

Early Origins of Genetic Systems

George E. Fox

Abstract—Much of the early development of genetic machinery was accompanied by two-way exchange of components between early cellular systems possibly mediated in some cases by lateral gene transfer. A search for structural similarities between ribosomal proteins and other proteins revealed likely examples of such transfer.

IT IS GENERALLY BELIEVED THAT living organisms, as we know them, likely emerged from an RNA world, first with an RNA based genetic system and later with DNA as the genetic material.¹⁻³ The development of sophisticated translation machinery and its integration with RNA level regulation of transcription may have been a major driving force in the early history of life. With these two core processes in place, early organisms could then have expanded their repertoire of capabilities leading to the discovery of improved information storage (DNA) followed by a second major addition of functionality. If this scenario is correct, it is likely that the initial period would have been followed by a cross fertilization period that occurred after these central cellular processes were well established. Hence, we would expect to find that the earliest protein components of these processes might have been incorporated into later evolving cellular systems.

In order to test this hypothesis, we focused our efforts on ribosomal proteins seeking to identify individual proteins or protein domains that may have been moved between and among the major cellular systems. Ribosomal machinery is ideal for such a study for two reasons. First, because crystals of the 30S and 50S ribosomal subunits have been studied,^{4,7} the structures of most of the ribosomal proteins (*r*-proteins) associated with the protein synthesis machinery are now known at atomic resolution. Second, prior examination of the ribosome assembly process⁸ and sequence conservation patterns suggests an early origin for certain proteins, e.g., L2, L3, L4, and L24, etc. Therefore, the folds they contain are likely to be among the oldest. Consistent with this, the core fold seen in L3 has also been found in EF-Tu, EF-G, initiation factors (IF2 and eIF2), riboflavin synthase, ferredoxin reductase, NADPH-cytochrome p450 reductase, and L-fucose isomerase.



Dr. George E. Fox

Methodology

In addition to the ribosomal proteins, the structure of many cellular proteins including many of those involved in DNA replication has been studied at atomic resolution. There is an ongoing project known as SCOP⁹ in which all known protein structures have been examined and classified by the folds they contain. This database is publicly available at <<http://scop.mrc-lmb.cam.ac.uk/scop/>>. In general, proteins that contain similar folds have a significant potential to be historically related whereas those that are not so similar are far less likely to be related. We use this database as a starting point to identify non-ribosomal proteins (and other ribosomal proteins, as well) that shared similar folds. We then compared these possible analogs.

Results

An initial examination of the distribution of ribosomal proteins in the SCOP database revealed that the mostly single domain ribosomal proteins, whose structures are known, fall into 46 different categories. Thus, it is clear that most of the ribosomal proteins do not share a common early history. There are, however, several examples of *r*-proteins that are related by insertion, fusion, and/or duplication events. In those cases, we are interested in which protein is the predecessor. Interest focuses on which proteins are recruited to the ribosome and which are recruited from the ribosome at different times.

One key example is the very ancient protein L2 which 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 subsequent 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.

Another partial duplication event (Wang and Fox, *manuscript in preparation*) has been 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 missing in L18e. The binding site for L18e in the Archaeal ribosomal RNA includes an inserted loop that is not present in bacterial *r*RNAs. Hence, the newer protein interacts with an added/newer feature in the RNA.

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.

Among proteins that have likely been recruited from elsewhere to the ribosome are S6 and S10 which resemble the ferredoxins likely among the first proteins. Other recent additions are *r*-proteins that contain Zn-binding motifs. These include L337e, L37ae, L44e, and S27e. None of these proteins are universal.

Discussion

Although SCOP focuses exclusively on proteins whose three-dimensional structure is known, this does not exclude the inclusion of other proteins in future work. Once folds are identified, they can be sought in proteins that have not yet been crystallized by looking for similar domains that can be identified at the primary sequence level or by comparing predicted secondary structures to known three-dimensional structures. Thus, for example, we examined⁶ a fold known as the S1 domain (originally discovered in ribosomal protein S1). Sequence comparisons resulted in the discovery of a similar fold in large numbers of proteins. The proteins containing S1 domains can be broadly grouped into three main functional groups: RNA processing, involvement in transcription or translation, and chromatin or septum regulation. Although the S1 motif is found in all three domains of life, only the IF-1/eIF1A types are universally distributed, suggesting that this might be the original source of the fold. It is likely that ribosomal protein S1, itself, is a late addition to the ribosome, probably derived from the initiation machinery. This example and the earlier L3 example illustrate a common theme that may be largely unique to the earlier history of life on Earth, mainly the lateral transfer of domains between genes as opposed to the lateral transfer of genes between genomes.

Conclusions

Results obtained here illustrate a key role for the ribosomal machinery as a source and sink of useful proteins or protein domains throughout early evolution. The immediate goal in the research effort will be to add DNA replication and repair proteins to the analysis mix. Funding for this purpose will be sought from NASA's Exobiology Program.

References

- ¹G. J. Olsen, "The History of Life," *Nat Genet.* 28 (2001): 197-98.
- ²J. K. Harris, S. T. Kelley, G. B. Spiegelman, and N. R. Pace, "The Genetic Core of the Universal Ancestor," *Genome Res.* 13 (2003): 407-12.
- ³E. V. Koonin, "Comparative Genomics, Minimal Gene-sets and the Last Universal Common Ancestor," *Nat. Rev. Microbiol.* 1 (2003): 127-36.
- ⁴N. Ban, P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz, "The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution," *Science* 289 (2000): 905-20.
- ⁵J. Harms, F. Schluenzen, R. Zarivach, A. Bashan, S. Gat, I. Agmon, H. Bartels, F. Franceschi, and A. Yonath, "High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium," *Cell* 107 (2001): 679-88.
- ⁶D. E. Brodersen, W. M. Clemons, Jr., A. P. Carter, B. T. Wimberly, and V. Ramakrishnan, "Crystal Structure of the 30 S Ribosomal Subunit from *Thermus thermophilus*: Structure of the Proteins and their Interactions with 16S RNA," *J. Mol. Biol.* 316 (2002): 725-68.
- ⁷D. J. Klein, P. B. Moore, and T. A. Steitz, "The Roles of Ribosomal Proteins in the Structure Assembly, and Evolution of the Large Ribosomal Subunit," *J. Mol. Biol.* 340 (2004): 141-77.
- ⁸G. E. Fox and A. K. Naik, "The Evolutionary History of the Ribosome," in *The Genetic Code and the Origin of Life*. Ed. L. Ribas de Poplana. Georgetown, TX: Landes Bioscience, 2004. 92-105.
- ⁹A. Andreeva, D. Howorth, S. E. Brenner, T. J. P. Hubbard, C. Chothia, and A. G. Murzin, "SCOP Database in 2004: Refinements Integrate Structure and Sequence Family Data," *Nucl. Acid Res.* 32 (2004): D226-29.

Publications

- Hury, J., U. Nagaswamy, M. Larios-Sanz, and G. E. Fox. "Ribosome Origins: The Relative Age of 23S *r*RNA Domains," *Origins Life & Evol. Biosphere*. (In press.)

Presentations

- Dasgupta I., Y. Liu, J. Wang, H.-C. Huang, U. Nagaswamy, and G. E. Fox. "Conservation and Clustering of Translation Related Genes," Cold Springs Harbor Meeting on Genome Informatics, Cold Springs Harbor, NY, Oct. 28–Nov. 1, 2005.
- Fox, G. E. "EF-G, A Key Historical Advance in Early Genetic Systems," Origin of Life Gordon Conf., Bates College, Lewiston, ME, July 23–28, 2006. (Invited to speak.)
- . "Inferring Evolutionary History from Multiple Data Sets: Insights to the Origins of the Translation Machinery," Computational Molecular Biology: The Future, University of Houston, Houston, TX, April 4, 2005. (Invited speaker.)
- . "Origins of the Translation Machinery," Planetary Protection Group, Jet Propulsion Laboratory, Pasadena, CA, Dec. 8, 2005. (Invited seminar speaker.)
- . "Unraveling the History of the Ribosome," Houston Society for Engineering in Medicine and Biology 23rd Annual Conf. on Biomedical Engineering Research, Houston, TX, Feb. 9–10, 2006. (Invited symposium speaker.)

(Continued on page 95.)



Courtesy NASA-Kennedy Space Center



Early Origins of Genetic Systems

(Continued from page 65.)

Funding and Proposals

Fox, G. E. "The Origins of Translation and Early Evolution of Life," NASA Exobiology Program, Aug. 15, 2005–Aug. 14, 2008, \$288,268. *(Funded.)*

Travisano, M. T., G. E. Fox, et al. "Shared Genomic Resources in Prokaryotic Evolution," NSF-Frontiers in Biology Program, Oct. 1, 2005–Sept. 30, 2010, \$7,612,1140. *(Not funded.)*