substrate molecule to its 5′ end. Upon reverse transcription, the resulting DNA template is suitable for further transcription. An inactive ribozyme is unable to append the substrate molecule and cannot complete the replication cycle. (C) Despite lacking ribozyme activity, the RNA Z parasite is able to replicate. The key features of RNA Z are a hairpin region and a sequence that is the reverse complement of the T7 promoter. Upon reverse transcription, the resulting DNA molecule folds back on itself to prime second-strand synthesis, which completes the T7 promoter, leading to additional RNA Z transcription. This cycle of reverse transcription and transcription can occur continuously.

intermediate). Biochemical fractionation of bacteria infected with MS2 bacteriophage led to the purification of an RNA-dependent RNA polymerase or "replicase". The replicase was shown to synthesize RNA using the MS2 RNA genome as its template, preferring the genome by 100-fold over other RNA or DNA. Similar work was done with the related Qβ replicase, and it was further shown that the replicated RNA was still capable of forming phage particles and that mutant templates led to mutant progeny. Starting with template RNA concentrations lower than that needed to saturate the enzyme, exponential increase in the amount of nucleic acid synthesized with respect to time (another hallmark of biological systems) was observed.

Such exponential replication was a necessary prerequisite for evolution, in which allelic variants of a parental sequence can eventually displace the parent from a population. In this case, serial transfers of the replicating RNA into fresh mixtures of nucleoside triphosphates and Qβ replicase led to increasing fitness *in vitro* but quickly decreasing fitness *in vivo* (since there was no longer any constraint to infect or replicate inside of cells). Indeed, after 74 transfers, the rate of replication was ~15 times that of the starting RNA genome. Further evolution of the shortened RNA "mini-monster" population led to a 218 nt variant (a loss of 94% of the genome[26]). By carrying out *in vitro* replication of the RNA template under different conditions, variants were evolved that could replicate in low substrate conditions or in the presence of inhibitors.

Such RNA parasites were not confined to Qβ replicase. Indeed, any system of continuous molecular replication is potentially subject to invasion by parasites. For example, a parasite, termed "RNA Z" emerged following an attempt by the Joyce laboratory to select for ribozyme function by continuous evolution.[27] The original experimental design is based on the self-sustained sequence replication (3SR) cycle[28] (Figure 3A). Unlike 3SR, however, the primers used do not append the promoter sequence upon reverse transcription. Joyce and co-workers hoped to select an RNA molecule that could only complete the replication cycle by ligating a promoter sequence to its 5′ end (Figure 3B). However, the RNA pool had other ideas, and RNA Z1 and RNA Z2 were born. In these parasites, the randomized region contained the reverse complement of the T7 RNA polymerase promoter followed by a primer binding site (Figure 3C). The parasites amplified by being reverse transcribed into a cDNA copy that contained a hairpin element at its 3′ end. The short hairpin was further

extended to create a double-stranded T7 RNA polymerase promoter. Transcription from this promoter regenerated the RNA Z species, and autoamplification continued.

One way to avoid the accumulation of molecular parasites during continuous amplification is to increase the number of specificity hurdles that must be leaped for replication to occur. For example, cooperative amplification of templates by cross-hybridization (CATCH) is in the same vein as 3SR, except it involves two molecular species that act cooperatively for their mutual amplification.[29] CATCH is more resistant to parasitism than 3SR, because there are more constraints on the species to be replicated. However, nonspecific products form, even under optimized conditions. Under conditions of low initial template or low enzyme, the reaction devolves into parasitism at the expense of the desired products. Similarly, loop-mediated isothermal amplification of DNA (LAMP) is a continuous process that utilizes a combination of hairpin formation, primer extension, and strand displacement to generate long concatemers of a target sequence.[30] LAMP utilizes four primers that recognize six specific sequences, thus constraining templates much more than 3SR and thereby avoiding parasites.

Discontinuous amplification is less prone to parasitic infestation. This is illustrated by the relatively small fraction of parasitic amplicons that arise in any given polymerase chain reaction (PCR[31]). Although iterative rounds of primer binding and extension allow DNA to be amplified by many orders of magnitude, each thermal cycle contains a discrete "extension" phase. By separating the periods of amplification and thereby avoiding continuous amplification, the advantages of quickly replicating parasitic species are greatly reduced.[32] In consequence, the robustness of PCR has led to its ready adaptation in schemes for molecular evolution, such as the selection of nucleic acid binding species, aptamers, and ribozyme catalysts.

Cellular Replicators

While amplification methods can be chosen so as to avoid parasitism, molecular parasites are still common in abiotic replication, where cell walls cannot constrain their spread or invasion. Nucleic acid replicators might have begun to escape parasites by cellularization,[2] but when and how this occurred is a mystery. Fortunately, cells may have already been available for the nascent replicators. Further stretching the definition of a living system, the Luisi laboratory generated lipid replicators that had semidefined compositions, rather than defined sequences or structures. They developed a system based on the self-replication of micelles.[33] Ethyl-caprylate slowly hydrolyzed in alkaline solution, yielding ethanol and sodium-caprylate, which is amphipathic and formed micelles. Micelles catalyzed hydrolysis, thereby slowly increasing the rate of micelle formation. Once a critical concentration of micelles was reached, micelle concentration increased exponentially. A similar scheme was used to generate a self-replicating bilayered vesicle composed of either oleic or caprylic acid in a buffer near the pK_a of the fatty acid.[34] The equimolar mixture of protonated and anionic fatty acids stabilized the biologically relevant, bilayered arrangement.

In an effort to show the feasibility of bringing together nucleic acid replicators and protocells, the Qβ genome was shown to replicate inside of autoreplicating vesicles.[35] Additionally, the use of water-in-oil emulsions (described below) confined parasites to a given compartment, while allowing exponential amplification of the full-length genome to occur in nonparasitized compartments[36] (Figure 4A). This finding has important implications for early life, because it is an experimental instantiation of the theoretical prediction that molecular parasitism can be effectively mitigated by compartmentalization.[2]

The ability to couple cellularization and template replication is the hallmark of modern living systems, and a number of synthetic variations on modern cells have been developed as a means of facilitating the evolution of ribozymes and proteins. The use of *in vitro* compartmentalization (IVC) began with the work of Tawfik and Griffiths. They were able to mix bacterial lysate, oil, and surfactants to generate water-in-oil compartments roughly the size of bacteria.[37] If an appropriate number of DNA templates is added before emulsifying the reaction, then each compartment is expected to contain only one template. This allows the genotype of the template and the phenotype for which it encodes to be linked to each other, while being separated from other templates. As an example, the gene encoding *Hae*III methyltransferase was emulsified with bacterial lysate such that each gene was transcribed and translated within its own compartment (Figure 4B). Active methyltransferases then methylated the DNA in their respective compartments. The DNA that encoded for an active methyltransferase would be, itself, methylated. After breaking the emulsion, methylated templates were protected from *Hae*III restriction endonuclease digestion. Undigested templates were then preferentially recovered. Starting with a mixture of active *Hae*III encoding template and an excess of irrelevant templates,

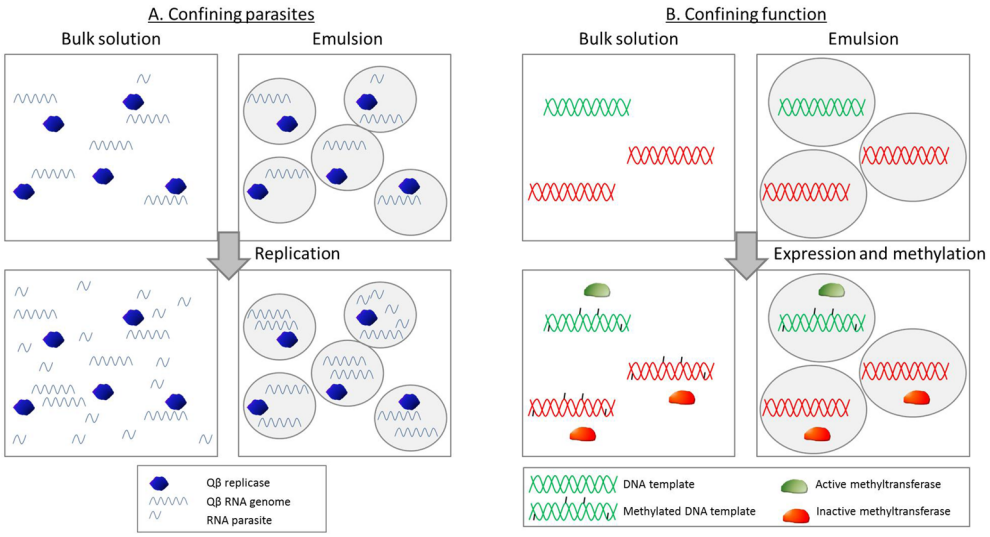


FIGURE 4. Compartmentalization of parasites and function. (A) In a continuous aqueous phase (bulk), a short parasite is preferentially replicated by the Qβ replicase, at the expense of the full-length RNA genome of Qβ. The use of water-in-oil emulsion confines existing or emerging parasites to a given compartment, allowing productive genome replication in the other compartments. (B) In bulk cell-free lysate, a mixture of active and inactive methyltransferase DNA templates are transcribed and translated. The active methyltransferase proteins are able to methylate all DNA templates in solution. In the ideal emulsion, only genes encoding active methyltransferase are, themselves, methylated.

they demonstrated a 1000-fold enrichment of the active templates. Thus, IVC confines the benefits of enzymatic function to the gene that codes for said function. This simple principle has now been exploited in the selection of genes encoding proteins and RNAs with numerous functions[38] and may also have implications for the early evolution of biopolymers.

Finally, replication cycles of increasing complexity are being demonstrated in compartments, suggesting that a cellularized nascent genome could have begun to elaborate its functionality. In work performed by the Holliger laboratory, the gene for Taq DNA polymerase was expressed in *Escherichia coli*, and individual bacteria were compartmentalized in a water-in-oil emulsion.[39] The compartment contained all the necessary reagents for the PCR amplification of the gene itself, except the Taq DNA polymerase that it encodes. Upon thermal cycling, the bacteria burst apart and released their contents into the broader compartment. Those bacteria harboring functional Taq polymerases imbued the compartment with the ability to amplify the gene. The Taq DNA polymerase gene is replicated in the same sense as any PCR amplicon; however, unlike other PCR amplicons, it is subject to the Darwinian selection. Similarly, a DNA construct composed of the T7 promoter controlling expression of the T7 RNA polymerase gene, a so-called "autogene", has been adapted to IVC and shown to autocatalytically amplify its RNA and protein components.[40] Like the Holliger-Taq system, the RNA is passively replicated by an enzyme for which it codes, thus allowing for *in vitro* Darwinian evolution.

Conclusion

Chemistry can be thought of as the hard-wired rule set that at some level governs more contingent biological evolution. There is an inevitability to the types of chemical templating that can lead to biology, with linear polymers that can mutate in a quantized way being the most likely to be successful, with mutualistic and conformational replicators lacking the ability to effectively compete. Interestingly, this may lead to the conclusion that all living systems that rely on Darwinian evolution must also rely on nucleic acid-like polymers, since the universe of compounds that can mutate in a quantized way is seemingly quite small. Once nucleic acid-like polymers can replicate, the problem of efficient self-replication comes to the fore, with both product inactivation via hybridization and parasitism being large barriers to success. These problems may be overcome by the parallel development of replication (and associated metabolism) via gene-like pieces, leading inevitably to metabolism, and by cellularization, leading inevitably to modern biology.

This work was funded by the Welch Foundation (F-1654) and the Defense Threat Reduction Agency (HDTRA-08-1-0052). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Defense Threat Reduction Agency or Welch Foundation.

BIOGRAPHICAL INFORMATION

Adam Meyer received a B.A. in Molecular and Cell Biology from the University of California, Berkeley, and is currently pursuing a Ph.D. under the supervision of Professor Ellington. He endeavors to engineer systems and evolve protein function.

Jared Ellefson received his B.S. in Ecology and Evolutionary biology from the University of Arizona. His interests are to create tools to manipulate fundamental biological processes.

Andrew Ellington received a B.S. in Biochemistry from Michigan State University and a Ph.D. in Biochemistry from Harvard University. He is currently a Professor at the University of Texas, Austin. His laboratory is focused on molecular programming and directed evolution.

FOOTNOTES

*To whom correspondence should be addressed. E-mail address: andy.ellington@mail.utexas.edu.

The authors declare no competing financial interest.

REFERENCES

1 Szostak, J. W. Attempts to define life do not help to understand the origin of life. *J. Biomol. Struct. Dyn.* **2012**, *29*, 599–600.

2 Tabor, J. J.; Levy, M.; Simpson, Z. B.; Ellington, A. D. Parasitism and protocells: The tragedy of the molecular commons. In *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; pp 367−384

3 Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738.

4 von Kiedrowski, G. A Self-Replicating Hexadeoxynucleotide. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 932–935.

5 Zielinski, W. S.; Orgel, L. E. Autocatalytic Synthesis of a Tetranucleotide Analogue. *Nature* **1987**, *327*, 346–347.

6 Goodwin, J. T.; Lynn, D. G. Template-Directed Synthesis: Use of a Reversible Reaction. *J. Am. Chem. Soc.* **1992**, *114*, 9197–9198.

7 Kiedrowski, G. V. Minimal Replicator Theory I: Parabolic Versus Exponential Growth Conclusions. *Bioorg. Chem.* **1993**, 113–146.

8 Sievers, D.; von Kiedrowski, G. Self-Replication of Complementary Nucleotide-Based Oligomers. *Nature* **1994**, *369*, 221–224.

9 Benner, S. A.; Hutter, D. Phosphates, DNA, and the Search for Nonterrean Life: A Second Generation Model for Genetic Molecules. *Bioorg. Chem.* **2002**, *30*, 62–80.

10 Lee, D. H.; Severin, K.; Yokobayashi, Y.; Ghadiri, M. R. Emergence of Symbiosis in Peptide Self-Replication through a Hypercyclic Network. *Nature* **1997**, *390*, 591–594.

11 Dawkins, R. *The Selfish Gene*; Oxford University Press: New York, 1976; Vol. 8, p 368.

12 Bartel, D.; Szostak, J. Isolation of New Ribozymes from a Large Pool of Random Sequences. *Science* **1993**, *261*, 1411–1418.

13 Ellington, A. D.; Szostak, J. W. In Vitro Selection of RNA Molecules That Bind Specific Ligands. *Nature* **1990**, *346*, 818–822.

14 Ekland, E.; Szostak, J.; Bartel, D. Structurally Complex and Highly Active Rna Ligases Derived from Random RNA Sequences. *Science* **1995**, *269*, 364–370.

15 Ekland, E. H.; Bartel, D. P. RNA-Catalysed RNA Polymerization Using Nucleoside Triphosphates. *Nature* **1996**, *382*, 373–376.

16 Johnston, W. K.; Unrau, P. J.; Lawrence, M. S.; Glasner, M. E.; Bartel, D. P. RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science* **2001**, *292*, 1319–1325.

17 Eigen, M.; Schuster, P. The Hypercycle. *Naturwissenschaften* **1977**, *64*, 541–565.

18 Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science* **2011**, *332*, 209–212.

19 Rogers, J.; Joyce, G. F. The Effect of Cytidine on the Structure and Function of an RNA Ligase Ribozyme. *RNA* **2001**, *7*, 395–404.

20 Paul, N.; Joyce, G. F. A Self-Replicating Ligase Ribozyme. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12733–12740.

21 Kim, D. Cross-Catalytic Replication of an RNA ligase Ribozyme. *Chem. Biol.* **2004**, *11*, 1505–1512.

22 Kim, K.-S.; Oh, S.; Yea, S. S.; Yoon, M.-Y.; Kim, D.-E. Amplification of an RNA Ligase Ribozyme under Alternating Temperature Conditions. *FEBS Lett.* **2008**, *582*, 2745–2752.

23 Lincoln, T. a; Joyce, G. F. Self-Sustained Replication of an RNA Enzyme. *Science* **2009**, *323*, 1229–1232.

24 Doudna, J. A.; Couture, S.; Szostak, J. W. A Multisubunit Ribozyme That Is a Catalyst of and Template for Complementary Strand RNA Synthesis. *Science* **1991**, *251*, 1605–1608.

25 Hayden, E. J.; Lehman, N. Self-Assembly of a Group I Intron from Inactive Oligonucleotide Fragments. *Chem. Biol.* **2006**, *13*, 909–918.

26 Kacian, D. L.; Mills, D. R.; Kramer, F. R.; Spiegelman, S. A Replicating RNA Molecule Suitable for a Detailed Analysis of Extracellular Evolution and Replication. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 3038–3042And references therein.

27 Breaker, R. R.; Joyce, G. F. Emergence of a Replicating Species from an in Vitro RNA Evolution Reaction. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 6093–6097.

28 Guatelli, J. C.; Whitfield, K. M.; Kwoh, D. Y.; Barringer, K. J.; Richman, D. D.; Gingeras, T. R. Isothermal, in Vitro Amplification of Nucleic Acids by a Multienzyme Reaction Modeled after Retroviral Replication. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7797.

29 Ehricht, R.; Ellinger, T.; McCaskill, J. S. Cooperative Amplification of Templates by Cross-Hybridization (CATCH). *Eur. J. Biochem.* **1997**, *243*, 358–364.

30 Notomi, T.; Okayama, H.; Masubuchi, H.; Yonekawa, T.; Watanabe, K.; Amino, N.; Hase, T. Loop-Mediated Isothermal Amplification of DNA. *Nucleic Acids Res.* **2000**, *28*, E63.

31 Saiki, R.; Scharf, S.; Faloona, F.; Mullis, K.; Horn, G. Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science* **1985**, *230*, 1350–1354.

32 Bull, J. J.; Pease, C. M. Why Is the Polymerase Chain Reaction Resistant to In Vitro Evolution? *J. Mol. Evol.* **1995**, 1160–1164.

33 Bachmann, P. A.; Luisi, P. L.; Lang, J. Autocatalytic Self-Replicating Micelles As Models for Prebiotic Structures. *Nature* **1992**, *357*, 57–59.

34 Walde, P.; Wick, R.; Fresta, M.; Mangone, A.; Luisi, P. L. Autopoietic Self-Reproduction of Fatty Acid Vesicles. *J. Am. Chem. Soc.* **1994**, *116*, 11649–11654.

35 Oberholzer, T.; Wick, R.; Luisi, P. L.; Biebricher, C. K. Enzymatic RNA Replication in Self-Reproducing Vesicles: An Approach to a Minimal Cell. *Biochem. Biophys. Res. Commun.* **1995**, *207*, 250–257.

36 Urabe, H.; Ichihashi, N.; Matsuura, T.; Hosoda, K.; Kazuta, Y.; Kita, H.; Yomo, T. Compartmentalization in a Water-in-Oil Emulsion Repressed the Spontaneous Amplification of RNA by Qβ Replicase. *Biochemistry* **2010**, *49*, 1809–1813.

37 Tawfik, D. S.; Griffiths, A. D. Man-Made Cell-Like Compartments for Molecular Evolution. *Nat. Biotechnol.* **1998**, *16*, 652–656.

38 Griffiths, A. D.; Tawfik, D. S. Miniaturising the Laboratory in Emulsion Droplets. *Trends Biotechnol.* **2006**, *24*, 395–402and references therein.

39 Ghadessy, F. J.; Ong, J. L.; Holliger, P. Directed Evolution of Polymerase Function by Compartmentalized Self-Replication. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4552–4557.

40 Davidson, E. A.; Meyer, A. J.; Ellefson, J. W.; Levy, M.; Ellington, A. D. An in Vitro Autogene. *ACS Synth. Biol.* **2012**, *1*, 190–196.