RNA catalysis—is that it?

TIMOTHY J. WILSON and DAVID M.J. LILLEY

Cancer Research UK Nucleic Acid Structure Research Group, MSI/WTB Complex, The University of Dundee, Dundee DD1 5EH, United Kingdom

RNA catalysis

Ribozymes are RNA molecules that accelerate chemical reactions, enzymes that happen to be made of RNA rather than protein. Aside from the chemical challenge of trying to explain how a nucleic acid can do that, why should anyone care? Although these species created a lot of interest when the *RNA* journal began 20 years ago, they have slipped out of fashion more recently, and some may think RNA catalysis is only a niche activity restricted to a few slimy things in ponds. Yet this is far from the truth. Two of the cell's most important reactions are catalyzed by RNA. The condensation of amino acids in the peptidyl transferase center of the ribosome (arguably THE most important reaction in the cell!) is catalyzed not by protein, but by the major RNA component of the large subunit. Similarly, the splicing of mRNA in eukaryotes is catalyzed by the U2-U6 snRNA. So mechanistically both the ribosome and the spliceosome are ribozymes. RNase P processes tRNA in all domains of life, yet it is the RNA component that carries out the cleavage reaction. Splicing of eukaryotic mRNA is mechanistically closely related to the group II intron ribozymes, and both these and the group I introns are widespread. The nucleolytic ribozymes are a diverse group of smaller RNA species that generate site-specific cleavage and ligation of RNA, frequently to process replication intermediates, or to control gene expression. Three examples (hammerhead, HDV, and twister ribozymes) have been found to be widely dispersed in genome sequences. Some are expressed, and their positions suggest genetic control functions, although the biological function is unclear in some cases and will require in depth investigation. Not all members of the class are widely dispersed however. The Varkud satellite ribozyme (on which the authors have invested considerable time and intellectual effort!) is the largest of the group, yet was isolated just once from a *Neurospora* isolate from a small village in India. Nevertheless, this ribozyme is now probably the best understood example of a ribozyme operating by a nucleobase-mediated general acid-base catalytic mechanism (see below), so this effort was not wasted.

While the pace of discovery of new ribozymes had slowed to about one per decade for a while, this has recently accelerated greatly, with the Breaker lab developing a bioinformatic pipeline that is generating a steady flow of new ribozymes.

Another reason why RNA catalysis is important lies in its postulated importance in the early development of life on the planet. It is the key to the RNA world hypothesis proposed by Crick, Orgel, and others that potentially solves a massive chicken-and-egg problem that results from the requirement for the coincident emergence of nucleic acids and proteins with a division of labor between genetic encoding and catalysis. While no-one can rewind the tape to prove such a scenario, the discovery of the peptidyl transferase ribozyme is the closest we get to a "smoking gun" proof of this concept. A viable RNA world would have required many more chemical activities than are represented by the currently-known ribozymes, and it is conceivable that "molecular fossils" still exist. We expand on this idea below.

How does RNA catalyze reactions?

It may seem remarkable that a nucleic acid can function as an enzyme. Its components occupy a small fraction of the chemical space of a protein. Yet ribozymes can accelerate reactions a million fold or more, providing a challenge to the mechanistic chemist to explain how this is possible. With the exception of peptidyl transferase, all known natural ribozymes catalyze phosphoryl transfer reactions, either transesterification or hydrolysis of phosphate esters. Protein enzymes have evolved two solutions to the catalysis of phosphoryl transfer, and the ribozymes exhibit a similar divergence of strategy.

In general RNA and DNA polymerases and many nucleases act as metalloenzymes, using hydrated metal ions (usually Mg^{2+}) to position reactants, activate nucleophiles, and stabilize transition states. The self-splicing introns adopt a similar strategy. X-ray crystallography has provided the structures of both group I from the Golden and Strobel labs and group II intron from the Pyle lab ribozymes, with metal ions observed

bound at the catalytic centers by inner-sphere-coordination. While structural studies are highly instructive of course, we cannot observe transition states by crystallography; only chemistry can report on these species that by definition have a fleeting existence. An in-depth dissection of the transition state of the group I intron by the Herschlag and Piccirilli labs combining metal ion exchange with atomic substitution of putative ligands showed that three metal ions are bound to the transition state. Long-standing evidence indicates a role for metal ion binding to the U6 snRNA stem–loop in spliceosomal catalysis, consistent with a strong connection with the group II intron ribozyme. RNase P also seems to act as a metalloenzyme.

The second catalytic strategy used by proteins is general acid-base catalysis. RNase A uses the imidazole side chains of two histidines in the cleavage reaction, one as a general base (in its deprotonated form) to remove the proton from the O2$'$ nucleophile and a second as a general acid (in its protonated form) to protonate the oxyanion leaving group. Much the same strategy is used by the nucleolytic ribozymes to carry out essentially the same reaction (and its reverse, ligation). The structures of the six main nucleolytic ribozymes are now known, and the last decade has seen detailed mechanistic investigation. A common theme is the use of nucleobases as general acid and/or base, yet within the group the detailed mechanism of catalysis and the functional groups involved are somewhat variable. Guanines are involved in the mechanisms of all the ribozymes except the HDV ribozyme. Work in our lab has shown that the VS and hairpin ribozymes use a combination of guanine and adenine nucleobases, and despite having very different overall architectures, the topological arrangement of the active site components is very similar, a probable case of convergent evolution. By contrast, the HDV ribozyme employs a cytosine nucleobase as a general acid in cleavage together with a hydrated metal ion to activate the nucleophile acting either as a general base or Lewis acid. In GlmS the general acid is provided by the glucosamine-6-phosphate ligand (see below). Lastly, for the hammerhead ribozyme, crystallographic and mechanistic evidence suggests that a 2$'$-hydroxyl group may act as the general acid. The overall similarities suggest that perhaps there are relatively few solutions to the problem of how to construct an efficient nucleolytic ribozyme, and evolution has converged upon similar endpoints. In each case it is likely that other processes contribute to the overall rate enhancement achieved, including transition state stabilization by charge neutralization (e.g., by metal ions, not necessarily site-bound) and hydrogen bonding—RNA catalysis is undoubtedly multifactorial.

The observed rates of cleavage of the nucleolytic ribozymes are typically of the order 1–10 per minute. Given that these are self-cleaving ribozymes this is probably as fast as they require, but compared to protein nucleases it is very slow. A significant limitation is imposed by the pK_a values of the nucleobases which, in contrast to histidine, can be relatively far from neutrality. For example, the VS ribozyme uses the combination of nucleobases of relatively low (adenine) and high (guanine) pK_a so that only a few ribozyme molecules in 10,000 are in the required state of protonation at a given moment. If we correct for this we find that the intrinsic rate of cleavage by active ribozyme molecules is comparable with that for RNase A.

Prospects for new ribozymes

Are there more ribozymes waiting to be discovered in cells? For a while it began to look as if a relatively small group of natural ribozymes might all have been identified, but the recent flurry of new nucleolytic ribozymes suggests there could still be many more to be found. While these new ribozymes may each be mechanistically distinct in one way or another, they are all carrying out essentially variations on the same transesterification and hydrolysis reactions. A viable RNA world would have required many more RNA-catalyzed activities, the most critical of which would have been a template-dependent RNA polymerase since evolution cannot begin to work until a means of replication (however crude) is in hand. While no natural RNA-based polymerase has been discovered, in vitro selection has been used to evolve such activities in the Bartel and Joyce labs. Holliger and coworkers have demonstrated RNA-based-templated synthesis for hundreds of nucleotides within an ice lattice. RNA species have also been selected in vitro with ligase activity.

Thus, phosphoryl transfer reactions seem well within the compass of RNA. But could RNA participate in a wider range of chemistry? Peptidyl transferase differs from the other natural ribozymes in that a nitrogen nucleophile attacks the sp^2-hybridized carbon atom, proving that the chemical repertoire of contemporary ribozymes can extend beyond phosphoryl transfer. A viable RNA world would have required the catalyzed synthesis of a range of metabolites, including the rather more-demanding formation of carbon-carbon bonds. This has been achieved by in vitro selection of RNA that catalyze Michael addition, aldol condensation, and Diels-Alder reactions, although the latter cycloaddition reaction probably requires little in the way of catalysis beyond the correct juxtaposition of the diene and dienophile. Protein enzymes can bring a variety of chemical resources to bear on catalysis, including imidazole (perfectly suited for general acid-base catalysis at neutral pH), a primary amine (that can be acylated, or can form a Schiff base as in the aldol condensation for example), carboxylate, primary alcohol, and sulfhydryl groups. RNA lacks most of these functionalities, being limited to the nucleobases (proficient in general acid-base catalysis as we have seen), phosphate and associated hydrated metal ions and the 2$'$OH (a secondary alcohol, less reactive than a primary alcohol such as serine). Yet even proteins frequently use "bolt-on" components to expand their range of chemistry in the form of coenzymes. Most contain heterocyclic bases, and some (e.g., SAM, NAD$^+$, FAD) are based on nucleosides.

Their ubiquitous occurrence in all cells suggests that they were present in the LUCA, and all considered this suggests that they could trace back to the RNA world itself, as previously separately proposed by Benner and Yarus, and thus could conceivably have formerly acted as coenzymes in conjunction with ribozymes. And perhaps they still might be doing so, if we but knew where to look.

Recruitment of small molecules

The structure of RNA lends itself very well to the selective binding of ligands with high affinity. In vitro selection has generated RNA species that will bind all manner of ligands with great selectivity. But the natural species that really drive home this point in nature are of course the riboswitches. These form in the untranslated regions of mRNA, to bind metabolites so as to bias a see-saw-like bistable conformation in order to control gene expression via transcriptional termination, translational initiation, mRNA splicing or stability. Riboswitches have been identified that bind a number of coenzymes (e.g., TPP, FMN, SAM, SAH, THF, Ado-cobalamine), purines (e.g., adenine, guanine, cyclic-di-GMP), amino acids (e.g., lysine, glycine), and other metabolites.

Perhaps it is not so much of a leap from binding a coenzyme to using it in chemical catalysis. Or indeed the reverse could also be true, with some riboswitches descended from former coenzyme-dependent ribozymes that have lost the catalytic function. One could imagine that a SAM riboswitch might be elaborated into a methyl transferase ribozyme using S-adenosyl methionine as methyl donor, perhaps in combination with an RNA related to a box C/D snoRNA. The specificity of ligand binding by RNA suggests that reactions could be carried out with great regio- and stereoselectivity. The GlmS riboswitch provides a precedent for the use of coenzymes by RNA, where the available evidence strongly suggests that the amine of the bound glucosamine-6-phosphate molecule acts as general acid to protonate the 5′-oxyanion leaving group in the cleavage of the mRNA. The participation of guanosine in the first step of the splicing reaction catalyzed by the Group I intron is another example of participation by a small molecule, although the bound molecule is a reactant rather than a catalytic component. Proof of principle of the concept is also offered by the in vitro selection of a DNA species that uses a bound histidine molecule as a participant in RNA cleavage by Breaker and coworkers. In addition to SAM, tetrahydrofolate (another riboswitch ligand) could also be used in single-carbon transfer reactions. In contemporary cells thymidylate synthase uses this coenzyme in the conversion of uracil to thymine, a reaction that would have been required as DNA became adopted as the genetic repository.

TPP riboswitches form the single-most abundant class of riboswitches (>2000 occurrences), having been found in bacteria and plants. Thiamine pyrophosphate is a very versatile coenzyme, involved in the formation and breakage of carbon-carbon bonds, as in transketolase for example. The key functional element is the central thiazole ring; the positively-charged nitrogen stabilizes an adjacent carbanion that is a powerful nucleophile, and can furthermore stabilize negative charge in a variety of adducts formed at that position. In both plant and bacterial TPP riboswitches a k-junction juxtaposes two loop-containing helices to create the ligand binding site, whereby the pyrimidine is bound to one, and the pyrophosphate to the other. The net effect is that the key functional thiazole ring is suspended in an open cleft on the major groove side of the RNA junction, in principle accessible to diffusion of molecules from the solvent, thus suggesting how a ribozyme using TPP as a coenzyme might interact with substrate molecules.

The widespread (∼1000 occurrences) FMN riboswitches bind riboflavin mononucleotide; this coenzyme can undergo one- and two-electron transfers and is therefore used in enzymatic redox reactions. Perhaps this too could conceivably function catalytically in an RNA context? Suga and colleagues have selected an RNA that uses NAD^+ and Zn^{2+} ions to oxidize benzyl alcohol linked to the RNA, and the reduction of the corresponding benzaldehyde using NADH. These experiments demonstrate that RNA can in principle perform reversible redox reactions using a bound cofactor. Interestingly Liu and coworkers have isolated natural NAD^+-linked RNA in bacteria.

Pyridoxal phosphate is an extremely versatile coenzyme that forms a Schiff base with amino acids to allow a series of chemical conversions to be catalyzed. Although there are no known pyridoxal riboswitches at the present time, the functionalized pyridine ring could undoubtedly be specifically bound in a suitable cavity within RNA structure.

Searching for new ribozyme activities

In principle it is clear that by recruiting coenzymes, RNA could potentially greatly expand its range of chemical catalysis. Such new ribozyme activities are undetected to date, but has anyone really looked? The search requires considerable insight, knowing what to look for, and where and how to look. This would not be easy, and could still be a fool's errand. It is not a simple matter to find a black cat in a dark room … especially if there is no cat! However a large fraction of transcribed RNA is of unknown function at the present time and who knows what might be lurking there.

The majority of natural ribozyme reactions operate in *cis*, but most protein enzymes carry out reactions on substrates in *trans*. That complicates the search greatly. In order to find potential *trans*-acting ribozymes some educated guesswork into their potential substrates is required in order to set up a suitable assay. The process might be simplified by an initial search for natural RNA sequences that bind a probable substrate, and thereafter looking for chemical transformation. Alternatively, as the amount of sequence data becomes ever

greater, particularly with the plethora of new long non-coding RNA species, perhaps bioinformatics can be used to search for orphan sequences sharing similarity with known riboswitches. A related approach is already generating a steady stream of new ribozymes in the Breaker lab.

It seems highly probable that chemically-diverse ribozymes existed during an RNA-world era, but whether or not any remnants of these species still exist in contemporary cells is not known at present. We hope that a serious effort will be made to seek such novel RNA species in the third decade of this journal, because the impact would be tremendous!

Acknowledgments