Cell

Crawling Out of the RNA World

Thomas R. Cech[1,*]
[1]Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA
*Correspondence: thomas.cech@colorado.edu
DOI 10.1016/j.cell.2009.02.002

Comparison of phylogenetically diverse ribonucleoprotein (RNP) enzymes and information about their biochemistry have stimulated hypotheses about their evolution. Instead of the canonical view, in which catalysis proceeds from ribozyme to RNP enzyme to protein enzyme, RNP enzymes and proteins are seen to share contemporary catalysis. Furthermore, the RNA components of RNP enzymes show no evidence of fading out but instead, in some cases, have elaborated new functions.

Although catalytic machines comprised entirely of proteins outnumber those that contain essential RNA subunits, the latter also play a central role in biology. Of the four most fundamental processes that maintain and express genetic systems, two—DNA replication and transcription—are accomplished by purely protein polymerases. The other two—mRNA splicing and protein synthesis—are catalyzed by RNP complexes (the spliceosome and ribosome). Other RNP enzymes make essential contributions to RNA processing, protein translocation across membranes, and addition of telomeric DNA repeats to chromosome ends.

What is the history of contemporary catalysis? Many find it attractive to envision a primordial RNA world, where RNA provided both information and function, genotype and phenotype. Certainly RNA is an excellent informational molecule, as evidenced by messenger RNAs and by viruses that have RNA genomes (such as, the influenza virus). RNA is also a versatile and powerful catalyst: the *Tetrahymena* ribozyme has reached a sort of "catalytic perfection" (Herschlag and Cech, 1990), and ribozymes discovered by in vitro selection perform a wide range of reactions, even carbon-carbon bond formation (Helm et al., 2005). Certainly self-replicating RNA would provide a solution to the chicken-and-egg problem of early evolution: which came first, the informational molecule or the catalyst capable of replicating that informational molecule?

If indeed such an ancient RNA world existed, by what pathways did it evolve to provide the ensemble of biological catalysts found in current organisms? Evidence for the pathways comes from two sources: comparative analysis of noncoding RNAs that perform the same function in phylogenetically disparate organisms and biochemical and structural analysis of the separate contributions of the RNAs and proteins in various RNP enzymes.

Diverse Pathways of RNP Evolution

RNase P

We begin with RNase P, which cleaves a specific phosphodiester bond of tRNA precursors to generate the mature 5′ end of tRNAs and processes other RNA substrates as well. RNase P is composed of a catalytic RNA and one or more proteins. A number of bacterial RNase P RNAs are ribozymes; that is, they are able to perform accurate pre-tRNA processing in vitro in the complete absence of protein, albeit under conditions of a high divalent cation concentration. Each bacterial RNase P holoenzyme also contains a small basic protein that stabilizes the correct folded structure of the RNA under physiological salt conditions. This small basic protein also may extend the substrate range of the enzyme, may help to discriminate pre-tRNA substrate from tRNA product, and may mediate RNase P dimerization. Because all three domains of life (Bacteria, Archaea, and Eukarya) have structurally related RNase P RNAs but different protein components, the RNA component is likely to be the most ancient and to be already present in the common ancestor of extant life (Evans et al., 2006). Later (and certainly after the advent of message-directed protein synthesis), the small basic protein became an essential subunit of RNase P in Bacteria. In the evolutionary line that led to Eukarya and Archaea, at least four proteins were added (Evans et al., 2006; Walker and Engelke, 2006). The functions of these RNase P enzymes became more equally shared between RNA and protein, as evidenced by the fact that their RNA subunits have a very limited ability to catalyze tRNA processing in the absence of their protein subunits.

The recent finding that human mitochondrial RNase P is entirely a protein enzyme (Holzmann et al., 2008) might at first glance appear to provide evidence for the evolution of RNP enzymes to protein enzymes. However, the three protein subunits of the mitochondrial RNase P are unrelated to any known protein subunits of RNase P RNPs. Instead, the mitochondrial RNase P appears to be cobbled together from pre-existing protein enzymes. Thus, this interesting discovery appears to uncover a new pathway for tRNA maturation, rather than a new mode of RNP evolution.

Group I Introns

These present a situation similar to that of bacterial RNase P: in many cases, the group I introns are ribozymes, carrying out self-splicing under physiological conditions in vitro, but many of them require protein components for activity in vivo. For example, the CYT-18 protein of the mold *Neurospora* binds directly to a number of group I introns and facilitates their splicing. CYT-18 is a mitochondrial tyrosyl tRNA synthetase with idiosyncratic insertions within its nucleotide-binding domain that contribute to

stabilization of the catalytically active structure of the intron. Details by which the protein stabilizes the folded intron RNA have recently been revealed by an X-ray crystal structure (Paukstelis et al., 2008). Other group I introns rely on protein helpers completely unrelated to CYT-18. A distinction between the group I intron-associated proteins and those of RNase P is that the former may have additional functions unrelated to RNA splicing and represent enormous variety, whereas the latter are dedicated RNase P subunits.

Functional stabilization of a ribozyme need not require a large protein. For example, two peptides 17 and 19 amino acids in length and connected by a peptide linker can rescue the splicing of a mutated group I intron (Atsumi et al., 2001), providing an experimental model for the early evolution of a catalytic RNP. Furthermore, the ligation of short peptides into larger active molecules can itself be catalyzed by RNA templates (Kashiwagi et al., 2007).

Group II Introns and snRNPs

Group II introns have an RNA secondary structure and an "intron lariat" splicing mechanism distinct from those of group I; they are found in Bacteria, Archaea, and the organelles of Eukarya. Many of the group II introns are not very efficient self-splicers in vitro because they work naturally as RNPs. Their protein components may be host splicing factors, or they may be "maturase" proteins encoded by open reading frames that protrude from the intron RNA. Some group II introns are mobile genetic elements, encoding reverse transcriptases that contribute to their transposition. It has been argued that the evolution of these group II intron RNPs can be traced back to a common ancestral RNP consisting of a bacterial group II RNA structure and a compact reverse transcriptase, after which the RNA and protein coevolved (Toor et al., 2001). This is as far back as the ancestry can be traced; an RNA world origin for the ribozyme component remains an attractive but unproven hypothesis.

The spliceosome, the five-RNA multiprotein complex that catalyzes mRNA splicing (see Review by M.C. Wahl, C.L. Will, and R. Luhrmann in this issue of *Cell*), presumably arose in Eukarya after the division of the three domains. The possibility that the spliceosome evolved from the group II self-splicing introns is intriguing. More specifically, the model is that small nuclear RNAs (snRNAs) form the RNA catalytic center for the two chemical steps of mRNA splicing, and that once they have assembled with the RNA the local structure is homologous to that formed "in *cis*" by intramolecular folding of a group II self-splicing intron. Mechanistic and structural similarities establish that group II intron self-splicing and spliceosomal mRNA splicing are analogous (Villa et al., 2002; Toor et al., 2008). Yet, it is still not possible to be certain that their similarity represents a common evolutionary origin rather than convergent evolution.

Ribosomes

The structural conservation of ribosomal RNA in all three domains of life has long supported the argument that the ribosome is an ancient RNA machine. More recently, the crystal structure of the ribosome large subunit has cemented the view that the ribosome is fundamentally a peptidyl-transferase ribozyme supported by proteins that contribute to the correct folding of the RNA structure and improve the efficiency and accuracy of translation (Noller, 2006). Because many of the ribosomal proteins are also evolutionarily conserved, the last common ancestor of extant life already synthesized proteins using an RNP ribosome.

Since the adoption of this ancient ribosome, both protein and RNA components have continued to evolve. Many bacterial ribosomal proteins do not have homologs in Archaea, and although a ribosomal protein may occupy the same position in the structure, detailed analysis reveals convergent evolution from different starting points (Klein et al., 2004). Some of these less-conserved ribosomal proteins are observed simply to fill in the cracks between RNA structural elements, presumably stabilizing the folded structure (Brodersen et al., 2002; Klein et al., 2004). The ribosome contains multiple examples of glycine-, lysine-, and arginine-rich protein extensions that penetrate deep into the ribosomal RNA and assume specific structures only upon interaction with the RNA, providing credence to the idea that primordial ribosomes and other ribozymes could have benefited from even rather short basic peptides produced by random condensation of amino acids. RNA-binding peptides synthesized by a primitive ribosome could enhance its assembly, structure, and function, and thereby its evolution (Noller, 2006). In terms of the rRNA, the phylogenetically conserved core structure has been elaborated by the addition of internal "expansion segments" in eukaryotic nuclear-encoded rRNAs and the loss of structural elements in mitochondrial rRNAs.

Signal Recognition Particle

The signal recognition particle (SRP) binds to the signal sequence of secreted proteins and directs them to the endoplasmic reticulum of eukaryotic cells or to bacterial plasma membranes. SRP is an RNP in all three domains of life; a portion of the RNA structure and one of the protein subunits (a GTPase called SRP54 or Ffh) are conserved, thereby establishing SRP as an ancient RNP enzyme (Poritz et al., 1990). The SRP of Bacteria is the most primitive, consisting of just a 4.5S RNA and the GTPase. Other SRPs have additional protein subunits and larger RNAs. Some of the well-studied roles of the RNA subunit include its direct contribution to signal sequence recognition and its facilitation of the binding of SRP to its receptor; but these are the sorts of activities that proteins are fully capable of performing. The most fundamental reasons for the maintenance of RNA in this RNP enzyme instead may lie in the ability of RNA to span large distances, to impart a special sort of conformational flexibility (bending at internal loops and bulges between helices), and to bind to the ribosome by RNA-RNA interactions.

Clearly, the SRP is a very different sort of RNP enzyme than the ribosome. The SRP contains an extended structural RNA and a protein enzyme, whereas the ribosome is a ribozyme supported by structural proteins. Making matters even more interesting, the SRP that targets light-harvesting chlorophyll proteins to the thylakoid membrane in plant chloroplasts contains a homolog of the SRP54 GTPase, but no RNA (Schuenemann et al., 1998). This system provides a paradigm for evolution from the RNP world to the world of protein enzymes.

Telomerase

Telomerase, which adds telomeric DNA repeats to chromosome ends, is composed of essential RNA and protein

subunits. Biochemical studies have demonstrated that telomerase is not a ribozyme with an essential protein stabilizer, nor is it simply a protein reverse transcriptase that copies a portion of its own internal RNA template. Instead, catalysis depends on an intimate collaboration of proteins and RNAs, making telomerase a true RNP enzyme (e.g., Miller and Collins, 2002; Qiao and Cech, 2008).

Furthermore, it seems incorrect to think of telomerase as an RNP enzyme whose RNA is progressively "shrinking" in terms of size or importance. For example, yeast telomerase RNAs have long protein-binding arms that bring accessory subunits such as Est1, Ku, and the Sm proteins into the complex (Zappulla and Cech, 2004). It is parsimonious to conceive of these arms as having been elaborated in yeast, as they have not been found in other branches of the Eukarya. Vertebrate telomerase RNAs, on the other hand, have added an RNA domain found in small nucleolar RNPs that contributes to their biogenesis.

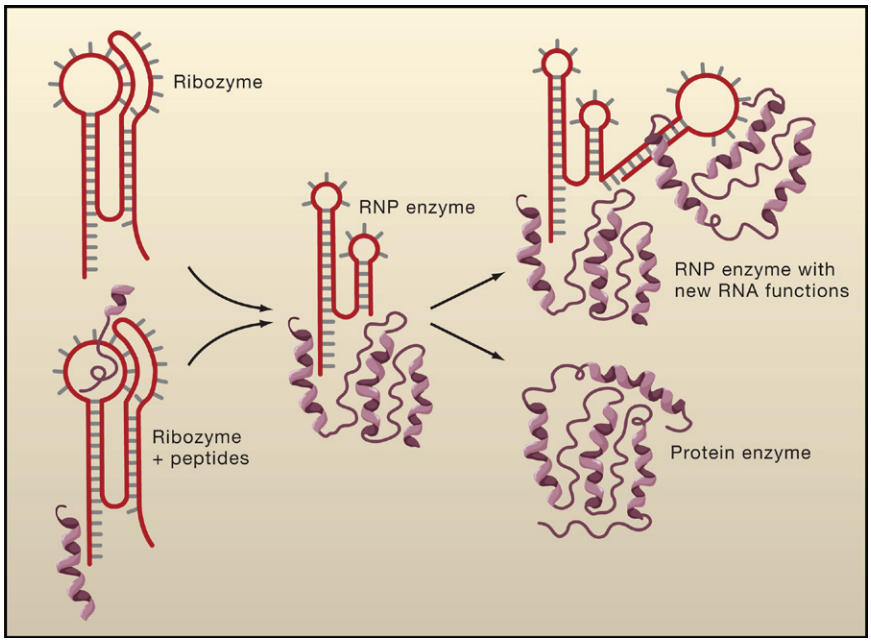

Figure 1. Model for the Evolution of Biological Catalysis

(Upper left) Primordial ribozymes were capable of purely RNA catalysis, but (lower left) given random peptides in their environment, selection for fitness may have occurred at the level of the RNP.
(Middle) A more advanced stage in which systems reproducibly synthesized peptides and polypeptides that enhanced RNP stability and function, ultimately including translation of mRNAs on ribosomes.
(Upper right) RNP enzymes have continued to evolve by adding (and subtracting) RNA elements that either contribute directly at the RNA level or bind to additional protein subunits. (Lower right) In most cases, catalysis has been taken over by protein enzymes.

Model for Evolution of Biocatalysis

The new experimental findings described above provide the basis for an alternative to "the standard assumption" that enzymatic activity evolved by the transfer of catalysis from RNA → RNP → protein (Poole et al., 1998). Instead, as shown in Figure 1, RNA self-replication and ribozyme catalysis in a primordial RNA world might have already had to deal with, and in some cases benefited from, peptides of random sequence and mixed chirality generated in their environment. In other words, although RNA alone is clearly capable of catalyzing reactions, it may not have had the opportunity to act as a pure ribozyme if peptides and other small molecules were present. Peptides and other small molecules binding to RNA would expand its structural repertoire, and thereby its functional repertoire (Noller, 2004).

Why do we think that any environment rich in nucleic acid precursors would necessarily contain amino acids and peptides? The landmark prebiotic simulation experiments of Miller and Urey readily produced high concentrations of natural amino acids. However, the highly reducing environment used in those experiments is now thought to be improbable in the prebiotic world. More recent studies have shown substantial production of amino acids from nonreducing mixtures that included N_2 and CO_2, especially if oxidation is inhibited (Cleaves et al., 2008).

Assuming that short peptides of random sequence were constituents of the first "RNP world," then the key evolutionary breakthrough (Figure 1, middle) would have involved the reproducible synthesis of useful peptides. Noncoded synthesis of specific peptides could be catalyzed by ribozymes, analogous to the action of modern enzymes such as D-Ala-D-Ala ligase involved in bacterial cell wall peptidoglycan synthesis. Even more powerful would be message-encoded peptide synthesis on primitive ribosomes composed of RNA but already stabilized and enhanced by short peptides in their environment (Noller, 2004).

The next evolutionary steps can be sketched with more confidence because we see evidence for them in modern biology. RNP enzymes can add new RNA elements, which either contribute directly to catalytic function or bring new proteins into the complex (Figure 1, upper right). Examples are seen in RNase P, SRP, and telomerase. In other cases, where there is insufficient advantage to having RNA in the catalyst, protein enzymes take over the function (Figure 1, lower right). The major incentive for evolution to move toward protein enzymes may be the greater functional repertoire provided by 20 amino acids, which contribute hydrophobic and hydrophilic side chains and groups well suited to general acid-base catalysis. Other factors may also contribute to proteins replacing RNAs for catalysis. The incentive for evolution to retain some RNP enzymes is presumably that there are some things RNA does very well, and perhaps more easily than proteins. These include recognition of RNA substrates by intermolecular base pairing and the ability to orchestrate large-scale movements. Thus, protein enzymes and RNP enzymes are likely to continue to share the job of biological catalysis for the foreseeable future.

Future Research Prospects

There are excellent opportunities for exploring the present world of RNP enzymes and, in the process, gleaning clues about their history. We have a very incomplete picture of the separate functions of RNA and proteins in known RNPs; perhaps the most interesting situations occur where the collaboration is so intimate that the RNA and protein functions are not separable. Increased understanding will come from determination of molecular structures coupled with mechanistic biochemistry. Furthermore, although studying the same RNP in multiple species can be a thankless task, it remains a powerful approach for inferring evolutionary history. On another front, new large noncoding RNAs are being identified and shown to be involved in transcriptional regulation in mammals (for example, Rinn et al., 2007; Wang et al., 2008; see Review by C.P. Ponting, P.L. Oliver, and W. Reik in this issue of *Cell*). There may be several mechanistic classes; some of these RNAs may bind to chromatin-modifying enzymes and thereby qualify as RNP enzymes, others may act as nonenzymatic RNPs, and still others may exert their function at the RNA level, for example, as antisense RNAs. The process of silencing of one of the two X chromosomes in female mammals by *Xist* RNA remains incompletely understood. It will be exciting to determine whether its mechanism resembles that of some of the newly discovered noncoding RNAs that silence gene expression more locally.

ACKNOWLEDGMENTS

I thank R. Batey, A. Lambowitz, H. Noller, N. Pace, A. Pyle, and J. Szostak for helpful discussions.

REFERENCES

Atsumi, S., Ikawa, Y., Shirasishi, H., and Inoue, T. (2001). EMBO J. *20*, 5453–5460.

Brodersen, D.E., Clemons, W.M., Carter, A.P., Wimberly, B.T., and Ramakrishnan, V. (2002). J. Mol. Biol. *316*, 725–768.

Cleaves, H.J., Chambers, J.H., Lazcano, A., Miller, S.L., and Bada, J.L. (2008). Orig. Life Evol. Biosph. *32*, 105–115.

Evans, D., Marquez, S.M., and Pace, N.R. (2006). Trends Biochem. Sci. *31*, 333–341.

Helm, M., Petermeier, M., Ge, B., Fiammengo, R., and Jaeschke, A. (2005). J. Am. Chem. Soc. *127*, 10492–10493.

Herschlag, D., and Cech, T.R. (1990). Biochemistry *29*, 10159–10171.

Holzmann, J., Frank, P., Loffler, E., Bennett, K.L., Gerner, C., and Rossmanith, W. (2008). Cell *135*, 462–474.

Kashiwagi, N., Furuta, H., and Ikawa, Y. (2007). Nucleic Acids Symp. Ser. *51*, 387–388.

Klein, D.J., Moore, P.B., and Steitz, T.A. (2004). J. Mol. Biol. *340*, 141–177.

Miller, M.C., and Collins, K. (2002). Proc. Natl. Acad. Sci. USA *99*, 6585–6590.

Noller, H.F. (2004). RNA *10*, 1833–1837.

Noller, H.F. (2006). Evolution of ribosomes and translation from an RNA World. In The RNA World, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 287–307.

Paukstelis, P.J., Chen, J.H., Chase, E., Lambowitz, A.M., and Golden, B.L. (2008). Nature *451*, 94–97.

Poritz, M.A., Bernstein, H.D., Strub, K., Zopf, D., Wilhelm, H., and Walter, P. (1990). Science *250*, 1111–1117.

Poole, A.M., Jeffares, D.C., and Penny, D. (1998). J. Mol. Evol. *46*, 1–17.

Qiao, F., and Cech, T.R. (2008). Nat. Struct. Mol. Biol. *15*, 634–640.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Jelms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y. (2007). Cell *129*, 1311–1323.

Schuenemann, D., Gupta, S., Persello-Cartieaux, F., Klimyuk, V.I., Jones, J.D., Nussaume, L., and Hoffman, N.E. (1998). Proc. Natl. Acad. Sci. USA *95*, 10312–10316.

Toor, N., Hausner, G., and Zimmerly, S. (2001). RNA *7*, 1142–1152.

Toor, N., Rajashankar, K., Keating, K.S., and Pyle, A.M. (2008). Nat. Struct. Mol. Biol. *15*, 1221–1222.

Villa, T., Pleiss, J.A., and Guthrie, C. (2002). Cell *109*, 149–152.

Walker, S.C., and Engelke, D.R. (2006). Crit. Rev. Biochem. Mol. Biol. *41*, 77–102.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Nature *454*, 126–130.

Zappulla, D., and Cech, T.R. (2004). Proc. Natl. Acad. Sci. USA *101*, 10024–10029.