Coevolution of ribosomal proteins and RNA: evidence against the RNA world hypothesis

Gustavo Caetano-Anollés (gca@illinois.edu); Evolutionary Bioinformatics Laboratory, Department of Crop Sciences, University of Illinois; Urbana, IL, USA

The evolutionary study of thousands of RNA molecules and millions of proteins reveals the coordinated evolution (coevolution) of protein and RNA components of the ribosome. Remarkably, patterns of evolutionary accretion suggest the ribosome originated in structures necessary for ribosomal mechanics and not for protein synthesis. This phylogenomic evidence is, therefore, incompatible with the appearance of the ribosome in an ancient "RNA world." Instead, a growing body of evidence suggests that modern biochemistry originated gradually in a "world" with constituents that were quite similar to our own.

There is still much to learn about how modern biochemistry originated and evolved. The widely embraced "RNA world" model of origin of life is based on the premise that "genetics" preceded "metabolism" and that RNA molecules originated prior to proteins (Gilbert 1986; Orgel 1992). However, the existence of a thriving RNA world is troublesome and violates the principle of continuity treasured by evolutionary biologists. The RNA world lacks: (1) *Persistency:* It relinquished most of its crucial catalytic and replication abilities to proteins (e.g., metabolic enzymes, synthetases, polymerases), and most of its remnants cannot function without them (e.g., ribosomes, RNase P, tRNA), and exceptions to these rules, such as riboswitches and viroid ribozymes, are of mostly specialized utility; (2) *Ubiquity:* Ancient RNA vanished from metabolic networks, whereas newly uncovered RNA is fundamentally regulatory and recent in origin; (3) *Specificity:* The product of the RNA world, the genetic code, is defined and maintained by crucial aminoacyl-tRNA synthetase (aaRS) proteins interpreters, not RNA counterparts; and (4) *Evolvability:* The drivers of the RNA-to-protein transition are expected to operate in the large subunit of ribosomal RNA (rRNA) in the absence of selective pressures that would favor protein encoding and links to synthesis, while, at the same time, ancient replicative functions are not sustained by regulatory control networks, which are recent, not ancient.

The lack of persistence of an ubiquitous catalytic pure RNA world is particularly troublesome. Persistence is a necessary consequence of evolution; it manifests itself in molecular structure, from highly dynamic conformers to essentially immutable folds. Take, for example, the rRNA molecule; its structure is fluid. Base pairs quickly associate and disassociate to form a multitude of alternative structural conformations. Evolution, however, has reduced the number of possible structures by increasing the stability and average lifetime of only a few (Fontana 2002; Schultes et al. 2005). This "structurally canalized" repertoire harbors useful and durable functions in robust structures that are carefully optimized as the molecules change by mutation and interaction with their environment. Canalization is a powerful and important concept that ensures persistence and durability. It was introduced by Waddington to describe the robustness of an epigenetic

landscape (Waddington 1942). It explains the existence of evolutionary constraints acting on molecules, cells, and organisms and the influence of the environment. It is genetic and epigenetic and embodies "coevolution" – a bias in the components of a biological system as components interact and change.

In the case of rRNA, the components of the system in question are nucleotides that interact with each other and with other molecules in various ways. In ribosomal proteins (r-proteins), coevolution may express itself as coordinated changes in sequence, structure, or both. For example, computational analyses revealed that coevolutionary relationships between PFAM protein domains led to structural and functional constraints (Yeang and Haussler 2007). These constraints manifest tendencies of spatial coupling and occur in functionally important sites of proteins. For example, 43 coevolving amino acid residues were identified in 10 r-protein domains of the small ribosomal subunit, all of which are close to tRNA binding sites. This includes sites in the S12 r-protein, which, as I will describe below, is the most ancient of the ensemble. Remarkably, not all physical interactions are coevolved, meaning, again, that only a selected few are crucial and constrained.

If canalization and coevolution are inescapable structuring forces, we can make use of these principles to find deep phylogenetic imprints in molecular structure and identify an "arrow of time" capable of defining evolutionary origins. Under these basic assumptions, we searched for the roots of modern biochemistry without invoking assumptions drawn from origins of life models. Instead, we focused on the structure of molecules and conducted a survey of protein domain structures in genomes and substructures in rRNA molecules. In search of direct answers, we used this information to build phylogenies and timelines of ribosomal history with standard bioinformatics tools of phylogenetic analysis (Caetano-Anollés 2002; Sun and Caetano-Anollés 2009; Harish and Caetano-Anollés 2012). Tendencies towards increases in domain abundance or decreases in structural conformations were used to root phylogenomic statements. The relative ages of structures were derived directly from phylogenies, indexed with functional and molecular contact information, and finally mapped (by color) onto three-dimensional models of the ribosome (Figure 1A). The outcome of our most recent studies was unexpected: (1) Subunit RNA and proteins coevolved tightly, starting with interactions between the oldest proteins (S12 and S17) and the oldest rRNA helix in the small subunit (the ribosomal ratchet responsible for ribosomal dynamics) and ending with the rise of a modern multisubunit ribosome; (2) A major transition in evolution ca. 3.1 billion years ago (Gya) brought independently evolving ribosomal subunits together by unfolding inter-subunit (bridge) contacts and interactions with full cloverleaf tRNA structures; (3) During this transition, a fully-fledged peptidyl transferase center (PTC) responsible for protein synthesis appeared by duplication of local helical structures, supporting an appealing model of PTC origin (Agmon et al. 2006); and (4) A second evolutionary transition occurred almost concurrently with the "great oxidation event" (ca. 2.4 Gya) and involved the discovery of the L7/L12 protein complex that stimulates the GTPase activity of elongation factor G. This second transition would likely have notably enhanced ribosomal efficiency.

A “pure” RNA world is incompatible with the existence of coevolutionary patterns in ribosomal molecules. It is also incompatible with evolutionary timelines of domains (e.g., Caetano-Anollés et al. 2012) and molecular functions (Kim and Caetano-Anollés 2010) derived from a genomic census of domain structures and gene ontologies (Figure 1B). These timelines show congruently that metabolic enzymes appeared prior to RNA-binding proteins. Phylogenomics, therefore, provides evidence against a pure RNA world hypothesis. These results also call into question the validity of assumptions used in studies of ribosomal structure that simulate the evolution of the large rRNA subunit (Bokov and Steinberg 2009; Hsiao et al. 2009; Fox 2010). In these studies, it is assumed that helical stacking interactions recapitulate molecular growth in RNA, and that structures grow in concentric shells from an ancient core that embeds the PTC. Although we find that the majority of helices evolved before the corresponding adenosine stack in A-minor interactions and that, in general, the ribosomal core is more ancient than peripheral regions (matching patterns of sequence conservation), structural canalization in our experiments does not place the origin of the ribosome in the PTC. Thus, ribosomal components were recruited for protein synthesis from structures that were performing other functions. Remarkably, the ancient “processive” ribosomal core we identified showed homologies to *in vitro* evolved RNA replicase ribozymes and proteins structures in extant replication machinery (Harish and Caetano-Anollés 2012).

The corollary is that a fully functional biosynthetic mechanism responsible for primordial peptides (and ancient r-proteins) must have existed that was superseded by the ribosome. Given canalization, then the ancient putative mechanism must be operational today. Indeed, catalytic domains of aaRSs and aminoacylating modules of non-ribosomal protein synthetases (NRPSs) harbor peptide biosynthetic functions either as small standalone enzymes or in large, assembly line ensembles (reviewed in Caetano-Anollés et al. 2012). Remarkably, their domain structures appear earlier than r-proteins (Figure 1), suggesting that they embody ribosomal predecessors.

An important question for the future is how proteins assembled and retained memory prior to the appearance of modern genetics. A hint comes from organisms with faulty aaRS enzymes and statistical proteomes (e.g., Boniecki and Martinis 2012), the observation that random protein sequences fold into defined structures (LaBean et al. 2012), and evolutionary outcomes in mRNA display experiments (Seelig 2011) that point to remarkable structural tendencies. I will just mention the promise of statistical proteomes for synthetic biology. Cells maintain viability despite loss of crucial aaRS hydrolytic editing activities, which cause widespread and presumably lethal proteomic mutations. Default to primordial mechanisms by destruction of coding specificities showcases the unprecedented resilience of the structures and functions of proteins. An evolutionary scenario of gradual emergence of structural and genetic codes from stochastic behavior is therefore feasible and worthy of careful experimental and theoretical exploration.

References:

Agmon, I., A. Bashan, and A. Yonath (2006). “On ribosome conservation and evolution.” Israel Journal of Ecology & Evolution **52**: 359–374.

Bokov, K. and S.V. Steinberg (2009). “A hierarchical model for evolution of 23S ribosomal RNA.” Nature **457**: 977–980.

Boniecki, M.T. and S. A. Martinis (2012). “Coordination of tRNA synthetase active sites for chemical fidelity.” Journal of Biological Chemistry **287**: 11285-11289.

Caetano-Anollés, G. (2002). “Tracing the evolution of RNA structure in ribosomes.” Nucleic Acids Research **30**: 2527-2587.

Caetano-Anollés, G., K.M. Kim and D. Caetano-Anollés (2012). “The phylogenomic roots of modern biochemistry: Origins of proteins, cofactors and protein biosynthesis.” Journal of Molecular Evolution **74**: 1-34

Fontana, W. (2002). “Modelling ‘evo-devo’ with RNA.” BioEssays **24**: 1164–1177.

Fox, G.E. (2010). “Origin and evolution of the ribosome.” Cold Spring Harbour Perspectives in Biology **2**: a003483.

Gilbert, W. (1986). ‘The RNA world.” Nature **31:** 618.

Harish, A. and G. Caetano-Anollés (2012). “Ribosomal History Reveals Origins of Modern Protein Synthesis.” PLoS ONE **7**(3): e32776.

Hsiao, C., S. Mohan, B.K. Kalahar, and L.D. Williams (2009). “Peeling the onion: Ribosomes are ancient molecular fossils.” Molecular Biology & Evolution **26**: 2415–2425.

Kim, K.M. and G. Caetano-Anollés (2010). “Emergence and evolution of modern molecular functions inferred from phylogenomic analysis of ontological data.” Molecular Biology & Evolution **27**: 1710-1733.

LaBean, T.H., T.R. Butt, S.A. Kauffman and E.A. Schultes (2011). “Protein folding absent selection.” Genes **2**: 608-626.

Orgel, L. E. (1992). “Molecular Replication.” Nature **358**: 203-209.

Schultes, E.A., A. Spasic, U. Mohanty, and D.P. Bartel (2005). “Compact and ordered collapse of randomly generated RNA sequences.” Nature Structural & Molecular Biology **12**:1130–1136.

Seelig, B. (2011). “mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries.” Nature Protocols **6**: 540-552.

Sun, F.-J. and G. Caetano-Anollés (2009). “The evolutionary history of the structure of 5S ribosomal RNA.” Journal of Molecular Evolution **69**: 430-443.

Waddington, C.H. (1942). "Canalization of development and the inheritance of acquired characters." Nature **150**: 563–565.

Yeang, C.H. and D. Haussler (2007). “Detecting coevolution in and among protein domains.” PLoS Computational Biology **3**: e211.

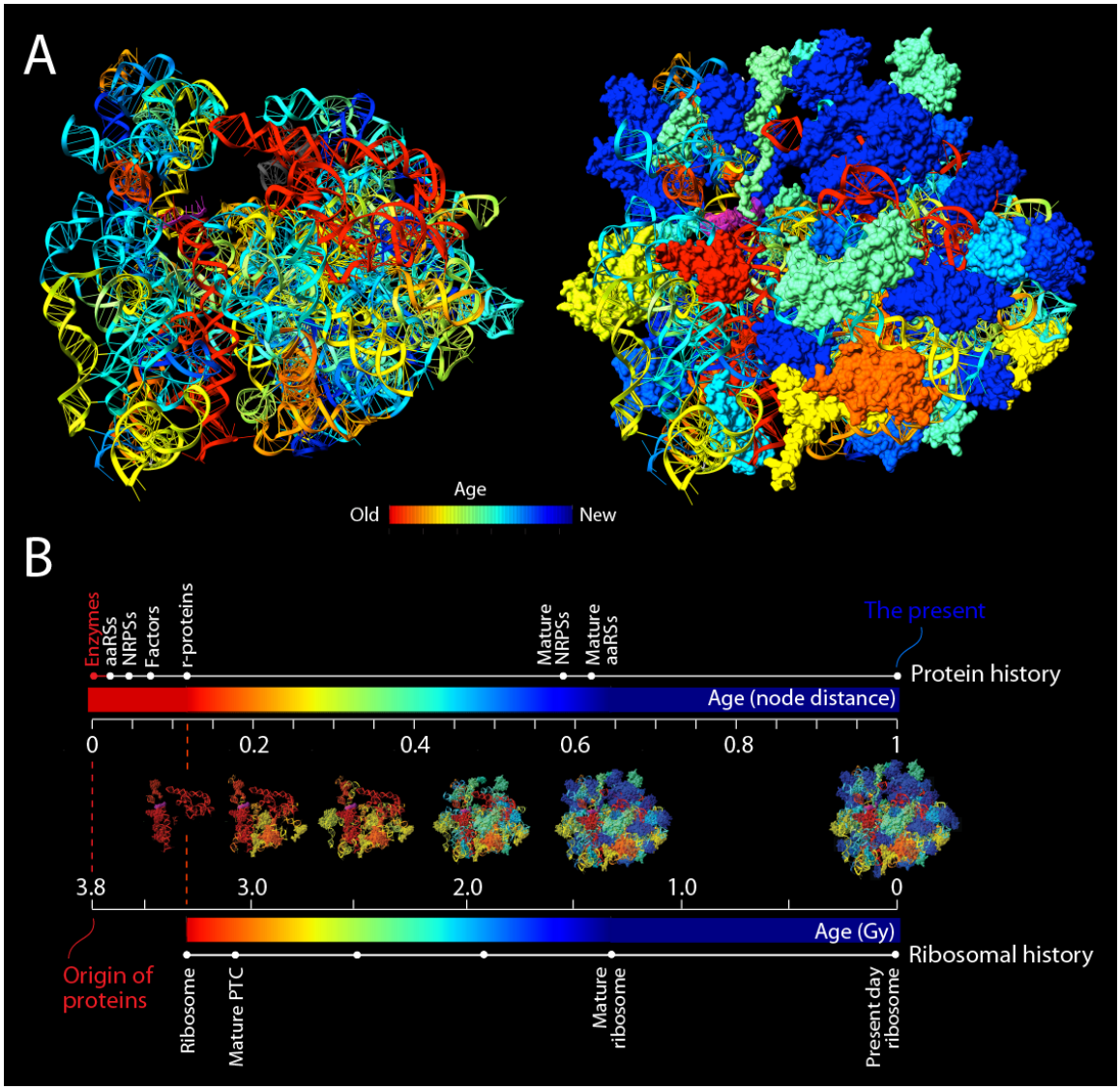

Figure 1: A. Evolutionary history of rRNA (left) and rRNA and r-proteins (right) traced onto the three-dimensional structure of the core ribosome, with ages of RNA and protein domains colored with hues from red (ancient) to blue (recent). B. Phylogenomic history of protein domains shows that metabolic enzymes preceded RNA-protein interactions and the ribosome. The protein timeline (top line with time flowing from left to right in a relative 0-1 scale) was derived from a universal phylogenetic tree of protein domain structure at family level of complexity. The short segment of the timeline that is colored in red (~100 million years of evolution) depicts metabolic domain families appearing in evolution before the first domains known to interact with RNA, the catalytic domains of aaRSs. aaRSs and NRPSs are responsible for non-ribosomal protein synthesis, factors are effective switches and transporters, and r-proteins associate with the ribosome. Their first appearances are labeled with dots. A ribosomal timeline (bottom line) shows that the ribosome emerged late in evolution following metabolic enzymes, aaRSs, factors, and NRPSs.