View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by NASA Technical Reports Server

NASA-CR-194124

GRANT
IN-51-CR
10 CIT
185598
7P

Final Technical Report
"Evolution of Catalytic RNA in the Laboratory"

NASA Innovative Research Program, NAGW-1671
April 1, 1989 – March 31, 1992

Gerald F. Joyce

The Scripps Research Institute
10666 N. Torrey Pines Road
La Jolla, California 92037

(NASA-CR-194124) EVOLUTION OF CATALYTIC RNA IN THE LABORATORY Final Technical Report, 1 Apr. 1989 - 31 Mar. 1992 (Scripps Institution of Oceanography) 7 p

N94-13210

Unclas

G3/51 0185598

Background

One of the principal objectives of NASA's research program is to understand the origin, evolution, and distribution of life on earth and throughout the universe.[1] Since the time of the Viking mission to Mars in 1976 this objective has been addressed by focusing on the problem of how life originated on earth. The rationale is that by better understanding the origins of life on earth, we will be better equipped to carry out the search for life elsewhere in the solar system. In particular we must learn to recognize those chemical events that signify that the transition from inanimate matter to living systems has taken place.

Living systems are characterized by several attributes, chief among which are self-replication, metabolic function, and the capacity to evolve. One of the major difficulties in understanding how life arose is to determine how genetic properties (the ability to store and replicate genetic information) become intertwined with catalytic properties (the ability to perform specific metabolic tasks) to produce a system that is capable of evolving. Put more simply: "Which came first, the chicken or the egg (metabolism or genetics)?" The recent discovery of RNA enzymes (ribozymes) has shed new light on this problem.[2-4] For the first time we have a single molecule that has both genetic and catalytic properties, suggesting that the "chicken" and the "egg" may have arose together.

The discovery of RNA enzymes by no means solves the problem of the origins of life. There are several reasons to believe that life did not begin with RNA and the identity of the first genetic molecule is not known.[5-8] However, there is abundant evidence to suggest that an RNA-based life form preceded the DNA/protein-based life form that is common to all known terrestrial biology. We know very little about this postulated RNA-based life form except what can be inferred by examining the role of RNA in contemporary organisms and by studying the behavior of RNA in the laboratory. In recent years our understanding of the chemistry, biochemistry, and molecular biology of RNA has advanced to the point that many questions concerning RNA-based life can be approached experimentally.

Research Objectives

We are interested in the biochemistry of existing RNA enzymes and in the development of RNA enzymes with novel catalytic function. The focal point of our research program has been the design and operation of a laboratory system for the controlled evolution of catalytic RNA. This system serves as working model of RNA-based life and can be used to explore the catalytic potential of RNA.

Evolution requires the integration of three chemical processes: amplification, mutation, and selection. Amplification results in additional copies of the genetic material. Mutation operates at the level of genotype to introduce variability, this variability in turn being expressed as a range of phenotypes. Selection operates at the level of phenotype to reduce variability by excluding those individuals that do not conform to the prevailing fitness criteria. These three processes must be linked so that only the selected individuals are amplified, subject to mutational error, to produce a progeny distribution of mutant individuals.

We devised techniques for the amplification, mutation, and selection of catalytic RNA, all of which can be performed rapidly *in vitro* within a single reaction vessel.[9] We integrated these techniques in such a way that they can be performed iteratively and routinely. This allowed us to conduct evolution experiments in response to artificially-imposed selection constraints. Our objective was to develop novel RNA enzymes by altering the selection constraints in a controlled manner. In this way we were able to expand the catalytic repertoire of RNA. Our long-range objective is to develop an RNA enzyme with RNA replicase activity. If such an enzyme had the ability to produce additional copies of itself, then RNA evolution would operate autonomously and the origin of life will have been realized in the laboratory.

Progress During Grant Period

In August 1988 I submitted a proposal to NASA's Innovative Research Program describing our plan to construct a laboratory system for the controlled evolution of RNA enzymes. We encountered surprisingly little difficulty in implementing this plan. During the first year of NASA support (4/89 - 3/90) we demonstrated an *in vitro* method for the selective amplification of catalytic RNA.[10] We also completed a comprehensive deletion analysis of a group I ribozyme, defining the minimum secondary structure requirements for its catalytic function.[11] During the second year (4/90 - 3/91) we improved our ability to select rare advantageous mutants from a large, heterogeneous population of RNAs and began to operate RNA evolution in a continuous manner. During the third year (4/91 - present) we took our first evolutionary footsteps, directing a population of 2×10^{13} RNA enzymes toward the expression of a novel catalytic behavior. We have characterized this evolutionary transition in detail, noting changes in genotype and phenotype over successive generations.

a) Selective amplification of an RNA enzyme

We are able to amplify virtually any RNA using a combination of two polymerase enzymes.[12,13] RNA is copied to complementary DNA (cDNA) using reverse transcriptase and the resulting cDNA is transcribed back to RNA using T7 RNA polymerase. Amplification occurs at the level of transcription due to the ability of T7 RNA polymerase to generate 200-1200 copies of RNA transcript per copy of cDNA template.[14] The amplification reaction can be carried out in a single test tube at a constant temperature of 37 °C, resulting in 10^3 - 10^6-fold amplification of the input RNA after one hour.[15,16]

We can carry out amplification in a selective manner by requiring that individual RNAs in the population catalyze a particular chemical reaction in order to become eligible for amplification. The selection scheme is based on the ability of group I ribozymes to catalyze a sequence-specific phosphoester transfer reaction involving an oligonucleotide (or oligonucleotide analogue) substrate. The product of the reaction is a molecule that contains the 3′ portion of the substrate attached to the 3′ end of the ribozyme. Selection occurs when an oligodeoxynucleotide primer is hybridized across the ligation junction and used to initiate cDNA synthesis. The primer does not bind to unreacted starting materials ($< 10^{-9}$ compared to reaction products) and thus leads to selective amplification of the catalytically active RNAs.

We first tested this selective amplification scheme using a set of structural variants of the *Tetrahymena* ribozyme. This enzyme catalyzes cleavage/ligation reactions involving RNA substrates, but was thought to be incapable of performing comparable reactions involving DNA substrates.[17] We found that the *Tetrahymena* ribozyme is able to cleave a target DNA substrate[10], although the reaction is almost undetectable unless one employs conditions of high temperature (50 °C) and/or high salt (50 mM $MgCl_2$). Selecting for DNA cleavage activity under high-temperature, high-salt conditions, we obtained a particular structural variant of the ribozyme (the ΔP9 mutant) that cleaves DNA more efficiently than does the wild-type.[10] This work demonstrated the feasibility of directed evolution techniques for the development of RNAs with novel catalytic function.

b) Structural requirements for catalytic activity of a self-splicing group I intron

The *Tetrahymena* ribozyme is a self-splicing group I intron derived from the large ribosomal RNA precursor of *Tetrahymena thermophila*. It consists of 413 nucleotides and assumes a well-defined secondary and tertiary structure that is responsible for its catalytic activity. A secondary structural model of the molecule has been developed based on phylogenetic comparison with other group I introns.[18] The model suggests that there is a catalytic center comprised of conserved sequence elements, supported by a number of stem-loop structures that are less highly conserved.

We carried out a comprehensive deletion analysis of the *Tetrahymena* ribozyme, showing that nearly all of the supporting stem-loop structures can be deleted in a piecewise fashion without loss of activity.[11] This extended previous work that had been conducted along these lines[19-23] and defined the minimum secondary structural requirements for catalytic activity of a self-splicing group I intron. While it was not been possible to combine *all* of the deletions to produce a naked reaction center, a variety of combined deletions, totaling as many as 201 nucleotides, were shown to retain activity.[11] Having defined those regions of the molecule that are not required for catalytic activity, we were then able to direct random mutations to the remaining areas that are most likely to influence catalytic function.

c) Enhanced selective amplification techniques

The selective power of an *in vitro* evolution system is determined by three factors:

1) sensitivity – the ability to select very rare individuals that have some desirable catalytic property, i.e., the ability to amplify a faint *signal;*

2) specificity – the ability to reject large numbers of individuals that lack the desired property, i.e., the ability to exclude background *noise;*

3) generation time – the amount of time required to complete a round of selective amplification.

We made considerable progress in improving the "signal-to-noise" ratio of the system while minimizing the generation time. We are now able to detect a signal of less than 10^{-8} pmol ($\approx 10^3$ molecules) while excluding a background of 20 pmol ($\approx 10^{13}$ molecules). Thus our signal-to-noise ratio is about 10^{-10}. The selective amplification procedure can be performed in 1 hour. Allowing for set-up time and accompanying analytic work, we can carry out 1-2 generations per day.

We also were concerned with maximizing the absolute number of molecules that can be perpetuated by the *in vitro* evolution system. The greater the population size, the greater the number of potentially desirable mutants that can be surveyed. By integrating the RNA amplification technique (described in section **a**, above) with the polymerase chain reaction (PCR)[24], we were able to amplify a signal of 10^3 molecules by a factor of 10^{10} - 10^{11}, providing roughly 100 pmols of the selected RNAs. The products of isothermal amplification were then used directly to initiate the PCR. Typically we would perform RNA amplification using a primer that binds selectively to reaction products, followed by the PCR using a nonselective primer that restores the 3′ terminus of the ribozyme. The products of the PCR were then transcribed, using T7 RNA polymerase, to produce RNA which were used to begin the next generation.

The PCR had proven useful in two other respects. First, it simplified the process of subcloning individuals from the evolving population. Normally only a small portion of the DNA present in the RNA amplification mixture is fully double-stranded, but with the PCR the amount of double-stranded DNA is greatly increased. Second, the PCR allowed us to introduce random mutations at a frequency of up to 1% per position per generation. This was done by performing the PCR under mutagenic conditions[25,26] to generate variant copies of the selected individuals. In this way we were able to integrate mutation with selective amplification to produce an evolving system that operates in a continuous manner.

d) Directed evolution of an RNA enzyme

As mentioned previously, the wild-type *Tetrahymena* ribozyme is able to cleave a target DNA substrate under high-temperature, high-salt conditions (e.g. 50 °C, 50 mM $MgCl_2$) (ref. 10,27), but has very little activity when tested under physiologic conditions (e.g. 37 °C, 10 mM $MgCl_2$). We used *in vitro* evolution techniques to develop a family of RNA enzymes that cleave DNA efficiently under physiologic conditions. Beginning with the wild-type *Tetrahymena* ribozyme, we introduced random mutations at a frequency of 5% per position over 140 nucleotide positions that encompass the catalytic center of the molecule. We prepared a population of 2×10^{13} ribozyme variants, including all possible 1-, 2-, 3-, and 4-error mutants. Selective amplification was

carried out over ten successive generations with a mutation rate of 10^{-3} per position per generation. This resulted in a population of ribozymes whose aggregate activity is 27-fold greater than that of the wild type when tested with a DNA substrate under physiologic conditions. Individuals isolated from the population (at generation 9) were found to have activities ranging up to 78-fold greater than that of the wild type.

Having, for the first time, evolved an enzyme with novel catalytic function, we were then in a position to provide a detailed description of evolution at the molecular level. We used shotgun cloning techniques to obtain individuals from each of the ten generations. We determined the complete nucleotide sequence of 25 subclones from both the 3rd and 6th generations and 50 subclones from the 9th generation. This provided an overall picture of how genotype changes over the course of evolutionary history. We prepared RNA from 14 of the subclones from the 9th generation and studied their catalytic properties on an individual basis. We then conducted a detailed kinetic analysis using three of the most successful individual RNAs.

References

1. DeVincenzi, D.L. (1984) "NASA's exobiology program" *Origins of Life* **14**, 793-799.
2. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. & Cech, T.R. (1982) "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence" *Cell* **31**, 147-157.
3. 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.
4. Cech, T.R. (1986) "A model for the RNA-catalyzed replication of RNA" *Proc. Natl. Acad. Sci. USA* **83**, 4360-4363.
5. Shapiro, R. (1984) "The improbability of prebiotic nucleic acid synthesis" *Origins of Life* **14**, 565-570.
6. Joyce, G.F., Schwartz, A.W., Miller, S.L. & Orgel, L.E. (1987) "The case for an ancestral genetic system involving simple analogues of the nucleotides" *Proc. Natl. Acad. Sci. USA* **84**, 4398-4402.
7. Joyce, G.F. (1989) "RNA evolution and the origins of life" *Nature* **338**, 217-224.
8. Orgel, L.E. (1989) "Was RNA the first genetic polymer?" In *Evolutionary Tinkering in Gene Expression*, Grunberg-Manago, M., Clark, B.F.C. & Zachau, H.G. (eds.), Plenum, London, pp. 215-224.
9. Joyce, G.F. (1989) "Amplification, mutation and selection of catalytic RNA" *Gene* **82**, 83-87.
10. Robertson, D.L. & Joyce, G.F. (1990) "Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA" *Nature* **344**, 467-468.
11. Beaudry, A.A. & Joyce, G.F. (1990) "Minimum secondary structure requirements for catalytic activity of a self-splicing group I intron" *Biochemistry* **29**, 6534-6539.
12. Kwoh, D.Y., Davis, G.R., Whitfield, K.M., Chappelle, H.L., DiMichele, L.J. & Gingeras, T.R. (1989) "Transcription-based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format" *Proc. Natl. Acad. Sci. USA* **86**, 1173-1177.
13. Joyce, G.F. (1989) "Building the RNA world: evolution of catalytic RNA in the laboratory" In *Molecular Biology of RNA: UCLA Symposia on Molecular and Cellular Biology*, Cech, T.R. (ed.), Alan R. Liss, New York, pp. 361-371.
14. Chamberlin, M. & Ryan, T. (1982) "Bacteriophage DNA-dependent RNA polymerases" In *The Enzymes*, Boyer, P. (ed.), Academic Press, New York, pp. 87-108.

15. Guatelli, J.C., Whitfield, K.M., Kwoh, D.Y., Barringer, K.J., Richman, D.D. & Gingeras, T.R. (1990) "Isothermal, *in vitro* amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication" *Proc. Natl. Acad. Sci. USA* **87**, 1874-1878.

16. Joyce, G.F. "Selective amplification techniques for optimization of ribozyme function" In *Antisense RNA and DNA*, Murray, J.A.H. (eds.), Alan R. Liss, New York, in press.

17. Cech, T.R. (1987) "The chemistry of self-splicing RNA and RNA enzymes" *Science* **236**, 1532-1539.

18. Burke, J.M., Belfort, M., Cech, T.R., Davies, R.W., Schweyen, R.J., Shub, D.A., Szostak, J.W. & Tabak, H.F. (1987) "Structural conventions for group I introns" *Nucl. Acids Res.* **15**, 7217-7221.

19. Price, J.V., Kieft, G.L., Kent, J.R., Sievers, E.L. & Cech, T.R. (1985) "Sequence requirements for self-splicing of the *Tetrahymena* pre-ribosomal RNA" *Nucl. Acids Res.* **13**, 1871-1889.

20. Szostak, J.W. (1986) "Enzymatic activity of the conserved core of a group I self-splicing intron" *Nature* **322**, 83-86.

21. Joyce, G.F. & Inoue, T. (1987) "Structure of the catalytic core of the *Tetrahymena* ribozyme as indicated by reactive abbreviated forms of the molecule" *Nucl. Acids Res.* **15**, 9825-9840.

22. Doudna, J.A. & Szostak, J.W. (1989) "RNA-catalysed synthesis of complementary-strand RNA" *Nature* **339**, 519-522.

23. Joyce, G.F., van der Horst, G. & Inoue, T. (1989) "Catalytic activity is retained in the *Tetrahymena* group I intron despite removal of the large extension of element P5" *Nucl. Acids Res.* **17**, 7879-7889.

24. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1988) "Primer-directed enzymatic amplificiation of DNA with a thermostable DNA polymerase" *Science* **239**, 487-491.

25. Sinha, N.K. & Haimes, M.D. (1981) "Molecular mechanisms of substitution mutagenesis: an experimental test of the Watson-Crick and Topal-Fresco models of base mispairings" *J. Biol. Chem.* **256**, 10671-10683.

26. Leung, D.W., Chen, E. & Goeddel, D.V. (1989) "A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction" *Technique* **1**, 11-15.

27. Herschlag, D. & Cech, T.R. (1990) "DNA cleavage catalyzed by the ribozyme from *Tetrahymena*" *Nature* **344**, 405-409.

Publications During Grant Period

Joyce, G.F. (1989) "RNA evolution and the origins of life" *Nature* **338**, 217-224.

Joyce, G.F., van der Horst, G. & Inoue, T. (1989) "Catalytic activity is retained in the *Tetrahymena* group I intron despite removal of the large extension of element P5" *Nucl. Acids Res.* **17**, 7879-7889.

Joyce, G.F. (1989) "Amplification, mutation, and selection of catalytic RNA" *Gene* **82**, 83-87.

Robertson, D.L. & Joyce, G. F. (1990) "Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA" *Nature* **344**, 467-468.

Beaudry, A.A. & Joyce, G.F. (1990) "Minimum secondary structure requirements for catalytic activity of a self-splicing group I intron" *Biochemistry* **29**, 6534-6539.

Robertson, D.L. & Joyce, G.F. (1991) "The catalytic potential of RNA" In *Molecular Evolution on Rugged Landscapes: Proteins, RNA and the Immune System*, A.S. Perelson & S.A. Kauffman (eds.), Santa Fe Institute Studies in the Sciences of Complexity, vol. 8, Addison-Wesley, Redwood City, pp. 265-277.

Joyce, G.F. (1991) "The rise and fall of the RNA world" *The New Biologist* **3**, 1-9.

Joyce, G.F. (1992) "Selective amplification techniques for optimization of ribozyme function" In *Antisense RNA and DNA*, J.A.H. Murray (ed.), Modern Cell Biology Series, Alan R. Liss, New York, pp. 353-372.

Beaudry, A.A. & Joyce, G.F. (1992) "Directed evolution of an RNA enzyme" *Science* **257**, 635-641.

Cadwell, R.C. & Joyce, G.F. (1992) "Randomization of genes by PCR mutagenesis" *PCR Methods Appl.* **2**, 28-33.