In FOCUS

ACS chemical biology

What's So Great about RNA?

Andrew D. Ellington*

Department of Chemistry and Biochemistry, Institute for Cell and Molecular Biology, University of Texas at Austin, Austin, Texas 78712

RNA proved to be an amazing early building material for both origins and in modern biotechnology, primarily because of the great simplicity with which Watson–Crick base pairs can be engineered by evolution or humans. The remarkable feats that can be performed by RNA are shown off by recent work that adapts a synthetic riboswitch to function as a control element in bacterial chemotaxis. However, RNA is ultimately limited by its paucity of chemical groups, and thus it comes as no surprise that it finds a lasting role as a "bit" player that relies mainly on sequence recognition, both in nature and in engineered genetic circuitry.

Two parallel RNA worlds can be conceptualized: the ancient, hypothetical RNA world in which ribozymes were the primary organismal catalysts for metabolic reactions, and the more modern RNA world in which functional nucleic acids are engineered for biotechnology applications. The comparison of these two RNA worlds is in many ways instructive, primarily because of the question that they both beg: what's so great about RNA? Obviously, many of the functional RNA molecules in the ancient RNA world were eventually replaced by proteins after the advent of translation, so functional RNAs apparently didn't possess any enduring superiority. Given that we already have proteins in the modern world, what drives the use of RNA in engineering applications? Why not just use proteins? In both the ancient and modern worlds, the answer is similar: RNA is easy to use but far from optimal. Thus, the adoption of functional RNAs can be viewed as a simple stopgap on the way to more optimal molecules.

In a way, it is odd to question the potential of functional RNAs now. The utility of RNA as a regulatory molecule in both natural and synthetic circuits is just now becoming manifest. The number of microRNAs (miRNAs) and other small RNAs that regulate gene expression continues to grow, and it has been estimated that base-pairing interactions with miRNAs may control upwards of 30% of human genes (*1*). Similarly, Breaker and Tucker (*2*) have discovered a wide variety of so-called riboswitches that mediate the control of gene expression by small organic metabolites. The use of RNAs in gene regulation is particularly attractive, because functional sequences are short enough to arise *de novo* (*3*) and thus may be among the first tools to be used during adaptation or development.

A number of engineered examples of regulatory RNAs exist. Werstuck and Green (*4*) originally showed that ligand binding to an aptamer inserted into a 5′ UTR could lead to translational repression in *Escherichia coli* and eukaryotic tissue culture cells. The Group I self-splicing intron has been adapted to function as an allosteric ribozyme (or "aptazyme") that responds to that favorite proof-of-principle analyte, theophylline (*5*). Theophylline-regulated thymidine kinase splicing can act as a growth switch, leading thymidine-starved *E. coli* that are unable to grow in the absence of theophylline to produce thymidine in its presence. Bayer and Smolke (*6*) recently invented small, conformation-switching anti-

***Corresponding author, andy.ellington@mail.utexas.edu.**

Published online July 20, 2007

10.1021/cb700139t CCC: $37.00

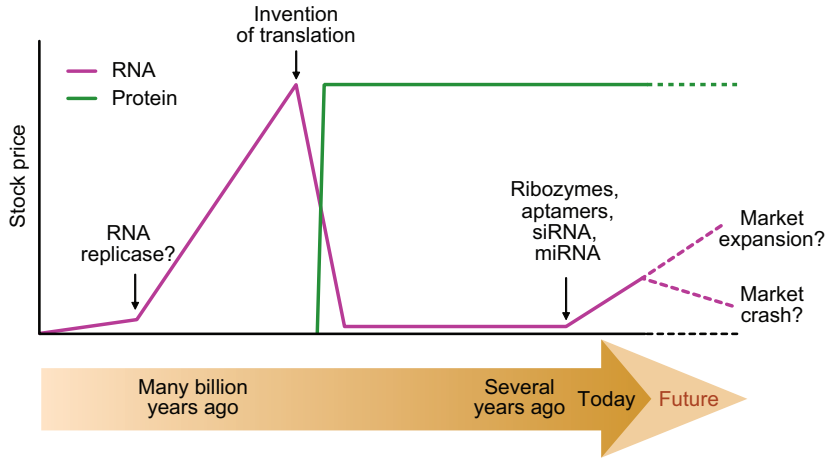

Figure 1. The changing value of RNA.

sense RNAs (antiswitches). When expressed in yeast, these RNAs respond to theophylline by extruding a short single-stranded region that can bind to and inhibit the translation of messenger RNAs (mRNAs). Selected ligand-binding RNAs have also been used to control transcriptional initiation (*7*), small interfering RNA (siRNA) processing (*8*), and mRNA splicing (*9*).

The recent work of Gallivan and coworkers (*13*) highlights RNA's potential as an engineering tool and its limitations. Building on the model originally set forth by Werstuck and Green (*4*) (and subsequently expanded upon by a number of other researchers, notably Grate and Wilson (*10*), Harvey *et al.* (*11*), and Suess *et al.* (*12*)), these researchers attempted to engineer theophylline-dependent repression of β-galactosidase. Paradoxically, upstream insertion of the aptamer led to theophylline-dependent activation of translation. Undaunted, Lynch *et al.* (*14*) used a genetic selection that improved the original 12-fold activation to 36-fold. The resultant riboswitch was then mounted in front of the CheZ gene, and this led to control of bacterial motility in the presence of theophylline (*15*). The engineered motility system differs from the normal control mechanism in a variety of ways. For example, regulation is based on relatively slow changes in gene expression rather than an enzymatic phosphorylation cascade, and the artificial chemotractant causes cells to stop tumbling at low concentrations rather than to swim toward the chemotractant through the interplay of sensing and reversing the polarity of the flagellar motor.

Although these examples are impressive and even arresting, much of their impact derives from the observation that we didn't previously recognize how useful RNA might be. The early days of biochemistry and molecular biology focused on proteins as the primary actors in metabolism, and RNA was relegated to a bit role as an information-carrying macromolecule. Crick (*16*), Orgel (*17*), and others suggested that there might be more to RNA than was apparent at first glance. However, it wasn't until Cech *et al.* (*18*) studied intron splicing in a somewhat obscure organism that the potential of RNA as a catalyst was truly appreciated (Figure 1).

It is true that RNA has been somewhat overlooked over the years, but a more dispassionate view suggests that this emphasis was not misplaced: protein catalysts and protein-mediated regulation of gene expression were first discovered, most thoroughly studied, and best exploited by both nature and biotechnology precisely because proteins are better functional molecules than RNA. Many proteins bind their ligands into the femtomolar to picomolar range; RNAs generally bind in the nanomolar to micromolar range. Proteins frequently catalyze reactions at up to thousands of turnovers per second; RNAs generally catalyze reactions at one turnover per minute. Proteins function in real-time signaling cascades; RNAs function in more slowly developing gene expression paradigms.

The question thus becomes: if organisms largely abandoned RNA as a catalyst and regulatory molecule many billions of years ago, why are we attempting to revive its biotechnological utility in organisms today? Why make RNA switches or circuits (a pastime we and others have frequently advocated (*19, 20*))?

The answer, of course, is that it is relatively easy to do so. RNA engineering is based largely on the fundamentals of Watson–Crick base-pairing, and thus RNA constructs that have some functionality can be readily designed by rational means. Examples of "engineering in Flatland", with only the 2D secondary structure of RNA molecules used as a guide, abound. Indeed, it has proven possible to design RNA-based switches that undergo programmed conformational changes *via* only computational methods (*21, 22*). Moreover, even in the absence of a complete knowledge of how secondary structure abets function, the advent of selection methods allows RNA molecules to be quickly optimized, including *in vivo* as Gallivan and coworkers have shown (*14*). Finally, between rational design of base pairing and irrational selection for function, nucleic acids are almost perfect modular parts for the construction of synthetic circuits. Stojanovic and Stefanovic (*23*) most clearly showed this by adapting deoxyribozymes to "play tic-tac-toe". In bacteria, *trans* RNAs have been engineered that base pair with one another and thereby activate gene expression, forming a modular signal transduction path based solely on RNA (*24*).

In one advanced view, such nucleic-acid-based switches, circuits, and automata may someday function as autonomous drug delivery devices (*25*).

Ultimately, though, even with these engineering advantages (which would have been as important billions of years ago in the course of natural selection as they are today in the laboratory), RNA just does not have the same potential that protein catalysts have. Perhaps the best cautionary tale in this regard involves one of the fastest known RNA catalysts, the Bartel class I ligase. This complex ribozyme was originally selected from a long, random sequence population (*26*), and it was found to be capable of catalyzing the turnover of multiple ligation substrates in *trans* at a rate of ~100 min^{-1}. Various attempts to further optimize the ligase for catalysis by directed evolution have succeeded in changing the sequence and purpose of the enzyme, but its speed has not been greatly increased (*27–29*). This is significant because the class I ligase is still orders of magnitudes slower than most protein enzymes. This comparison is somewhat daunting because the class I ligase is arguably the most complex RNA catalyst to ever emerge from an *in vitro* selection experiment, with an information content such that it probably should have only been found once every 10,000 times the experiment was run (*30*). The apparent optimality of this rare catalyst suggests that fast RNA catalysts are relatively rare in RNA sequence space, fast RNA catalysts are by and large isolated from one another in sequence space, and/or there cannot be fast RNA catalysts because of the physical limitations inherent in RNA chemistry.

Thus, "early adopter" advantages may exist when RNA is used as a substrate for cellular engineering. However, it really is only a matter of time until protein engineering catches up and displaces these advantages, just as ancient proteins displaced ancient ribozymes. In this regard, it is noteworthy that computational engineering of protein function (*31, 32*) and even allosteric activation (*33*) is becoming more commonplace, and thus any advantages of the ability to calculate nucleic acid secondary structures are dwindling. Also of note, emulsion technologies (*34*) are allowing protein populations to be sieved that are almost as large (10^{10} variants) as the nucleic acid populations typically plumbed by *in vitro* selection (10^{14} variants), and synthetic genetic circuits that revolve solely around protein regulatory components are being engineered with increasing speed (*35*).

This is not to say that all RNA engineering efforts are fruitless. Far from it. RNAs are proving to be remarkable drugs and may have definitive advantages in terms of delivery and pharmacokinetics (*36*). Nonetheless, although RNA parts and perhaps even RNA circuits may be useful for organismal engineering and synthetic biology applications, the fact is that predicting the functions of designed RNAs is frequently difficult (as Desai and Gallivan (*13*) found), and selected RNAs are often less robust than their protein counterparts. As we lurch toward refactoring entire genomes (*37*), it seems likely that RNA will again be relegated to its ancient and proper role as a bit player: a molecule that does not act as a machine but that instead holds and deciphers information based on the almost unique capability of the natural nucleobases to faithfully pair with one another.

Acknowledgment: This work was supported by a grant from the Welch Foundation (no. F-1654).

REFERENCES

1. Nilsen, W. (2007) Mechanisms of microRNA-mediated gene regulation in animal cells, *Trends Genet. 23*, 243–249.
2. Tucker, B. J., and Breaker, R. R. (2005) Riboswitches as versatile genetic control elements, *Curr. Opin. Struct. Biol. 15*, 342–348.
3. Salehi-Ashtiani, K., Luptak, A., Litovchick, A., and Szostak, J. W. (2006) A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene, *Science 313*, 1788–1792.
4. Werstuck, G., and Green, M. R. (1998) Controlling gene expression in living cells through small molecule–RNA interactions, *Science 282*, 296–298.
5. Thompson, K. M., Syrett, H. A., Knudsen, S. M., and Ellington, A. D. (2002) Group I aptazymes as genetic regulatory switches, *BMC Biotechnol. 2*, 21.
6. Bayer, T. S., and Smolke, C. D. (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression, *Nat. Biotechnol. 23*, 337–343.
7. Buskirk, A. R., Landrigan, A., and Liu, D. R. (2004) Engineering a ligand-dependent RNA transcriptional activator, *Chem. Biol. 11*, 1157–1163.
8. An, C.-I., Trinh, V. B., and Yokobayashi, Y. (2006) Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction, *RNA 12*, 710–716.
9. Kim, D.-S., Gusti, V., Pillai, S. G., and Gaur, R. K. (2006) An artificial riboswitch for controlling pre-mRNA splicing, *RNA 11*, 1667–1677.
10. Grate, D., and Wilson, C. (2001) Inducible regulation of the S. cerevisiae cell cycle mediated by an RNA aptamer-ligand complex, *Bioorg. Med. Chem. 9*, 2565–2570.
11. Harvey, I., Garneau, P., and Pelletier, J. (2002) Inhibition of translation by RNA–small molecule interactions, *RNA 8*, 452–463.
12. Suess, B., Hanson, S., Berens, C., Fink, B., Schroeder, R., and Hillen, W. (2003) Conditional gene expression by controlling translation with tetracycline-binding aptamers, *Nucleic Acids Res. 31*, 1853–1858.
13. Desai, S. K., and Gallivan, J. P. (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation, *J. Am. Chem. Soc. 126*, 13247–13254.
14. Lynch, S. A., Desai, S. K., and Sajja, H. K., J. P. (2007) A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function, *Chem. Biol. 14*, 173–184.
15. Topp, S., and Gallivan, J. P. (2007) Guiding bacteria with small molecules and RNA, *J. Am. Chem. Soc. 129*, 6807–6811.
16. Crick, F. H. (1968) The origin of the genetic code, *J. Mol. Biol. 38*, 367–379.
17. Orgel, L. E. (1968) Evolution of the genetic apparatus, *J. Mol. Biol. 38*, 381–393.
18. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena, *Cell 31*, 147–157.
19. Isaacs, F. J., Dwyer, D. J., and Collins, J. J. (2006) RNA synthetic biology, *Nat. Biotechnol. 24*, 545–554.
20. Davidson, E. A., and Ellington, A. D. (2007) Synthetic RNA circuits, *Nat. Chem. Biol. 3*, 23–28.
21. Penchovsky, R., and Breaker, R. R. (2005) Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes, *Nat. Biotechnol. 23*, 1424–1433.
22. Hall, B., Hesselberth, J. R., and Ellington, A. D. (2007) Computational selection of nucleic acid biosensors via a slip structure model, *Biosens. Bioelectron. 15*, 1939–1947.
23. Stojanovic, M. N., and Stefanovic, D. (2003) A deoxyribozyme-based molecular automaton, *Nat. Biotechnol. 21*, 1069–1074.

24. Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., and Collins, J. J. (2004) Engineered riboregulators enable post-transcriptional control of gene expression, *Nat. Biotechnol. 22*, 841–847.
25. Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., and Shapiro, E. (2004) An autonomous molecular computer for logical control of gene expression, *Nature 429*, 423–429.
26. Bartel, D. P., and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences, *Science 261*, 1411–1418.
27. Wright, M. C., and Joyce, G. F. (1997) Continuous in vitro evolution of catalytic function, *Science 276*, 614–617.
28. Vaish, N. K., Jadhav, V. R., Kossen, K., Pasko, C., Andrews, L. E., McSwiggen, J. A., Polisky, B., and Seiwert, S. D. (2003) Zeptomole detection of a viral nucleic acid using a target-activated ribozyme, *RNA 9*, 1058–1072.
29. Levy, M., Griswold, K. E., and Ellington, A. D. (2005) Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization, *RNA 11*, 1555–1562.
30. Ekland, E. H., Szostak, J. W., and Bartel, D. P. (1995) Structurally complex and highly active RNA ligases derived from random sequences, *Science 269*, 364–370.
31. Dwyer, M. A., Looger, L. L., and Hellinga, H. W. (2004) Computational design of a biologically active enzyme, *Science 304*, 1967–1971.
32. Ashworth, J., Havranek, J. J., Duarte, C. M., Sussman, D., Monnat, R. J., Stoddard, B. L., and Baker, D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity, *Nature 441*, 656–659.
33. Shulman, A. I., Larson, C., Mangelsdorf, D. J., and Ranganathan, R. (2004) Structural determinants of allosteric ligand activation in RXR heterodimers, *Cell 116*, 417–429.
34. Griffiths, A. D., and Tawfik, D. S. (2006) Miniaturising the laboratory in emulsion droplets, *Trends Biotechnol. 24*, 395–402.
35. Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy, M., Davidson, E. A., Scouras, A., Ellington, A. D., Marcotte, E. M., and Voigt, C. A. (2005) Synthetic biology: engineering Escherichia coli to see light, *Nature 438*, 441–442.
36. Que-Gewirth, N. S., and Sullenger, B. A. (2007) Gene therapy progress and prospects: RNA aptamers, *Gene Ther. 14*, 283–291.
37. Lartigue, C., Glass, J. I., Alperovich, N., Pieper, R., Parmar, P. P., Hutchinson, C. A., Smith, H. O., and Venter, J. C. (2007) Genome transplantation in bacteria: changing one species into another, *Science* [Online early access]. DOI: 10.1126/science.1144622. Published online: June 28, 2007.