

## Summary report on previous projects: Early origins of genetic systems and remnants of the RNA world

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. A detailed examination of modern genetic systems is being conducted, and it has been found to support the notion that RNA-catalyzed protein synthesis likely predates DNA replication and is at least as old as transcription. Thus, we conclude 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 a sophisticated translation machinery and its integration with RNA-level regulation of transcription was therefore likely a major driving force in the early history of life. A detailed study of the ribosomal proteins (r-proteins) and ribosomal RNAs (rRNAs) associated with the translation machinery is being carried out, and many remnants of early evolution have been identified.

### METHODOLOGY

We sought to compare the relative age of major cellular processes to translation as well as to the relative age of components of the translation machinery. In doing this, we attempted to take advantage of the fact that atomic resolution structures are available for almost all the r-proteins. The structures of these and many other cellular proteins 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. Using this resource, the r-proteins and rRNAs associated with the large ribosomal subunit of archaeal and bacterial ribosomes were intercompared, with particular attention paid to features that were both unique to one or the other type of ribosome or shared between kingdoms. The universal components are likely older than kingdom-specific components. Likewise, those that are central to ribosome assembly are more likely older than those that are added at the end. These hypotheses about age were also evaluated within the context of gene organization. Genome data were used to assess the conservation of gene order of all translation-related proteins in both Archaea and Bacteria. Components intuitively thought to be older by other indicators were consistently found in the most conserved gene clusters.

### RESULTS

Initial studies revealed multiple examples of r-proteins that are related by insertion, fusion, and/or duplication events. Moreover, many non-universal r-proteins have analogs rather than homologs when the Archaea and Bacteria are compared. These analogs occupy equivalent positions in the ribosomes. 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 the various RNA and protein changes in the cluster are of similar age.

In our more recent work, the organization of r-protein genes has been examined in detail. There are 10 well-characterized r-protein operons in Bacteria, five of which are also present in the Archaea. In examining all the r-proteins in detail, it was discovered that additional highly conserved clusters are present in both the Bacteria and Archaea. A scoring system was developed to rank the relative conservation of each cluster. The 17 most conserved clusters include only r-proteins and translation-related proteins such as Ef-Ts and *frr*. The only significant exceptions are core polymerase genes involved in transcription. It is only among the less conserved clusters that one begins to see genes associated with DNA replication and other cellular processes, again supporting the notion that translation and transcription have been present and coupled since the very earliest times.

The assembly of the 30S subunit was also re-examined to see if patterns of the sort seen previously with the large subunit could be found. Previous efforts in this regard had not been fruitful, as almost all the proteins are universal. The most notable new observation was the binding pattern of the proteins involved in the S7 branch of 30S ribosomal assembly. R-protein S7 and subsequent joining proteins (S9-S13-S19, S14-S10-S3-S2) all bind to the 3' major domain of the 16S rRNA, which hosts the functionally indispensable decoding region of the 30S subunit, which one would therefore expect to be older than other areas of the 30S subunit. It was observed that all the proteins on the S7 assembly path also belong to universal and highly conserved gene clusters. In contrast, most of the proteins associated with the other assembly paths are members of non-universal r-protein gene clusters. These observations

strongly support the notion that the decoding region is the oldest part of the 30S subunit.

## DISCUSSION AND CONCLUSIONS

The results obtained from the studies funded by these two mini-grants have made it clear that it is possible to determine the relative age of many cellular components. It is generally held that, if two proteins or RNAs are similarly distributed among the various taxa, it is impossible to assess their relative age. However, it has been found that by examining the conservation of genomic associations, this limitation can be over-

come. This is the case despite the fact that many components are completely universal. 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 will be sought from the National Science Foundation in the summer of 2008. At the last Gordon Conference on the Origin of Life, representatives from NSF indicated the agency had an interest in funding work relating to the biological aspects of astrobiology and exobiology.