

## EVOLUTION

## The lost language of the RNA World

James W. Nelson<sup>1\*</sup> and Ronald R. Breaker<sup>1,2,3†</sup>

The possibility of an RNA World is based on the notion that life on Earth passed through a primitive phase without proteins, a time when all genomes and enzymes were composed of ribonucleic acids. Numerous apparent vestiges of this ancient RNA World remain today, including many nucleotide-derived coenzymes, self-processing ribozymes, metabolite-binding riboswitches, and even ribosomes. Many of the most common signaling molecules and second messengers used by modern organisms are also formed from RNA nucleotides or their precursors. For example, nucleotide derivatives such as cAMP, ppGpp, and ZTP, as well as the cyclic dinucleotides c-di-GMP and c-di-AMP, are intimately involved in signaling diverse physiological or metabolic changes in bacteria and other organisms. We describe the potential diversity of this “lost language” of the RNA World and speculate on whether additional components of this ancient communication machinery might remain hidden though still very much relevant to modern cells.

## Introduction

Initial proposals (1–3) that modern forms of life might have emerged from an RNA World that predated the evolutionary emergence of proteins (4, 5) have fostered much speculation on the roles that RNA might have played in primitive cells (6–9). Some apparent remnants of the RNA World are components from which the most fundamental biochemical pathways of all extant organisms are built. Over the last 35 years, the experimental validation of a total of 14 natural ribozyme classes (10, 11) has demonstrated that RNA catalysts of ancient origin promote fundamental reactions involved in protein synthesis (12) and RNA processing (13). Furthermore, it has been proposed that many nucleotide-like coenzymes might be relics of an era when ribozymes catalyzed all the reactions necessary for life (1, 6, 7). Because natural ribozymes and nucleotide-derived coenzymes are intimately involved in central aspects of modern biological catalysis, it seems likely that a rich diversity of catalytic processes could have existed long before protein enzymes emerged.

Similarly, evidence from modern cells suggests that some of the most widespread signaling molecules, particularly those derived from RNA, are also ancient relics that have persisted since their first use by RNA World organisms. For example, numerous classes of riboswitches (14–19) sense and respond to fundamental metabolites, demonstrating that such RNA sensors and regulatory elements could have guided complex metabolic processes long before the emergence of proteins (17–19). Consistent with their proposed ancient origin, many of the most common riboswitch classes discovered to date selectively bind target ligands that are also derived from RNA nucleotides (Fig. 1A). Riboswitches and their ligands might therefore provide a means to look backward in evolutionary time and make predictions about the origins of signaling in modern cells.

There are now five riboswitch classes known to be responsive to bacterial signaling compounds that are directly made from RNA monomers or their biosynthetic intermediates. Specifically, there are two riboswitch classes that respond to c-di-GMP (3',5'-cyclic di-guanosine monophosphate) (20, 21), a class that responds to c-di-AMP (3',5'-

cyclic di-adenosine monophosphate) (22), a collection of variants of a c-di-GMP riboswitch class that recognize c-AMP-GMP (23, 24), and another class that binds the purine biosynthetic intermediate and alarmone ZTP (5-aminoimidazole-4-carboxamide riboside 5'-triphosphate) (25). The cyclic dinucleotide compounds alert cells to changes in metabolic status or environmental conditions that require the cells to make various physiological changes (20–27). In contrast, the alarmone ZTP, which is a derivative of a purine biosynthetic intermediate, signals the need to adapt to a deficiency in folate cofactor metabolism (25, 28). Although many other pairings between RNA-derived ligands and riboswitch sensors are known to exist (Fig. 1B), these other ligands also have direct metabolic roles and therefore are not produced by cells for the sole purpose of signaling.

The existence of true signal-receptor pairs, all built entirely from RNA, highlights the possibility that these and other RNA-derived molecules represent the descendants of ancient signaling systems used in the RNA World. Here, we provide an overview of the known RNA-derived signaling molecules in modern cells and speculate on how these and related molecules might have formed a sophisticated molecular language that allowed RNA World organisms to attain great complexity in both metabolic and physiological processes. Because modern cells were built upon this existing architecture, we discuss the prospects for finding additional remnants of this lost RNA World language among current forms of life.

## Communication in the RNA World

If the first simple self-replicating RNAs relied on abiotically produced nucleotide-like monomers, it seems self-evident that there would have been little need at this stage for signaling systems. However, as soon as the first biochemical pathways began to emerge, early life-forms would have found great utility in controlling these pathways to direct limited resources along specific paths and to regulate increasingly complex life cycles. To achieve such regulation, early RNA-based life-forms would have needed to form chemical signals from their limited RNA components and develop sophisticated RNA structures that could selectively recognize and respond to these signals. Can sufficient chemical and structural diversity be found among RNA building blocks and RNA polymers to perform complex signaling tasks?

The structural and functional complexity of ribozymes, like self-splicing RNAs (29, 30), the catalytic RNA core of ribosomes (12), and ribonuclease P (13), is striking. Undoubtedly, many other such complex RNAs were lost after protein enzymes arrived on the evolutionary

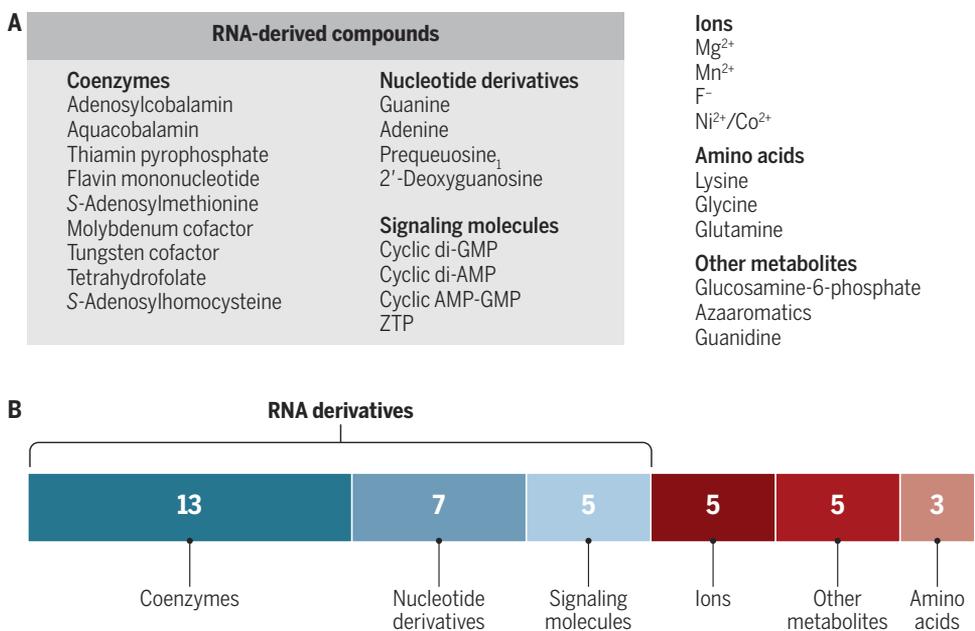
Copyright © 2017  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim  
to original U.S.  
Government Works.

Downloaded from <http://stke.sciencemag.org/> on October 31, 2020

<sup>1</sup>Department of Molecular, Cellular and Developmental Biology, Yale University, P.O. Box 208103, New Haven, CT 06520, USA. <sup>2</sup>Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208103, New Haven, CT 06520, USA. <sup>3</sup>Howard Hughes Medical Institute, Yale University, New Haven, CT 06520, USA.

\*Present address: Howard Hughes Medical Institute and Department of Chemistry and Chemical Biology, Harvard University, 13 Oxford Street, Cambridge, MA 02138, USA.

†Corresponding author. Email: [ronald.breaker@yale.edu](mailto:ronald.breaker@yale.edu)



**Fig. 1. Biological ligands for the known natural riboswitch classes are predominantly derived from RNA.** (A) Comprehensive list of ligands sensed by riboswitches, grouped according to general chemical types. (B) Chart depicting the total number of riboswitch classes organized according to their ligand types. Some compounds act as ligands for multiple riboswitch classes. For example, there are five distinct classes of riboswitches for *S*-adenosylmethionine: three for prequeuosine-1, two for 2'-deoxyguanosine, two for *c*-di-GMP, two for Mg<sup>2+</sup>, and three for guanidine.

scene. Before the arrival of proteins, RNA-mediated metabolic processes must have been very complex to have given rise to all the machinery needed to produce the diverse metabolites required to build RNA and then eventually to produce the building blocks for DNA and proteins. Thus, the complex RNA functions we see in modern cells must be just a small sampling of the sophisticated repertoire that existed in the RNA World. At its apex, just before synthesis of encoded proteins became the predominant route for building catalytic biopolymers, the RNA World would have had a great need for signaling systems involving small communication molecules and their corresponding molecular sensors. However, much of this complex RNA World machinery presumably has been lost, which makes it challenging to fully evaluate the functional potential of RNA.

What might RNA World signaling molecules and their sensors have looked like? Perhaps some of these early signaling molecules were retained by modern cells because they regulated fundamental processes that also have persisted. Retention of certain RNA World signaling molecules would be analogous to the persistence through evolution of some of the common ribozymes, coenzymes, and riboswitches. If true, then we can examine the known RNA-like signaling molecules present in existing organisms to gain understanding about the potential “words” of the language of the RNA World, as well as some of the mechanisms used to read this language.

### Examples of Modern RNA-based Signaling Molecules

Among the most widespread signaling molecules in biology today are the cyclic ribonucleotide *c*AMP (3',5'-cyclic adenosine monophosphate) (31, 32) and its guanosine analog *c*GMP (Fig. 2) (33, 34). Both compounds are essential signaling molecules in eukaryotes, and, whereas *c*GMP is used less commonly in eubacteria, *c*AMP is widely important

in Gram-negative bacteria (35, 36). Growing evidence suggests that the similar monophosphorylated cyclic ribonucleotides *c*MP (3',5'-cyclic cytosine monophosphate) and *c*UMP (3',5'-cyclic uracil monophosphate) might also serve as biological second messengers (37, 38). Formation of these cyclic phosphate products can easily proceed through nucleophilic attack of the 3'-oxygen atom on the  $\alpha$ -phosphorus of any high-energy nucleoside 5'-triphosphate (NTP; where N is any of the four common RNA bases). The starting material for biosynthesis of cNMPs is thus highly abundant in cells. Moreover, once the signal is no longer needed, cNMPs can be easily hydrolyzed to generate 5'-NMP, which can be phosphorylated to regenerate NTPs.

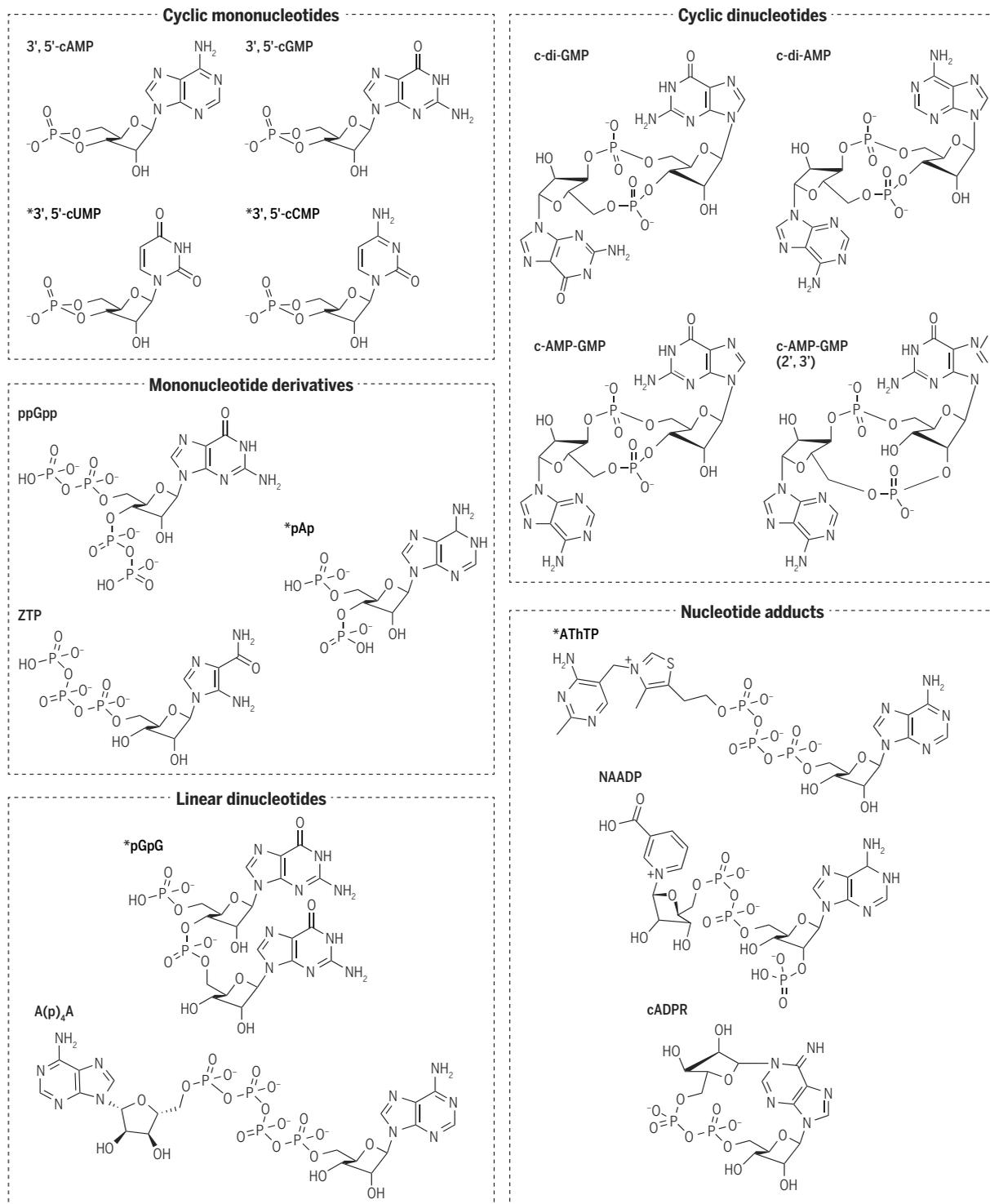
It is not surprising that biology has embraced such an economical means of producing and degrading signaling molecules. What is surprising is that, although the same economy can be had by exploiting the 2'-deoxy versions of these compounds, these signaling molecules are not widely used by cells for any purpose. This strongly suggests that cNMP signaling molecules predate the evolutionary emergence

of DNA. Thus, to meet their earliest signaling needs, RNA World creatures might have exploited simple derivatives of the very building blocks needed to produce RNA polymers.

Other more complex RNA-based signaling compounds are also widely used in biological systems. This is presumably because the 3',5'-cNMPs noted above allow only limited communications to take place, unless they are further derivatized. Modification of the base or ribose structures seems to be a chemically expensive process that would need to be carefully reversed, or else cells would generate various nucleotide modifications that could interfere with accurate genome replication. Although the biological relevance of the 2',3'-cNMP analogs as natural signaling compounds is being investigated (39, 40), these compounds might be less desirable to organisms than their more common 3',5'-cNMP isomers. 2',3'-cNMPs can be generated through the normal enzymatic or spontaneous breakdown of RNA polymers, which could lead to inappropriate signaling. Unless cells have a need to exploit these degradation products to report on the progression of RNA breakdown, it seems unlikely that 2',3'-cNMPs would be widely used as signaling molecules.

Perhaps to avoid such erroneous and potentially detrimental signaling events when expanding the diversity of signaling compounds, biology has further decorated mononucleotides with additional phosphates (Fig. 2). Again, the advantages of using these compounds as signaling molecules are clear. New phosphoester and phosphoanhydride linkages can easily be formed by nucleophilic attack on a phosphorus center. As noted earlier, such bonds are easily broken by hydrolysis to regenerate the starting material. Compounds such as ppGpp, also called guanosine-3',5'-bis(diphosphate) (41), and ZTP (25, 28) are widely used by bacteria as “alarmons”, or metabolite derivatives that signal adaptation to particular metabolic stresses.

Other compounds such as pAp (3'-phosphoadenosine-5'-phosphate) (42) and AThTP (the adenylated derivative of thiamin triphosphate) (43)



**Fig. 2. Known natural signaling compounds derived from RNA nucleotides or their precursors.** The compound ppGpp, which carries a pyrophosphate both at the 5' and 3' positions, is derived from a pentaphosphate precursor (pppGpp) that carries a triphosphate on the 5' position and is not shown. ZTP is AICA ribonucleoside 5'-triphosphate; AThTP is adenosine thiamine triphosphate. Asterisks denote putative signaling compounds for which biological roles and phylogenetic distributions have not been well established.

have been implicated in signaling processes, but their roles and the distribution of their use throughout the tree of life have not yet been well established. More detailed information is available for the signaling compounds cADPR [cyclic adenosine 5'-diphosphate (ADP)

ribose] (44) and NAADP (nicotinic acid adenine dinucleotide phosphate) (45), although the existence of these compounds as signaling molecules among the three domains of life is not yet clear. Regardless, some of these rarer compounds demonstrate how easily nature can

modify a single common nucleotide, adenosine, to create a diversity of possible signaling molecules.

Another simple mechanism to expand the language of RNA is to fuse two ribonucleotide monomers to create RNA dimers. This is manifested in the widespread signaling molecule A(p)<sub>4</sub>A (diadenosine tetraphosphate) (46, 47), which can be thought of simply as two ADP molecules joined by an extra phosphoanhydride linkage (Fig. 2). Considerable evidence suggests that this compound could be just one member of a great diversity of dinucleoside polyphosphate signaling molecules (48), each composed of two nucleotides of either identical or mixed nucleobase types joined through two or more phosphates spanning the 5'-oxygen atoms of the linked monomers. However, it is much more common for cells to exploit a simpler connection between two monomers made through the formation of a normal 3',5'-phosphodiester linkage, and this is the subject of the following two sections.

### Communicating with Linear and Cyclic Dinucleotides

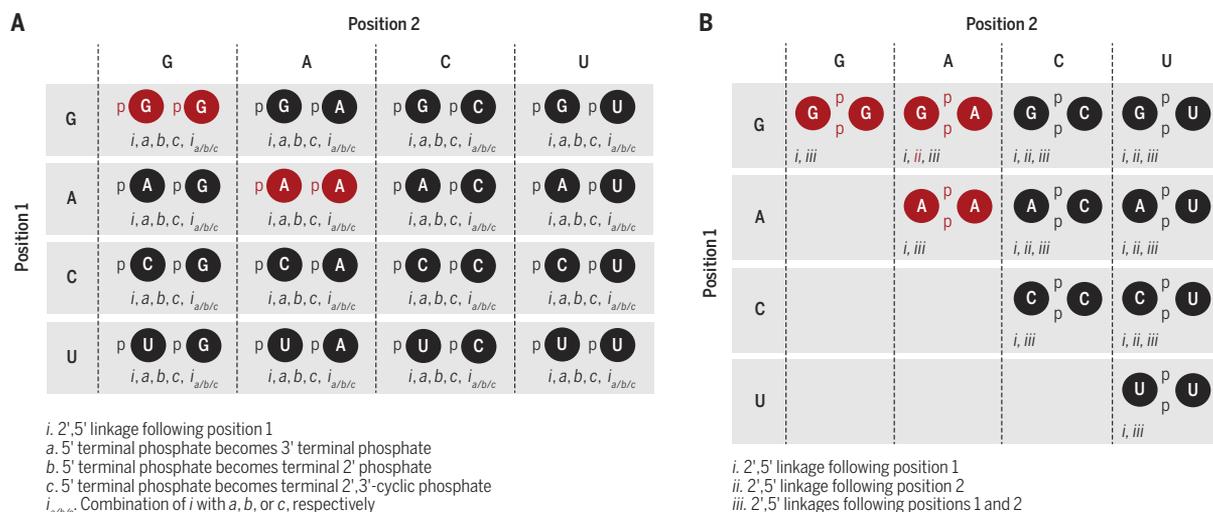
There are a tremendous number of RNA words that can be written using just one of the four standard nucleotides. However, a very simple mechanism to further expand this RNA language is to fuse two nucleotides by a 3',5'-phosphodiester linkage to create a combinatorial collection of all possible linear dinucleotides. Thus, the four words formed by GMP, AMP, CMP, and UMP can be expanded by fusing two monomers to create 16 additional distinct words (Fig. 3A).

Compared to chemically modifying the sugar or base, there is a considerable advantage to using unmodified dinucleotides to expand the number of RNA words. First, generating RNA phosphodiester linkages is a simple chemical reaction that can be easily achieved by both natural (49, 50) and engineered (51, 52) ribozymes. Thus, no additional evolution of specialized active sites for generating more challenging chemical linkages is required. Second, no novel ribozymes with complex functions are needed to reverse any chemical modifications placed on

monomers. The dinucleotide signaling molecules can simply be hydrolyzed to generate the starting monomers in unaltered form. Third, additional chemical variations could readily be made to expand the number of chemical words produced. For instance, simply by varying the connectivity of the phosphates at the 5' or 3' terminus or by varying the connectivity of the phosphodiester joining the two nucleotides, a total of 128 discrete RNA words can be generated by linear dinucleotides (Fig. 3A).

Despite these noted advantages, using linear 3',5'-linked dinucleotides can still be problematic. How can the host organism tell the difference between a dinucleotide signaling molecule and the random-sequence dinucleotides that might be generated through the enzymatic or spontaneous degradation of longer RNA polymers? Of course, RNA strand scission by direct hydrolysis or by internal phosphoester transfer, respectively, will generate distinctive 5'-monophosphate or 2',3'-cyclic monophosphate groups. Unless these chemical differences are exploited by their receptors, these RNA degradation mechanisms could generate dinucleotide products that could be confused with signaling dinucleotides that carry similar chemical features. Furthermore, 3',5'-linked dinucleotides might also be produced by RNA polymerase "false starts" during transcription. Therefore, a distinctive chemical configuration for signaling dinucleotides would be advantageous, unless the cell's only goal is to monitor these biosynthetic and degradative processes.

Cyclic dinucleotides, such as c-di-GMP (26) or c-di-AMP (27) (Fig. 2), offer an easy solution to avoid confusion with RNA degradation products. Chemical and enzymatic mechanisms for RNA degradation do not produce cyclic dinucleotides, so the presence of these compounds indicates that they have been purposefully synthesized for their signaling functions. Modern cells have dedicated synthase proteins to generate specific cyclic dinucleotides (53, 54). Joining two nucleotides by two 3',5'-phosphodiester linkages also allows these molecules to avoid degradation by exoribonucleases that otherwise would rapidly degrade RNAs with a free 5' or 3' terminus. Again, dedicated phosphodiesterase enzymes



**Fig. 3. The possible diversity of cyclic dinucleotide RNAs and their immediate degradation products.** (A) Matrix depicting all 16 possible linear dinucleotides that can be directly generated by degradation of cyclic dinucleotide RNAs. Compounds are schematically represented by depicting the nucleoside (circled letter) joined by phosphate backbone atoms (p). Annotations *i*, *a*, *b*, *c*, and *i<sub>a/b/c</sub>* represent all possible combinations of isomers with variation in the phosphodiester linkage connectivity and variation in the location and connectivity of the terminal phosphate. (B) Matrix depicting all 10 possible cyclic dinucleotide RNAs formed by the four common ribonucleotides joined by 3',5'-phosphodiester linkages. Annotations *i*, *ii*, and *iii* represent all possible combinations of derivatives wherein 2',5'-phosphodiester linkages are present. The compounds for which there is evidence of biological roles in modern cells are highlighted in red.

that selectively degrade cyclic dinucleotides have been identified (26, 54). The dedicated synthases and phosphodiesterases for cyclic dinucleotides create an isolated circuitry for biological signaling by these compounds.

However, there is one disadvantage to using cyclic dinucleotides. Upon circularization, some information content is lost because there is no distinction between the first and second nucleotide. For example, in a linear dinucleotide, pGpA (written from 5' to 3') is distinct from pApG. In contrast, the circular dinucleotide c-GMP-AMP is the same as c-AMP-GMP. Thus, there are only 10 distinct cyclic dinucleotides formed when joining the four common ribonucleotides (Fig. 3B). This diversity can again be augmented by varying the connectivity of the phosphodiester linkage. Although there are only two additional isomers that result for homodimers (a single 2',5' linkage, or two), there are three additional isomers for each mixed dimer (2',5' linkage after position one, after position two, or both). As a result, the potential number of words formed by cyclic dinucleotides expands from 10 to 36 (Fig. 3B).

### Nature's Exploitation of Dinucleotide Signaling Molecules

Cyclic dinucleotides are among the most widespread signaling molecules currently used by living systems. The first cyclic dinucleotide to be identified was c-di-GMP, which was found to function as an activator of cellulose synthase in *Acetobacter xylinum* 30 years ago (55). C-di-GMP has since been shown to be widely involved in the control of numerous fundamental bacterial processes, including biofilm formation, chemotaxis, differentiation, virulence, and antibiotic production (26). It is also found in eukaryotes, where it has been demonstrated to induce cell differentiation in *Dictyostelium discoideum* (56). Even in some organisms that do not produce c-di-GMP endogenously, such as humans, it is sensed, among other cyclic dinucleotides, as a marker for bacterial infection (57). In bacteria, c-di-GMP is extremely widespread, to the point of being nearly universally used.

It took almost 20 years for another cyclic dinucleotide to be identified in a biological system. The compound c-di-AMP is analogous to c-di-GMP, except that two AMP monomers are used to form the cyclic dinucleotide. It was serendipitously identified when it cocrystallized within the binding pocket of a protein later shown to function as a diadenylate cyclase (58). Although not as widespread as c-di-GMP, c-di-AMP is found throughout Gram-positive bacteria, where it is implicated in the control of sporulation and germination, as well as in certain Deltaproteobacteria, where it is involved in the response to osmotic shock (59). Also, c-di-AMP is involved in some central metabolic processes (60), indicating that it is intimately integrated into signaling metabolic changes in certain cells. Much more work remains to be done to fully flesh out the roles of this signaling molecule.

The remaining two cyclic dinucleotides discovered in biological systems to date are even more striking because they exploit some of the variations in nucleotide sequence and in linkage diversity as noted above (Fig. 3B) to generate distinct signaling compounds. The first to be identified, c-AMP-GMP, was found to be produced by the enzyme DncV in *Vibrio cholerae* and was implicated in the control of bacterial virulence (61). Also, c-AMP-GMP was identified in certain species of Deltaproteobacteria, where it is involved in the control of bacterial exoelectrogenesis, the export of electrons (23, 24). Although barely the first chapter has been written on the study of c-AMP-GMP, we have speculated that it is involved in the control of specialized adhesion processes that occur at the interface between a bacterial biofilm and the surface to which it adheres (24).

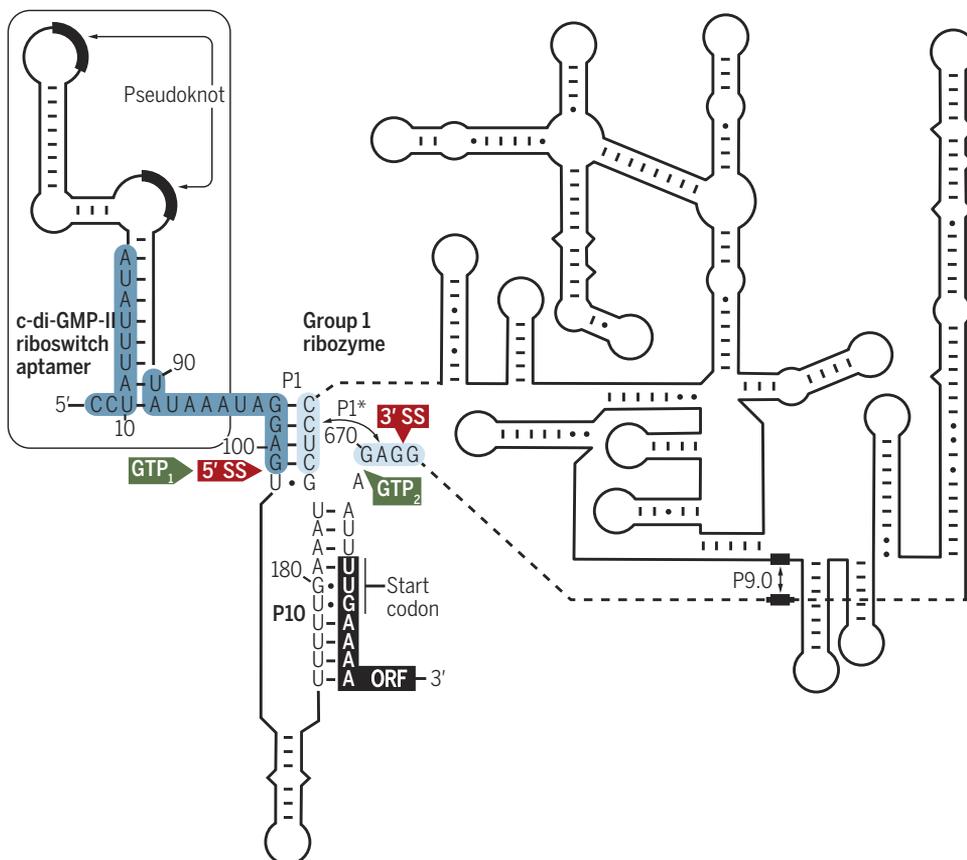
Several years after the discovery of c-AMP-GMP, a structural isomer of this molecule was discovered and named cGAMP (62, 63). Whereas all the other known naturally occurring cyclic dinucleotides are composed of two 3',5' linkages, cGAMP contains one 3',5' and one 2',5' linkage (64–66). Similarly, all the other cyclic dinucleotides are found in bacteria, but cGAMP is exclusively found in higher eukaryotes, where it is produced in response to the presence of foreign DNA (62, 63). Cyclic dinucleotides with mixed sequence (c-AMP-GMP) and mixed phosphodiester linkages (cGAMP) serve as modern examples of the greater diversity of RNA words harnessed by present cells that could also have been exploited billions of years ago in the RNA World. Finally, evidence is emerging for the biological roles of the linear dinucleotides pGpG and pApA (67–69). These compounds can be formed by the degradation of c-di-GMP and c-di-AMP, respectively, suggesting that the c-di-GMP and c-di-AMP signaling networks may extend to sensing of the linear degradation products of these cyclic nucleotides.

All of the cyclic dinucleotides known to be present in bacteria are detected by at least one class of riboswitch (20–24). Furthermore, these RNAs have extraordinarily high affinities for their cyclic dinucleotide ligands and have been shown to be easily tunable to recognize different ligands (23, 24, 70, 71), suggesting that evolution of RNA aptamers for these compounds in the RNA World was certainly possible. During that early evolutionary time, these compounds could have functioned to trigger or repress allosteric ribozymes, allowing for both genetic and metabolic control.

Rare natural examples of such cyclic dinucleotide-controlled allosteric ribozymes have been identified (21, 72). By searching bacterial genomes for riboswitch sequences that are located in proximity to sequences encoding ribozymes, we discovered a modern version of what we would expect to find in a real RNA World creature. Two functional RNA molecules, a riboswitch for c-di-GMP and a group I self-splicing ribozyme that uses GTP (guanosine 5'-triphosphate) as a substrate, are juxtaposed (Fig. 4) such that proper pre-mRNA (messenger RNA) splicing (21) and gene expression (72) require both c-di-GMP and GTP. When c-di-GMP is bound by the riboswitch aptamer, the adjoining ribozyme can access the 5' splice site and properly undergo splicing to yield a translation-ready mRNA. In contrast, when c-di-GMP is absent, the riboswitch sequesters the normal 5' splice site, leading to ribozyme-mediated processing at a site that removes nucleotides necessary to permit translation of the adjoining open reading frame. This mechanism of splicing regulation is similar to those observed for eukaryotic TPP (thiamin pyrophosphate) riboswitches (73–77), although splicing in those organisms is triggered by an apparent RNA World coenzyme (TPP) and catalyzed by the ribozyme core of spliceosomes (78).

### Why are RNA Signaling Molecules Predominantly Purine-based?

Curiously, whereas the second messengers cAMP and cGMP are frequently used by cells, scant evidence has been presented for the existence of pyrimidine-based signaling molecules, such as cCMP or cUMP (79). Moreover, c-di-GMP and c-di-AMP are widely exploited by bacteria as second messengers, but there are no known natural examples of homodimers or heterodimers that include C or U nucleotides. Pyrimidine nucleotides such as CTP (cytidine 5'-triphosphate) and UTP (uridine 5'-triphosphate) are present at high concentrations in cells and therefore could readily have been used as building blocks for pyrimidine-based signaling compounds. Why is there



**Fig. 4. A complex all-RNA signal-receptor system.** A natural allosteric ribozyme from *Clostridium difficile* uses a c-di-GMP-II riboswitch (21) to sense an RNA-based signaling molecule, whereupon ligand binding triggers proper self-splicing by a GTP-dependent group I ribozyme. Formation of the ribozyme P1 stem permits GTP (here designated GTP<sub>1</sub>) to attack the 5' splice site (5' SS). The newly formed 3'-hydroxyl group of G101 serves as a nucleophile to attack the 3' splice site (3' SS) that follows G667 of the ribozyme. Joining the exons creates a strong ribosome binding site that promotes translation of the adjacent open reading frame (ORF) (72). Without c-di-GMP, the riboswitch aptamer (boxed structure including the pseudoknot) and the ribozyme reorganize to permit the complementary regions highlighted in dark blue to base-pair, and the complementary regions shown in light blue to base-pair, which creates alternative stem P1\*. This promotes GTP (here designated GTP<sub>2</sub>) to attack the riboswitch after G670, which precludes proper splicing. The resulting GTP<sub>2</sub> attack product lacks nucleotides that otherwise could serve as a ribosome binding site, which prevents translation of the adjacent open reading frame. The solid line represents the RNA chain, with only specific nucleotides critical for the mechanism depicted. Dashed lines indicate zero-length connections.

such a strong bias in favor of purines in RNA-derived signaling molecules?

We speculate that this bias was likely established early in the RNA World and that this favoritism is based on the chemical properties of the compounds involved, rather than any accident of evolutionary history. In part, this could be due to the lower solubility of purines relative to pyrimidines, leading either to aggregation or to association with RNA polymers. A more formal way to view this advantage for purines is to recognize that they have greater potential for forming stronger binding interactions than pyrimidines, which increases the probability that receptors will have emerged to recognize compounds carrying these moieties. C-di-GMP is known to self-aggregate (80) and is sometimes bound as a dimeric complex (81, 82).

In contrast to the single ring of pyrimidines, purines contain a fused two-ring system, which makes stacking interactions with their biological receptors more favorable. This has been empirically observed by examining helical RNA (83, 84) and DNA (85, 86) struc-

tures that are made more stable by single dangling nucleotides that extend base stacking along the helical axis. Generally, in these studies, stacking strengths decline in the order of A > G > C ≈ U or T. Furthermore, the purines have more potential for forming hydrogen-bonding contacts compared to the pyrimidines. Again, this chemical property supports the preferential use of signaling molecules with guanine and adenine moieties simply because they can form more hydrogen bonds with receptor pockets compared to their pyrimidine competitors.

This bias in favor of purine-containing compounds has also been observed with substrates and ligands for engineered populations of deoxyribozymes and allosteric ribozymes. For example, an evolving population of self-phosphorylating deoxyribozymes adapted over generations of directed evolution to triphosphate substrates in the order GTP > ATP > CTP > UTP (87). Similarly, an evolving population of allosteric ribozymes adapted to recognize cyclic mononucleotides in the order cGMP > cCMP > cAMP (88). In this latter study, cUMP completely failed to be recognized by any members of the population, which began with one quadrillion ( $10^{15}$ ) distinct RNA sequences.

These test-tube evolution experiments strongly suggest that RNA sequence space is much more sparsely populated with receptors for pyrimidine moieties compared to those that can bind purine derivatives. Likewise, in the early RNA World, this increased probability of evolving receptors for purine moieties could have given signaling molecules carrying guanine and adenine moieties a huge advantage. Once biological roles for purine-based com-

pounds were established, evolutionary history, as well as the less favorable chemical properties of pyrimidines, would make it even more difficult for widespread signaling networks to adapt to any new pyrimidine-based signaling competitors.

### The Complex Architectures of Ancient RNA Signaling Systems

Could ancient signaling systems made entirely of RNA have achieved the structural and functional sophistication needed to manage a complex metabolic state before proteins were available? We think that it is well within the capability of RNA signaling compounds and their RNA receptors to have sensed and responded appropriately to the chemical and biological needs of complex RNA World organisms. As already noted above, dozens of unique signals can be generated using only the four standard RNA nucleotides and simple phosphate ester chemistry, and derivatization easily yields hundreds more. Also, RNA has proven to be very adept at forming selective receptors for

some of these compounds in the form of riboswitches for cyclic dinucleotides (20–24) and engineered aptamers for cyclic mononucleotides (88, 89). Notably, many natural riboswitches easily form ligand-binding pockets that accommodate negatively charged phosphate groups (14–19), usually by exploiting  $Mg^{2+}$  bridges between phosphates of the ligand and the riboswitch aptamer. This means that many of the phosphate-rich signaling molecules discussed above would make easy targets for binding pockets made of RNA and divalent metal ions. Thus, molecular recognition of signaling molecules by RNA is unlikely to be a limitation.

RNA receptors must also be able to convert recognition of the ligand into a regulatory response. Again, the allosteric self-splicing ribozyme described earlier (Fig. 4) provides a natural example of how ligand binding can alter important biochemical processes such as RNA splicing (21) and gene expression (72). One possible route for evolution of this complex RNA architecture is the chance integration of a selfish group I ribozyme between a *c*-di-GMP-II riboswitch and the open reading frame this riboswitch controls. Normally, such an event would not alter the biological function of the original genetic arrangement. The self-splicing ribozyme would simply remove itself from the pre-mRNA, and the fused exons would therefore function just like its unaltered predecessor. However, fortuitous base-pairing between the riboswitch aptamer and the 5' splice site of the ribozyme would interfere with normal splicing, unless *c*-di-GMP binding to the aptamer were to liberate this site.

Similarly, by appending or colocalizing natural aptamers with other functional RNA domains, a huge collection of more sophisticated regulatory RNA devices could have accumulated in advanced RNA World organisms. In extant organisms, we see abundant evidence of mixing and matching of riboswitch aptamers to create cooperative (90–92) or otherwise more “digital” (93, 94) genetic switches. Combining aptamers or riboswitches can also create two-input logic gates (95, 96) that respond to two different chemical signals. The vast majority of known riboswitches harness metabolite-mediated structural rearrangements to regulate gene expression by manipulating how RNA polymerase interacts with its DNA template or how a ribosome interacts with an mRNA (19). However, other mechanisms exist wherein a glucosamine-6-phosphate ligand is used as a cofactor (97) or GTP is used as a substrate (21) for RNA processing reactions. These natural examples of RNA-based signaling systems in modern cells provide only a small sampling of the diverse possibilities that might have been used at the height of the RNA World.

### Prospects for Discovering Additional Lost Words from the RNA World Language

The potential for forming additional RNA words by subtle chemical variation of the known classes of signaling compounds is enormous. Certainly, the signaling molecules noted above form only an incomplete list, and many more likely remain to be discovered. How might we discover more of these compounds, or entirely distinct types of RNA-derived compounds, that might be used by modern cells? Straightforward, albeit labor-intensive, methods would involve conducting isotopic labeling or biochemical separation techniques to identify novel signaling compounds in a manner similar to what was used to discover compounds like ppGpp (98), ZTP (28), and *c*-di-GMP (55). Modern analytical techniques that are useful for metabolomics analyses such as coupled liquid chromatography–mass spectroscopy could help accelerate the discovery process. However, these purification and analysis

methods require cells to be exposed to conditions that favor the production of the signaling molecule—and preferably large amounts of it—which is a difficult task when the specific conditions that might induce the production of a such a molecule are unknown.

Alternatively, bioinformatics methods might be useful for revealing the existence of additional signaling molecules and their receptors. We (24) and others (23) identified receptors for the newly discovered signaling molecule *c*-AMP-GMP by examining unusual variants of a known class of riboswitches that respond to *c*-di-GMP. Nucleotides within or near the ligand-binding site of the riboswitch deviated from the consensus, which suggested that these riboswitches had undergone a shift in specificity. These RNAs associated with a collection of target genes distinct from that of consensus *c*-di-GMP riboswitches, which also provided a strong clue that they had changed their ligand specificity. Moreover, we have discovered a series of additional, albeit rare, variants of two *c*-di-GMP riboswitch classes that carry mutations in key binding site nucleotides (99). Perhaps other riboswitch classes will be discovered for novel signaling compounds, and their variants might lead to additional expansions of the known RNA language.

A similar approach based on the bioinformatics and biochemical analyses of proteins that synthesize signaling molecules should be effective. Synthase enzymes that promiscuously form *c*-di-GMP, *c*-AMP-GMP, and *c*-di-AMP that tune their product output based on the relative concentrations of the substrates GTP and ATP (adenosine 5'-triphosphate) have been identified (100). Are there other variants of known synthases for RNA-derived signaling molecules whose substrate-binding or catalytic activities have been altered? If new compounds can be identified among the products of these enzymes, then perhaps new signaling pathways will be identified.

There also seem to be more opportunities to expand the known types of riboswitches for the current RNA signaling compounds. Are there riboswitches for cyclic mononucleotides? These compounds are widely exploited by many cells, and therefore, cognate riboswitches might still exist in some cells to receive these signals. Also, it seems reasonable to expect that there are riboswitches to be discovered for ppGpp, given the diversity of signaling roles this molecule has in many bacterial species (101). Certainly, riboswitches for this compound must have once existed. If discovered, then more complete signaling networks for these compounds can be established because most riboswitches are directly linked to the genes they regulate.

### Concluding Remarks

The intriguing link between riboswitches and RNA-derived signaling molecules was first noted (102) after the discoveries of *c*-di-GMP riboswitches (20, 21). After these initial examples of partnerships between riboswitches and a putative RNA World signaling molecule were reported, three additional riboswitch-signal relationships have been published (22–25). Although the era of signaling pathway discovery is not yet over, it is likely that the general nature of the most fundamental signaling molecules is already well represented by the known collection. If true, it is abundantly clear that the most widespread and ancient signaling compounds are based on RNA, and not on DNA, protein, or another form of biochemical media. For example, the predominant cyclic mono- and dinucleotide signaling compounds, such as cAMP and *c*-di-GMP, could just as likely have been represented by their DNA analogs, had these signaling molecules emerged after biology established metabolic processes to generate and polymerize DNA. Thus, the trend favoring RNA can be easily explained if the most

common signaling compounds we see in today's organisms emerged first in an RNA World.

Although not known today, one might also imagine that DNA analogs of these signaling molecules might have been more sparingly used in the RNA World to serve as longer-lived signals that resisted degradation owing to their greater chemical stability. Early organisms might have needed a more permanent chemical message to be sent, and evolved the ability to synthesize small deoxyribonucleotide-based signals for that purpose. This chemical step might then have served as the seed from which the DNA genomes of today were grown. Ironically, if true, it would suggest that DNA first served as messenger and RNA as genome, in contrast to their roles today.

Likewise, various amino acid derivatives could have just as easily been adapted for major cellular communication tasks had their biosynthesis and polymerization pathways been in place before these needs arose. Although amino acids and their derivatives are used for important signaling processes, these appear to be involved in more derived processes, rather than for regulation of fundamental processes that are shared by all organisms on the evolutionary tree. For instance, short peptides are used in quorum signaling in Gram-positive bacteria, but acyl homoserine lactones are more commonly used for the same purpose in Gram-negative bacteria (103).

Finally, if the RNA-derived signaling molecules that still exist in cells today originated in the RNA World, they would represent yet another reminder that the biology of today is still intertwined with the processes of the past. Several RNA-like compounds used as signals by modern cells appear to be just as ancient as the RNA-based coenzymes, the common ribozymes, the widespread riboswitches, and some other structured noncoding RNAs. Through each new discovery, we gain the opportunity to identify more molecules that originated in the RNA World, which allows us to peer back in time to speculate on the nature of our earliest ancestors.

## REFERENCES AND NOTES

1. C. Woese, *The Genetic Code: The Molecular Basis for Genetic Expression* (Harper & Row, 1967).
2. F. H. C. Crick, The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
3. L. E. Orgel, Evolution of the genetic apparatus. *J. Mol. Biol.* **38**, 381–393 (1968).
4. W. Gilbert, Origin of life: The RNA world. *Nature* **319**, 618 (1986).
5. G. F. Joyce, The rise and fall of the RNA world. *New Biol.* **3**, 399–407 (1991).
6. H. B. White III, Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, 101–104 (1976).
7. S. A. Benner, A. D. Ellington, A. Tauer, Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7054–7058 (1989).
8. D. C. Jeffares, A. M. Poole, D. Penny, Relics from the RNA World. *J. Mol. Evol.* **46**, 18–36 (1998).
9. H. S. Bernhardt, The RNA world hypothesis: The worst theory of the early evolution of life (except for all the others). *Biol. Direct* **7**, 23 (2012).
10. T. J. Wilson, Y. Liu, D. M. J. Lilley, Ribozymes and the mechanisms that underlie RNA catalysis. *Front. Chem. Sci. Eng.* **10**, 178–185 (2016).
11. R. M. Jimenez, J. A. Polanco, A. Lupták, Chemistry and biology of self-cleaving ribozymes. *Trends Biochem. Sci.* **40**, 648–661 (2015).
12. P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
13. J. C. Ellis, J. W. Brown, The RNase P family. *RNA Biol.* **6**, 362–369 (2009).
14. A. Roth, R. R. Breaker, The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.* **78**, 305–334 (2009).
15. R. R. Breaker, Prospects for riboswitch discovery and analysis. *Mol. Cell* **43**, 867–879 (2011).
16. A. Serganov, E. Nudler, A decade of riboswitches. *Cell* **152**, 17–24 (2013).
17. A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, M. S. Gelfand, Riboswitches: The oldest mechanism for the regulation of gene expression? *Trends Genet.* **20**, 44–50 (2004).
18. R. R. Breaker, Riboswitches: From ancient gene-control systems to modern drug targets. *Future Microbiol.* **4**, 771–773 (2009).
19. R. R. Breaker, Riboswitches and the RNA World. *Cold Spring Harb. Perspect. Biol.* **4**, a003566 (2012).
20. N. Sudarsan, E. R. Lee, Z. Weinberg, R. H. Moy, J. N. Kim, K. H. Link, R. R. Breaker, Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **321**, 411–413 (2008).
21. E. R. Lee, J. L. Baker, Z. Weinberg, N. Sudarsan, R. R. Breaker, An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* **329**, 845–848 (2010).
22. J. W. Nelson, N. Sudarsan, K. Furukawa, Z. Weinberg, J. X. Wang, R. R. Breaker, Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat. Chem. Biol.* **9**, 834–839 (2013).
23. C. A. Kellenberger, S. C. Wilson, S. F. Hickey, T. L. Gonzalez, Y. Su, Z. F. Hallberg, T. F. Brewer, A. T. Iavarone, H. K. Carlson, Y.-F. Hsieh, M. C. Hammond, GEMM-I riboswitches from *Geobacter* sense the bacterial second messenger cyclic AMP-GMP. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5383–5388 (2015).
24. J. W. Nelson, N. Sudarsan, G. E. Phillips, S. Stav, C. E. Lünse, P. J. McCown, R. R. Breaker, Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5389–5394 (2015).
25. P. B. Kim, J. W. Nelson, R. R. Breaker, An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol. Cell* **57**, 317–328 (2015).
26. U. Römling, M. Y. Galperin, M. Gomelsky, Cyclic di-GMP: The first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* **77**, 1–52 (2013).
27. F. M. Commichau, A. Dickmanns, J. Gundlach, R. Ficner, J. Stülke, A jack of all trades: The multiple roles of the unique essential second messenger cyclic di-AMP. *Mol. Microbiol.* **97**, 189–204 (2015).
28. B. R. Bochner, B. N. Ames, ZTP (5-amino 4-imidazole carboxamide riboside 5'-triphosphate): A proposed alarmone for 10-formyl-tetrahydrofolate deficiency. *Cell* **29**, 929–937 (1982).
29. J. L. Hougland, J. A. Piccirilli, M. Forconi, J. Lee, D. Herschlag, How the group I intron works: A case study of RNA structure and function, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, ed. 3, 2006), pp. 133–205.
30. A. M. Lambowitz, S. Zimmerley, Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* **3**, a003616 (2011).
31. T. W. Rall, E. W. Sutherland, Formation of a cyclic adenine ribonucleotide by tissue particles. *J. Biol. Chem.* **232**, 1065–1076 (1958).
32. P. B. Daniel, W. H. Walker, J. F. Habener, Cyclic AMP signaling and gene regulation. *Annu. Rev. Nutr.* **18**, 353–383 (1998).
33. A. Y. Kots, E. Martin, I. G. Sharina, F. Murad, A short history of cGMP, guanylyl cyclases, and cGMP-dependent protein kinases. *Handb. Exp. Pharmacol.* **191**, 1–14 (2009).
34. R. B. Pilz, D. E. Casteel, Regulation of gene expression by cyclic GMP. *Circ. Res.* **93**, 1034–1046 (2003).
35. K. A. McDonough, A. Rodriguez, The myriad roles of cyclic AMP in microbial pathogens: From signal to sword. *Nat. Rev. Microbiol.* **10**, 27–38 (2012).
36. S.-Q. An, K.-H. Chin, M. Febrer, Y. McCarthy, J.-G. Yang, C.-L. Liu, D. Swarbreck, J. Rogers, J. Maxwell Dow, S.-H. Chou, R. P. Ryan, A cyclic GMP-dependent signalling pathway regulates bacterial phytopathogenesis. *EMBO J.* **32**, 2430–2438 (2013).
37. K. Y. Beste, R. Seifert, cCMP, cUMP, cTMP, cIMP and cXMP as possible second messengers: Development of a hypothesis based on studies with soluble guanylyl cyclase  $\alpha_1\beta_1$ . *Biol. Chem.* **394**, 261–270 (2013).
38. R. Seifert, cCMP and cUMP: Emerging second messengers. *Trends Biochem. Sci.* **40**, 8–15 (2015).
39. E. K. Jackson, The 2',3'-cAMP-adenosine pathway. *Am. J. Physiol. Renal Physiol.* **301**, F1160–F1167 (2011).
40. E. K. Jackson, Discovery and roles of 2',3'-cAMP in biological systems. *Handb. Exp. Pharmacol.* **238**, 229–252 (2017).
41. V. Haurlyuk, G. C. Atkinson, K. S. Murakami, T. Tenson, K. Gerdes, Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat. Rev. Microbiol.* **13**, 298–309 (2015).
42. G. M. Estavillo, P. A. Crisp, W. Pornsiriwong, M. Wirtz, D. Collinge, C. Carrie, E. Giraud, J. Whelan, P. David, H. Javot, C. Brearley, R. Hell, E. Marin, B. J. Pogson, Evidence for a SAL1-PAP chloroplast retrograde pathway that functions in drought and high light signaling in *Arabidopsis*. *Plant Cell* **23**, 3992–4012 (2011).
43. L. Bettendorff, B. Wirtzfeld, A. F. Makarchikov, G. Mazzucchelli, M. Frédérix, T. Gigliobianco, M. Gangolf, E. De Pauw, L. Angenot, P. Wins, Discovery of a natural thiamine adenine nucleotide. *Nat. Chem. Biol.* **3**, 211–212 (2007).
44. A. H. Guse, Biochemistry, biology, and pharmacology of cyclic adenosine diphosphoribose (cADPR). *Curr. Med. Chem.* **11**, 847–855 (2004).
45. A. H. Guse, I. M. A. Ernst, R. Fliegert, NAADP signaling revisited. *Curr. Top. Med. Chem.* **13**, 2978–2990 (2013).
46. A. Guranowski, Metabolism of diadenosine tetraphosphate ( $A_2P_4$ ) and related nucleotides in plants; review with historical and general perspective. *Front. Biosci.* **9**, 1398–1411 (2004).
47. P. Plateau, S. Blanquett, Dinucleoside oligophosphates in micro-organisms. *Adv. Microb. Physiol.* **36**, 81–109 (1994).

48. V. Jankowski, M. van der Giet, H. Mischak, M. Morgan, W. Zidek, J. Jankowski, Dinucleoside polyphosphates: Strong endogenous agonists of the purinergic system. *Br. J. Pharmacol.* **157**, 1142–1153 (2009).
49. J. L. Houghland, J. A. Piccirilli, M. Forconi, J. Lee, D. Herschlag, How the group I intron works: A case study of RNA structure and function, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, ed. 3, 2006), pp. 133–205.
50. B. A. McNeil, C. Semper, S. Zimmerly, Group II introns: Versatile ribozymes and retroelements. *WIREs RNA* **7**, 341–355 (2016).
51. E. H. Ekland, J. W. Szostak, D. P. Bartel, Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364–370 (1995).
52. D. P. Horning, G. F. Joyce, Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9786–9791 (2016).
53. C. G. Whiteley, D.-J. Lee, Bacterial diguanylate cyclases: Structure, function and mechanism in exopolysaccharide biofilm development. *Biotechnol. Adv.* **33**, 124–141 (2015).
54. C. Opoku-Temeng, J. Zhou, Y. Zheng, J. Su, H. O. Sintim, Cyclic dinucleotide (c-di-GMP, c-di-AMP, and cGAMP) signalings have come of age to be inhibited by small molecules. *Chem. Commun.* **52**, 9327–9342 (2016).
55. P. Ross, H. Weinhouse, Y. Aloni, D. Michaeli, P. Weinberger-Ohana, R. Mayer, S. Braun, E. de Vroom, G. A. van der Marel, J. H. van Boom, M. Benziman, Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325**, 279–281 (1987).
56. Z.-h. Chen, P. Schaap, The prokaryote messenger c-di-GMP triggers stalk cell differentiation in *Dictyostelium*. *Nature* **488**, 680–683 (2012).
57. D. L. Burdette, K. M. Monroe, K. Sotelo-Troha, J. S. Iwig, B. Eckert, M. Hyodo, Y. Hayakawa, R. E. Vance, STING is a direct innate immune sensor of cyclic di-GMP. *Nature* **478**, 515–518 (2011).
58. G. Witte, S. Hartung, K. Büttner, K.-P. Hopfner, Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol. Cell* **30**, 167–178 (2008).
59. R. M. Corrigan, A. Gründling, Cyclic di-AMP: Another second messenger enters the fray. *Nat. Rev. Microbiol.* **11**, 513–524 (2013).
60. K. Sureka, P. H. Choi, M. Precit, M. Delince, D. A. Pensinger, T. N. Huynh, A. R. Jurado, Y. A. Goo, M. Sadilek, A. T. Iavarone, J.-D. Sauer, L. Tong, J. J. Woodward, The cyclic dinucleotide c-di-AMP is an allosteric regulator of metabolic enzyme function. *Cell* **158**, 1389–1401 (2014).
61. B. W. Davies, R. W. Bogard, T. S. Young, J. J. Mekalanos, Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholera* virulence. *Cell* **149**, 358–370 (2012).
62. L. Sun, J. Wu, F. Du, X. Chen, Z. J. Chen, Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).
63. J. Wu, L. Sun, X. Chen, F. Du, H. Shi, C. Chen, Z. J. Chen, Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339**, 826–830 (2013).
64. X. Zhang, H. Shi, J. Wu, X. Zhang, L. Sun, C. Chen, Z. J. Chen, Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol. Cell* **51**, 226–235 (2013).
65. P. Gao, M. Ascano, Y. Wu, W. Barchet, B. L. Gaffney, T. Zillinger, A. A. Serganov, Y. Liu, R. A. Jones, G. Hartmann, T. Tuschl, D. J. Patel, Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153**, 1094–1107 (2013).
66. E. J. Diner, D. L. Burdette, S. C. Wilson, K. M. Monroe, C. A. Kellenberger, M. Hyodo, Y. Hayakawa, M. C. Hammond, R. E. Vance, The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep.* **3**, 1355–1361 (2013).
67. M. W. Orr, G. P. Donaldson, G. B. Severin, J. Wang, H. O. Sintim, C. M. Waters, V. T. Lee, Oligoribonuclease is the primary degradative enzyme for pGpG in *Pseudomonas aeruginosa* that is required for cyclic-di-GMP turnover. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5048–E5057 (2015).
68. D. Cohen, U. Mechold, H. Nevenzal, Y. Yarmiyhu, T. E. Randall, D. C. Bay, J. D. Rich, M. R. Parsek, V. Kaever, J. J. Harrison, E. Banin, Oligoribonuclease is a central feature of cyclic diguanylate signaling in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11359–11364 (2015).
69. L. Bowman, M. S. Zeden, C. F. Schuster, V. Kaever, A. Gründling, New insights into the cyclic di-adenosine monophosphate (c-di-AMP) degradation pathway and the requirement of the cyclic-dinucleotide for acid stress resistance in *Staphylococcus aureus*. *J. Biol. Chem.* **291**, 26970–26986 (2016).
70. K. D. Smith, S. A. Strobel, Interactions of the c-di-AMP riboswitch with its second messenger ligand. *Biochem. Soc. Trans.* **39**, 647–651 (2011).
71. C. A. Kellenberger, S. C. Wilson, J. Sales-Lee, M. C. Hammond, RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. *J. Am. Chem. Soc.* **135**, 4096–4909 (2013).
72. A. G. Y. Chen, N. Sudarsan, R. R. Breaker, Mechanism for gene control by a natural allosteric group I ribozyme. *RNA* **17**, 1967–1972 (2011).
73. M. T. Cheah, A. Wächter, N. Sudarsan, R. R. Breaker, Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**, 497–500 (2007).
74. A. Wächter, M. Tunc-Ozdemir, B. C. Grove, P. J. Green, D. K. Shintani, R. R. Breaker, Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell* **19**, 3437–3450 (2007).
75. M. T. Croft, M. Moulin, M. E. Webb, A. G. Smith, Thiamine biosynthesis in algae is regulated by riboswitches. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20770–20775 (2007).
76. S. Bocoza, A. Adato, T. Mandel, M. Shapira, E. Nudler, A. Aharoni, Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes Dev.* **21**, 2874–2879 (2007).
77. S. Li, R. R. Breaker, Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Res.* **41**, 3022–3031 (2013).
78. S. M. Fica, N. Tuttle, F. Novak, N.-S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley, J. A. Piccirilli, RNA catalyzes nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013).
79. M. Desch, E. Schinner, F. Kees, F. Hofmann, R. Seifert, J. Schlossmann, Cyclic cytidine 3',5'-monophosphate (cCMP) signals via cGMP kinase I. *FEBS Lett.* **584**, 3979–3984 (2010).
80. Z. Zhang, B. L. Gaffney, R. A. Jones, c-di-GMP displays a monovalent metal ion-dependent polymorphism. *J. Am. Chem. Soc.* **126**, 16700–16701 (2004).
81. N. De, M. V. A. S. Navarro, R. V. Raghavan, H. Sondermann, Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. *J. Mol. Biol.* **393**, 619–633 (2009).
82. N. Tschowri, M. A. Schumacher, S. Schlimpert, N. B. Chinnam, K. C. Findlay, R. G. Brennan, M. J. Buttner, Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. *Cell* **158**, 1136–1147 (2014).
83. S. M. Freier, D. Alkema, A. Sinclair, T. Neilson, D. H. Turner, Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGCCp, XCCGGp, XGGCCYp, and XCCGGYp helices. *Biochemistry* **24**, 4533–4539 (1985).
84. S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, D. H. Turner, Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9373–9377 (1986).
85. S. Bommarito, N. Peyret, J. SantaLucia Jr., Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.* **28**, 1929–1934 (2000).
86. K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, D. C. Tahmassebi, E. T. Kool, Factors contributing to aromatic stacking in water: Evaluation in the context of DNA. *J. Am. Chem. Soc.* **122**, 2213–2222 (2000).
87. Y. Li, R. R. Breaker, Phosphorylating DNA with DNA. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2746–2751 (1999).
88. M. Koizumi, G. A. Soukup, J. N. Q. Kerr, R. R. Breaker, Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nat. Struct. Biol.* **6**, 1062–1071 (1999).
89. M. Koizumi, R. R. Breaker, Molecular recognition of cAMP by an RNA aptamer. *Biochemistry* **39**, 8983–8992 (2000).
90. M. Mandal, M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo, R. R. Breaker, A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**, 275–279 (2004).
91. A. Gao, A. Serganov, Structural insights into recognition of c-di-AMP by the *ydaO* riboswitch. *Nat. Chem. Biol.* **10**, 787–792 (2014).
92. J. J. Trausch, P. Ceres, F. E. Reyes, R. T. Batey, The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer. *Structure* **19**, 1413–1423 (2011).
93. R. Welz, R. R. Breaker, Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*. *RNA* **13**, 573–582 (2007).
94. H. Zhou, C. Zheng, J. Su, B. Chen, Y. Fu, Y. Xie, Q. Tang, S.-H. Chou, J. He, Characterization of a natural triple-tandem c-di-GMP riboswitch and application of the riboswitch-based dual-fluorescence reporter. *Sci. Rep.* **6**, 20871 (2016).
95. N. Sudarsan, M. C. Hammond, K. F. Block, R. Welz, J. E. Barrick, A. Roth, R. R. Breaker, Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314**, 300–304 (2006).
96. C. D. Stoddard, R. T. Batey, Mix-and-match riboswitches. *ACS Chem. Biol.* **1**, 751–754 (2006).
97. W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins, R. R. Breaker, Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281–286 (2004).
98. M. Cashel, J. Gallant, Two compounds implicated in the function of the *RC* gene of *Escherichia coli*. *Nature* **221**, 838–841 (1969).
99. Z. Weinberg, J. W. Nelson, C. E. Lünse, M. E. Sherlock, R. R. Breaker, Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2077–E2085 (2017).
100. Z. F. Haliber, X. C. Wang, T. A. Wright, B. Nan, O. Ad, J. Yeo, M. C. Hammond, Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3',3'-cGAMP). *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1790–1795 (2016).

101. K. Potrykus, M. Cashel, (p)ppGpp: Still magical? *Annu. Rev. Microbiol.* **62**, 35–51 (2008).
102. R. R. Breaker, RNA second messengers and riboswitches: Relics from the RNA World? *Microbe* **5**, 13–20 (2010).
103. C. M. Waters, B. L. Bassler, Quorum sensing: Cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **21**, 319–346 (2005).

**Acknowledgments:** We thank A. Roth, N. Sudarsan, and M. Sherlock for helpful discussions and comments on the manuscript. **Funding:** This work was supported by NIH grant GM022778 to R.R.B., whose research on RNA is also supported by the Howard Hughes Medical

Institute. **Author contributions:** Both authors contributed to writing this manuscript and generating figures. **Competing interests:** The authors declare that they have no competing interests.

Submitted 30 January 2017

Accepted 9 May 2017

Published 13 June 2017

10.1126/scisignal.aam8812

**Citation:** J. W. Nelson, R. R. Breaker, The lost language of the RNA World. *Sci. Signal.* **10**, eaam8812 (2017).

## The lost language of the RNA World

James W. Nelson and Ronald R. Breaker

*Sci. Signal.* **10** (483), eaam8812.  
DOI: 10.1126/scisignal.aam8812

### ARTICLE TOOLS

<http://stke.sciencemag.org/content/10/483/eaam8812>

### RELATED CONTENT

<http://stke.sciencemag.org/content/sigtrans/10/464/eaam8636.full>  
<http://stke.sciencemag.org/content/sigtrans/10/475/eaal3011.full>  
<http://stke.sciencemag.org/content/sigtrans/8/371/ec84.abstract>  
<http://stke.sciencemag.org/content/sigtrans/9/441/ra81.full>  
<http://stke.sciencemag.org/content/sigtrans/9/450/ra102.full>  
<http://stke.sciencemag.org/content/sigtrans/9/450/fs16.full>  
<http://stke.sciencemag.org/content/sigtrans/8/386/ec195.abstract>  
<http://stke.sciencemag.org/content/sigtrans/8/380/ra57.full>  
<http://stke.sciencemag.org/content/sigtrans/8/372/ra36.full>  
<http://stke.sciencemag.org/content/sigtrans/8/380/fs12.full>

### REFERENCES

This article cites 100 articles, 32 of which you can access for free  
<http://stke.sciencemag.org/content/10/483/eaam8812#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science Signaling* (ISSN 1937-9145) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Signaling* is a registered trademark of AAAS.

Copyright © 2017 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works.