

Published in final edited form as:

*Biochim Biophys Acta*. 2014 October ; 1839(10): 900–907. doi:10.1016/j.bbagr.2014.02.011.

## Computational analysis of riboswitch-based regulation

Eric I. Sun<sup>1</sup> and Dmitry A. Rodionov<sup>2,3,\*</sup>

<sup>1</sup>Department of Molecular Biology, Division of Biological Sciences, University of California at San Diego, La Jolla, California 92093, USA

<sup>2</sup>Sanford-Burnham Medical Research Institute, La Jolla, California 92037, USA

<sup>3</sup>A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia

### Abstract

Advances in computational analysis of riboswitches in the last decade have contributed greatly to our understanding of riboswitch regulatory roles and mechanisms. Riboswitches were originally discovered as part of sequence analysis of the 5'-untranslated region of mRNAs in the hope of finding novel gene regulatory sites, and the existence of structural RNAs appeared to be a spurious phenomenon. As more riboswitches were discovered, they illustrated diversity and adaptability of these RNA regulatory sequences. The fact that a chemically monotonous molecule like RNA can discern a wide range of substrates and exert a variety of regulatory mechanisms was subsequently demonstrated in diverse genomes and has hastened the development of sophisticated algorithms for their analysis and prediction. In this review, we focus on some of the computational tools for riboswitch detection and secondary structure prediction. The study of this simple yet efficient form of gene regulation promises to provide a more complete picture of a world that RNA once dominated and allows rational design of artificial riboswitches. This article is part of a Special Issue entitled: Riboswitches.

### Keywords

RNA regulatory motif; Riboswitch; Regulon; Gene Function; Comparative genomics; Bacteria

## 1. Introduction

Prior to discovery and systemic analysis of metabolite-sensing riboswitches, scientists had noticed the presence of possible RNA regulatory elements at the 5'-untranslated regions (UTRs) of several operons encoding the biosynthetic enzymes for amino acids and co-factors, and some of these leader regions have been further investigated by functional and genetic experiments [1–4]. Specifically, these RNA elements appeared to alter their

© 2014 Elsevier B.V. All rights reserved.

\*Corresponding author: rodionov@burnham.org (D. Rodionov), Phone: +1 858 646 3100.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

conformation in the presence of metabolites associated with the end products of the regulated operons. Experiments associated with the discovery of these bimodal structural elements include in-line probing, which confirmed conformational changes that abolish base pairing upon metabolite binding [5–10], and nuclease cleavage assay, which showed loss of binding from a DNA probe to a complementary RNA sequence upon introduction of a specific metabolite [11–13]. The binding of metabolites to specific RNA sequences were confirmed using equilibrium dialysis assay [8, 9, 14], and the induced structural changes were measured via fluorescence quenching of a natural metabolite upon binding to a specific RNA sequence [13, 15]. Importantly, all the experiments demonstrated conformational changes can be achieved without any protein factors and confirms the existence of riboswitches.

However, a comprehensive picture of riboswitches did not truly emerge until researchers were able to conduct cross-genomic analyses of riboswitches. Initial analyses were carried out by analyzing conserved intergenic regions across diverse bacterial genomes, and riboswitch functions were putatively assigned based on contents of regulated operons [16–18]. The detection effort was partly facilitated by the large sequence conservation in the aptamer, or ligand-binding domain of a riboswitch, as ribonucleotide conservation is required for proper substrate recognition. Such conservation is especially important for the core ribonucleotides that are in direct contact with substrate [19–21] and has allowed detection of riboswitches in diverse genomes. The analysis of riboswitches has truly progressed significantly with rapid expansion in genomic database and advances in computer processing power in the last decade.

## 2. Riboswitch identification tools based on folding energy and motif conservation

Riboswitches were originally identified as structural elements with potential regulatory functions at the 5'-UTR of mRNAs, and the computational analysis of riboswitch sequences was facilitated by taking advantage of this fact [14, 16, 22–27]. Although not traditionally regarded as riboswitches (i.e. regulatory RNAs that recognize small molecules utilizing non-Watson-Crick interactions), the sequence and structure regularity of T-boxes in Gram-positive bacteria nevertheless lay the foundation to the early discovery effort on riboswitches [28–31]. In addition to their genomic location, highly conserved aptamer domains within riboswitch sequences can be used to identify individual riboswitch classes. To avoid false positive detection, computational methods can be applied to aid the elimination of weak structural RNAs, repeat regions and putative intrinsic terminators (terminator sequences without upstream structural regions) in intergenic regions [32]. As a result, a large number of putative riboswitches have been discovered through comparative sequence analysis [16, 33–35]. Regulatory roles, if not the riboswitch ligand, can usually be assigned based on functional context of the regulated operons [2, 23, 24, 26, 27, 36, 37]. All the riboswitch characteristics described here are not only suitable for analyzing the genomic distribution of known sequences but also useful for predicting novel riboswitches.

On top of distinguishing conserved primary sequence in the aptamer domain, detection of riboswitches usually relies on the ability of a single stranded RNA to naturally form

intramolecular interaction via base pairing to minimize the overall free energy in solution. One of the earliest tools developed and made available to the public is the Mfold algorithm [38–40], which predicts the folding pattern for a single stranded RNA using the minimum free energy (MFE) principle that takes into consideration the energies involved in base stacking and closing base pairs. The values of free energies for various ribonucleotide configurations were based on values obtained from applied mathematics as well as experimental observations [41–43]. Mfold also significantly shortens the time for structure prediction by introducing thermodynamic constraints that limit the number of possible base pairings in a given sequence length. If available, enzyme accessibility and chemical modification data can be incorporated into the Mfold algorithm to produce more realistic structure prediction. Although Mfold produces a fairly accurate picture of the overall fold of a RNA molecule, the algorithm does not take the kinetics of transcription into account in its structure prediction, and phylogenetic information is left out of structure prediction.

Another popular suite of computational tools, RNAfold and RNAalifold [44, 45], was introduced shortly after and was built on the concept introduced by Mfold. Like Mfold, RNAfold and RNAalifold calculate base pair probabilities and secondary structures of RNA using the MFE principle. In addition, RNAalifold takes a multiple sequence alignment file for consensus secondary structure prediction. These programs can also generate a list of possible structures within a given energy level as defined by the user. The newest implementation of RNAalifold provides a more realistic treatment of alignment gaps to optimize MFE-based structure prediction and utilizes RIBOSUM-like scoring matrices [44]. As will be shown next, these changes make RNAalifold an attractive alternative to computationally intensive tools based on stochastic context-free grammars. However, RNAfold and RNAalifold do not predict complex structures like pseudoknots, and thermodynamic-based structure prediction on which these tools were based may still oversimplify conditions that contribute to final RNA configuration in the complex cellular environment.

The RNA-PATTERN program was developed to conduct genome-wide searches of conserved RNA motifs including *B12* element (B12 riboswitch), *RFN* element (FMN riboswitch), *THI* element (TPP riboswitch), *LYS* element (lysine riboswitch), S-box (SAM riboswitch) and T-box regulatory motifs [23, 24, 26–28, 36, 37]. The program uses a set of input RNA sequences characterized by conserved sequence motifs and established models of RNA secondary structures. The RNA-PATTERN algorithm describes RNA secondary structures as a set of the following parameters: the number of helices, the length of each helix, the loop lengths, and the topological description of helix pairs. The overall structure arrangement described by RNA-PATTERN includes: (i) consecutive helices, (ii) nested helices, and (iii) pseudoknots. The program not only looks at individual helices but also their relationship with other structure elements within the sequence to measure degree of conservation.

Another program, RNAMotif [46], was built on earlier concept of structural RNA introduced by RNAMOT, Palingol and PatScan motif search tools [47–49]. Like RNA-PATTERN, the RNAMotif algorithm combines both sequence and structure information and includes pseudoknots, triplexes, and quadruplexes for detection of riboswitches with

complex folds. With a motif descriptor provided by the user, the tool helps detect structural RNAs that may constitute riboswitch sequences in a genome or sequence database. In cases where sequence constraints are not specified as part of the descriptor, filters that incorporate thermodynamic stability and sequence complexity may be introduced to reduce false positive detection. Unlike RNA-PATTERN, RNAMotif added context-based search to allow riboswitch annotation. Despite the versatility, RNAMotif is built on a pure pattern language, and certain complex structure motifs may still need to be approximated. As with RNA-PATTERN program, intimate knowledge of a conserved riboswitch template is required for accurate riboswitch detection in diverse genomes.

Similar to RNAMotif, Riboswitch Finder [50] takes into account both primary sequences and secondary structures of RNAs. The program utilizes a fast pattern-matching routine of Stiegler and Zuker [51] to make *de novo* prediction of novel riboswitches. Additionally, like the Mfold algorithm [38–40], energy consideration and RNA folding are taken into account in the construction of a set of possible structures. Pseudoknot is allowed at the strictest level of consensus while mismatches in loop regions are tolerated at the loose consensus level. Despite the accessibility of web implementation of Riboswitch Finder, query search is restricted to a few well-characterized riboswitches (i.e. purine riboswitch), and search parameters are only optimized for *Bacillus subtilis*. As a consequence, the program is of limited use in predicting novel riboswitches. Further, the program does not take advantage of sequence alignment, and evolutionary significance of linked nucleotides is unaccounted for.

Another web-based tool for detection of riboswitches, RibEx (Riboswitch Explorer) [52, 53], expanded the list of riboswitches collected in Riboswitch Finder to 16 known riboswitches as well as 341 predicted riboswitch-like RNA elements. Unlike Riboswitch Finder, RibEx relies on sequence conservation in the aptamer domain with motif identification via the Motif Alignment and Search Tool (MAST) [54]. By default, a predicted riboswitch sequence needs to be present in at least five non-redundant genomes to qualify as a genuine riboswitch. Once a prediction is made, genomic context of the discovered riboswitches is analyzed with an associated gene context tool (GeConT) [55, 56]. Thus, unlike Riboswitch Finder, a discovered riboswitch is placed into context of any neighboring transcriptional terminator and open reading frames, allowing tentative prediction of regulatory mechanism and functional assignment of detected riboswitches. Despite improved functionality over Riboswitch Finder, riboswitch discovery via RibEx is limited by the absence of secondary structure prediction; as a result, RibEX is highly susceptible to indel in homologous sequences as well as riboswitches that are short and low in complexity [57].

### 3. Genome-wide riboswitch identification using covariance models

A revolution in riboswitch prediction came when implementation of hidden Markov model (HMM) allowed comprehensive description of sequence-diverse yet structurally conserved riboswitches and greatly improved investigators' ability to conduct comparative genome analysis in diverse groups of organisms. The Infernal algorithm [58] is based upon stochastic context-free grammar, which is an alternative implementation of HMM and takes

into account base interactions in a given linear sequence. In addition to the sequence-based prediction of HMM, the Infernal algorithm takes advantage of (non-pseudoknotted) secondary structure conservation for the creation of a riboswitch profile termed a covariance model (CM) [59–61]. The combined sequence and structural approach of CM is highly sensitive and specific to homology with either sequence or structural conservation and, importantly, tolerates variations in non-conserved regions of riboswitches. Despite the ability of the Infernal algorithm to detect sequence-diverse riboswitches, prediction accuracy is highly dependent on the quality of input multiple sequence alignment, and user inspection for effective profile creation is essential. If conservation of primary sequence is low, user knowledge of conserved motifs is required for proper sequence alignment; on the other hand, a multitude of diverse yet homologous sequences have to be well-represented to cover all possible sequence covariation and increase predictive power of the model. Construction of a riboswitch CM is also computationally expensive especially in the database search and e-value calibration steps; however, these steps can be skipped at the expense of sensitivity. A newer version of Infernal was introduced recently to speed up the rate of discovery with little penalty to sensitivity for large database searches [62].

Similar to Infernal, CMfinder [63] takes advantage of CMs of existing riboswitch profiles and combines them with motif-based analysis. The algorithm is built in a Bayesian framework that incorporates both mutual information and folding energy model. As with Infernal, CMfinder generally outperforms other programs when sequence conservation is low or when aligned motif is short; detection of riboswitch sequences is also robust against noise introduced by flanking sequences. By using unaligned sequences as input, CMfinder allows large-scale screening of putative riboswitches, and the process can be conducted without user intervention. Riboswitches detected via CMfinder in large scale genome searches include the SAH and *ydaO-yuaA* riboswitches [64, 65].

#### 4. Databases of riboswitch families and regulons

With the rapid accumulation of riboswitch sequences, a database that houses these sequences and group them based on homology become necessary in order to conduct comprehensive comparative genome analysis. Based on concept established by the protein family database Pfam [66], the Rfam database [67, 68] was created shortly after, and the database now houses the largest collection of predicted and functionally validated riboswitches on the web. Detection of riboswitch family members is facilitated via construction of multiple sequence alignments (curated as ‘seed’ alignments), conserved secondary structure predictions, and CMs [58, 68]. Confirmation of newly detected riboswitch sequences allows expansion of seed multiple sequence alignments for increased sensitivity and specificity in iterative searches. Although the ability for riboswitch discovery in diverse genomes is greatly expanded via the collection of seed multiple sequence alignments, searches using CMs can be computationally expensive (see previous section), and any riboswitch not in the 5'-UTR region (for example, riboswitch that functions as ribozyme [16]) may not be recognized. Specifically for small riboswitches, non-functional (‘pseudo’) riboswitches may be indiscernible from genuine ones without further genome context analysis.

Seeking to combine CM-based riboswitch discovery of Rfam with genome context analysis, RegPrecise [69, 70] was introduced to improve annotation of the existing riboswitches. Originally the RegPrecise database was intended for comparative genome analysis of operons that are regulated by functionally orthologous transcription factors in bacterial genomes. The database had since expanded to include not only transcription factor-mediated regulons but also regulons controlled by riboswitches and other RNA regulatory motifs [71]. As with Rfam, detection of riboswitches in diverse genomes is greatly improved via iterative search and expansion of existing riboswitch profiles. Once all possible riboswitch sequences are detected using the Infernal tool, another web-based tool, RegPredict [72], is used for the comparative genomics analysis of the respective riboswitch regulons and annotation of operon content before data deposition into the RegPrecise database. A riboswitch regulon reconstructed in a given bacterial genome by using the RegPredict tool represents a collection of operons regulated by riboswitches from the same Rfam family. By combining analysis of operon content and its upstream regulatory sequence, the RegPredict tool facilitates the construction of a regulon [71]. The riboswitch regulon analysis allows a comprehensive analysis of the metabolic capacity of a single regulator and facilitates gene annotation where analysis based on sequence conservation is not feasible [73]. A collection of orthologous regulons in several related genomes from a narrow taxonomic group of species comprises a regulog, which ultimately permits comparison of regulatory networks between different organisms.

## 5. Computational analysis of riboswitch distribution

From initial estimate, approximately 10 to 15% of bacterial genomes are comprised of non-coding DNAs that have regulatory roles when transcribed into RNAs [74]. In the case of *Bacillus subtilis*, more than 4% of genes are predicted to be under the regulation of riboswitches and assorted *cis*-regulatory elements [75]. Our recent results [71], as well as the results of other previous studies [19, 35], showed that despite the proposed ancient origins of many known metabolite-responsive riboswitches, many other riboswitches and RNA regulatory motifs with yet unknown mechanisms and effectors are much more restricted in their phylogenetic distributions and implied their origins from just a few lineages of bacteria. Some widely distributed riboswitches are regulated by essential cellular cofactors such as thiamin pyrophosphate (TPP) and cobalamin (Table 1). These riboswitches likely have originated from the last common bacterial ancestor. Still, the wide distribution of the predicted *ybp-ykoY* riboswitch points to its possible role in basal metabolism, and this putative riboswitch may respond to ligands associated with pH stress [76, 77]. Other riboswitches such as the tetrahydrofolate (THF) riboswitch and the predicted *ylbH* riboswitch (Table 1) are limited to a select few lineages in Firmicutes where they likely originated [71].

It is worth noting that the distribution of riboswitches does not appear to show an inverse correlation with their lengths (Table 1). Instead, the overall distribution seems to be subjected to strong evolutionary pressure, with only those riboswitches that demonstrate structural stability and strong ligand specificity (both characteristics are more likely in longer sequences) are selected for [77]. So even though a short riboswitch-like element may arise by random mutation, the chance of that sequence producing a stable and functionally



specific structure is highly unlikely and it will be quickly eliminated. Such a phenomenon may account for the limited phylogenetic distribution of SAM-SAH riboswitch despite the ubiquity of its ligands across bacterial genomes [35, 71]. More riboswitches are likely to be discovered in the future as the genome database expands, and diverse sequences in the same riboswitch family are discovered.

## 6. Prediction of riboswitch regulatory mechanism

Even though the expression platforms of riboswitches show little sequence conservation, regulatory mechanism of a riboswitch can usually be predicted based on the spatial relationship of the platform with terminator and ribosomal binding site (RBS) sequences [19]. A common theme that occurred when analyzing the regulatory mechanisms of various riboswitches is that regulation via transcription termination occurs more often in Gram-positive bacteria while the control of translation initiation (RBS occlusion) occurs more in Gram-negative bacteria [19, 78]. It is currently unknown what accounts for this discrepancy in diverse bacterial lineages although it may be related to fundamental metabolic differences between different bacterial lineages. In addition, gene regulation may occur as a result of transcript cleavage via formation of ribozyme such as the *glmS* sequence [18, 79] or recruitment of RNaseP [80]. It may also be possible for the riboswitch-regulated transcript to exert some regulatory effect through an antisense mechanism [19, 36, 81–83]. An unusual phenomenon exists in the occurrence of tandem riboswitches, which appear frequently in glycine riboswitches but much less in others [17, 22, 77, 84]. The occurrence of this tandem arrangement may allow ligand sensing to occur in a wider range of substrate concentration although it is sometimes difficult to predict the physiological response of such arrangement [17, 19, 76, 77]. The potential for riboswitches to form complex regulatory architectures from simple modular components imply great flexibility of structural RNAs in adapting to different modes of gene regulation.

## 7. Current state of riboswitch computational analysis

Given a set of physical constraints such as sequence lengths and ionic environment, the ability of a RNA sequence to fold in a predictable manner has allowed development of algorithms for accurate prediction of RNA structure. Although thermodynamic-based RNA structure prediction allows fast and somewhat realistic prediction of actual riboswitch structure, algorithms taking advantage of multiple sequence alignments with simultaneous construction of CMs remain the gold standard in predicting riboswitch structure in all of their phylogenetic variations. Indeed, the application of CMs, when used in conjunction with functional evidence, can be very powerful in detecting relevant regulatory sequences in highly divergent genomes where primary sequence conservation can be beyond the detection limit of algorithms that only analyze primary sequences. Even though construction and application of CMs can be computationally expensive, several filtering techniques had been developed to make large-scale database searches feasible [85, 86]. The latest advancement in HMM implementation by Infernal promises to make *de novo* prediction even faster with little sacrifice to sensitivity [62].

Although most of RNA structure prediction algorithms, including those that utilize CMs, are not equipped to identify tertiary structure elements that are comprised of non-Watson-Crick and non-nested base pairs [74], comparative sequence analysis allows identification of evolutionarily important elements that may otherwise be difficult to predict using existing mathematical models [19]. Given an ever expanding genome database and metagenomic collection, it has become possible to make an accurate structure prediction of a riboswitch (albeit in the absence of the ligand) before any crystallographic data is available. When coupled to operon context and comparative genome analysis such as those utilized by RegPrecise, it is also possible for researchers to make accurate functional prediction of putative riboswitches to certain extent. Conversely, a functionally characterized riboswitch can be utilized to annotate previously uncharacterized genes when evidence based on sequence homology of the genes is lacking to support their functional characterization [69, 70, 73]. In particular, comparative genomics analysis of riboswitch regulons provided for substantial progress in functional annotation of hypothetical transporters including candidate uptake transporters for amino acids lysine LysW and methionine MetT in *Shewanella oneidensis*, and for vitamins riboflavin RibU and thiamin ThiT in *Bacillus subtilis*. All these predictions were based on co-regulation with the respective amino acid or vitamin biosynthetic genes by a specific metabolite-responsive riboswitch [23, 24, 26, 36].

## 8. Existing challenges in computational analysis of riboswitches

Given that riboswitches exhibit great specificity to their native ligands and can discern a wide range of substrates with large differences in sizes and chemical properties (from a single cation such as magnesium to a large macromolecule like adenosylcobalamin), it is probable that nature has retained a large repertoire of these regulatory elements, and many more riboswitches remained undiscovered [77]. Even though computational analysis of riboswitch has progressed significantly in the past decade, accurate detection of short sequences less than 100 nucleotides long (PreQ1 and mini-*ykkC* riboswitches being notable examples) may still present a challenge for highly divergent genomes. The problem may be alleviated as more riboswitch sequences from diverse genomes are added to the database to capture all possible sequence variations of existing riboswitch families while minimizing false positive detection. For instance, structural RNAs composed of simple repetitive elements may be rejected if there is little evidence of covariation existing between different genomes [33]. Comparative sequence analysis also helps in the detection of indel sequences that are weakly conserved between species and maintaining the fidelity of the CMs [74]. And if possible, genome context information should always be incorporated to validate the existence and regulatory potentials of putative riboswitch sequences [16, 19].

Currently, riboswitch ligand prediction still relies mostly on context of the riboswitch-associated operon, and reliable assignment may prove difficult for genes encoding transporters or proteins of unknown function. The process of ligand validation is made more difficult for sequences that bind to signaling molecules such as cyclic di-GMP [76], which are present only in minute amount under physiological condition. Currently, there is no reliable way to predict ligand specificity purely through the analysis of aptamer sequences as the actual number of ribonucleotides making physical contact with a ligand is usually very small and highly influenced by slight differences in the overall fold of the aptamer domain.



Advancement in ligand prediction will likely come from a high-throughput system that screens *in vitro* transcribed RNAs against a possible pool of ligand candidates that have been vetted through genome analysis.

The knowledge gained by studying examples of riboswitches in living system will hopefully lead to design of novel artificial riboswitches with desirable properties. A riboswitch-based gene regulatory system would provide more flexibility than existing systems that depend on the presence of a proteinaceous mediator, allowing change in gene expression to occur directly upon introduction of a metabolite. Current challenge is to modify aptamer region of a riboswitch to allow sensing of a wide range of molecules and produce a predictable regulatory outcome. One promising approach involves *in vitro* evolution [87–90]. A regulatory outcome that involves formation of a self-cleaving ribozyme is also well-characterized [6, 77, 91–95]. So far, induction of certain artificial riboswitch systems have been exploited to elicit cell motility [96], probe the productivity of certain metabolic pathway [97], and aide in high-throughput enzyme evolution [98]. As more sequence data become available, researchers can better define the parameters and shorten the development cycle to produce a novel riboswitch with desirable ligand and regulatory property [99, 100].

## 9. Conclusions

The study of riboswitch has advanced greatly due to advances in applying existing computational methods to the prediction of structural RNAs. Besides taking into consideration thermodynamic constraints in RNA folding, prediction of riboswitch sequences was aided by the construction of CMs that take into account phylogenetic variations in multiple sequence alignments. Computational analysis of the distribution of riboswitches revealed that while some riboswitches originated in the last common ancestor of bacteria, others arose in a few select lineages. The analysis of riboswitch sequences also reveal that they are mechanistically diverse, and basic building blocks of a riboswitch regulatory system may be modified to allow additional regulatory control. The computational study of riboswitch demonstrated great flexibility of RNAs as ligand recognition and gene regulatory elements and deserves greater attention for future experimental application.

## Acknowledgments

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area. Additional funding was provided by National Institute of General Medical Sciences (R01GM077402) and by the Russian Academy of Sciences via the Molecular and Cellular Biology program.

## References

1. Nudler E, Mironov AS. The riboswitch control of bacterial metabolism. *Trends Biochem Sci.* 2004; 29:11–17. [PubMed: 14729327]
2. Gelfand MS, Mironov AA, Jomantas J, Kozlov YI, Perumov DA. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet.* 1999; 15:439–442. [PubMed: 10529804]

3. Kochhar S, Paulus H. Lysine-induced premature transcription termination in the *lysC* operon of *Bacillus subtilis*. *Microbiology*. 1996; 142(Pt 7):1635–1639. [PubMed: 8757727]
4. Nou X, Kadner RJ. Adenosylcobalamin inhibits ribosome binding to *btuB* RNA. *Proc Natl Acad Sci U S A*. 2000; 97:7190–7195. [PubMed: 10852957]
5. Mandal M, Breaker RR. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol*. 2004; 5:451–463. [PubMed: 15173824]
6. Soukup GA, Breaker RR. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA*. 1999; 5:1308–1325. [PubMed: 10573122]
7. Soukup GA, DeRose EC, Koizumi M, Breaker RR. Generating new ligand-binding RNAs by affinity maturation and disintegration of allosteric ribozymes. *RNA*. 2001; 7:524–536. [PubMed: 11345431]
8. Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR. Genetic control by a metabolite binding mRNA. *Chem Biol*. 2002; 9:1043. [PubMed: 12323379]
9. Winkler W, Nahvi A, Breaker RR. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*. 2002; 419:952–956. [PubMed: 12410317]
10. Winkler WC, Cohen-Chalamish S, Breaker RR. An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A*. 2002; 99:15908–15913. [PubMed: 12456892]
11. Epshtein V, Mironov AS, Nudler E. The riboswitch-mediated control of sulfur metabolism in bacteria. *Proc Natl Acad Sci U S A*. 2003; 100:5052–5056. [PubMed: 12702767]
12. McDaniel BA, Grundy FJ, Artsimovitch I, Henkin TM. Transcription termination control of the *S* box system: direct measurement of *S*-adenosylmethionine by the leader RNA. *Proc Natl Acad Sci U S A*. 2003; 100:3083–3088. [PubMed: 12626738]
13. Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*. 2002; 111:747–756. [PubMed: 12464185]
14. Winkler WC, Nahvi A, Sudarsan N, Barrick JE, Breaker RR. An mRNA structure that controls gene expression by binding *S*-adenosylmethionine. *Nat Struct Biol*. 2003; 10:701–707. [PubMed: 12910260]
15. Wickiser JK, Winkler WC, Breaker RR, Crothers DM. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell*. 2005; 18:49–60. [PubMed: 15808508]
16. Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I, Wickiser JK, Breaker RR. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A*. 2004; 101:6421–6426. [PubMed: 15096624]
17. 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]
18. Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*. 2004; 428:281–286. [PubMed: 15029187]
19. Barrick JE, Breaker RR. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol*. 2007; 8:R239. [PubMed: 17997835]
20. Dambach MD, Winkler WC. Expanding roles for metabolite-sensing regulatory RNAs. *Curr Opin Microbiol*. 2009; 12:161–169. [PubMed: 19250859]
21. Roth A, Breaker RR. The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem*. 2009; 78:305–334. [PubMed: 19298181]
22. Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell*. 2003; 113:577–586. [PubMed: 12787499]
23. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Comparative genomics of thiamin biosynthesis in prokaryotes. New genes and regulatory mechanisms. *J Biol Chem*. 2002; 277:48949–48959. [PubMed: 12376536]

24. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Res.* 2003; 31:6748–6757. [PubMed: 14627808]
25. Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* 2003; 17:2688–2697. [PubMed: 14597663]
26. Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS. Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res.* 2002; 30:3141–3151. [PubMed: 12136096]
27. Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS. Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA.* 2003; 9:1084–1097. [PubMed: 12923257]
28. Vitreschak AG, Mironov AA, Lyubetsky VA, Gelfand MS. Comparative genomic analysis of T-box regulatory systems in bacteria. *RNA.* 2008; 14:717–735. [PubMed: 18359782]
29. Grundy FJ, Henkin TM. tRNA as a positive regulator of transcription antitermination in *B. subtilis*. *Cell.* 1993; 74:475–482. [PubMed: 8348614]
30. Grundy FJ, Winkler WC, Henkin TM. tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. *Proc Natl Acad Sci U S A.* 2002; 99:11121–11126. [PubMed: 12165569]
31. Gutiérrez-Preciado A, Henkin TM, Grundy FJ, Yanofsky C, Merino E. Biochemical features and functional implications of the RNA-based T box regulatory mechanism. *Microbiology and Molecular Biology Reviews.* 2009; 73:36–61. [PubMed: 19258532]
32. Lesnik EA, Fogel GB, Weekes D, Henderson TJ, Levene HB, Sampath R, Ecker DJ. Identification of conserved regulatory RNA structures in prokaryotic metabolic pathway genes. *Biosystems.* 2005; 80:145–154. [PubMed: 15823413]
33. Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, Sudarsan N, Neph S, Tompa M, Ruzzo WL, Breaker RR. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.* 2007; 35:4809–4819. [PubMed: 17621584]
34. Corbino KA, Barrick JE, Lim J, Welz R, Tucker BJ, Puskarz I, Mandal M, Rudnick ND, Breaker RR. Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biol.* 2005; 6:R70. [PubMed: 16086852]
35. Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.* 2010; 11:R31. [PubMed: 20230605]
36. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory systems. *Nucleic Acids Res.* 2004; 32:3340–3353. [PubMed: 15215334]
37. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J Biol Chem.* 2003; 278:41148–41159. [PubMed: 12869542]
38. Zuker M. Computer prediction of RNA structure. *Methods Enzymol.* 1989; 180:262–288. [PubMed: 2482418]
39. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003; 31:3406–3415. [PubMed: 12824337]
40. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* 1981; 9:133–148. [PubMed: 6163133]
41. Tinoco I Jr, Uhlenbeck OC, Levine MD. Estimation of secondary structure in ribonucleic acids. *Nature.* 1971; 230:362–367. [PubMed: 4927725]
42. Gralla J, Crothers DM. Free energy of imperfect nucleic acid helices. II. Small hairpin loops. *J Mol Biol.* 1973; 73:497–511. [PubMed: 4715014]
43. Uhlenbeck OC, Borer PN, Dengler B, Tinoco I Jr. Stability of RNA hairpin loops: A 6 -C m -U 6. *J Mol Biol.* 1973; 73:483–496. [PubMed: 4715013]

44. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics*. 2008; 9:474. [PubMed: 19014431]
45. Hofacker IL, Fekete M, Stadler PF. Secondary structure prediction for aligned RNA sequences. *J Mol Biol*. 2002; 319:1059–1066. [PubMed: 12079347]
46. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res*. 2001; 29:4724–4735. [PubMed: 11713323]
47. Laferriere A, Gautheret D, Cedergren R. An RNA pattern matching program with enhanced performance and portability. *Comput Appl Biosci*. 1994; 10:211–212. [PubMed: 7517334]
48. Billoud B, Kontic M, Viari A. Palingol: a declarative programming language to describe nucleic acids' secondary structures and to scan sequence database. *Nucleic Acids Res*. 1996; 24:1395–1403. [PubMed: 8628670]
49. Pesole G, Liuni S, D'Souza M. PatSearch: a pattern matcher software that finds functional elements in nucleotide and protein sequences and assesses their statistical significance. *Bioinformatics*. 2000; 16:439–450. [PubMed: 10871266]
50. Bengert P, Dandekar T. Riboswitch finder--a tool for identification of riboswitch RNAs. *Nucleic Acids Res*. 2004; 32:W154–159. [PubMed: 15215370]
51. Zuker M. Calculating nucleic acid secondary structure. *Curr Opin Struct Biol*. 2000; 10:303–310. [PubMed: 10851192]
52. Abreu-Goodger C, Merino E. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res*. 2005; 33:W690–692. [PubMed: 15980564]
53. Abreu-Goodger C, Ontiveros-Palacios N, Ciria R, Merino E. Conserved regulatory motifs in bacteria: riboswitches and beyond. *Trends Genet*. 2004; 20:475–479. [PubMed: 15363900]
54. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*. 1998; 14:48–54. [PubMed: 9520501]
55. Ciria R, Abreu-Goodger C, Morett E, Merino E. GeConT: gene context analysis. *Bioinformatics*. 2004; 20:2307–2308. [PubMed: 15073003]
56. Martinez-Guerrero CE, Ciria R, Abreu-Goodger C, Moreno-Hagelsieb G, Merino E. GeConT 2: gene context analysis for orthologous proteins, conserved domains and metabolic pathways. *Nucleic Acids Res*. 2008; 36:W176–180. [PubMed: 18511460]
57. Singh P, Bandyopadhyay P, Bhattacharya S, Krishnamachari A, Sengupta S. Riboswitch detection using profile hidden Markov models. *BMC Bioinformatics*. 2009; 10:325. [PubMed: 19814811]
58. Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. *Bioinformatics*. 2009; 25:1335–1337. [PubMed: 19307242]
59. Eddy SR. A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. *BMC Bioinformatics*. 2002; 3:18. [PubMed: 12095421]
60. Eddy SR, Durbin R. RNA sequence analysis using covariance models. *Nucleic Acids Res*. 1994; 22:2079–2088. [PubMed: 8029015]
61. Sakakibara Y, Brown M, Hughey R, Mian IS, Sjolander K, Underwood RC, Haussler D. Stochastic context-free grammars for tRNA modeling. *Nucleic Acids Res*. 1994; 22:5112–5120. [PubMed: 7800507]
62. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013; 29:2933–5. [PubMed: 24008419]
63. Yao Z, Weinberg Z, Ruzzo WL. CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics*. 2006; 22:445–452. [PubMed: 16357030]
64. Block KF, Hammond MC, Breaker RR. Evidence for widespread gene control function by the *ydaO* riboswitch candidate. *J Bacteriol*. 2010; 192:3983–3989. [PubMed: 20511502]
65. Edwards AL, Reyes FE, Heroux A, Batey RT. Structural basis for recognition of S-adenosylhomocysteine by riboswitches. *RNA*. 2010; 16:2144–2155. [PubMed: 20864509]
66. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. The Pfam protein families database. *Nucleic Acids Res*. 2012; 40:D290–301. [PubMed: 22127870]

67. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* 2013; 41:D226–232. [PubMed: 23125362]
68. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. Rfam: updates to the RNA families database. *Nucleic Acids Res.* 2009; 37:D136–140. [PubMed: 18953034]
69. Novichkov PS, Brettin TS, Novichkova ES, Dehal PS, Arkin AP, Dubchak I, Rodionov DA. RegPrecise web services interface: programmatic access to the transcriptional regulatory interactions in bacteria reconstructed by comparative genomics. *Nucleic Acids Res.* 2012; 40:W604–608. [PubMed: 22700702]
70. Novichkov PS, Laikova ON, Novichkova ES, Gelfand MS, Arkin AP, Dubchak I, Rodionov DA. RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res.* 2010; 38:D111–118. [PubMed: 19884135]
71. Sun EI, Leyn SA, Kazanov MD, Saier MH, Novichkov PS, Rodionov DA. Comparative genomics of metabolic capacities of regulons controlled by cis-regulatory RNA motifs in bacteria. *BMC genomics.* 2013; 14:597. [PubMed: 24060102]
72. Novichkov PS, Rodionov DA, Stavrovskaya ED, Novichkova ES, Kazakov AE, Gelfand MS, Arkin AP, Mironov AA, Dubchak I. RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res.* 2010; 38:W299–307. [PubMed: 20542910]
73. Gutiérrez-Preciado A, Merino E. Elucidating Metabolic Pathways and Digging for Genes of Unknown Function in Microbial Communities: The Riboswitch Approach. *Clin Microbiol Infect.* 2012; 18:35–39. [PubMed: 22647046]
74. Westhof E. The amazing world of bacterial structured RNAs. *Genome Biol.* 2010; 11:108. [PubMed: 20236470]
75. Winkler WC. Metabolic monitoring by bacterial mRNAs. *Arch Microbiol.* 2005; 183:151–159. [PubMed: 15750802]
76. Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N, Breaker RR. Challenges of ligand identification for riboswitch candidates. *RNA Biol.* 2011; 8:5–10. [PubMed: 21317561]
77. Breaker RR. Prospects for riboswitch discovery and analysis. *Mol Cell.* 2011; 43:867–879. [PubMed: 21925376]
78. 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]
79. Klein DJ, Ferre-D'Amare AR. Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science.* 2006; 313:1752–1756. [PubMed: 16990543]
80. Altman S, Wesolowski D, Guerrier-Takada C, Li Y. RNase P cleaves transient structures in some riboswitches. *Proc Natl Acad Sci U S A.* 2005; 102:11284–11289. [PubMed: 16061811]
81. Mellin JR, Tiensuu T, Becavin C, Gouin E, Johansson J, Cossart P. A riboswitch-regulated antisense RNA in *Listeria monocytogenes*. *Proc Natl Acad Sci U S A.* 2013; 110:13132–13137. [PubMed: 23878253]
82. Groisman EA, Cromie MJ, Shi Y, Latifi T. A Mg<sup>2+</sup>-responding RNA that controls the expression of a Mg<sup>2+</sup> transporter. *Cold Spring Harb Symp Quant Biol.* 2006; 71:251–258. [PubMed: 17381304]
83. Andre G, Even S, Putzer H, Burguiere P, Croux C, Danchin A, Martin-Verstraete I, Soutourina O. S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res.* 2008; 36:5955–5969. [PubMed: 18812398]
84. Welz R, Breaker RR. Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*. *RNA.* 2007; 13:573–582. [PubMed: 17307816]
85. Weinberg Z, Ruzzo WL. Sequence-based heuristics for faster annotation of non-coding RNA families. *Bioinformatics.* 2006; 22:35–39. [PubMed: 16267089]
86. Weinberg Z, Ruzzo WL. Exploiting conserved structure for faster annotation of non-coding RNAs without loss of accuracy. *Bioinformatics.* 2004; 20(Suppl 1):i334–341. [PubMed: 15262817]
87. Toulme JJ, Di Primo C, Boucard D. Regulating eukaryotic gene expression with aptamers. *FEBS Lett.* 2004; 567:55–62. [PubMed: 15165893]



88. Suess B, Hanson S, Berens C, Fink B, Schroeder R, Hillen W. Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Res.* 2003; 31:1853–1858. [PubMed: 12655001]
89. Grate D, Wilson C. Inducible regulation of the *S. cerevisiae* cell cycle mediated by an RNA aptamer-ligand complex. *Bioorg Med Chem.* 2001; 9:2565–2570. [PubMed: 11557344]
90. Werstuck G, Green MR. Controlling gene expression in living cells through small molecule-RNA interactions. *Science.* 1998; 282:296–298. [PubMed: 9765156]
91. Breaker RR. Engineered allosteric ribozymes as biosensor components. *Curr Opin Biotechnol.* 2002; 13:31–39. [PubMed: 11849955]
92. Roth A, Breaker RR. Selection in vitro of allosteric ribozymes. *Methods Mol Biol.* 2004; 252:145–164. [PubMed: 15017047]
93. Seetharaman S, Zivarts M, Sudarsan N, Breaker RR. Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nat Biotechnol.* 2001; 19:336–341. [PubMed: 11283591]
94. Thompson KM, Syrett HA, Knudsen SM, Ellington AD. Group I aptazymes as genetic regulatory switches. *BMC Biotechnol.* 2002; 2:21. [PubMed: 12466025]
95. Yen L, Svendsen J, Lee JS, Gray JT, Magnier M, Baba T, D'Amato RJ, Mulligan RC. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature.* 2004; 431:471–476. [PubMed: 15386015]
96. Sinha J, Reyes SJ, Gallivan JP. Reprogramming bacteria to seek and destroy an herbicide. *Nat Chem Biol.* 2010; 6:464–70. [PubMed: 20453864]
97. Fowler CC, Brown ED, Li Y. Using a riboswitch sensor to examine coenzyme B(12) metabolism and transport in *E. coli*. *Chem Biol.* 2010; 17:756–765. [PubMed: 20659688]
98. Michener JK, Smolke CD. High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metab Eng.* 2012; 14:306–16. [PubMed: 22554528]
99. Dohno C, Kohyama I, Kimura M, Hagihara M, Nakatani K. A synthetic riboswitch that operates using a rationally designed ligand-RNA pair. *Angew Chem Int Ed Engl.* 2013; 52:9976–9. [PubMed: 24030897]
100. Ceres P, Trausch JJ, Batey RT. Engineering modular 'ON' RNA switches using biological components. *Nucleic Acids Res.* 2013; 41:10449–61. [PubMed: 23999097]
101. Argaman L, Hershsberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol.* 2001; 11:941–950. [PubMed: 11448770]
102. Kim JN, Roth A, Breaker RR. Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2'-deoxyguanosine. *Proc Natl Acad Sci U S A.* 2007; 104:16092–16097. [PubMed: 17911257]
103. Roth A, Winkler WC, Regulski EE, Lee BW, Lim J, Jona I, Barrick JE, Ritwik A, Kim JN, Welz R, Iwata-Reuyl D, Breaker RR. A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat Struct Mol Biol.* 2007; 14:308–317. [PubMed: 17384645]
104. Ames TD, Rodionov DA, Weinberg Z, Breaker RR. A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate. *Chem Biol.* 2010; 17:681–685. [PubMed: 20659680]
105. Weinberg Z, Regulski EE, Hammond MC, Barrick JE, Yao Z, Ruzzo WL, Breaker RR. The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. *RNA.* 2008; 14:822–828. [PubMed: 18369181]
106. Fuchs RT, Grundy FJ, Henkin TM. The S(MK) box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nat Struct Mol Biol.* 2006; 13:226–233. [PubMed: 16491091]
107. Winkler WC, Breaker RR. Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol.* 2005; 59:487–517. [PubMed: 16153177]
108. Tripp HJ, Schwalbach MS, Meyer MM, Kitner JB, Breaker RR, Giovannoni SJ. Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. *Environ Microbiol.* 2009; 11:230–8. [PubMed: 19125817]



### Highlights

- We review computational methods and tools for prediction of riboswitch sequences and functions.
- Cross-genome sequence analyses allow to find novel riboswitches and to predict their structure and function.
- A high level of conservation among riboswitches is useful for their computational identification in genomic sequences.
- Computational tools for riboswitch identification and RNA secondary structure prediction are available on the Web.

Table 1

Functional overview and distribution of riboswitches and riboswitch-like elements.

Rfam ID	Name	Effector	Length <sup>1</sup>	Total number <sup>2</sup>	Representative lineages and riboswitch numbers <sup>2</sup>	Functional categories of riboswitch regulons	References <sup>3</sup>
RF00230	T-box	uncharged tRNA	564	1134	<i>Deinococcus</i> (10), <i>Bacillales</i> (313), <i>Lactobacillales</i> (434), <i>Clostridia</i> (337), <i>Actinomycetales</i> (15), <i>Chloroflexi</i> (25)	Amino acid (AA) biosynthesis, AA transporters, Amino acyl-tRNA synthetases	[28] (b), [31] (e), [36] (a/b)
RF00059	TPP ( <i>THI</i> element)	thiamin pyrophosphate	318	564	<i>Proteobacteria</i> : $\alpha$ (64), $\beta$ (24), $\gamma$ (128), $\delta$ (21), $\epsilon$ ; <i>Deinococcales</i> (10), <i>Bacillales</i> (77), <i>Lactobacillales</i> (59), <i>Clostridia</i> (62), <i>Actinomycetales</i> (45), <i>Cyanobacteria</i> (14), <i>Chlorobiales</i> (11), <i>Bacteroidales</i> (19), <i>Thermotogales</i> (16), <i>Chloroflexi</i> (14), <i>Fusobacteriales</i> , <i>Spirochaetales</i> , Archaea, Eukaryotes	Thiamin biosynthesis, Thiamin & precursor transporters	[23] (a/b), [52] (g)
RF00174	Cobalamin ( <i>B12</i> element)	adenosylcobalamin	441	536	<i>Proteobacteria</i> : $\alpha$ (112), $\beta$ (45), $\gamma$ (96), $\delta$ (23), $\epsilon$ ; <i>Deinococcales</i> (12), <i>Bacillales</i> (18), <i>Lactobacillales</i> (1), <i>Clostridia</i> (50), <i>Actinomycetales</i> (27), <i>Cyanobacteria</i> (20), <i>Chlorobiales</i> (34), <i>Bacteroidales</i> (65), <i>Thermotogales</i> (20), <i>Chloroflexi</i> (13), <i>Acidobacteria</i> , <i>Chlamydia</i> , <i>Fusobacteriales</i> , <i>Spirochaetales</i>	B12 biosynthesis, B12 & precursor transporters, Cobalt transporters, Isozymes of B12-dependent enzymes	[19] (d), [27], [35] (e/g/h), [37] (a/b)

Rfam ID	Name	Effector	Length <sup>1</sup>	Total number <sup>2</sup>	Representative lineages and riboswitch numbers <sup>2</sup>	Functional categories of riboswitch regions	References <sup>3</sup>
RF00504	Glycine	glycine	220	324	<i>Proteobacteria</i> : $\alpha$ (78), $\beta$ (42), $\gamma$ (69), $\delta$ (6), $\epsilon$ ; <i>Bacillales</i> (30), <i>Lactobacillales</i> (18), <i>Clostridiales</i> (52), <i>Actinomycetales</i> (23), <i>Chloroflexi</i> (6), <i>Fusobacteriales</i>	Glycine metabolism, Serine metabolism, Glycine transporters	[17], [52] (g), [108]
RF00050	FMN (RFN element)	flavin mononucleotide	221	233	<i>Proteobacteria</i> : $\alpha$ (20), $\beta$ (15), $\gamma$ (52), $\delta$ (10), $\epsilon$ ; <i>Deinococcales</i> (5), <i>Bacillales</i> (34), <i>Lactobacillales</i> (38), <i>Clostridia</i> (38), <i>Actinomycetales</i> (3), <i>Chlorobiales</i> (2), <i>Thermotogales</i> (8), <i>Chloroflexi</i> (8), <i>Fusobacteriales</i>	Riboflavin biosynthesis, Riboflavin transporters	[26] (a/b), [52] (g)
RF00080	<i>yypP-ykoY</i>	?	221	232	<i>Proteobacteria</i> : $\alpha$ (8), $\beta$ (20), $\gamma$ (73), $\delta$ (7); <i>Deinococcales</i> (5), <i>Bacillales</i> (23), <i>Lactobacillales</i> (28), <i>Clostridiales</i> (35), <i>Actinomycetales</i> (23), <i>Cyanobacteria</i> (4), <i>Thermotogales</i> (1), <i>Chloroflexi</i> (5)	Miscellaneous	[52] (g), [101] (a)
RF00162	SAM (S-box)	S-adenosylmethionine	231	257	<i>Proteobacteria</i> : $\alpha$ , $\gamma$ , $\delta$ (4), $\epsilon$ ; <i>Deinococcales</i> (9), <i>Bacillales</i> (138), <i>Lactobacillales</i> (2), <i>Clostridia</i> (72), <i>Actinomycetales</i> (1), <i>Cyanobacteria</i> (2), <i>Chlorobiales</i> (14), <i>Bacteroidales</i> , <i>Thermotogales</i> (1), <i>Chloroflexi</i> (14), <i>Acidobacteria</i> , <i>Fusobacteriales</i>	Methionine biosynthesis, Methionine & SAM recycling, Cysteine biosynthesis, Methionine transporters	[35] (e/g/h), [36] (a/b)

Rfam ID	Name	Effector	Length <sup>1</sup>	Total number <sup>2</sup>	Representative lineages and riboswitch numbers <sup>2</sup>	Functional categories of riboswitch regions	References <sup>3</sup>
RF00168	Lysine (L-box)	lysine	274	186	<i>Proteobacteria</i> : $\gamma$ (72); <i>Bacillales</i> (38), <i>Lactobacillales</i> (29), <i>Clostridia</i> (38), <i>Thermotogales</i> (9), <i>Acidobacteria</i> , <i>Fusobacteriales</i>	Lysine biosynthesis, Lysine transporters	[24] (a/b), [52] (g)
RF00167	Purine (G-box)	guanine, adenine	113	141	<i>Proteobacteria</i> : $\gamma$ (13), $\delta$ , $\epsilon$ ; <i>Bacillales</i> (56), <i>Lactobacillales</i> (29), <i>Clostridia</i> (41), <i>Thermotogales</i> (2), <i>Fusobacteriales</i>	Purine metabolism, Purine & precursor transporters	[52] (g), [102] (d)
RF01051	GEMM	cyclic di-GMP	136	89	<i>Proteobacteria</i> : $\beta$ (1), $\gamma$ (36), $\delta$ (5); <i>Deinococcales</i> (3), <i>Bacillales</i> (9), <i>Clostridia</i> (29), <i>Cyanobacteria</i> (6)	Polysaccharide degradation	[33] (a/d/e/h)
RF00522	PreQ1	pre-queuosine <sub>1</sub>	70	72	<i>Proteobacteria</i> : $\alpha$ (1), $\beta$ , $\gamma$ (9), $\delta$ , $\epsilon$ ; <i>Bacillales</i> (28), <i>Lactobacillales</i> (11), <i>Clostridia</i> (23), <i>Fusobacteriales</i>	Queuosine biosynthesis, Queuosine & precursor transporters	[103] (d)
RF01068	mini- <i>ykkC</i>	?	56	67	<i>Proteobacteria</i> : $\alpha$ (26), $\beta$ (11), $\gamma$ (28), $\delta$ (1); <i>Cyanobacteria</i> (1)	Urea and agmatine utilization, Multidrug resistance transporter	[33] (a/d/e/h)
RF01055	MOCO	molybdenum or tungsten cofactor?	255	62	<i>Proteobacteria</i> : $\alpha$ (2), $\gamma$ (50); <i>Deinococcales</i> (1), <i>Clostridia</i> (6), <i>Chlorobiales</i> (1), <i>Thermotogales</i> (2)	Molybdenum cofactor biosynthesis, Molybdenum & tungsten transporters	[33] (a/d/e/h)
RF00379	<i>ydaO-yuaA</i>	ATP	335	59	<i>Proteobacteria</i> : $\delta$ (1); <i>Bacillales</i> (18), <i>Clostridia</i> (18), <i>Actinomycetales</i> (14), <i>Cyanobacteria</i> (8)	Potassium transporters	[16], [52] (g)
RF00442	<i>ykkC-ykkD</i>	?	181	58	<i>Proteobacteria</i> : $\alpha$ (10), $\beta$ (7), $\gamma$ (7); <i>Bacillales</i> (12), <i>Lactobacillales</i> (2), <i>Clostridia</i> (11), <i>Actinomycetales</i> (3), <i>Cyanobacteria</i> (6)	Urea and agmatine utilization, Multidrug resistance transporter	[16], [33] (a/d/e/h), [35] (e/g/h)

Rfam ID	Name	Effector	Length <sup>1</sup>	Total number <sup>2</sup>	Representative lineages and riboswitch numbers <sup>2</sup>	Functional categories of riboswitch regions	References <sup>3</sup>
RF00380	<i>ykoK</i> (M-box)	magnesium	278	48	<i>Proteobacteria</i> : $\gamma$ (6); <i>Bacillales</i> (7), <i>Lactobacillales</i> (15), <i>Clostridia</i> (10), <i>Actinomycetales</i> (10)	Magnesium transporters,	[16]
RF00234	<i>glnS</i>	glucosamine-6-phosphate	236	44	<i>Deinococcales</i> (3), <i>Bacillales</i> (18), <i>Lactobacillales</i> (1), <i>Clostridiales</i> (17), <i>Actinomycetales</i> , <i>Chloroflexi</i> (5), <i>Fusobacteriales</i>	Aminosugar biosynthesis	[18], [52] (g), [79]
RF01057	SAH	S-adenosylhomocysteine	172	27	<i>Proteobacteria</i> : $\beta$ (14), $\gamma$ (7); <i>Actinomycetales</i> (6)	Methionine biosynthesis, Methionine & SAM recycling,	[33] (a/d/e/h)
RF00521	SAM-Alpha (SAM-II)	S-adenosylmethionine	85	39	<i>Proteobacteria</i> : $\alpha$ (39), $\beta$ , $\gamma$ ; <i>Bacteroidales</i>	Methionine biosynthesis	[34], [35] (e/g/h)
RF00517	<i>serC</i>	?	55	32	<i>Proteobacteria</i> : $\alpha$ (32)	Serine metabolism	[34]
RF01831	THF	tetrahydrofolate	179	23	<i>Lactobacillales</i> (12), <i>Clostridia</i> (11)	Folate biosynthesis, Folate transporters	[35] (e/g/h), [104] (a/d/h)
RF01054	PreQ1-II	pre-queuosine <sub>1</sub>	132	17	<i>Lactobacillales</i> (17)	Queuosine & precursor transporters	[33] (a/d/e/h)
RF01070	<i>sucA</i>	?	102	14	<i>Proteobacteria</i> : $\beta$ (14)	Citric acid cycle	[33] (a/d/e/h)
RF01739	<i>glnA</i>	glutamine	274	13	<i>Cyanobacteria</i> (13)	Nitrogen metabolism	[17], [35] (e/g/h)
RF00634	SAM-IV	S-adenosylmethionine	152	13	<i>Actinomycetales</i> (13)	Methionine biosynthesis,	[33] (a/d/e/h), [105] (a/d/e/h)
RF01767	Smk-box (SAM-III)	S-adenosylmethionine	148	13	<i>Lactobacillales</i> (13)	Methionine & SAM recycling	[106]
RF01727	SAM-SAH	S-adenosylmethionine, S-adenosylhomocysteine	54	13	<i>Proteobacteria</i> : $\alpha$ (13)	Methionine & SAM recycling	[35] (e/g/h)
RF00518	<i>speF</i>	?	426	12	<i>Proteobacteria</i> : $\alpha$ (12)	Ornithine degradation	[34]
RF01724	SAM-Chlorobi	S-adenosylmethionine	113	11	<i>Chlorobiales</i> (11)	Methionine & SAM recycling	[35] (e/g/h)
RF00516	<i>yibH</i>	?	139	10	<i>Bacillales</i> (10)	Miscellaneous	[16]
RF00520	<i>ybhL</i>	?	92	8	<i>Proteobacteria</i> : $\alpha$ (8)	Miscellaneous	[34]
RF01056	Mg sensor	magnesium	119	5	<i>Proteobacteria</i> : $\gamma$ (5)	Miscellaneous	[82] (a)

The majority of the table is adopted from the previous comparative genomic analysis of riboswitches by Sun et al. [71].

<sup>1</sup> Nucleotide length of riboswitches is based on the 'seed' alignments in the Rfam database [67].

<sup>2</sup> Total number of riboswitch sites and numbers riboswitches detected in each lineage (given in parenthesis) are according to Sun et al. [71]. Representative lineages<sup>2</sup>: incorporates data from Winkler & Breaker (2005) [107] and Barrick & Breaker (2007) [19].

<sup>3</sup> References are given for papers that have used the following computational tools and databases for riboswitch identification: (a) Mfold, RNAfold and/or RNAalifold; (b) RNA-PATTERN; (c) RNAMotif; (d) Infernal; (e) Rfam; (f) Riboswitch Finder; (g) RibEX; (h) CMfinder.