

# Early Origins of Genetic Systems and Remnants of the RNA World

By George E. Fox

**ABSTRACT—These two projects focused on different aspects of the hypothesis that DNA based Earth life arose from an earlier RNA World: (1) Early Origins of Genetic Systems and (2) Remnants of the RNA World: RNA Structures Associated with Gene Regulation.” Researchers conducted a detailed examination of modern genetic systems and found support for the notion that RNA catalyzed protein synthesis predates DNA replication and is at least as old as transcription. Thus, we concluded that living organisms as we know them likely emerged first with an RNA-based genetic system and only later with DNA as the genetic material. The development of sophisticated translation machinery and its integration with RNA level regulation of transcription was therefore likely to have been a major driving force in the early history of life. A detailed study of the RNAs and proteins associated with translation machinery identified many remnants of early evolution.**

**W**E SOUGHT TO COMPARE THE RELATIVE AGE OF MAJOR cellular processes to translation as well as the relative age of components of the translation machinery. To this end, the *r*-proteins and RNAs associated with the large ribosomal subunit of archaeal and bacterial ribosomes were intercompared with particular attention paid to features unique to one or the other, as universal components are likely older than kingdom specific components. Likewise, those that are central to ribosome assembly are more likely older than those which are added at the end. These hypotheses about age were also evaluated within the context of gene organization. Components thought to be older were consistently found in more conserved gene clusters. The structures of many other cellular proteins are also known and these have been classified by the folds they contain, as summarized in the SCOP (structural classification of proteins) database at <http://scop.mrc-lmb.cam.ac.uk/scop/>. We used this database as a starting point to identify non-*r*-proteins that share similarity with *r*-proteins.

## Results

Multiple examples of *r*-proteins are related by insertion, fusion, and/or duplication events; in those cases, an interest is which protein is the predecessor. One key example is the very ancient protein L2 which is universal, early in the assembly process and in one of the most conserved *r*-protein gene clusters. This protein has two domains. One has an OB fold and the other an SH3 fold. These folds differ by the insertion/deletion of a single alpha helical element. Thus, a likely scenario is that L2 began as a single domain protein with an SH3 fold

which allowed it to interact with RNA. A subsequently duplication event followed by a second insertion event would then create a new second domain. The resulting OB fold, which may have originated with L2, is found in many modern membrane associated proteins.

Among the most interesting findings is the observation that elongation factor G (EF-G) is largely a composite of several *r*-protein domains. EF-G has five structural domains. Domain II which is also shared with EF-Tu is seen in one of the oldest ribosomal proteins, L2. Domains III and V have the same ferredoxin-like fold seen in *r*-proteins S6 and S10. Domain IV has an alpha/beta structure as found in *r*-protein S9 and the central domain of *r*-protein S5. EF-G is involved in GTP cleavage and is a key component of ribosomal bioenergetics. In its absence, the rate of translation is dramatically slowed but not eliminated. In total, the evidence suggests that it is a relatively recent addition to the ribosomal machinery. Given its dramatic role in the rate of protein synthesis, its introduction may have been a major transition in the history of living systems.

Another far more recent partial duplication event was detected for L15 and L18e. The former is a universal protein and therefore likely to be older, since the latter is not found in bacteria. The two proteins share significant sequence homology as well as structural similarity, but L15 has an extension that is missing in L18e. The binding site for L18e in the archaeal ribosomal RNA includes an inserted loop not present in bacterial *r*RNAs. Hence, the newer protein interacts with an added/newer feature in the RNA.

There are, in fact, many such minor changes in the RNAs when kingdoms are compared. Some of these, like the L18e example, are associated with *r*-proteins unique to a kingdom. However, a more detailed examination revealed three large clusters of changes in which the variant RNA regions interact with either one another or with novel proteins. Thus, it is likely that the various components of each cluster co-evolved suggesting that various RNA and protein changes in the cluster are of a similar age.

## Discussion and Conclusions

Results obtained from these studies have made it clear that it is possible to determine the relative age of many cellular components. It is generally thought that if two proteins or RNAs are similarly distributed among the various taxa that it would not be possible to show that characteristic despite the fact that many components are completely universal. One can nevertheless use data from multiple sources to assess the relative age of even these early components.

In conclusion, a multifaceted study of protein and RNA that takes into account sequence structure, genetic regulation, and functional positioning can provide meaningful insight into the earliest history of cellular organization and evolution. Funding for this purpose was sought and obtained from NASA's Exobiology program.

These preliminary data also yield information on RNA structure that, in part, led to a funded project from the Texas Advanced Research Program. In addition, multiple publications are now in preparation.

## Publications

Hury J., U. Nagaswamy, M. Larios-Sanz, and G. E. Fox, "Ribosome Origins: The Relative Age of 23S rRNA Domains," *Origins Life & Evol. Biosphere* 36 (2006): 421-29.  
Wang J., "From Genome to Structure: Comparative Studies of Archaeal Unique Ribosomal Proteins," Ph. D. dissertation, University of Houston, Houston, TX, 2006.

## Presentations

Fox, G. E., "Inferring Evolutionary History from Multiple Data Sets: Insights to the Origins of the Translation Machinery," Invited Symposium Speaker, Computational Molecular Biology: The Future; University of Houston, Houston, TX, April 4, 2005.  
Dasgupta I., Y. Liu, J. Wang, H-C. Huang, U. Nagaswamy, G. E. Fox, "Conservation and Clustering of Translation Related Genes," Cold Springs Harbor Meeting on Genome Informatics, Oct. 28–Nov. 1, 2005, Cold Springs Harbor, NY.  
Fox, G. E., "Origins of the Translation Machinery," Invited Seminar Speaker, Planetary Protection Group, Jet Propulsion

Laboratory, Pasadena, CA. Dec. 8, 2005.

Fox, G. E., "Unraveling the History of the Ribosome," Invited Symposium Speaker, Houston Society for Engineering in Medicine and Biology 23rd Annual Conf. on Biomedical Engineering Research, Houston, TX, Feb. 9–10, 2006.  
Fox, G. E., "EF-G, A Key Historical Advance in Early Genetic Systems," Invited speaker, Origin of Life Gordon Conf., Bates College, Lewiston, MN, July 23–28, 2006.  
Fox, G. E. "The Ribosome as a Model Nanomachine," Invited seminar, Dept. of Chemical Engineering, Syracuse University, Syracuse, NY, April 20, 2007.

## Funding and Proposals

Travisano, M. T., Fox, G. E. et al., "Shared Genomic Resources in Prokaryotic Evolution," NSF-Frontiers in Biology Program, October 1, 2005–September 30, 2010. Total Costs: \$7,612,114. (*Not funded.*)  
NASA Exobiology Program; "The Origins of Translation and Early Evolution of Life," Aug. 15, 2005–Aug. 14, 2008. Total costs: \$288,268. (*Funded.*)  
Texas ARP, "Artificial Stable RNA Sequestration of Heavy Metals," Jan. 1, 2006–Aug. 31, 2008. Total costs: \$176,076.



**RESEARCH CAMPUS**—At the heart of any great research university is an extensive library. UH's M.D. Anderson Library, (*far l.*) which underwent a major expansion in 2005, provides the UH System necessary support for advanced study in the 21st century.