LESSONS FROM CATALYSIS BY RNA ENZYMES

DANIEL HERSCHLAG[*] AND RAGHUVIR SENGUPTA[*,†]

[*]*Departments of Biochemistry and Chemical Engineering and Chemistry, Stanford ChEM-H (Chemistry, Engineering, and Medicine for Human Health), Stanford University, Stanford, CA 94305, United States of America*

[†]*Beckman Center, 279 Campus Drive B400, Stanford, CA 94305, United States of America*

Chemical, historical, and evolutionary perspectives on research of RNA catalysis

We can set the stage for discussion of RNA catalysis by considering the inception of this field less than 25 years ago and asking the following question: Why was catalysis by RNA a surprise to the scientific community? At the time, 1982, there was strong opposition to Cech's conclusion that an RNA sequence, and not a protein, was responsible for cleaving and joining RNA regions in a process he referred to as 'self-splicing' (Figure 1A)[1]. The objections ranged from 'there must be a protein contaminant' to 'it's not a real enzyme because it doesn't carry out a multiple turnover reaction [2].' But the data were strong, the original group I self-splicing intron carried out self-splicing, the intron was readily converted into an RNA enzyme (or 'ribozyme;' Figure 1B)[3], and Pace and Altman demonstrated the following year that an RNA is the catalytic component of RNase P, a multi-turnover enzyme that processes tRNAs to allow them to function in protein synthesis [4].

While biological RNA catalysis was a surprising result, why was it so difficult to accept? From a chemical perspective it would seem reasonable that RNA could be a catalyst [5]. The field of biomimetic chemistry was in full swing, and there were examples of catalytic lipids (in the form of micelles) [6], and there were even examples of template-directed oligonucleotide synthesis [7, 8]. While 'real' or biological enzymes were much better catalysts, these results demonstrated that you didn't have to be a protein to perform catalysis, a straightforward point from a chemical perspective.

It seems that the reticence to view Cech's results on the face of the data arose from dogma, and in this case not just any dogma, but the 'Central Dogma of Biology:'

DNA (information) → RNA → protein (function)

Crick's model, especially when presented in 1958 [9], was enormously useful for describing and teaching biology, organizing knowledge, and framing much needed research for the coming decades. In it, RNA is an intermediary, carrying information from the code, DNA, to the protein synthesis machinery that assembles proteins to perform the encoded functions. However hard we try, our thoughts, opinions, and analyses are affected by what we already know –and what we already *think* we know, and so it is likely that scientists were influenced by the dogma of the day, as well as RNA's known, seemingly uninteresting role. Indeed, 'dogma' can take on lives of its own, and Crick later regretted using this appellation, noting that he didn't at the time know the word's definition [10].

It is informative to consider, in this environment, how three scientists, including Crick, independently proposed RNA as a biological catalyst, more than a decade prior to Cech's discovery, in three single-author papers published in the late 1960s in the Journal of Molecular Biology[11, 12, 13]. First, rather than accepting a dogma as fact, Crick, Orgel, and Woese were thinking about biology –how biological systems might operate and how they might have evolved. Second, whereas mRNA in the central dogma was (and is) often drawn as a squiggly line, tRNA was known to be structured and these authors recognized 'where there's structure, function can follow.' Their central idea, later dubbed the 'RNA World' by Wally Gilbert [14], stemmed from the need for information to be stored and that information to be copied, or replicated, for life to occur and the corresponding 'chicken and egg' problem of which came first, molecules for information storage or functional molecules to copy that information; RNA could solve this problem by serving both functions.

As implied above, this idea was not widely discussed until after Cech's 1982 discovery, though now the RNA World might be considered a dogma of its own. Nevertheless, learning from the insights of Crick, Orgel, and Woese, we were inspired to think further about the molecular features of RNA –in particular the ability to form highly stable local structure with a small number of monomeric units– how these features may have been particularly amenable to early life forms, and how these properties may have also set the stage for the later takeover by proteins as the predominant bio-catalysts in modern-day life on Earth [15, 16].

Progress in RNA catalysis research and contemporary questions

Given the above, it is not surprising that it took the field some time to demystify the behavior, properties, and catalytic mechanisms of RNA enzymes, but nevertheless, overall progress has been rather rapid. Key contributions came from developing pre-steady state kinetic approaches for RNA to build kinetic and thermodynamic frameworks that allowed individual reaction steps to be investigated [17, 18]; the extension of metal ion rescue that Cohn and Eckstein had used with protein enzymes to quantitative 'thermodynamic fingerprint analysis' that allowed individual metal ions could be functionally assessed among the sea of metal ions bound in the RNA ion atmosphere [19-25]; the identification of ribozymes that acted without a requirement for metal ions and thus the necessity to consider additional catalytic strategies [26-28]; an illuminating exposition on the ability of nucleic acid side chains to act in general acid-base catalysis despite their non-optimal pK_a values relative to protein side chains from Bevilacqua [29] ; and, also building on work from protein enzymology, a clear exposition of the factors that can contribute to RNA cleavage from Breaker [30]. With these foundational tools, a combination of structural studies and very clever chemical biology approaches to manipulate properties of groups potentially involved in catalysis has led to reasonable models for catalytic interactions for nearly all known ribozymes [31-34].

Consider the catalytic interactions and active site architecture depicted in Figure 1C-E for the group I ribozyme, which are supported by a substantial interplay of functional and structural results [34-36]. One Mg^{2+} ion activates the guanosine (G) nucleophile and the other Mg^{2+} ion stabilizes charge development on the leaving group oxygen, and both Mg^{2+} ions interact with a nonbridging phosphoryl oxygen atom, likely making favorable electrostatic interactions and providing a template for the transition state. The nucleophile and leaving groups are in binding sites, held in position by multiple interactions that position the substrates and greatly increase their reaction probability and the reaction's specificity. The metal ions are themselves held in place by RNA as well as substrate ligands, and networks of interactions to accomplish their positioning have been demarcated. Overall the reaction is catalyzed by ~10^{11} – 10^{14} (depending on the chosen comparison state), well within the range of catalysis by protein enzymes [17, 37].

The above and related research has placed RNA on roughly similar footing to protein enzymes. And these studies have also taught us much about RNA folding, dynamics, and structure, by using catalysis as a convenient and powerful readout. Here I focus on unique opportunity for deep insights provided by the discovery of catalytic RNA. Specifically, comparing and contrasting two distinct macromolecules that have been used by Nature to carry out catalysis with high rate enhancements and high specificity can help reveal what is common –and fundamental. Conversely, the differences help us better understand each macromolecule, its behaviors, capacities, and limitations. Indeed, our research has been

greatly enriched by these comparisons and will continue to be, and I briefly highlight important examples and current challenges below.

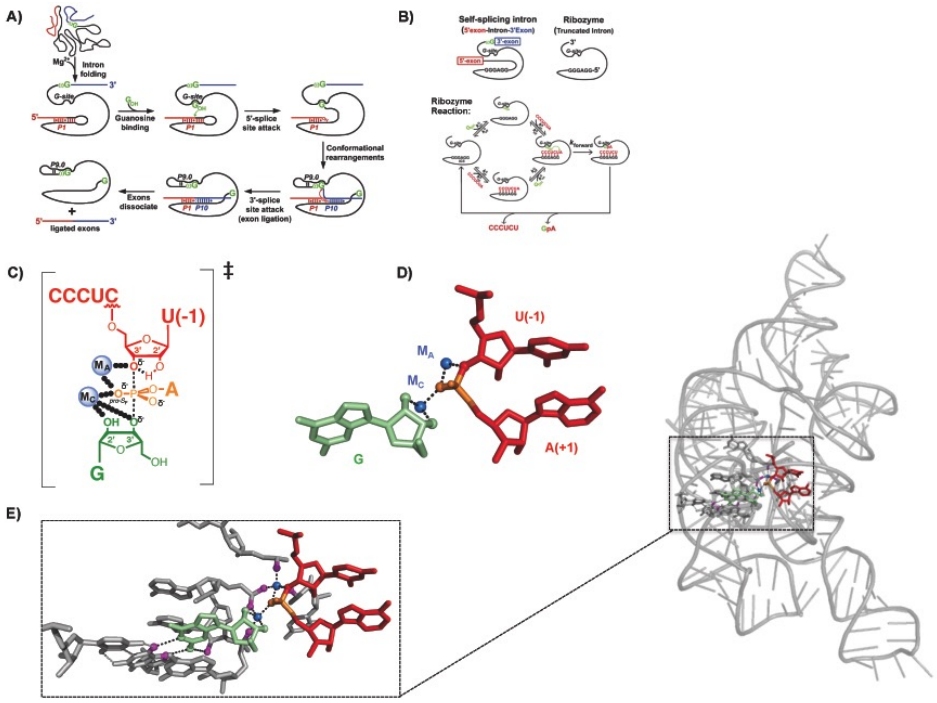

Figure 1. Group I intron catalysis. A) Cartoon representation of self-splicing. B) Comparison between the self-splicing group I intron and the truncated form ("ribozyme"). Images from (A) and (B) were taken with permission from [34]. C-E) Model of the group I ribozyme active site. C) Transition state model derived from biochemical and structural data (see [36] and references therein). Closed circles and hatched lines represent metal ion interactions and hydrogen bonds, respectively. Partial-negative charges in are represented by "δ^-." D) Model of the ground state E•S•G complex [36]. E) Model of the network of interactions within the active site of the E•S•G complex of the group I ribozyme [36]. Atoms highlighted in magenta contact M_A, M_C, and G. For The location of this network within the overall structure of the *Azoarcus* group I ribozyme is shown on the right.

Most broadly, RNA and protein enzymes share the ability to fold into distinct, globular structures with indentations or pockets that serve as binding sites for substrates. Groups involved in carrying out chemical catalysis, such as general acids and bases, are positioned near to where they function, greatly increasing the reaction probability, as are groups with charge complementarity to the reactions' transition states. These features fortify foundational principles of enzymology, and the examples below extend the value of these comparisons to deepen our conceptual understanding of enzyme energetics.

A major conundrum facing early investigators of enzymatic catalysis and energetics was the special ability of enzymes to accelerate reaction of just the 'right' substrate. It was particularly baffling that *smaller* substrates would be excluded –e.g., how is hydrolysis of activated compounds like ATP prevented and group transfer allowed? Jencks provided strong support for usage of binding interactions for catalysis and a coupling of their energetics –*i.e.*, reaction of the correct substrate is favored because it's binding is used for catalysis [38]. In many ways this concept is now self-evident –enzymes have specific binding sites and form Michaelis complexes with bound substrates. But the linkage to rate enhancement (*i.e.*, catalysis) and energetic consequences have been less clear to researchers and virtually absent from textbooks. Indeed, it has been common over the years to assign functions to enzyme residues in binding *or* catalysis rather than to seek to understand their interplay.

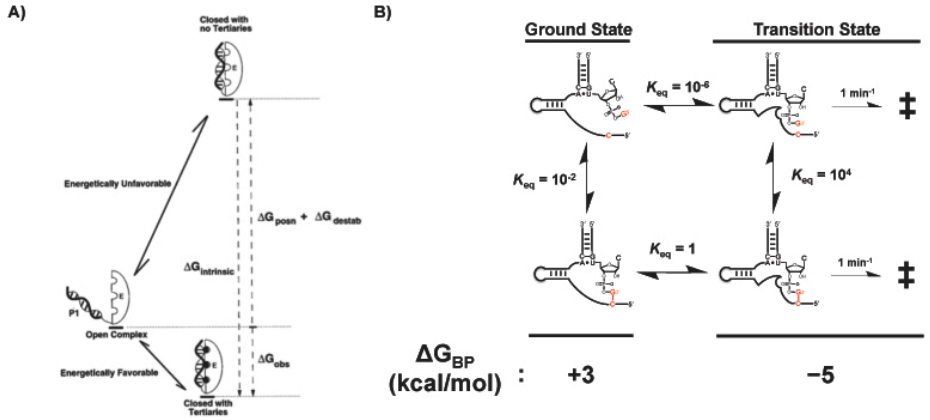

Figure 2. Use of intrinsic binding energy by the group I (A) and hammerhead (B) ribozymes. Figures taken from or adapted with permission from [5].

We have used the known energetics of base pair formation and the structural rigidity of RNA duplexes to probe the use of binding energy for catalysis with two RNA enzymes (Figure 2). For the group I ribozyme (Figure 2A), we were able to show that several substrate functional groups remote from the site of chemistry could contribute to catalysis within the Michaelis complex, thereby demonstrating a linkage of binding and catalysis. More directly, we could also show that those same functional groups, in the presence of a different constellation of functional groups elsewhere on the substrate helix, could instead contribute to binding [5, 39]. Thus, the same functional group can contribute to binding or catalysis, a dichotomy predicted by Jencks [38] and accounted for by a simple energetic model based on known properties of this ribozyme..

In a second example, we were able to show that a single canonical Watson-Crick base pair could contribute >5 kcal/mol to catalysis, whereas base pairs provide only 2-3 kcal/mol of binding stabilization [5, 40]. This difference again corresponds to 'intrinsic binding energy' as described by Jencks, where binding energy from base pair formation is used to align groups for reaction so that this binding energy does not show up (is not 'expressed') in a ground state complex but is expressed in the reaction's transition state. While Jencks presented compelling evidence for this concept for protein enzymes –e.g., the ability of remote side chain binding sites in elastase to aid catalysis and binding of transition state analogs but not affect substrate binding [38]– RNA's properties have allowed more direct demonstration of this energetic interplay.

Outlook to future developments of research on RNA catalysis

There are multiple exciting and important challenges in current ribozyme research. These are linked to evolutionary questions, including why RNA has remained active in certain roles in modern biology, and whether RNA is more or less adept at performing certain reactions and have these capabilities affected the course of evolution? On the functional side, can we use our growing knowledge of RNA folding and dynamics to understand and even predict RNA conformational states and functions? Further, are there additional comparisons between RNA and protein enzymes that can help evaluate mechanistic proposals? The fields of RNA and protein catalysis have, rightly so, focused on active sites; we are now ready for the next-generation question of how is an active site, RNA or protein, constructed, from its surrounding scaffold? We already have results that indicate that remote tertiary interactions have effects on distinct steps in the group I reaction [41], and new tools will allow us to deeply and comprehensively map and test these interactions and their origins. Thus, we are poised to enter a new structure-function era to elucidate the structural, dynamic, and energetic properties of RNA and protein enzymes, knowledge that will deepen our understanding of biology and enzyme function and our ability to manipulate and engineer biological molecules and systems.

References:

1.K. Kruger, P.J. Grabowski, A.J. Zaug, J. Sands, D.E. Gottschling, et al., *Cell* **31**, 147 (1982).
2. T.R. Cech, *Angew. Chem. Int. Edit.* **29**, 759 (1990).
3. A.J. Zaug, T.R. Cech, *Science* **231**, 470 (1986).
4. C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **35**, 849 (1983).
5. G.J. Narlikar, D. Herschlag. *Annu. Rev. Biochem.* **66**, 19 (1997).
6. J.H. Fendler, *Acc. Chem. Res.* **9**, 153 (1976).
7. R. Lohrmann, P.K. Bridson, L.E. Orgel, *Science* **208**, 1464 (1980).
8. T. Inoue, L.E. Orgel, *J. Am. Chem. Soc.* **103,** 7666 (1981).
9. F.H. Crick, *Symposia of the Society for Experimental Biology* **12**, 138 (1958).

10. H.F. Judson, *The Eighth Day of Creation: Makers of the Revolution in Biology. Expanded Ed.* (1996).
11. C. Woese. *The Genetic Code: The Molecular Basis for Genetic Expression.* (1967).
12. F.H.C. Crick, *J. Mol. Biol.* **38,** 367 (1968).
13. L.E. Orgel, *J. Mol. Biol.* **38**, 381 (1968).
14. W. Gilbert, *Nature* **319**, 618 (1986).
15. D. Herschlag, *J. Biol. Chem.* **270**, 20871 (1995).
16. Z. Tsuchihashi, M. Khosla, D. Herschlag, *Science* **262**, 99 (1993).
17. D. Herschlag, T.R. Cech, *Biochemistry* **29**, 10159 (1990).
18. K.J. Hertel, D. Herschlag, O.C. Uhlenbeck, *Biochemistry* **33**, 3374 (1994).
19. S.L. Wang, K. Karbstein, A. Peracchi, L. Beigelman, D. Herschlag, *Biochemistry* **38**, 14363 (1999).
20. S. Shan, A. Yoshida, S. Sun, J.A. Piccirilli, D. Herschlag, *Proc. Natl. Acad. Sci. U.S.A.* **96,** 12299 (1999).
21. S. Shan, A.V. Kravchuk, J.A. Piccirilli, D. Herschlag, *Biochemistry* **40**, 5161 (2001).
22. J.L. Hougland, A.V. Kravchuk, D. Herschlag, J.A. Piccirilli, *PLoS Biol.* **3**, e277 (2005).
23. M. Forconi, J. Lee, J.K. Lee, J.A. Piccirilli, D. Herschlag, *Biochemistry* **47**, 6883 (2008).
24. J.K. Fredericksen, N.S. Li, R. Das, D. Herschlag, J.A. Piccirilli, *RNA* **18**, 1123 (2012).
25. E.L. Christian, *Handbook of RNA Biochemistry* 319 (2005).
26. J.B. Murray, A.A. Seyhan, N.G. Walter, J.M. Burke, W.G. Scott, *Chemistry & Biology* **5**, 587 (1998).
27. S. Nakano, D.M. Chadalavada, P.C. Becilacqua, *Science* **287**, 1493 (2000).
28. J.A. Doudna, T.R. Cech, *Nature* **418**, 222 (2002).
29. P.C. Bevilacqua, *Biochemistry* **42**, 2259 (2003).
30. Y. F. Li, R.R. Breaker, *J. Am. Chem. Soc.* **121**, 5364 (1999).
31. M.J. Fedor, *Annu. Rev. Biophys.* **38**, 271 (2009).
32. D.M. Lilley, *Biochem. Soc. Trans.* **39**, 641 (2011).
33. M. Marcia, A.M. Pyle, *Cell* **151**, 497 (2012).
34. J. L. Hougland, J.A. Piccirilli, M. Forconi, J. Lee, D. Herschlag, *RNA World, Third Ed.* **43**, 133 (2006).
35. S.V. Lipchock, S.A. Strobel, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5699 (2008).
36. R.N. Sengupta, S.N. Van Schie, G. Giambasu, Q. Dai, J.D. Yesselman, *et al.*, *RNA* **22**, 32 (2016).
37. R. Wolfenden, M.J. Snider, *Acc. Chem. Res.* **34**, 938 (2001).
38. W.P. Jencks, *Adv. Enzymol. Relat. Areas Mol. Biol.* **43**, 219 (1975).
39. G.J. Narlikar, D. Herschlag, *Biochemistry* **37**, 9902 (1998).
40. K.J. Hertel, A. Peracchi, O.C. Uhlenbeck, D. Herschlag, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8497 (1997).
41. T.L. Benz-Moy, D. Herschlag, *Biochemistry* **50**, 8733 (2011).