# RNA Second Messengers and Riboswitches: Relics from the RNA World?

Some riboswitches and their ligands may be relics of signaling networks used when organisms relied on RNA instead of DNA and proteins

**Ronald R. Breaker**


## Summary

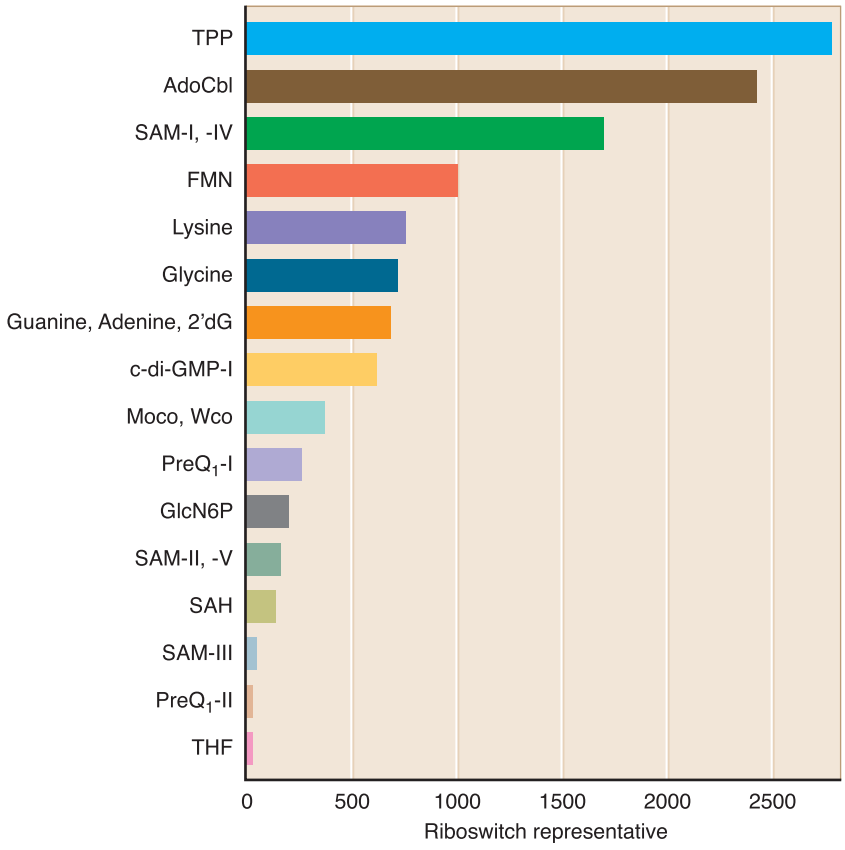
- Modern cells are widely believed to have descended from RNA World organisms that used functional RNA molecules to catalyze reactions and sense various metabolites.
- Riboswitches are structured RNA domains present in the mRNAs of many bacterial species that sense small molecules and control the expression of associated genes.
- Some modern riboswitches may have an RNA World heritage, and therefore may represent the types of structures and functions that were useful for ancient life forms.
- The bacterial second messenger c-di-GMP is sensed by riboswitches, and this partnership between a small RNA signaling molecule and a large RNA aptamer may constitute one of the earliest signaling networks.
- Additional second messenger and riboswitch interactions likely remain to be discovered, and these may expand our understanding of both modern and ancient forms of signaling with nucleic acids.


Compelling evidence supports the hypothesis that modern cells descended from organisms whose components were made entirely of RNA. Scientists who helped formulate this "RNA World" hypothesis for the evolution of life have identified many characteristics of modern cells that strongly suggest RNA once ruled the planet. For example, the existence of self-cleaving and self-splicing ribozymes that cut and ligate RNAs demonstrates that at least some RNA-processing reactions can be catalyzed without the need for proteins. Also, the existence of ribosomal RNAs that build all genetically encoded proteins emphasizes the fact that today's organisms depend on RNA to carry out fundamental biochemical processes. Even many nucleotide-like coenzymes that are near universal in all life forms likely first appeared in RNA World organisms. The RNA World hypothesis is appealing because it helps explain the arcane characteristics of genetic and biochemical processes in modern cells. Why do all cells use ribosomal RNAs, and not protein enzymes, to stitch together amino acids to make polypeptides? Most likely, the ribosome's peptidyl transferase center is an evolutionary descendent of an ancient ribozyme that first catalyzed this reaction before protein enzymes evolved. Why do so many coenzymes, such as adenosylcobalamin (AdoCbl), thiamin pyrophosphate (TPP), flavin mononucleotide (FMN), *S*-adenosylmethionine (SAM), molybdenum cofactor (Moco), and tetrahydrofolate (THF), all carry recognizable fragments of RNA nucleotides or derive from nucleotide precursors? Perhaps these compounds served as coenzymes for metabolic ribozymes of the RNA World, and they have been preserved as legacies from a time before protein enzymes existed.

The study of RNAs and RNA-derived small molecules may shed some light on the functions that RNA carried out before proteins emerged. We can use such insights to seek functional RNAs in modern cells that

*Ronald R. Breaker is Henry Ford II Professor in the Department of Molecular, Cellular and Developmental Biology; Department of Molecular Biophysics and Biochemistry; and is an Investigator with the Howard Hughes Medical Institute, Yale University, New Haven, Conn.*

**FIGURE 1**




Validated riboswitch classes are plotted in rank order relative to the number of known representatives. Riboswitch classes are named for the ligand they sense. The search for riboswitch representatives was conducted with RefSeq version 32, which has more than $5.1 \times 10^{-9}$ nucleotides and includes 764 bacteria whose genomes have been completely sequenced.

remain hidden, or to better harness the power of structured RNAs through molecular engineering efforts. Here I focus on the possibility that some natural metabolite-sensing RNAs called riboswitches may be primordial in origin. A newfound class of riboswitches that sense the bacterial second messenger cyclic-diguanylate-monophosphate (c-di-GMP) could be a modern manifestation of an RNA World signaling partnership.

## Riboswitches and the RNA World

Simple bacterial riboswitches are usually found in the 5′ untranslated region (UTR) of a messenger RNA, where they form a single RNA aptamer that selectively binds a target metabolite. Metabolite binding favors a folding change in the adjoining nucleotides, called the expression platform, which changes the amount of protein produced from the downstream open reading frame (ORF). Riboswitches usually exploit metabolite-regulated transcription termination or translation initiation to regulate ORF expression. However, bacteria (and some eukaryotes) also use other mechanisms such as mRNA self-cleavage, transcriptional interference, and alternative splicing.

There are numerous distinct classes of riboswitches (Fig. 1); the list of compounds that are sensed by these is nearly as long. What is it about a given riboswitch class that may imply an ancient origin? Certainly, not all expression platform mechanisms existed in a purely RNA World organism. Intrinsic transcription terminators, comprised of a strong stem followed by several U residues, interact with RNA polymerase to end transcription. However, protein-based RNA polymerase enzymes would not have existed until organisms began to produce polypeptides.

In contrast, the aptamers of many riboswitches are strikingly widespread and well conserved among numerous bacteria. Horizontal transfer could explain some of this distribution and apparent conservation, but it seems more likely that at least a few riboswitches were present very early in the evolutionary radiation of bacterial species. Indeed, TPP riboswitches are exceptionally common in bacteria, and representatives are found in at least some organisms from all three domains of life. This class is a candidate for an RNA World riboswitch, since its distribution suggests that it may have been present in the last common ancestor of all life.

TPP riboswitches also sense a coenzyme that may have been used by primordial ribozymes. All of the putative RNA World coenzymes have modern riboswitch partners that sense and respond to their changing cellular concentrations. None of these riboswitches is assuredly a direct descendant from the RNA World, but their existence in modern cells shows that RNA sequence-space could have offered sensing and

## Breaker: from Chicken Breeding to an Interest in Very Ancient Biology

As a youth, Ronald Breaker earned a dubious reputation from his informal efforts to breed chickens. "I knew I'd pushed my genetics tinkering too far when one year my father—seeing the season's young and very strange looking roosters stroll by—decided I needed to stop my experiments," he says. "Each new generation of roosters sprouted ever more interesting tufts of feathers on their bodies, and more elaborate combs on their heads. At the end of this evolutionary line was a rooster that was very large in size, visually terrifying, and [with] a nasty disposition that caused other small farm animals to flee from its path."

The line became extinct at his father's insistence, and by his pending departure for college. "Certainly I was doing what farmers have done for 10,000 years—although my selective breeding work was . . . not at all for increased crop yields or for improved domestication traits," Breaker says. Today, his studies are considerably less frightening, with potentially more valuable applications.

Breaker and his collaborators at Yale University, where he is a professor of Molecular, Cellular, and Developmental Biology and an investigator for the Howard Hughes Medical Institute, are studying natural and engineered nucleic acids. "In some ways, members of my laboratory are like molecular archaeologists," he says. "We are digging around the unexplored regions of genomes, looking for hidden treasures, or for clues about the nature of earth's earliest organisms. I . . . am intrigued by the likelihood that traces of our earliest ancestors are retained as useful components in modern cells."

Breaker, 45, was drawn to this research because of his longstanding interest in the origins of life. "One of the most compelling proposals, called the RNA World hypothesis, argues that the earliest life forms were made entirely of RNA," he says. "If . . . true, then RNA must be able to carry out far more biological functions than those we know exist from studies on modern organisms."

One goal is to determine whether RNA molecules act as molecular sensors and switches. "RNA World creatures must have had the ability to sense and respond to their surroundings, and so I felt this was not a high-risk project," he says. "When we validated the existence of the first riboswitches, we recognized that they control genes whose functions are sometimes essential for the survival of bacteria. Therefore, new forms of antibiotics can be created that target riboswitches and kill disease-causing bacteria." Recently, he and his research colleagues founded BioRelix, a New Haven biotech company that aims to develop just such antimicrobial compounds.

Breaker grew up in central Wisconsin, the only scientist among two brothers and a sister. His two brothers work in construction and manufacturing, and his sister is a traveling nurse who provides in-home care for the elderly. He comes from several generations of dairy farmers and lumberjacks. His father still owns the family farm, but gave up dairy herds to grow only crops. His mother recently retired from her job as a teller and loan officer for the local bank.

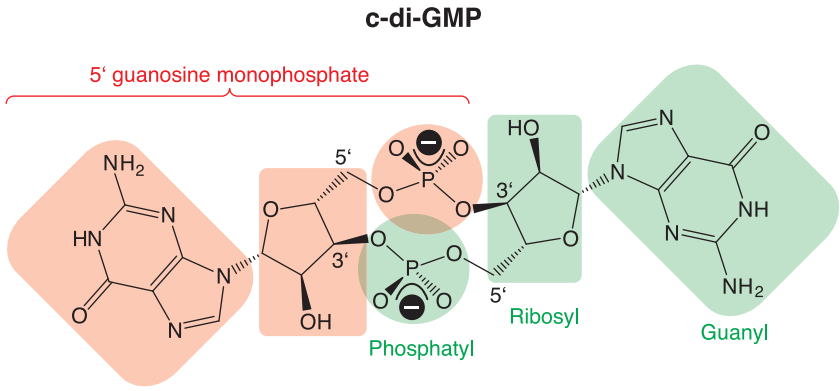
Breaker was a rare teenager who preferred farm work to partying. "I was surrounded by biology—the seasonal cycle of crop planting, growth and harvest, and the frequent births of calves or hatching chickens," he says. "As a high school student, I would sometimes volunteer to assist a cow giving birth, rather than spend a Saturday night in town. My friends put a stop to this by driving out to the farm and insisting I get in the car."

He received his B.S. degree in 1987 from the University of Wisconsin, Stevens Point, where his major was biology and his minor, chemistry. He earned a doctorate in biochemistry in 1992 from Purdue University, and did research as a postdoctoral fellow at the Scripps Research Institute in La Jolla, Calif. from 1992–1995. He joined the Yale faculty in 1995.

He and his wife Michelle, a math instructor at Gateway Community College, have two children. Their daughter is a freshman in the honors program at the University of Connecticut, with an interest in archaeology and biology. Their son is a high school sophomore, and a pitcher on the baseball team. Breaker also is interested in that sport, and helps coach Little League baseball. He also enjoys visits to Wisconsin to see friends, fish for trout, and pursue his interests as an amateur geologist. "I've begun to learn a little about the geology of glaciers that have left so much evidence of their existence in Wisconsin, including dropping billions of stones on farm fields—stones that we needed to remove by hand before each year's planting," he says.

**Marlene Cimons**
Marlene Cimons lives and writes in Bethesda, Md.

**FIGURE 2**




Chemical structure of c-di-GMP. The guanyl base, ribosyl and phosphatyl moieties for the two 5′-guanosine monophosphate substructures are similarly colored.

switching functions to the most primitive organisms. Moreover, not all presumably ancient compounds sensed by riboswitches are coenzymes, as exemplified by the curious case of c-di-GMP.

## Riboswitches That Sense a Bacterial Second Messenger

Bacteria usually do not come to mind when thinking of organisms with diverse signaling compounds and pathways. However, quorum sensing discoveries in recent years revealed numerous compounds that are involved in bacterial cell-to-cell communication. Moreover, nucleotide-based compounds such as cAMP and ppGpp serve as bacterial signaling compounds in a manner similar to eukaryotic second messengers. One of the most widespread bacterial second messengers, first encountered in the mid-1980s, is c-di-GMP (Fig. 2). Diguanylate cyclase enzymes use two GTP molecules and two phosphoester transfer reactions to form the circular RNA dinucleotide c-di-GMP linked via two 3′,5′-phosphodiester bonds.

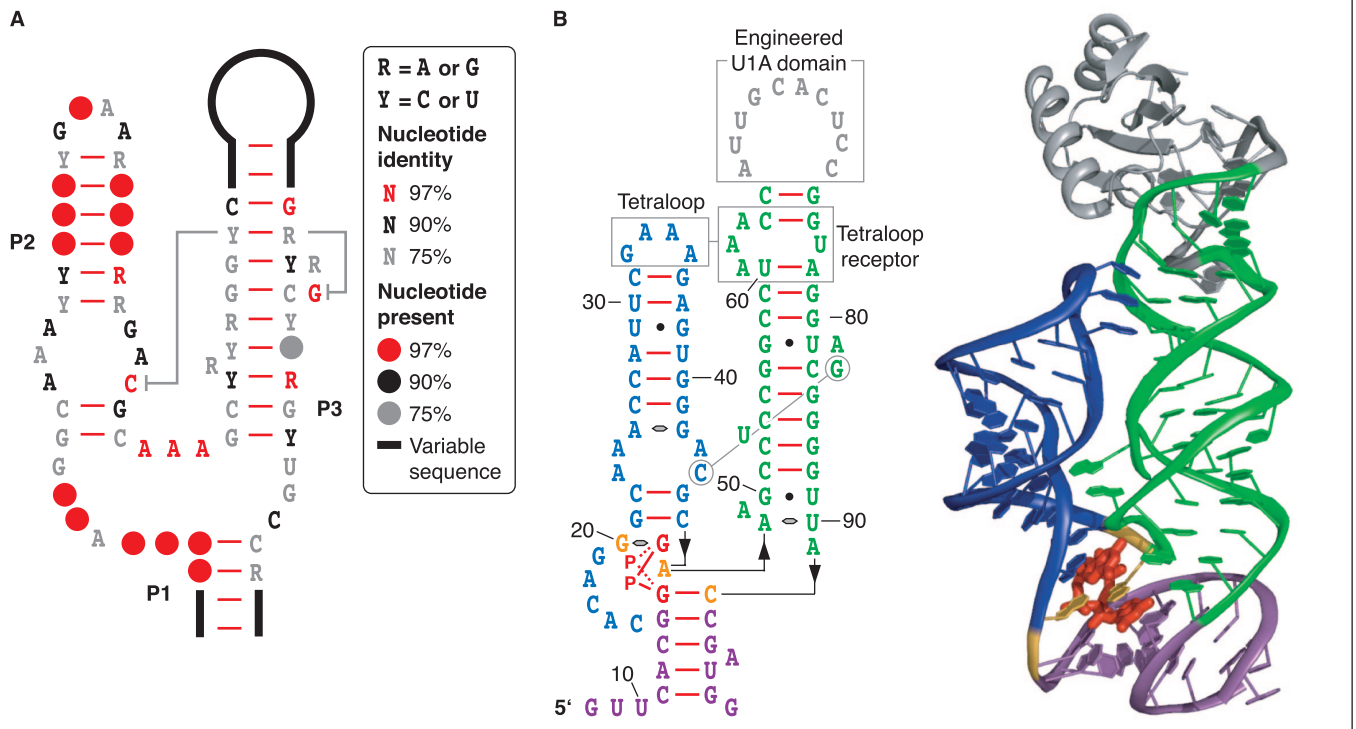
Homologs of the cyclase and nuclease enzymes can be found in a staggering diversity of both gram-positive and gram-negative bacteria, suggesting that c-di-GMP is an exceptionally widespread second messenger. This distribution implicates c-di-GMP as a key regulator of cellular physiology since a very early stage in bacterial evolution. The mechanisms by which this second messenger effects wide-ranging physiological changes remained puzzling for more than two decades. Although protein binding and regulation were identified, it was unclear how changing c-di-GMP concentrations controls genes. At least part of this mystery can be explained by the discovery of a new class of riboswitches that selectively binds c-di-GMP (Fig. 3).

As with many other riboswitches, the c-di-GMP aptamer domain has sequence and structural features that are highly conserved across a great diversity of bacteria (Fig. 3A). Each aptamer folds to form a three-stem junction, with conserved nucleotides sprinkled throughout the structure. An atomic-resolution structure (Fig. 3B) reveals that some of these conserved nucleotides serve as key components that hold the three-dimensional structure together. For example, the loop of stem P2 almost always forms a "GNRA" tetraloop (where R is either G or A), which commonly docks with a tetraloop receptor located near the loop of the adjacent P3 stem. These same two stems are also held closely by a surprising long-distance base pair formed by strictly conserved C and G nucleotides from bulges in the P2 and P3 stems, respectively.

These structural features reside at a distance from the c-di-GMP binding pocket, which is formed by conserved nucleotides in the regions that join the stems together. Since c-di-GMP is a circular version of an RNA dinucleotide, we speculated that riboswitch aptamers simply use Watson-Crick base pairing to recognize both G residues of the ligand. However, there were not two strictly conserved C residues that could base pair with c-di-GMP. The atomic-resolution structure again provided the answer—two seemingly imperfectly conserved nucleotides and a perfectly conserved A nucleotide form the highly selective binding pocket for c-di-GMP.

Watson-Crick base pairing does occur between one G of the ligand and a C residue of the aptamer (position 92 in Fig. 3B). Although the nucleotide at this position appears to be mutated in rare instances, in such cases a C residue usually resides at a nearby location. This C likely moves into the binding site to retain ligand-aptamer base pairing. The other G residue of the ligand is usually recognized by a nonstandard base pair with a G from the aptamer (position

FIGURE 3



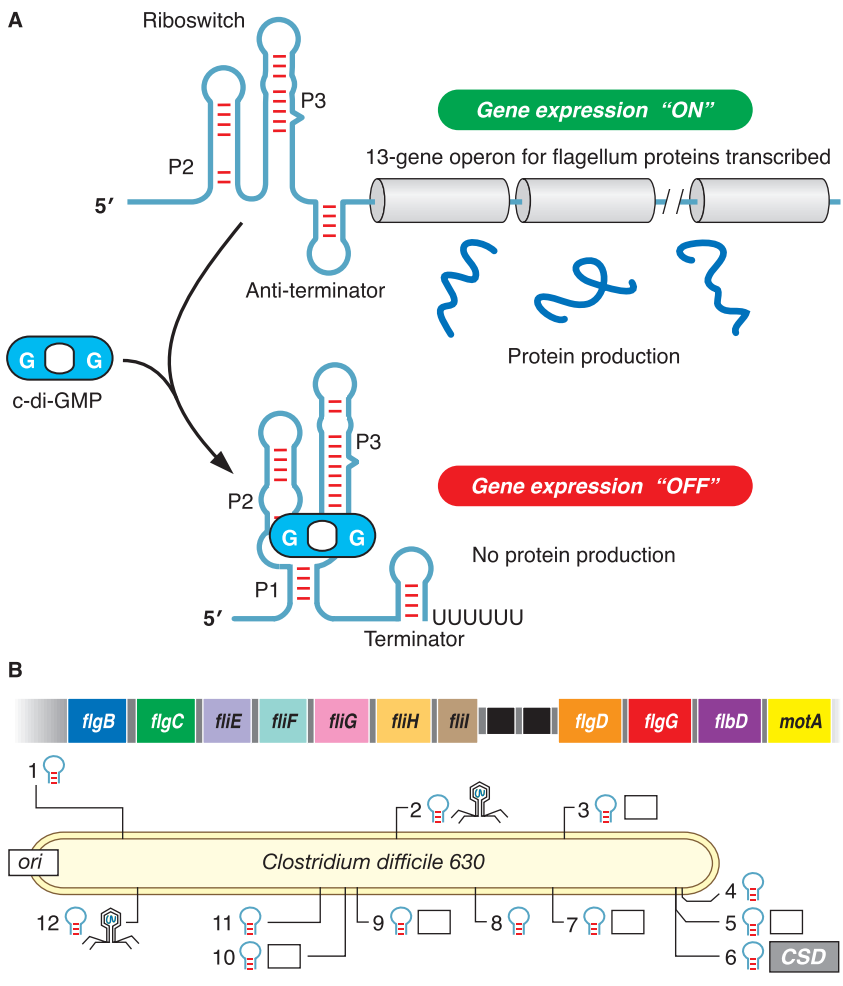

Sequence and structural features of c-di-GMP riboswitches. (A) Consensus sequence and structure of c-di-GMP riboswitches based on comparative sequence analysis and x-ray structure data. (B) Sequence and atomic-resolution structure model for Vc2 RNA, which represents one of two c-di-GMP riboswitches present in *Vibrio cholerae*. Note that the original sequence is engineered at the tip of P3 to carry a U1A protein binding site, which has been used to facilitate crystallization. G-U wobble base pairs are indicated with filled circles, and other non-standard base pairs are indicated with hexagons. The c-di-GMP ligand is depicted in red, and other nucleotides in the secondary structure model are colored to correspond to the similarly colored regions present in the tertiary structure model. Nucleotides are numbered according to the original nomenclature for Vc2. Tertiary structure model courtesy of Scott Strobel.

20 in Fig. 3B), although sometimes the aptamer nucleotide can be an A residue. The strictly conserved A residue (position 47 in Fig. 3B) intercalates between the two bases of c-di-GMP to help form a continuous base-stacking architecture involving aptamer and ligand nucleotides. Several additional ligand contacts are made by aptamer ribose residues or by divalent metal ions, which complete the local structure of the binding pocket. Such mixtures of familiar and unpredictable ways in which RNA can fold to grip a ligand are found with other riboswitch classes as well, indicating that RNA has a robust collection of folding tricks to form precision binding pockets for many different metabolites.

There are extensive differences in the structures of c-di-GMP aptamers when probed in the absence or presence of c-di-GMP. The atomic-resolution structural model reveals that c-di-GMP binding stabilizes an otherwise weak P1 stem, which is characteristic of many other riboswitch classes. Usually, some nucleotides involved in P1 formation could bind nucleotides in the expression platform to form alternative structures, such as terminator or antiterminator stems that influence gene expression.

The mutually exclusive formation of ligand-bound aptamer versus antiterminator structure is a mechanism proposed for a c-di-GMP riboswitch from the bacterial pathogen *Clostridium difficile* that terminates transcription when c-di-GMP levels are elevated (Fig. 4A). This riboswitch controls expression of a large operon that codes for many of the proteins needed to build flagellar organelles. The ligand-triggered transcription termination mechanism would help explain how increasing c-di-GMP concentra-

FIGURE 4




c-di-GMP riboswitches in *Clostridium difficile*. (A) Regulation of the 13-gene operon for flagellum biosynthesis by a c-di-GMP riboswitch. In the absence of second messenger, a portion of the aptamer forms an anti-terminator that permits mRNA transcription and expression of flagellar proteins. When c-di-GMP concentrations increase sufficiently, ligand binding by the aptamer prevents anti-terminator formation. Thus terminator stem formation causes transcription to halt before the coding regions of the mRNA are produced. (B) Map of 12 c-di-GMP riboswitches and associated genes in *C. difficile*. Open boxes designate genes of unknown function, and phage symbols indicate the riboswitch resides next to phage genes.

tions can be harnessed by *C. difficile* to change from a motile to a sessile life style.

*C. difficile* has at least 12 versions of this c-di-GMP riboswitch class encoded at various places in its genome (Fig. 4B). Since most of these riboswitches are associated with genes of unknown function, c-di-GMP can have multiple effects on bacterial physiology. For example, riboswitches 2 and 12 appear to be located closest to bacteriophage genes, which suggests that some bacteriophages can use riboswitches to read out c-di-GMP levels and spy on the physiological changes in its bacterial host. Also, riboswitches 4 and 11 appear to be unaffiliated with a coding region. Perhaps these unaffiliated riboswitches control the expression of noncoding RNAs that act *in trans* to regulate or participate in other biochemical processes.

## All RNA, All the Time

If this RNA ligand and RNA receptor partnership indeed is of ancient origin, predating even the emergence of DNA and protein in living systems, some RNA molecules would have had to function as c-di-GMP synthase ribozymes. Ribozyme-mediated biosynthesis of c-di-GMP from GTP, the same starting materials used by modern protein-based synthase enzymes, seems biochemically reasonable. Several research groups used test tube evolution methods to engineer ribozymes that form phosphodiester bonds by catalyzing the nucleophilic attack of a 3′ oxygen of an RNA on the triphosphate moiety from another RNA. Simply repeating this reaction twice with two GTP molecules as substrates would yield c-di-GMP.

Hypothetical RNA World organisms would also have found it useful to degrade c-di-GMP and reverse its biochemical effects on the cell, rather than simply disposing of it outside the cell. Model building suggests that c-di-GMP can be held in a geometry that would permit efficient nucleophilic attack by one of its 2′ oxygen atoms on the adjacent phosphorus center. This reaction will yield the linearized GpG>p dinucleotide (where the 3′ terminus is a 2′,3′-cyclic phosphate). To promote this reaction, a c-di-GMP phosphodiesterase ribozyme would need to hold the substrate in its reactive configuration, and perhaps hold a well-placed $Mg^{2+}$ ion or some other positively charged group, to accelerate c-di-GMP destruction to a speed that is biologically relevant.

The c-di-GMP aptamers examined to date do not alter the stability of c-di-GMP. How-

ever, RNAs that simply bind c-di-GMP have won half the battle. The X-ray model of the *Vibrio cholerae* aptamer reveals that a small shift in aptamer structure could produce the appropriate attack geometry; a nudge by a metal ion or some RNA functional group would yield a ribozyme that could selectively cleave c-di-GMP faster than spontaneous cleavage by a millionfold or more.

Thus, there appear to be no biochemical impediments to the existence of ribozymes that catalyze c-di-GMP formation and destruction. Even if such ribozymes never existed, it seems likely that RNA could be forced by skilled molecular engineers to carry out these reactions. Similar arguments could be made for the production and breakdown of other nucleotide-based signaling compounds in bacteria such as cAMP and ppGpp. Cells from the RNA World may have exploited ribozymes and receptors that participated in signaling networks based on these and other nucleotide second messengers, and their evolutionary descendants may await discovery in modern cells.

## Finding More Riboswitches for Second Messengers

Although only one riboswitch class for a second messenger has been identified, it seems plausible that additional riboswitch classes that sense bacterial second messengers will be found. Even though bacterial genomes are compact and the vast majority of their DNA codes for proteins, all sequenced genomes carry suspicious gaps. Some of these noncoding regions will only carry promoters and regulatory DNA domains that we have come to expect will play roles in gene regulation. Some other DNA gaps between genes might serve as templates for noncoding RNAs, including undiscovered riboswitches.

Comparative sequence analysis using bioinformatics has been a very effective strategy for riboswitch discovery. These algorithms seek out gaps between ORFs that carry conserved sequences and structural features. If the algorithm output presents an extensive consensus sequence with features characteristic of RNA secondary structure formation, then representatives may merit testing as riboswitch candidates. The best candidates tend to reside in the 5′ UTRs of genes from the same metabolic pathway, which makes guessing the ligand that triggers riboswitch action easier. In contrast, guessing the ligand becomes almost impossible if the candidates are distributed among a wide range of apparently unrelated genes, or genes of unknown function. Unfortunately, riboswitch candidates that bind second messengers are more likely to fit the latter category.

Another challenge is that not all these compounds have been identified. Second messengers that are narrowly distributed among bacterial species, or that control more obscure biological processes, may not be on the list of possible ligands to be tested for binding by newfound riboswitch candidates. The bacterial second messengers such as cAMP or ppGpp and related compounds are certainly widespread in bacteria and would be excellent candidates for detection by riboswitches. Since these compounds are also made from ribonucleotides and likely have a very old evolutionary origin, any riboswitches that sense these compounds also could be modern representatives of ancient signaling systems.

## Conclusions

Because the second messenger c-di-GMP is widespread in bacteria, it was likely present very early in the evolutionary diversification of bacteria. Furthermore, many bacteria sense this RNA signaling compound using a genetic switch composed entirely of RNA. Both these characteristics are precisely what one would expect if this signal and sensor system had its origin in an RNA World. If so, the structures and characteristics of c-di-GMP riboswitches provide information on the set of molecular signals used by some of the earliest forms of life.

However, there is much risk of error in such speculation. A circular RNA dinucleotide and its corresponding riboswitch RNA could have evolved long after proteins emerged on the scene. Some of the riboswitch RNAs we see today may be better suited to the roles they play than their protein-based counterparts. Riboswitches are small compared to the coding regions of protein genetic factors, are produced in the immediate vicinity of the genes they control, and directly interface with gene expression machinery. These and other features may be advantageous for some gene control tasks. Rather than a legacy system from the RNA World, c-di-GMP and its riboswitch sensor may be a product of modern organisms, one that is more

optimized for its roles than protein-based systems.

Regardless of whether c-di-GMP riboswitches are directly descended from RNA World components, or whether they are refined genetic switches with a more recent pedigree, they allow researchers to explore the true power of RNA as a functional polymer. Perhaps the discovery of new riboswitch classes that sense nucleotide-like second messengers will provide a richer view of how signaling circuitry looked in some of the earliest forms of life.

ACKNOWLEDGMENTS

RNA science in the Breaker laboratory is supported by grants from the NIH and by the Howard Hughes Medical Institute.

SUGGESTED READING

**Benner, S. A., Ellington, A. D., and Tauer, A.** 1989. Modern metabolism as a palimpsest of the RNA world. Proc. Natl. Acad. Sci. USA 86:7054–7058.

**Breaker, R. R.** 2006. Riboswitches and the RNA World, p. 89–107. In R. F. Gesteland, T. R. Cech, and J. F. Atkins (ed.): The RNA world. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, N.Y.

**Joyce, G. F.** 2002. The antiquity of RNA-based evolution. Nature 418:214–221.

**Romling, U., Gomelsky, M., and Galperin, M.Y.** 2005. C-di-GMP: the dawning of a novel bacterial signaling system. Mol. Microbiol. 57:629–639.

**Smith, K.D., Lipchock, S.V., Ames, T.D., Wang, J., Breaker, R.R., and Strobel, S.A.** 2009. Structural basis of ligand binding by a c-di-GMP riboswitch. Nat. Struct. Mol. Biol. (in press).

**Sudarsan, N., E. R. Lee, Z. Weinberg, R. H. Moy, J. N. Kim, K. H. Link, and R. R. Breaker.** 2008. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science 321: 411–413.

**Tamayo, R., J. T. Pratt, and A. Camilli.** 2007. Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. Annu. Rev. Microbiol. 61:131–148.

**Wilson, D. S., and J. W. Szostak.** 1999. In vitro selection of functional nucleic acids. Annu. Rev. Biochem. 68:611–647.

**White, H. B., III.** 1976. Coenzymes as fossils of an earlier metabolic state. J. Mol. Evol. 7:101–104.