



Published in final edited form as:

*Sci Signal*. ; 10(483): . doi:10.1126/scisignal.aam8812.

## The lost language of the RNA World

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

<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

### Abstract

The possibility of an RNA World is based on the notion that life on Earth passed through a primitive phase without proteins, at a time when all genomes and enzymes were composed of ribonucleic acid. Numerous apparent vestiges of this ancient RNA World remain today, including many nucleotide-derived coenzymes, self-processing ribozymes, metabolite-binding riboswitches, and even ribosomes. Intriguingly, many of the most common signaling molecules and second messengers used by modern organisms are also formed from RNA. 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. Herein 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

\*To whom correspondence should be addressed. Dr. Ronald R. Breaker, Tel: (203) 432-9389, Fax: (203) 432-0753, ronald.breaker@yale.edu.

<sup>†</sup>Present Address: Howard Hughes Medical Institute and Department of Chemistry and Chemical Biology, Harvard University, 13 Oxford, St., Cambridge, MA 02138

**Author contributions:** Both authors contributed to writing this manuscript and generating figures.

**Competing interests:** The authors declare that they have no competing interests.

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. 1). 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 guanosine monophosphate) (20, 21), a class that responds to c-di-AMP (3',5'-cyclic adenosine monophosphate) (22), a collection of variants of a c-di-GMP riboswitch class that recognize c-AMP-GMP (23, 24), and another 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 a various physiological changes (20–27). In contrast, the alarmone ZTP, which is a derivative of a purine biosynthesis intermediate, signals the need to adapt to a deficiency in folate cofactor metabolism (25, 28). Although numerous other pairings between RNA-derived ligands and riboswitch sensors are known to exist (Fig. 1), 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, highlight the possibility that these and other RNA-derived molecules represent the descendants of ancient signaling systems used in the RNA World. Below 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 lifeforms 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 the 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 RNase P (13) is striking. Undoubtedly, many other such complex RNAs were lost shortly after protein enzymes arrived on the evolutionary 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 ribonucleotides cAMP (3',5'-cyclic adenosine monophosphate) (31, 32) and its guanosine analog cGMP (33, 34) (Fig. 2). Both compounds are essential signaling molecules in eukaryotes, and, whereas cGMP is used less commonly in eubacteria, cAMP is widely important in Gram-negative bacteria (35, 36). Growing evidence suggests that the similar 3',5'-cyclic ribonucleotides cCMP and cUMP 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 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 these 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 'alarmones', 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) 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 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 manifest in the widespread signaling molecule A(p)<sub>4</sub>A (diadenosine tetraphosphate) (46, 47), which can be thought of simply as two adenosine-5'-diphosphate (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 comprised 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 formation of a normal 3',5'-phosphodiester linkage, and this is the subject of the following two sections.

## Communicating With Linear and Cyclic Dinucleotides

Already, 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 expand an RNA language is to fuse two nucleotides by a 3',5'-phosphodiester linkage to create a combinatorial collection of all possible linear dinucleotides. 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 has been proven to 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 (Fig. 3A).

Despite these noted advantages, using linear 3',5'-linked dinucleotides still can 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, and so the presence of these compounds indicates that they have been purposefully synthesized for their signaling functions. Indeed, 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 have been identified that selectively degrade cyclic dinucleotides (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. By 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* thirty 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* (56). Even in some organisms where c-di-GMP is not endogenously produced, 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 twenty years for another cyclic dinucleotide to be identified in biology. 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 less than a decade ago when it was co-crystallized 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 more 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 to date are even more striking because they begin to exploit 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 (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 it was discovered and named cGAMP (62, 63). Whereas all the other natural cyclic dinucleotides known 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 currently emerging for the biological roles of the linear dinucleotides pGpG and pApA (67–69). These compounds can result from the degradation of c-di-GMP and c-di-AMP, suggesting that their linear degradation products might also be sensed as an extension of the c-di-GMP or c-di-AMP signaling networks, respectively.

Intriguingly, all of the cyclic dinucleotides known to be present in bacteria, where riboswitches are prevalent, are detected by at least one corresponding riboswitch class (20–24). Furthermore, these RNAs have extraordinarily tight 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.

Indeed, rare natural examples of such cyclic dinucleotide-controlled allosteric ribozymes have been identified (21, 72). By searching through bacterial genomes for riboswitch sequences that are located in proximity to ribozyme sequences, 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 as a substrate, are juxtaposed (Fig. 4) such that proper pre-mRNA splicing (21) and gene expression (72) requires 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 messenger RNA. 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 riboswitches (73–77), although splicing in those organisms is triggered by an RNA World coenzyme 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 mixed or homodimers that include C or U nucleotides. Pyrimidine nucleotides such as CTP and UTP 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 such a strong bias in favor of purines in RNA-derived signaling molecules?

We speculate that this bias likely was 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. Indeed, c-di-GMP is known to self-aggregate (80), and is sometimes bound in the form of a dimer complex (81, 82).

Specifically, purines contain a fused two-ring system, as opposed to the single ring of pyrimidines, making stacking interactions with their biological receptors more favorable. This has been empirically observed by examining helical RNA (83, 84) and DNA (85, 86) structures 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  $A > G > C \approx U$  or  $T$ . Furthermore, the purines have more potential for forming hydrogen-bonding contacts compared to the pyrimidines. Again this chemical property supports our speculation that signaling molecules with guanine and adenine moieties are preferentially used, in this case simply because they can form more hydrogen bonds with receptor pockets compared to their pyrimidine competitors.

Interestingly, 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

compounds 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. Some support for this view is already noted above. Using only the four standard RNA nucleotides and simple phosphate ester chemistry, dozens of unique signals can be generated. Further 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 engineered aptamers for cyclic mononucleotides (88, 89) and riboswitches for cyclic dinucleotides (20–24). 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 co-localizing 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 (93, 94) more ‘digital’ genetic switches. Combining aptamers or riboswitches also can 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 a messenger RNA (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 employed 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 only form an incomplete list, and many more likely remain to be discovered. How might we discover more of these compounds, or entirely distinct types, 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). More modern analytical techniques useful for metabolomics analyses such as LC-MS 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.

Alternatively, bioinformatics methods might be useful to reveal 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 strange variants of a known riboswitch class for c-di-GMP. Nucleotides within or near the ligand-binding site of the riboswitch were observed to deviate from the consensus, which suggested they had undergone a shift in specificity. These RNAs associated with a collection of 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 protein syntheses of signaling molecules should be effective. Synthase enzymes that promiscuously form c-di-GMP, c-AMP-GMP, and c-di-AMP were found that tune their product output based on the relative concentrations of the substrates GTP and ATP (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 from the products of these enzymes, then perhaps new signaling pathways can be identified.

There also seem to be more opportunities to expand the known types of riboswitches for the current RNA signaling compounds. Where are the riboswitches for the 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 it holds 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, an additional three 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.

## Acknowledgments

We thank Dr. Adam Roth, Dr. Narasimhan Sudarsan, and Madeline 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.

## REFERENCES AND NOTES

1. Woese C. The genetic code: the molecular basis for genetic expression. Harper & Row. 1967
2. Crick FH. The origin of the genetic code. *J Mol Biol.* 1968; 38:367–379. [PubMed: 4887876]
3. Orgel LE. Evolution of the genetic apparatus. *J Mol Biol.* 1968; 38:381–393. [PubMed: 5718557]
4. Gilbert W. Origin of life: the RNA world. *Nature.* 1986; 319:618.
5. Joyce GF. The rise and fall of the RNA world. *New Biol.* 1991; 3:399–407. [PubMed: 1712228]
6. White HB III. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol.* 1976; 7:101–104. [PubMed: 1263263]
7. Benner SA, Ellington AD, Tauer A. Modern metabolism as a palimpsest of the RNA world. *Proc Natl Acad Sci USA.* 1989; 86:7054–7058. [PubMed: 2476811]
8. Jeffares DC, Poole AM, Penny D. Relics from the RNA World. *J Mol Evol.* 1998; 46:18–36. [PubMed: 9419222]
9. Bernhardt HS. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). *Biol Direct.* 2012; 7:23. [PubMed: 22793875]
10. Wilson TJ, Liu Y, Lilley DMJ. Ribozymes and the mechanisms that underlie RNA catalysis. *Front Chem Sci Eng.* 2016; 10:178–185.
11. Jimenez RM, Polanco JA, Lupták A. Chemistry and biology of self-cleaving ribozymes. *Trends Biochem Sci.* 2015; 40:648–661. [PubMed: 26481500]
12. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. *Science.* 2000; 289:920–930. [PubMed: 10937990]
13. Ellis JC, Brown JW. The RNase P family. *RNA Biol.* 2009; 6:362–369. [PubMed: 19738420]
14. Roth A, Breaker RR. The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem.* 2009; 78:305–334. [PubMed: 19298181]
15. Breaker RR. Prospects for riboswitch discovery and analysis. *Mol Cell.* 43:867–879. [PubMed: 21925376]
16. Serganov A, Nudler E. A decade of riboswitches. *Cell.* 2013; 152:17–24. [PubMed: 23332744]
17. Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS. Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.* 2004; 20:44–50. [PubMed: 14698618]
18. Breaker RR. Riboswitches: from ancient gene-control systems to modern drug targets. *Future Microbiol.* 2009; 4:771–773. [PubMed: 19722830]
19. Breaker RR. Riboswitches and the RNA World. *Cold Spring Harb Perspect Biol.* 2012; 4 pii: a003566.
20. Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science.* 2008; 321:411–413. [PubMed: 18635805]
21. Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science.* 2010; 329:845–848. [PubMed: 20705859]
22. Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX, Breaker RR. Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat Chem Biol.* 2013; 9:834–839. [PubMed: 24141192]
23. Kellenberger CA, Wilson SC, Hickey SF, Gonzalez TL, Su Y, Hallberg ZF, Brewer TF, Iavarone AT, Carlson HK, Hsieh YF, Hammond MC. GEMM-I riboswitches from *Geobacter* sense the bacterial second messenger cyclic AMP-GMP. *Proc Natl Acad Sci USA.* 2015; 112:5383–5388. [PubMed: 25848022]

24. Nelson JW, Sudarsan N, Phillips GE, Stav S, Lünse CE, McCown PJ, Breaker RR. Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proc Natl Acad Sci USA*. 2015; 112:5389–5394. [PubMed: 25848023]
25. Kim PB, Nelson JW, Breaker RR. An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol Cell*. 2015; 57:317–328. [PubMed: 25616067]
26. Römling U, Galperin MY, Gomelsky M. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev*. 2013; 77:1–52. [PubMed: 23471616]
27. Commichau FM, Dickmanns A, Gundlach J, Ficner R, Stülke J. A jack of all trades: the multiple roles of the unique essential second messenger cyclic di-AMP. *Mol Microbiol*. 2015; 97:189–204. [PubMed: 25869574]
28. Bochner BR, Ames BN. ZTP (5-amino 4-imidazole carboxamide riboside 5'-triphosphate): a proposed alarmone for 10-formyltetrahydrofolate deficiency. *Cell*. 1982; 29:929–937. [PubMed: 6185232]
29. Hougland, JL., Piccirilli, JA., Forconi, M., Lee, J., Herschlag, D. *The RNA World*. 3rd. Cold Spring Harbor Laboratory Press; 2006. How the group I intron works: a case study of RNA structure and function; p. 133-205.
30. Lambowitz AM, Zimmerley S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol*. 2011; 3:a003616. [PubMed: 20463000]
31. Rall TW, Sutherland EW. Formation of a cyclic adenine ribonucleotides by tissue particles. *J Biol Chem*. 1958; 232:1065–1076. [PubMed: 13549487]
32. Daniel PB, Walker WH, Habener JF. Cyclic AMP signaling and gene regulation. *Annu Rev Nutr*. 1998; 18:353–383. [PubMed: 9706229]
33. Kots AY, Martin E, Sharina IG, Murad F. A short history of cGMP, guanylyl cyclases, and cGMP-dependent protein kinases. *Hand Exp Pharmacol*. 2009; 191:1–14.
34. Pilz RB, Casteel DE. Regulation of gene expression by cyclic GMP. *Circ Res*. 2003; 93:1034–1046. [PubMed: 14645134]
35. McDonough KA, Rodriguez A. The myriad roles of cyclic AMP in microbial pathogens: from signal to sword. *Nat Rev Microbiol*. 2012; 10:27–38.
36. An S-Q, Chin KH, Febrer M, McCarthy Y, Yang JG, Liu CL, Swarbreck D, Rogers J, Maxwell Dow J, Chou SH, Ryan RP. A cyclic GMP-dependent signalling pathway regulates bacterial phytopathogenesis. *EMBO J*. 2013; 32:2430–2438. [PubMed: 23881098]
37. Beste KY, Seifert R. 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*. 2013; 394:261–270. [PubMed: 23087103]
38. Seifert R. cCMP and cUMP: emerging second messengers. *Trends Biochem Sci*. 2015; 40:8–15. [PubMed: 25435399]
39. Jackson EK. The 2',3'-cAMP-adenosine pathway. *Am J Physiol Renal Physiol*. 2011; 301:F1160–F1167. [PubMed: 21937608]
40. Jackson EK. Discovery and roles of 2',3'-cAMP in biological systems. *Handbook of Experimental Pharmacology*. 2015
41. Haurlyuk V, Atkinson GC, Murakami KS, Tenson T, Gerdes K. Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol*. 2015; 13:298–309.
42. Estavillo GM, Crisp PA, Pornsiriwong W, Wirtz M, Collinge D, Carrie C, Giraud E, Whelan J, David P, Javot H, Brearley C, Hell R, Marin E, Pogson BJ. Evidence for a SAL1-PAP chloroplast retrograde pathway that function in drought and high light signaling in *Arabidopsis*. *Plant Cell*. 2011; 23:3992–4012. [PubMed: 22128124]
43. Bettendorff L, Wirtzfeld B, Makarchikov AF, Mazzucchelli G, Frédéric M, Gigliobianco T, Gangolf M, De Pauw E, Angenot L, Wins P. Discovery of a natural thiamine adenine nucleotide. *Nat Chem Biol*. 2007; 3:211–212. [PubMed: 17334376]
44. Guse AH. Biochemistry, biology, and pharmacology of cyclic adenosine diphosphoribose (cADPR). *Curr Med Chem*. 2004; 11:847–855. [PubMed: 15078169]
45. Guse AH, Ernst IM, Fliegert R. NAADP signaling revisited. *Curr Top Med Chem*. 2013; 13:2978–2990. [PubMed: 24171770]

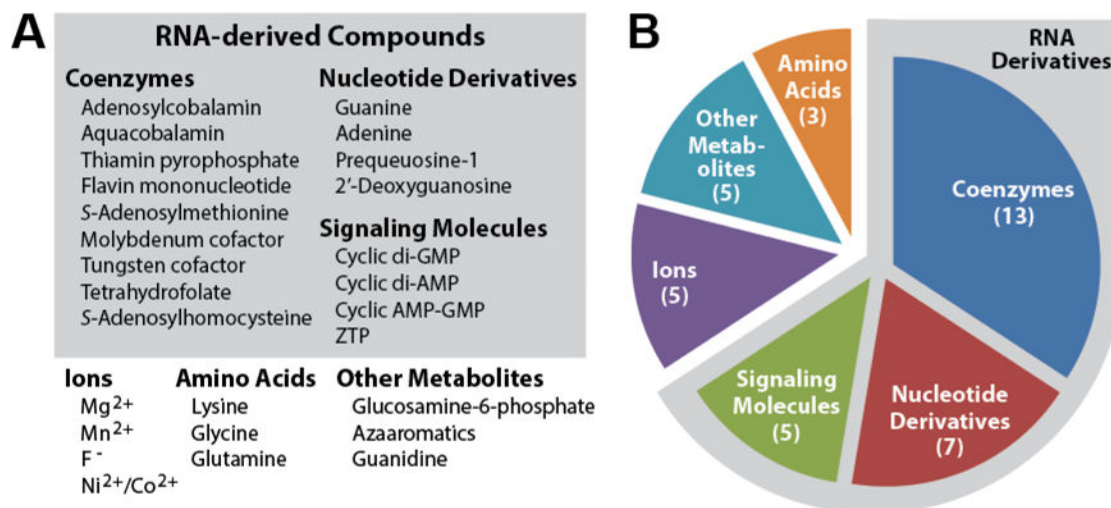
46. Guranowski A. Metabolism of diadenosine tetraphosphate (Ap<sub>4</sub>A) and related nucleotides in plants; review with historical and general perspective. *Front Biosci.* 2004; 9:1398–1411. [PubMed: 14977555]
47. Plateau P, Blanquest S. Dinucleoside oligophosphates in micro-organisms. *Adv Microb Phys.* 1994; 36:81–109.
48. Jankowski V, van der Giet M, Mischak H, Morgan M, Zidek W, Jankowski J. Dinucleoside polyphosphates: strong endogenous agonists of the purinergic system. *Br J Pharmacol.* 2009; 157:1142–1153. [PubMed: 19563527]
49. Houghland, JL., Piccirilli, JA., Forconi, M., Lee, J., Herschlag, D. *The RNA World*. 3rd. Cold Spring Harbor Laboratory Press; 2006. How the group I intron works: a case study of RNA structure and function.
50. McNeil BA, Semper C, Zimmerly S. Group II introns: versatile ribozymes and retroelements. *WIREs RNA.* 2016; 7:341–355. [PubMed: 26876278]
51. Ekland EH, Szostak JW, Bartel DP. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science.* 1995; 269:364–370. [PubMed: 7618102]
52. Horning DP, Joyce GF. Amplification of RNA by an RNA polymerase ribozyme. *Proc Natl Acad Sci USA.* 2016; 113:9786–9791. [PubMed: 27528667]
53. Whiteley CG, Lee DJ. Bacterial diguanylate cyclases: structure, function and mechanism in exopolysaccharide biofilm development. *Biotechnol Adv.* 2015; 33:124–141. [PubMed: 25499693]
54. Opoku-Temeng C, Zhou J, Zheng Y, Su J, Sintim HO. Cyclic dinucleotide (c-di-GMP, c-di-AMP, and cGAMP) signalings have come of age to be inhibited by small molecules. *Chem Comm.* 2016; 52:9327–9342. [PubMed: 27339003]
55. Ross P, Weinhouse H, Aloni Y, Michaeli D, Weinberger-Ohana P, Mayer R, Braun S, de Vroom E, van der Marel GA, van Boom JH, Benziman M. Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature.* 1987; 325:279–281. [PubMed: 18990795]
56. Chen Z, Schaap P. The prokaryote messenger c-di-GMP triggers stalk cell differentiation in *Dictyostelium*. *Nature.* 2012; 488:680–683. [PubMed: 22864416]
57. Burdette DL, Monroe KM, Sotelo-Troha K, Iwig JS, Eckert B, Hyodo M, Hayakawa Y, Vance RE. STING is a direct innate immune sensor of cyclic di-GMP. *Nature.* 2011; 478:515–518. [PubMed: 21947006]
58. Witte G, Hartung S, Buttner K, Hopfner KP. Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol Cell.* 2008; 30:167–178. [PubMed: 18439896]
59. Corrigan RM, Gründling A. Cyclic di-AMP: another second messenger enters the fray. *Nat Rev Microbiol.* 2013; 11:513–524. [PubMed: 23812326]
60. Sureka K, Choi PH, Precit M, Delince M, Pensinger DA, Huynh TN, Jurado AR, Goo YA, Sadilek M, Iavarone AT, Sauer JD, Tong L, Woodward JJ. The cyclic di-nucleotide c-di-AMP is an allosteric regulator of metabolic enzyme function. *Cell.* 2014; 158:1389–1401. [PubMed: 25215494]
61. Davies BW, Bogard RW, Young TS, Mekalanos JJ. Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholera* virulence. *Cell.* 2012; 149:358–370. [PubMed: 22500802]
62. Sun L, Wu J, Du F, Chen X, Chen ZJ. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science.* 2013; 339:786–791. [PubMed: 23258413]
63. Wu J, Sun L, Chen X, Du F, Shi H, Chen C, Chen ZJ. Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science.* 2013; 339:826–830. [PubMed: 23258412]
64. Zhang X, Shi H, Wu J, Zhang X, Sun L, Chen C, Chen ZJ. Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol Cell.* 2013; 51:226–235. [PubMed: 23747010]
65. Gao P, Ascano M, Wu Y, Barchet W, Gaffney BL, Zillinger T, Serganov AA, Liu Y, Jones RA, Hartmann G, Tuschl T, Patel DJ. Cyclic [G(2′,5′)pA(3′,5′)p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell.* 2013; 153:1094–1107. [PubMed: 23647843]

66. Diner EJ, Burdette DL, Wilson SC, Monroe KM, Kellenberger CA, Hyodo M, Hayakawa Y, Hammond MC, Vance RE. The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep.* 2013; 3:1355–1361. [PubMed: 23707065]
67. Orr MW, Donaldson GP, Severin GB, Wang J, Sintim HO, Waters CM, Lee VT. Oligoribonuclease is the primary degradative enzyme for pGpG in *Pseudomonas aeruginosa* that is required for cyclic-di-GMP turnover. *Proc Natl Acad Sci USA.* 2015; 112:E5048–E5057. [PubMed: 26305945]
68. Cohen D, Mechold U, Nevenzal H, Yarmiyhu Y, Randall TE, Bay DC, Rich JD, Parsek MR, Kaever V, Harrison JJ, Banin E. Oligoribonuclease is a central feature of cyclic diguanylate signaling in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA.* 2015; 112:11359–11364. [PubMed: 26305928]
69. Bowman L, Zeden MS, Schuster CF, Kaever V, Grundling A. 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.* 2016 (in press).
70. Smith KD, Strobel SA. Interactions of the c-di-AMP riboswitch with its second messenger ligand. *Biochem Soc Trans.* 2011; 39:647–651. [PubMed: 21428955]
71. Kellenberger CA, Wilson SC, Sales-Lee J, Hammond MC. RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. *J Am Chem Soc.* 2013; 135:4096–4909.
72. Chen AGY, Sudarsan N, Breaker RR. Mechanism for gene control by a natural allosteric group I ribozyme. *RNA.* 2011; 17:1967–1972. [PubMed: 21960486]
73. Cheah MT, Wachter A, Sudarsan N, Breaker RR. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature.* 2007; 447:497–500. [PubMed: 17468745]
74. Wachter A, Tunc-Ozdemir M, Grove BC, Green PJ, Shintani DK, Breaker RR. Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell.* 2007; 19:3437–3450. [PubMed: 17993623]
75. Croft MT, Moulin M, Webb ME, Smith AG. Thiamine biosynthesis in algae is regulated by riboswitches. *Proc Natl Acad Sci USA.* 2007; 104:20770–20775. [PubMed: 18093957]
76. Bocobza S, Adato A, Mandel T, Shapira M, Nudler E, Aharoni A. Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes Dev.* 2007; 21:2874–2879. [PubMed: 18006684]
77. Li S, Breaker RR. Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Res.* 2013; 41:3022–3031. [PubMed: 23376932]
78. Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. RNA catalyses nuclear pre-mRNA splicing. *Nature.* 2013; 503:229–234. [PubMed: 24196718]
79. Desch M, Schinner E, Kees F, Hofmann F, Seifert R, Schlossmann J. Cyclic cytidine 3',5' - monophosphate (cCMP) signals via cGMP kinase I. *FEBS Lett.* 2010; 584:3979–3984. [PubMed: 20691687]
80. Zhang Z, Gaffney BL, Jones RA. c-di-GMP displays a monovalent metal ion-dependent polymorphism. *J Am Chem Soc.* 2004; 126:16700–16701. [PubMed: 15612689]
81. De N, Navarro MVAS, Raghavan RV, Sondermann H. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. *J Mol Biol.* 2009; 393:619–633. [PubMed: 19695263]
82. Tschowri N, Schumacher MA, Schlimpert S, Chinnam NB, Findlay KC, Brennan RG, Buttner MJ. Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. *Cell.* 2014; 158:1136–1147. [PubMed: 25171413]
83. Freier SM, Alkema D, Sinclair A, Neilson T, Turner DH. Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helices. *Biochemistry.* 1985; 24:4533–4539. [PubMed: 4063336]
84. Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH. Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci USA.* 1986; 83:9373–9377. [PubMed: 2432595]
85. Bommarito S, Peyret N, SantaLucia J Jr. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.* 2000; 28:1929–1934. [PubMed: 10756193]

86. Guckian KM, Schweitzer BA, Ren RX, Sheils CJ, Tahmassebi DC, Kool ET. Factors contributing to aromatic stacking in water: evaluation in the context of DNA. *J Am Chem Soc.* 2000; 122:2213–2222. [PubMed: 20865137]
87. Li Y, Breaker RR. Phosphorylating DNA with DNA. *Proc Natl Acad Sci USA.* 1999; 96:2746–2751. [PubMed: 10077582]
88. Koizumi M, Soukup GA, Kerr JNQ, Breaker RR. Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nat Struct Biol.* 1999; 6:1062–1071. [PubMed: 10542100]
89. Koizumi M, Breaker RR. Molecular recognition of cAMP by an RNA aptamer. *Biochemistry.* 2000; 39:8983–8992. [PubMed: 10913311]
90. Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science.* 2004; 306:275–279. [PubMed: 15472076]
91. Gao A, Serganov A. Structural insights into recognition of c-di-AMP by the *ydaO* riboswitch. *Nat Chem Biol.* 2014; 10:787–792. [PubMed: 25086507]
92. Trausch JJ, Ceres P, Reyes FE, Batey RT. The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer. *Structure.* 2011; 19:1413–1423. [PubMed: 21906956]
93. Welz R, Breaker RR. Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*. *RNA.* 2007; 13:573–582. [PubMed: 17307816]
94. Zhou H, Zheng C, Su J, Chen B, Fu Y, Xie Y, Tang Q, Chou SH, He J. Characterization of a natural triple-tandem c-di-GMP riboswitch and application of the riboswitch-based dual-fluorescence reporter. *Sci Rep.* 2016; 6:20871. [PubMed: 26892868]
95. Sudarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, Breaker RR. Tandem riboswitch architectures exhibit complex gene control functions. *Science.* 2006; 314:300–304. [PubMed: 17038623]
96. Stoddard CD, Batey RT. Mix-and-match riboswitches. *ACS Chem Biol.* 2006; 1:751–754. [PubMed: 17240972]
97. Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. *Nature.* 2004; 428:281–286. [PubMed: 15029187]
98. Cashel M, Gallant J. Two compounds implicated in the function of the RC gene of *Escherichia coli*. *Nature.* 1969; 221:838–841. [PubMed: 4885263]
99. Weinberg Z, Nelson JW, Lünse CE, Sherlock ME, Breaker RR. Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. *Proc Natl Acad Sci USA.* (submitted).
100. Haliber ZF, Wang XC, Wright TA, Nan B, Ad O, Yeo J, Hammond MC. Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3',3'-cGAMP). *Proc Natl Acad Sci USA.* 2016; 113:1790–1795. [PubMed: 26839412]
101. Potrykus K, Cashel M. (p)ppGpp: still magical? *Annu Rev Microbiol.* 2008; 62:35–51. [PubMed: 18454629]
102. Breaker RR. RNA second messengers and riboswitches: relics from the RNA World? *Microbe.* 2010; 5:13–20.
103. Waters CM, Bassler BL. Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol.* 2005; 21:319–346. [PubMed: 16212498]

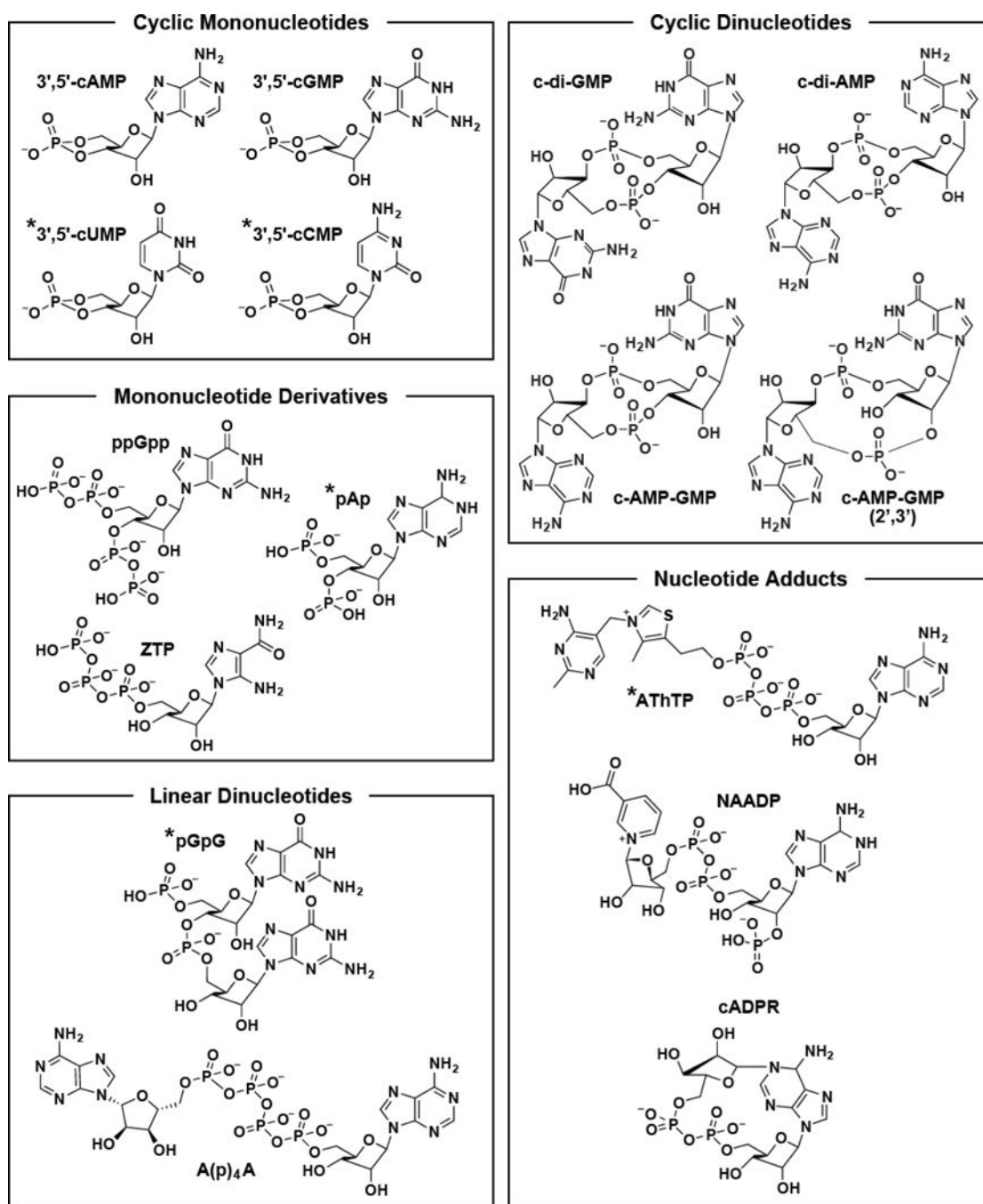
### Gloss

The discovery of numerous riboswitch classes reveal that many of these RNA structures regulate gene expression in response to the selective binding of coenzymes and signaling molecules derived from RNA monomers or their precursors. It has been proposed that many coenzymes might be of ancient origin, based on their universal distribution in biology and their RNA-like chemical composition. In this Review, which includes four figures and 100 references, we discuss the findings that support the hypothesis that common RNA-derived signaling compounds are ancient, and speculate on the possible complexity of the chemical language that might have been used by lifeforms long before proteins emerged.



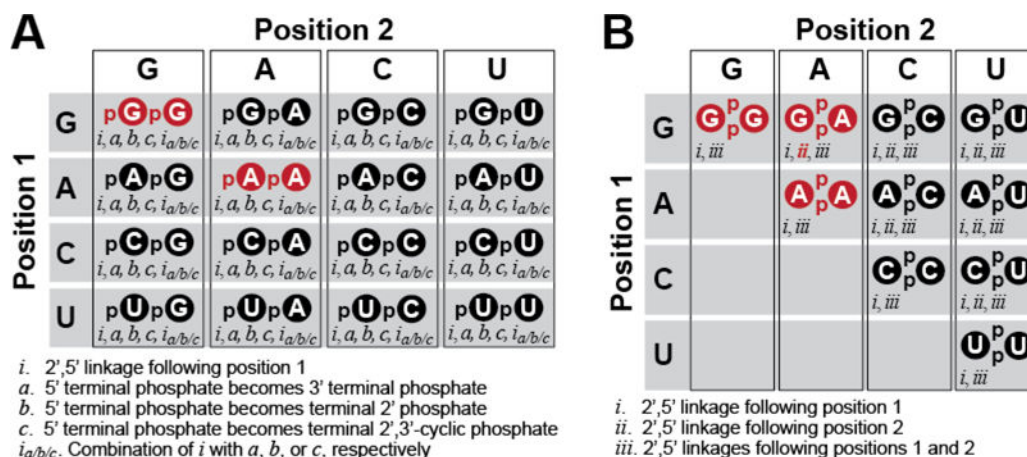
**Fig. 1. The biological ligands for the known natural riboswitch classes are predominantly derived from RNA**

(A) 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.



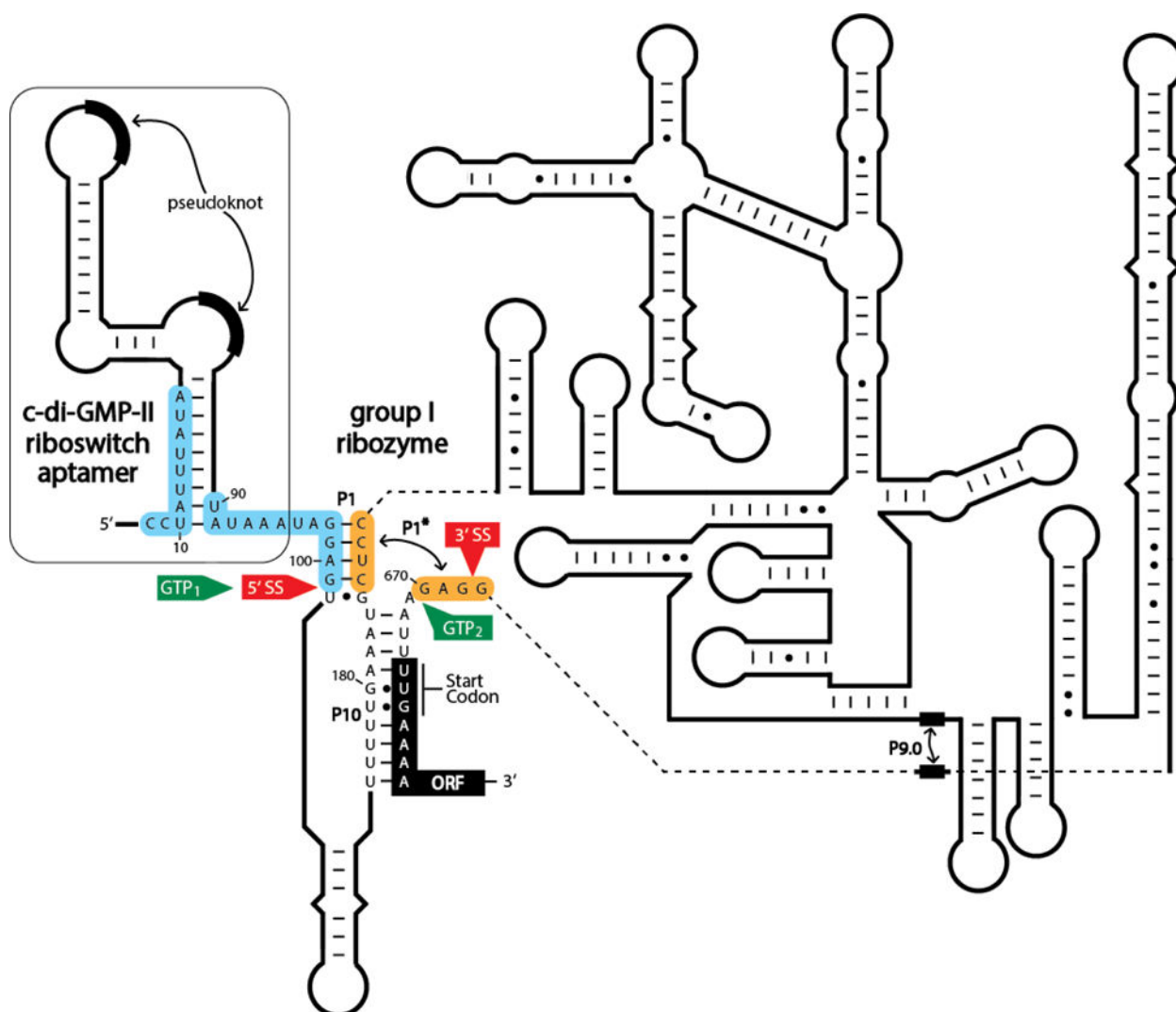
**Fig. 2. Known natural signaling compounds derived from RNA nucleotides or their precursors**

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 whose biological roles and phylogenetic distributions have not been well established.



**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. Red highlights the compounds pGpG (67) and pApA (68, 69), for which there is evidence of biological roles in modern cells. (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. Again, red identifies specific molecules that are known to be biologically active in modern cells.



**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 blue regions to base-pair, and the complementary orange regions 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 are zero-length connections.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript