Die Preisträger des Berichtsjahres 2009

(Die Preisträgervorträge wurden in einer Plenarsitzung
am 20. November 2009 vorgetragen)

Der **Dannie-Heineman-Preis 2009** wurde Herrn Gerald F. Joyce, La Jolla, USA, in Anerkennung seiner Arbeit „Self-sustained Replication of an RNA Enzyme“ verliehen.

Self-sustained Evolution of RNA

Gerald F. Joyce

Introduction

Gerald F. Joyce, Departments of Chemistry and Molecular Biology and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
Dannie-Heineman-Preisträger 2009

Living systems are characterized by several attributes, chief among which are self-replication, metabolic function, and the capacity to evolve. There is no agreed-upon definition of life[1], but most would regard a chemical system that undergoes Darwinian evolution without the aid of external evolved molecules as being at the threshold of life. Darwinian evolution provides the means to adapt to a changing environment, and generates a historical record of those adaptations as a lineage of genetic molecules. Replication of the genetic molecules must be supported by the conversion of high-energy starting materials to lower-energy products, and innovations such as compartmentalization and an increasingly sophisticated metabolism are likely to be necessary for the long-term survival of the system.

All life that is known to exist on Earth today, and all life for which there is evidence in the geological record, is based on DNA genomes and protein

enzymes. However, there are strong reasons to conclude that DNA- and protein-based life was preceded by a simpler form of life based primarily on RNA[2–7]. In that "RNA world"[8], RNA enzymes would have been responsible for catalyzing the replication of RNA genomes, thus enabling the evolution of RNA-based function. Recent work in our laboratory has led to the development of RNA enzymes that catalyze their own replication and have a limited capacity to evolve[9]. These molecules do not yet have sufficient genetic complexity to invent novel function, but they can adapt existing function to particular environmental conditions.

Our current research activities are focused on expanding the complexity of the RNA-based evolving system so that it has the capacity for inventive Darwinian evolution. Efforts are being directed toward maximizing the genetic information capacity of the synthetic evolving system so that it can develop complex functions, as would be needed to persist in a changing environment. This work aims to broaden our view of the chemical nature of life and to provide an experimental system to study life in a highly simplified form.

From Ligase to Replicase

There are now many examples of *in vitro* evolved RNA enzymes that catalyze the RNA-templated joining of RNA substrates[10–17]. Some of these enzymes catalyze a single joining reaction (ligation), while others catalyze multiple successive joinings (polymerization). These reactions have special relevance to the origins of life because the underlying chemistry is similar to what would be required by an RNA replicase. For many years, experimental efforts have focused on attempting to convert an RNA ligase RNA enzyme to a polymerase, and ultimately to a replicase. In one especially notable example, the class I RNA ligase[10,11] was converted to an RNA-dependent RNA polymerase that can extend an oligonucleotide primer by adding up to 20 successive NTPs[18]. This reaction is less general than first thought[19], but may offer the opportunity for further improvement, and suggests approaches that could be applied to other RNA ligases.

RNA-catalyzed RNA ligation typically involves the RNA-templated joining of an oligonucleotide 3′ hydroxyl and an oligonucleotide 5′-triphosphate, forming a 3′,5′-phosphodiester linkage and releasing inorganic pyrophosphate. Efforts to convert a ligase to a polymerase usually involve replacing the oligonucleotide 5′-triphosphate substrate by successive NTPs. An alternative approach is to treat oligonucleotide substrates as "mono-

mers", and attempt to progress from oligonucleotide ligation to oligonucleotide polymerization, and ultimately to RNA replication. There are two extreme versions of this alternative approach: one is to employ short oligonucleotides (2–6 residues) and require multiple successive oligonucleotide additions; the other is to employ long oligonucleotides (20–40 residues) and require only one or a few joining reactions to assemble a complete copy. The latter has recently been achieved in our laboratory, providing the first non-biological system that undergoes self-sustained exponential amplification[9]. This system also has been extended to allow for heritable mutation and survival of the fittest among a heterogeneous population of self-replicating RNAs.

The self-replicating RNA enzymes were derived from the "R3C" RNA ligase, developed previously in our laboratory[15]. This ligase has a simple three-way junction architecture, consisting of three stem-loops that are joined at a central location that contains the catalytic domain of the enzyme (Figure 1a). Nucleotides within the catalytic domain are highly conserved in sequence, but those within the pendant stem-loops are generic, as long as they form a stable duplex structure. Two of the stem-loop regions within the R3C ligase are involved in binding the RNA substrates. Because these regions are generic in sequence, they can be designed to accommodate two substrates (A and B) whose sequences correspond to the 5′ and 3′ portions, respectively, of the enzyme (E). When the substrates are ligated, they form another copy of the enzyme, allowing self-replication to occur (Figure 1b)[20]. This reaction does indeed proceed autocatalytically and is not limited by product dissociation, but replication is slow and does not reach a high maximum extent.

The next step was to devise two ligase enzymes (E and E′) that catalyze each other's synthesis from a total of four component substrates ($A' + B' \rightarrow E'$, catalyzed by E; and $A + B \rightarrow E$, catalyzed by E′)[21]. Compared to self-replication, cross-replication places fewer design constraints on the sequences of the replicating molecules. However, initial versions of the cross-replicating system also were inefficient and could barely generate as many new copies of the enzyme as were present at the start of the reaction. It thus became necessary to improve the rate and maximum extent of the cross-replicating RNA enzymes, which was accomplished using *in vitro* evolution[9]. The resulting optimized enzymes are able to achieve 100-fold amplification in 5 hours at a constant temperature of 42°C. Their replication can be continued indefinitely through a serial transfer procedure, in which a small aliquot is taken from a completed reaction mixture and transferred to a new reaction vessel that contains a fresh supply of substrates.

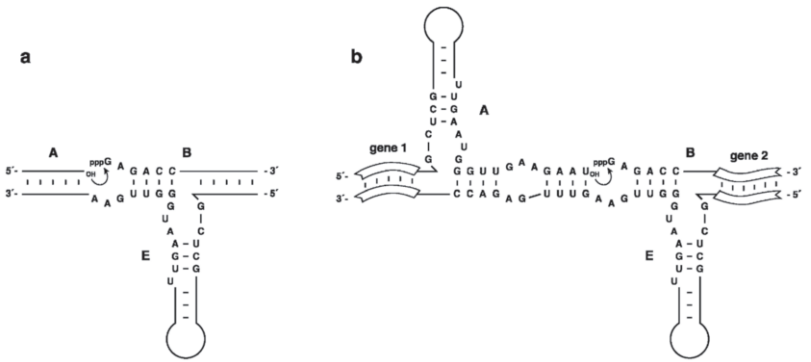

Figure 1: Standard and replicating forms of the R3C ligase RNA enzyme. **a,** The enzyme (E) adopts a three-way junction structure upon binding two substrates (A and B), which become ligated (curved arrow) to form the product. Nucleotides that are essential for catalytic function are shown. **b,** The self-replicating or cross-replicating enzyme ligates two substrates to yield a new copy of the enzyme or its cross-catalytic partner, respectively. Open boxes indicate the two genetic regions, which can have any complementary sequence.

Darwinian Evolution in a Synthetic Genetic System

Darwinian evolution requires many variants in a population, all of which can replicate, mutate, and compete for survival. This was achieved by constructing heterogeneous populations of cross-replicating RNA enzymes that undergo mutation through recombination, and selection based on their differential rates of replication[9]. Each enzyme contains two "genes", represented by the two regions of complementary pairing between the enzyme and its substrates (Figure 1b). Each gene can have many possible alleles, and each allele encodes a corresponding trait, which is the functional domain that is covalently linked to the allele. Recombination of the two genes occurs due to occasional incorporation of a mismatched substrate, resulting in recombinant enzymes that also can cross-replicate. Over time, recombinants can give rise to other recombinants, as well as revert back to non-recombinants.

As a test case, a population of 12 different pairs of cross-replicating RNAs were constructed, each pair containing different genetic sequences within the two allelic regions that encoded different functional sequences within the corresponding catalytic domains[9]. Together the 12 pairs of cross-replicators had the potential to give rise to 132 pairs of recombinants. A +serial transfer experiment was carried out, starting with 0.1 μM each of

the 12 starting replicators and 5 μM each of the various substrates. The population was subjected to 20 successive rounds of 20-fold amplification and 20-fold dilution ($\sim 10^{26}$-fold overall amplification) in a period of 100 hours. During this time novel recombinants arose and grew to dominate the population. Three recombinants in particular accounted for one-third of the evolved population. The basis for their selective advantage was shown to be their faster exponential growth rate in the complex mixture of substrates, and their propensity to support each other's production through preferred mutational pathways[9].

Replication Contingent on Other Functions

Although replication efficiency is the ultimate measure of fitness, other traits can confer selective advantage through their indirect effect on replication. So too in a synthetic genetic system, reproductive fitness can be made contingent on the execution of other functions. The cross-replicating RNA enzymes contain three generic stem-loops, two that are committed to substrate binding and a third that can contain a functional domain (Figure 1b). The functional domain could be an RNA aptamer that binds a specific ligand or a catalyst that brings about a particular chemical transformation.

It is straightforward to install an aptamer domain within the central stem-loop of the replicating enzymes so that they undergo exponential amplification in the presence, but not the absence, of the corresponding ligand[22]. Similar "aptazymes" have been developed in the laboratory for non-replicating RNA enzymes[23–30] and have been discovered within naturally occurring "riboswitches"[31]. Aptamers that specifically recognize either theophylline[32] or FMN[33] were installed within either or both members of a pair of cross-replicating enzymes, causing exponential amplification to be dependent on the presence of the corresponding ligand (Figure 2)[22]. In the absence of the ligand the aptamer is unstructured and cannot support the active structure of the enzyme, while in the presence of the ligand the aptamer adopts a well-defined structure that stabilizes and therefore activates the adjacent catalytic domain. Furthermore, the exponential growth rate of the cross-replicating aptazymes depends on the concentration of the ligand relative to the K_d of the aptamer domain. This provides a way for the replicators to sense the concentration of the ligand in their environment and to reflect this behavior in their reproductive fitness.

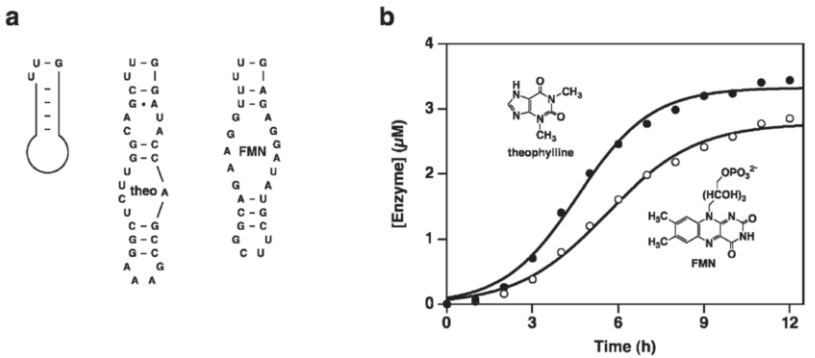

Figure 2: Ligand-dependent exponential amplification of RNA. **a.** The central stem-loop of the cross-replicating enzyme (*left*) was replaced by an aptamer domain that binds either theophylline (*center*) or flavin mononucleotide (*right*). **b.** Enzymes that contain either the theophylline aptamer (•) or FMN aptamer (○) undergo amplification in the presence (but not the absence) of 5 mM theophylline or FMN, respectively.

Information Capacity of the Synthetic Evolving System

The synthetic genetic system based on cross-replicating RNA enzymes has many of the properties of a living system, but thus far lacks the capacity for inventive Darwinian evolution. Genetic information within the system is represented by two regions of base pairing between the E and E′ enzymes of 7–8 nucleotides each. The sequence diversity available to the system is meager, limited to the $n \times m$ combinations of the two genetic loci. Sequence diversity in biology is much greater due to the 4^n possible combinations for a nucleic acid genome of length n. In the work described above, n and m both were chosen to be 12, resulting in 144 possible cross-replicating pairs. In principle, n and m each could be much larger, both on the order of 10^4–10^5, giving 10^8–10^{10} possible combinations. However, not all of these potential genotypes would be discriminated with high fidelity, especially those that involve subtle differences in sequence. In addition, it would be difficult for any replicator to find its corresponding substrates among a mixture of tens of thousands of potential substrates.

Current studies aim to maximize the information capacity of the synthetic genetic system so that it can provide the basis for the discovery of novel function. These efforts involve constructing populations of varying complexity, for example, randomizing three, four, or five nucleotides within each of the two allelic regions to give 64×64, 256×256, or $1{,}024 \times 1{,}024$ possible combinations. The various populations of enzymes will be allowed

to undergo self-sustained amplification. Sources of infidelity will be identified, perhaps indicating genetic sequences that are difficult to discriminate and should be excluded from the population.

It should be possible to maintain evolving populations with a complexity of at least 10^4 distinct replicators, and possibly as many as 10^7. Preliminary studies employing a population of 4,096 cross-replicators (64×64 combinations, all with the wild-type catalytic domain) demonstrated exponential growth starting from the initial combinatorial library, with 10-fold overall amplification in 7 hours. A serial transfer experiment was carried out involving six successive rounds of 10-fold amplification, starting with 0.2 nM of each E and E′ enzyme, and employing 0.6 μM of each A, A′, B, and B′ substrate throughout the experiment (Figure 3). The K_m of the enzyme for the A or A′ substrate is ∼0.4 μM and for the B or B′ substrate is ∼0.05 μM. Thus, in the complex mixture of 4,096 cross-replicators and their respective substrates, each enzyme is operating under near-saturating conditions with regard to its matching A or A′ substrate and under saturating conditions with regard to its matching B or B′ substrate, even though the matching substrates comprise only a small fraction of the total pool.

Despite the much higher concentration of non-matching compared to matching substrates, the kinetic properties of the enzyme may allow fine discrimination within the complex mixture. This is because the dissociation

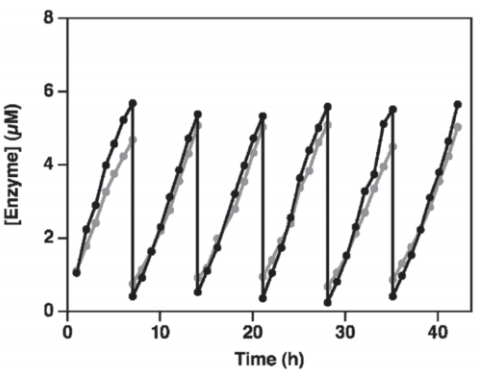

Figure 3: Self-sustained amplification of a population of 4,096 cross-replicating RNA enzymes. A serial transfer experiment was carried out, allowing ∼10-fold amplification before transferring 1/10th of the mixture to a new reaction vessel that contained a fresh supply of substrates. The aggregate yield of E (black) and E′ (gray) was measured at frequent intervals. Reaction conditions: 0.2 nM each starting E and E′, 0.6 μM each substrate, 25 mM $MgCl_2$, pH 8.5, 42°C.

rate of the enzyme-substrate complex is estimated to be $\sim$100 min^{-1} (based on a K_d of $\sim 10^{-7}$ M and duplex association rate of $\sim 10^9$ M^{-1} min^{-1}), while the catalytic rate is $\sim$1 min^{-1}. Thus, even for matched substrates, there is repeated sampling of the substrates before each ligation event. This could provide a kinetic proofreading mechanism that would discriminate against mismatched substrates, which are expected to have an even faster dissociation rate.

The maximum genetic complexity of the evolving population will determine the degree of functional sophistication that can be achieved within the system. Some of the simplest functions that might be attained involve ligand recognition, especially if the ligand is a compound that binds readily to RNA. There also are some nucleic-acid-catalyzed reactions, such as the divalent-metal-dependent cleavage of an RNA phosphodiester or the joining of activated oligonucleotides, that require only a low level of catalytic sophistication and likely could be brought about by a small catalytic motif. More complex functions will necessitate some means to provide much greater genetic complexity within the evolving system, for example, by requiring multiple ligation events to assemble a complete copy of the enzyme. Ultimately, the system should have the capacity to evolve functions as sophisticated as the replicase itself and to support an RNA-based metabolism. This would amount to the reinvention of the RNA world in the laboratory.

Acknowledgments. I am grateful to Tracey Lincoln for her dedicated efforts toward achieving the self-sustained replication of RNA, and to Michael Robertson for his studies involving increasingly complex populations of replicating RNAs. This work was supported by NASA grant NNX07AJ23G, NIH grant R01GM065130, and NSF grant MCB-0614614.

References

1. Cleland, C.E. & Chyba, C.F. (2002) Defining 'life'. *Orig. Life Evol. Biosph.* **32,** 387–393.
2. Woese, C. (1967) *The Genetic Code*. Harper & Row, New York, pp. 179–195.
3. Crick, F.H.C. (1968) The origin of the genetic code. *J. Mol. Biol.* **38,** 367–379.
4. Orgel, L.E. (1968) Evolution of the genetic apparatus. *J. Mol. Biol.* **38,** 381–393.
5. Joyce, G.F. (1989) RNA evolution and the origins of life. *Nature* **338,** 217–224.
6. Steitz, T.A. & Moore, P.B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.* **28,** 411–418.
7. Orgel, L.E. (2004) Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39,** 99–123.
8. Gilbert, W. (1986) The RNA world. *Nature* **319,** 618.

9. Lincoln, T.A. & Joyce, G.F. (2009) Self-sustained replication of an RNA enzyme. *Science* **323,** 1229–1232.
10. Bartel, D.P. & Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* **261,** 1411–1418.
11. Ekland, E.H., Szostak, J.W. & Bartel, D.P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269,** 364–370.
12. Hager, A.J. & Szostak, J.W. (1997) Isolation of novel ribozymes that ligate AMP-activated RNA substrates. *Chem. Biol.* **4,** 607–617.
13. Robertson, M.P. & Ellington, A.D. (1999) *In vitro* selection of an allosteric ribozyme that transduces analytes into amplicons. *Nature Biotechnol.* **17,** 62–66.
14. Jaeger, L., Wright, M.C. & Joyce, G.F. (1999) A complex ligase ribozyme evolved *in vitro* from a group I ribozyme domain. *Proc. Natl. Acad. Sci. USA* **96,** 14712–14717.
15. Rogers, J. & Joyce, G.F. (2001) The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA* **7,** 395–404.
16. Ikawa, Y., Tsuda, K., Matsumura, S. & Inoue, T. (2004) *De novo* synthesis and development of an RNA enzyme. *Proc. Natl. Acad. Sci. USA* **101,** 13750–13755.
17. Fujita, Y., Furuta, H. & Ikawa, Y. (2009) Tailoring RNA modular units on a common scaffold: a modular ribozyme with a catalytic unit for b-nicotinamide mononucleotide-activated RNA ligation. *RNA* **15,** 877–888.
18. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. & Bartel, D.P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292,** 1319–1325.
19. Zaher, H.S. & Unrau, P.J. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13,** 1017–1026.
20. Paul, N. & Joyce, G.F. (2002) A self-replicating ligase ribozyme. *Proc. Natl. Acad. Sci. USA* **99,** 12733–12740.
21. Kim, D.-E. & Joyce, G.F. (2004) Cross-catalytic replication of an RNA ligase ribozyme. *Chem. Biol.* **11,** 1505–1512.
22. Lam, B.J. & Joyce, G.F. (2009) Autocatalytic aptazymes enable ligand-dependent exponential amplification of RNA. *Nature Biotechnol.* **27,** 288–292.
23. Tang, J. & Breaker, R.R. (1997) Rational design of allosteric ribozymes. *Chem. Biol.* **4,** 453–459.
24. Soukup, G.A. & Breaker, R.R. (1999) Engineering precision RNA molecular switches. *Proc. Natl. Acad. Sci. USA* **96,** 3584–3589.
25. Koizumi, M., Soukup, G.A., Kerr, J.N.Q. & Breaker, R.R. (1999) Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nature Struct. Biol.* **6,** 1062–1071.
26. Robertson, M.P. & Ellington, A.D. (2000) Design and optimization of effector-activated ribozyme ligases. *Nucleic Acids Res.* **28,** 1751–1759.
27. Seetharaman, S., Zivarts, M., Sudarsan, N. & Breaker, R.R. (2001) Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nature Biotechnol.* **19,** 336–341.
28. Robertson, M.P. & Ellington, A.D. (2001) *In vitro* selection of nucleoprotein enzymes. *Nature Biotechnol.* **19,** 650–655.

29. Hartig, J.S., Najafi-Shoushtari, S.N., Grüne, I., Yan, A., Ellington, A.D. & Famulok, M. (2002) Protein-dependent ribozymes report molecular interactions in real time. *Nature Biotechnol.* **20,** 717–722.
30. Vaish, N.K., Dong, F., Andrews, L., Schweppe, R.E., Ahn, N.G., Blatt, L. & Seiwert, S.D. (2002) Monitoring post-translational modification of proteins with alllosteric ribozymes. *Nature Biotechnol.* **20,** 810–815.
31. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A. & Breaker, R.R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428,** 281–286.
32. Jenison, R.D., Gill, S.C., Pardi, A. & Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science* **263,** 1425–1429.
33. Burgstaller, P. & Famulok, M. (1994) Isolation of RNA aptamers for biological cofactors by *in vitro* selection. *Angew. Chemie* **33,** 1084–1087.