

GENE REGULATION BY RIBOSWITCHES

Maumita Mandal and Ronald R. Breaker

Riboswitches are complex folded RNA domains that serve as receptors for specific metabolites. These domains are found in the non-coding portions of various mRNAs, where they control gene expression by harnessing allosteric structural changes that are brought about by metabolite binding. New findings indicate that riboswitches are robust genetic elements that are involved in regulating fundamental metabolic processes in many organisms.

Expression of the numerous genes in a cell must be tailored to provide the appropriate levels of RNA and protein production at all times. Gene-control systems face an enormous challenge. They must coordinate numerous tasks that a typical cell carries out on an ever-changing cycle, and they must interpret many different chemical and physical signals. Even the simplest, single-celled organisms need to modulate the expression of hundreds of genes in response to a myriad of cellular needs and environmental cues. Gene-control systems, therefore, must have the ability to respond precisely to specific signals, rapidly bring about their intended genetic effect, and have **sufficient dynamic character to fine-tune the level of expression for hundreds of different genes.**

It has long been recognized that organisms make extensive use of protein-based control systems to modulate gene expression. The networks of protein signalling and gene-control factors are most complex in eukaryotes, where numerous factors typically work together to influence transcription, translation, mRNA processing/degradation and other mechanisms that control the levels of gene products in a cell¹. In addition, there is an emerging awareness of the role of RNA factors in gene control, which are of fundamental importance to such processes as developmental timing and other gene-silencing events^{2,3}. A variety of microRNAs (miRNAs) and related short-interfering RNAs (siRNAs) function by a series of protein-mediated processing events that eventually lead to the production of ~22-nucleotide-long fragments that form base-paired complexes with their target mRNAs. These

interactions cause inactivation of the targeted gene by subsequent nuclease processing or by other non-nucleolytic mechanisms^{4,5}. Although protein factors are required for these gene-silencing mechanisms to operate, they highlight the fact that many organisms rely on RNA molecules for critical regulatory tasks, and suggest that additional roles for RNA in gene-control systems might be possible.

Recent studies have demonstrated that the involvement of RNA in fundamental gene-control processes is indeed more extensive. Numerous mRNAs in prokaryotes carry complex folded domains, which are known as riboswitches^{6,7}, within the non-coding portions of their polynucleotide chains. Each riboswitch directly binds a specific metabolite, without the obligate involvement of a protein factor, and then controls gene expression by harnessing changes in RNA structure to influence transcription elongation, translation initiation, or other aspects of the process that leads to protein production^{7–12}. Here, we provide an overview of riboswitch mechanisms and we discuss the evidence which indicates that an expanded set of RNA switches might be of fundamental importance to the control of genes in many modern biological systems.

Allosteric RNAs and genetic switches

Roles for metabolite-binding RNAs in modern cells could not easily have been predicted from the earliest studies on protein enzymes and receptors. Polypeptide chains made of the standard 20 amino acids are an excellent medium for forming complex molecular architectures that are capable of serving as receptors or

Department of Molecular,
Cellular and Developmental
Biology, Yale University,
P.O. Box 208103,
New Haven, Connecticut
06520-8103, USA.
Correspondence to R.R.B.
e-mail:
ronald.breaker@yale.edu
doi:10.1038/nrm1403

Box 1 | The ligand-binding potential of RNA aptamers

Numerous RNA aptamers have been created by using *in-vitro*-evolution methods^{21–23}. Typically, pools of relatively short random-sequence nucleic acids are subjected to a selection and amplification process whereby molecules that bind to a particular target are selectively amplified. A large number of aptamers have been created for protein and small-molecule targets⁹², which indicates that RNA has the structural sophistication that is needed to serve as a sensor element. One of the first demonstrations of the precise molecular-recognition capability of RNA was with a theophylline-dependent aptamer, which discriminates against caffeine by more than 10,000 fold (REF. 93). Caffeine differs from theophylline by only a single methyl group.

In several instances, there are both engineered and natural aptamers for the same compound. For example, guanine-binding⁹⁴ and FMN-binding⁹⁵ aptamers have been isolated using *in vitro* evolution. In each instance, the engineered aptamers are smaller and carry fewer essential nucleotides and structural elements than their natural counterparts. Furthermore, the engineered aptamers are less able to discriminate against closely related analogues, and have poorer binding affinities for their respective targets. These observations imply that there is a higher demand for specificity and affinity with natural aptamers, which is achieved by the formation of more complex motifs.

ALLOSTERIC ENZYME

An enzyme that is triggered to alter its function in response to the binding of a target compound at a site that is distal from the active site of the enzyme.

RNA WORLD

A hypothetical time in early evolution, before the emergence of DNA and proteins, when biological processes were guided entirely by RNA molecules.

RIBOZYME

A nucleic-acid molecule that folds to form an active site and catalyzes a chemical reaction.

APTAMER

An RNA domain, either engineered or natural, that forms a precise three-dimensional structure and selectively binds a target molecule.

IN VITRO EVOLUTION

The use of various separation and amplification techniques that serve to mimic Darwinian evolution and create variants of proteins or nucleic acids that have new or improved functions.

UTR

(Untranslated region). Stretches of untranslated sequences located upstream and downstream of the coding region of an mRNA.

IN-LINE PROBING

An RNA-structure probing method that can be used to examine secondary-structure models and to determine whether RNAs undergo substantive structural rearrangements under different incubation conditions.

as biocatalysts that bind to the myriad of compounds in living systems. Furthermore, ALLOSTERIC ENZYMES that change shape and function in response to ligand binding have been known since the 1960s, and one could easily imagine how protein genetic switches might be made^{13–15}. **DNA and RNA seemed well suited for their roles in storing and transferring genetic information, whereas proteins were viewed as the molecules of biochemical action.** Therefore, the discovery of the first example of a protein genetic factor (the *lac* repressor) that responded to metabolites or other macromolecules, also in the early 1960s (REF. 16), fitted nicely with the idea that the division of labour among the biopolymers was clearly defined.

There was little need to invoke the existence of riboswitches when allosteric modulation of protein enzymes and protein genetic factors seemed sufficient to eventually satisfy most gene-control needs in even the most complex organisms. However, two questions remained to be addressed before RNA could be dismissed as a possible metabolite-sensing genetic element. First, how might early life forms have controlled molecular functions before the emergence of proteins? It is plausible that life passed through an 'RNA WORLD', in which genetic storage and primitive biological processes were thought to be carried out entirely by RNA^{17,18}. Support for this theory was delivered by the discovery of RIBOZYMES^{19,20}, which proved that RNA could form complex shapes and catalyze reactions of biological importance. If there were complex organisms in the RNA world, and particularly as such organisms began the transition into the 'DNA–RNA–protein world' of modern times, then it is likely that RNA switches would have been necessary to exert control over these biological processes.

Second, does RNA even have sufficient structural potential to serve as a highly specific receptor for small organic compounds? Indeed, numerous RNAs or APTAMERS^{21–23} (BOX 1) that bind to a diverse collection of targets are being generated in laboratories by using IN VITRO EVOLUTION. These engineered receptors can be

appended onto catalytic RNAs to create allosteric ribozymes, wherein ligand binding to an aptamer domain triggers a catalytic response in an adjoining ribozyme^{24–28}. These findings strongly indicate that RNA molecules have the structural and functional sophistication needed to function as natural genetic switches in ancient or even modern settings. This led a number of researchers to speculate, mostly in unpublished forums, that modern organisms might exploit ligand-binding RNAs for this very role^{29–34}.

Riboswitch discoveries

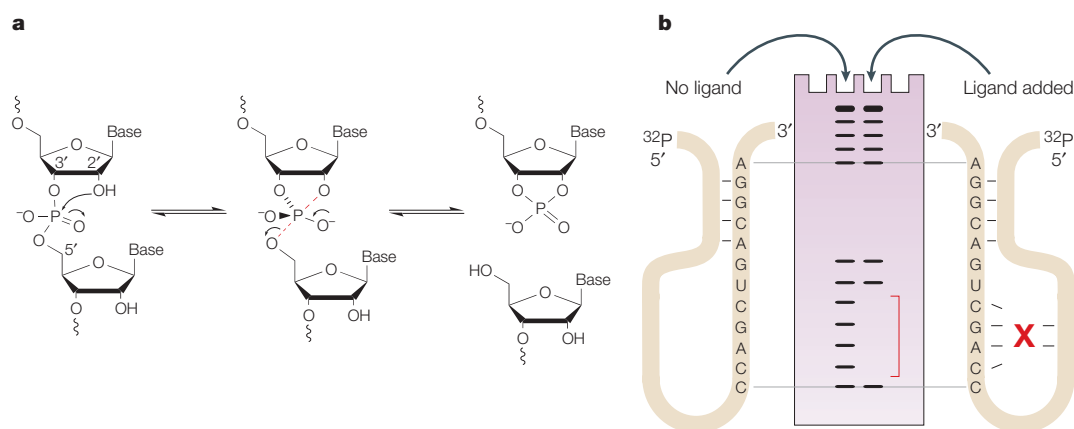
Beginning with the discovery of the *lac* repressor, an enormous body of data emerged that described a diversity of protein genetic factors that respond to various metabolites and signalling compounds. However, a number of puzzling instances of gene-control phenomena had been accumulating in the scientific literature for many years. Particularly intriguing to us was a series of reports that described the unique genetic-control characteristics of the *btuB* gene of *Escherichia coli* and the *cob* operon of *Salmonella typhimurium*. These genes are responsible for maintaining adequate levels of coenzyme B₁₂ in the cell either by importing or synthesizing this complex metabolite.

It was known for some time that coenzyme B₁₂ is involved in some form of feedback inhibition to control the expression of these genes, and that the 5'-untranslated region (5'-UTR) of each mRNA carries a conserved sequence element that is required for this control^{35–37}. However, repeated attempts to identify a protein factor that would convert a chemical signal (increasing coenzyme-B₁₂ concentrations) into a genetic response had been unsuccessful^{32,38}. Since a protein factor could not be implicated, it seemed likely that a direct interaction between the coenzyme and the mRNA was being made. Various attempts to show that the RNA bound the coenzyme directly, in the absence of proteins, were also unsuccessful^{32,36} and this hypothesis remained unproven.

We chose to examine the question of direct binding between the metabolite and the *btuB* RNA of *E. coli* using several approaches. First, we reasoned that natural metabolite-binding mRNAs would undergo some degree of structural reorganization on ligand binding, much like that which is observed with engineered allosteric RNAs. To search for metabolite-induced allosteric changes, we used a method known as IN-LINE PROBING, which takes advantage of the fact that the spontaneous rate of RNA cleavage is highly dependent on the local structure at each internucleotide linkage (BOX 2)^{39–41}. For example, a 315-nucleotide section of the *btuB* 5'-UTR has numerous changes in its pattern of spontaneous cleavage when coenzyme B₁₂ is present, which indicates that the RNA undergoes a substantial level of allosteric reorganization⁶ (FIG. 1). Similar results were found when in-line probing was conducted with 5'-UTR sections from the *btuB* and *cob* mRNAs of *S. typhimurium*^{6,42}.

Further examination of the RNAs using equilibrium dialysis, mutational analyses and gene-expression studies

Box 2 | RNA analysis by in-line probing



The dominant pathway by which RNA degrades involves an 'in-line' nucleophilic attack by the 2' oxygen on the adjacent phosphorus centre. Efficient cleavage takes place only if the attacking 2' oxygen, the phosphorus and the departing 5' oxygen of the phosphodiester linkage approach a linear configuration, hence the term 'in-line' (see figure, part a; red dashed line).

The speed at which spontaneous cleavage occurs⁴⁰ depends on the local structural context in which each RNA linkage exists^{39,41}. Linkages that reside in highly structured regions of a folded RNA (for example, a base-paired helix) typically resist cleavage because the relevant atoms are not held in an in-line configuration. However, if folding contacts do not restrict their structure, linkages occasionally take on in-line geometry through random motions and therefore are subject to an enhanced rate of spontaneous cleavage. In some instances, linkages within parts of a folded RNA molecule that have taken on a tertiary structure can have accelerated spontaneous cleavage because a linkage is fortuitously held in an optimal configuration for a nucleophilic attack³⁹. In part b of the figure, a radiolabelled RNA that is incubated in the absence of ligand (left) gives a distinctive pattern of spontaneous-cleavage products on separation by gel electrophoresis. This same RNA construct has an altered pattern of cleavage products (red bracket) when incubated with ligand X, which indicates that ligand binding alters the folding of these nucleotides.

confirmed that direct binding of coenzyme B₁₂ by mRNAs does occur and that this interaction is required for proper expression. The *btuB* riboswitch from *E. coli* binds coenzyme B₁₂ with an apparent DISSOCIATION CONSTANT (K_D) of ~300 nM, but rejects most other chemical analogues of the coenzyme that were examined⁶. The extraordinary light sensitivity of coenzyme B₁₂ (it has a half life of seconds in ambient light)^{6,43} resulted in technical challenges for these experiments, and might also have caused difficulties in previous attempts to prove direct binding. Nevertheless, coenzyme-B₁₂-dependent riboswitches are now recognized as a widespread genetic-control element in prokaryotes, as bioinformatics approaches have identified more than 200 variants of this element in diverse groups of bacteria^{42,44}. Furthermore, the resulting phylogenetic alignments of riboswitch variants are particularly useful when working to establish a model for the secondary structure of these complex folded RNAs.

There was also evidence in the literature for the metabolite-dependent control of several other genes, but in each case no protein factors had been identified. The list of metabolites included thiamin pyrophosphate (TPP)^{34,45}, flavin mononucleotide (FMN)^{31,46–48}, S-adenosylmethionine (SAM)⁴⁹, guanine^{50,51} and lysine^{52–56}. In each instance, these early reports have recently proven to correspond to known riboswitches that control the expression of adjoining genes in

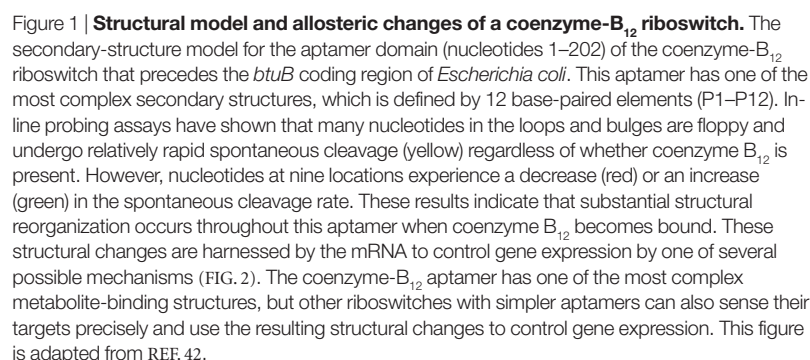
response to changing metabolite concentrations. For any riboswitch candidate, the challenge for researchers is to provide convincing evidence that the RNA serves as the molecular recognition element for the metabolite in the absence of any possible protein factor. In addition, evidence for a mechanism of gene control is needed to justify its classification as a riboswitch. This level of proof of the existence of TPP- and FMN-dependent riboswitches^{7,57,58} was generated almost simultaneously with that for the coenzyme-B₁₂ riboswitch. Some of the structural and functional characteristics of riboswitches revealed by these studies are outlined below.

Anatomy of a riboswitch

In prokaryotes, most riboswitches are located in the 5'-UTRs of mRNAs, and are typically composed of two functional domains¹⁰ (FIG. 2). The first structure that is encountered as the nascent mRNA emerges from the RNA polymerase during transcription is the aptamer domain. Like engineered aptamers, each natural aptamer serves as a molecular sensor embedded within the riboswitch, where it selectively recognizes its corresponding target molecule within the complex sea of other metabolites. The aptamer takes the place of a conventional protein factor that, otherwise, would serve as the sensory element. Like many protein factors, the aptamer must selectively recognize a metabolite with the appropriate affinity. Unlike protein, with 20

DISSOCIATION CONSTANT

The equilibrium constant for a ligand binding to its receptor, which, in the case of riboswitches, represents the concentration of ligand that is required to convert half of the aptamers that are present in a mixture to their ligand-bound form.



RIBOSOME-BINDING SITE
Also known as the Shine-Dalgarno sequence, it is a short stretch of conserved nucleotides that is situated several nucleotides upstream of the start codon in prokaryotic mRNAs. This sequence is recognized by the ribosome during translation initiation.

The second structure is the EXPRESSION PLATFORM⁵⁷, which is typically located immediately downstream from the aptamer domain, but in many instances the two domains overlap to some extent. The role of the expression platform is to transduce metabolite-binding events into gene-control consequences by allosteric modulation of the structure of the 5'-UTR. The expression platform varies substantially between even closely related organisms, which reflects the diversity in expression-platform function and the diversity of RNA structures that can achieve the intended goal. Riboswitches can modulate gene expression by controlling the efficiency of translation initiation^{6,57}, the transcription elongation of mRNA^{7,58–64}, and presumably even the stability and splicing of mRNA transcripts^{65,66}. Therefore, expression platforms that differ substantially in sequence permit riboswitch variants to control expression at a variety of levels.

Gene-control mechanisms of a riboswitch

In the case of the coenzyme- B_{12} riboswitch, two main mechanisms for gene control are apparent (FIG. 2a). The first mechanism regulates RNA transcription and involves the ligand-dependent formation of an INTRINSIC TERMINATOR stem. Intrinsic terminators are extended stem-loop structures, which are typically followed by a run of six or more U residues, that cause RNA polymerase to abort transcription before the coding portion of the mRNA has been made^{67,68}. When coenzyme B_{12} is not present in sufficient quantities, transcription of an mRNA that is associated with the coenzyme- B_{12} riboswitch produces a nascent mRNA wherein the aptamer domain remains uncomplexed with ligand. The unbound aptamer domain permits formation of an 'ANTI-TERMINATOR' stem, which precludes formation of the intrinsic terminator stem and thereby permits transcription of the complete mRNA. However, when coenzyme- B_{12} concentrations are adequate, the nascent mRNA binds to a coenzyme- B_{12} molecule and the allosteric change in structure permits the intrinsic terminator stem to form. Transcription termination results and gene expression is prevented because the coding region of the mRNA is not made. Although the only evidence for this mechanism with the coenzyme- B_{12} riboswitch comes from sequence analysis⁴⁴, direct experimental evidence for this mechanism exists for other riboswitch classes^{7,57-61}.

The second mechanism that is used by the coenzyme- B_{12} riboswitch functions at the level of translation initiation. Just as allosteric changes in aptamer structure can control the formation of intrinsic terminator and anti-terminator stems, coenzyme- B_{12} binding causes structural changes in full-length mRNAs to control access to the RIBOSOME-BINDING SITE (FIG. 2b). Specifically, it has been shown³⁵ that ribosomes are not able to form stable complexes with the *btuB* mRNA of *E. coli* when coenzyme B_{12} is present in an *in vitro* assay. This observation, coupled with sequence⁴⁴ and biochemical^{6,42} evidence for coenzyme-dependent alternative folding by the *btuB* 5'-UTR, supports this model for gene control.

In rare instances, it appears that both transcription and translation can be controlled simultaneously. This

would be possible if the transcription-terminator stem is formed by including base pairing with the ribosome-binding site. This combination of mechanisms would allow newly initiated mRNA transcripts to be aborted by the terminator stem, whereas transcripts whose synthesis had already passed the point of transcription termination could exploit the same structural change to prevent translation by occluding the ribosome-binding

site. Further experiments are needed to examine whether this combination of mechanisms is indeed exploited by some riboswitches.

The diversity of riboswitches

Riboswitches and the RNA world. The coenzyme- B_{12} riboswitch fits the hypothesis that allosteric RNAs could have emerged early in evolution. Sequence analyses^{42,44}

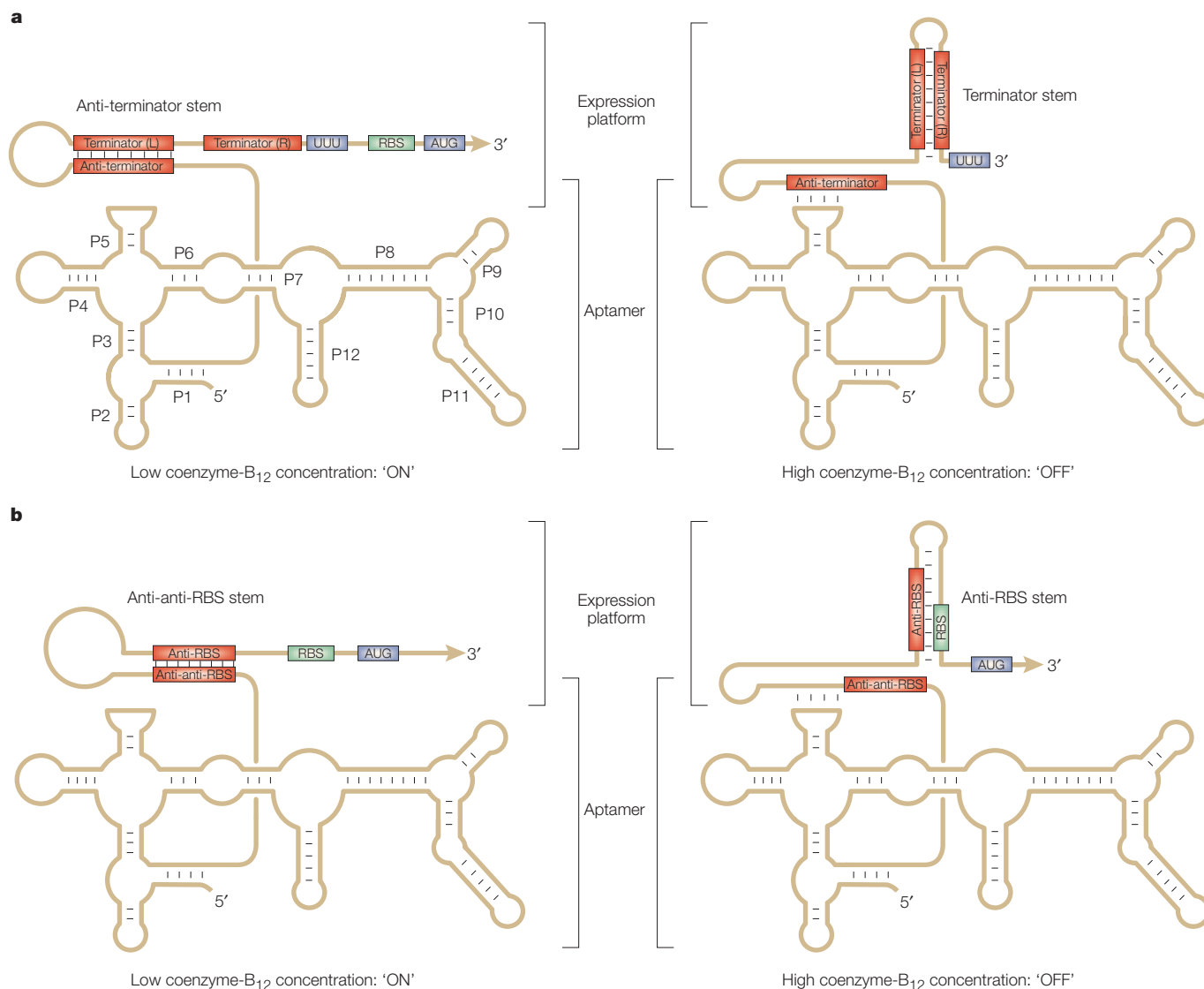


Figure 2 | Coenzyme- B_{12} -riboswitch structure and gene-control function. Each riboswitch is composed of an aptamer domain that remains largely conserved and an expression platform that varies more widely during evolution. The coenzyme- B_{12} aptamer controls gene expression by two main mechanisms, as dictated by the architecture of the expression platform. **a** | Transcription control is mediated by the mutually exclusive formation of an anti-terminator stem (left) or an intrinsic terminator stem (right). Ligand binding requires that base pairing be formed between the loop of P5 and a stretch of nucleotides that would otherwise form an anti-terminator. In its simplest form, metabolite binding favours the formation of the base pairs that comprise the intrinsic terminator stem, at the expense of the anti-terminator stem. The involvement of additional structures could provide more complex kinetics of RNA folding and, therefore, more diverse mechanisms for gene control. Red shadings identify the complementary sequences that comprise the terminator and anti-terminator stems, wherein L and R identify the left and right shoulders of the terminator stem. **b** | Translation control is mediated by the mutually exclusive formation of an anti-anti-RBS (ribosome-binding site) stem and an anti-RBS stem. This mechanism also involves the mutually exclusive formation of base-paired elements; however, pairing to the RBS prevents the ribosome from binding to the mRNA. In the absence of the target metabolite, the anti-anti-RBS yields a structurally unencumbered RBS and the ribosome can bind. AUG designates the translation start codon, and the boxed UUU represents a repetitive U sequence that is required for transcription termination in bacteria of the *Bacillus/Clostridium* group. P1–P12 represent base-paired elements.

indicate that the coenzyme- B_{12} riboswitch is widespread in prokaryotes — as would be expected if its origin were ancient. Furthermore, the chemical structure of coenzyme B_{12} carries an adenosyl moiety and another moiety that is strikingly similar to a nucleotide. This RNA-like character of portions of coenzyme B_{12} has led to speculation that this compound was important to organisms of the RNA world, as might be a number of other coenzymes that are essential to nearly

all modern organisms^{69–72}. From this viewpoint, it seemed reasonable to speculate that numerous riboswitches for many different compounds could have been present in ancient organisms. Although the eventual emergence of proteins probably displaced many of these RNA-based gene-control elements, it also seems possible that modern cells might still use some of these motifs to control genes that are relevant to the metabolites being sensed.

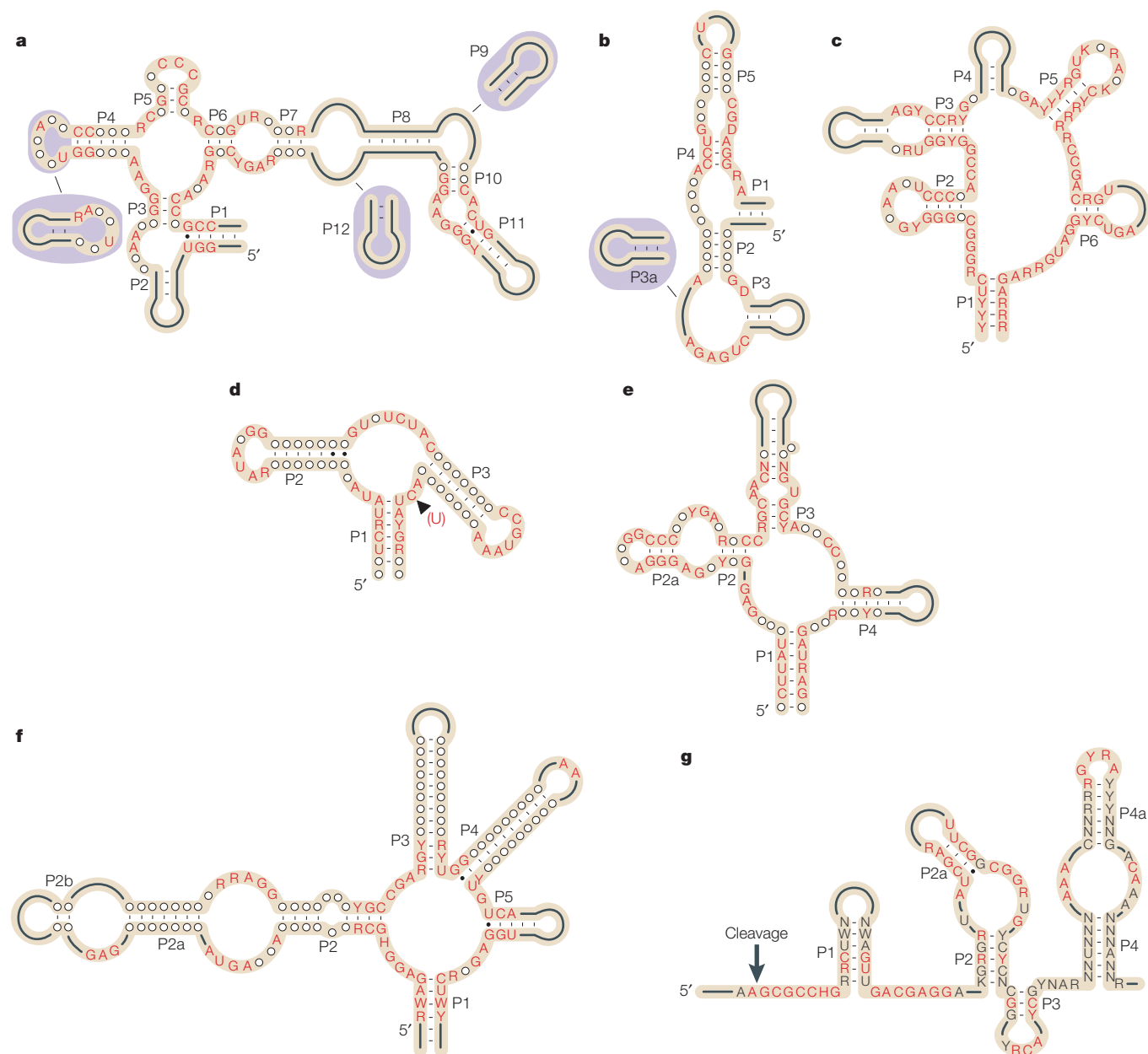


Figure 3 | Consensus sequences and structures for known riboswitch classes. For each riboswitch class (a–g), a consensus sequence and structural motif is shown. Nucleotides in red are conserved in at least 90% of the representatives examined, circles reflect the presence of a nucleotide of any base identity, thin lines identify base pairing, and thick lines represent non-conserved regions. Distinct pairing elements (P) are numbered. (d) The adenine riboswitch is a variant of the guanine riboswitch with a critical C-to-U mutation in the aptamer core (U). (g) The glucosamine-6-phosphate (GlcN6P) riboswitch is also a ribozyme that catalyzes a self-cleaving reaction at the site indicated by the arrow. Data supporting the consensus patterns for (a) coenzyme B_{12} (one of two main variants of the aptamer)^{6,42,44}, (b) thiamine pyrophosphate (TPP)^{7,34,65,73}, (c) flavin mononucleotide (FMN)⁴⁶, (d) guanine/adrenaline^{59,74}, (e) S-adenosylmethionine (SAM)^{49,60,62}, (f) lysine^{64,77} and (g) GlcN6P^{78,79} riboswitches have been published elsewhere.

How many modern riboswitches? Recent studies have made it clear that riboswitches are a diverse and widespread form of regulation. At the time of writing, there are reports of seven distinct classes of natural aptamers that recognize eight metabolites (FIG. 3). With one exception, differently folded RNAs are required to form distinct binding pockets for different metabolites. The exception is an aptamer class that was first shown to bind guanine, but can swap its molecular specificity to adenine by a single point mutation (see below). In this instance, no substantive change in the molecular architecture of the aptamer is required to swap binding specificity. Therefore, other riboswitch structural classes might also include variants whose purpose is to respond to chemically related targets. The potential for the discovery of new riboswitch classes, either as variants of known classes or as entirely new motifs, seems assured. We anticipate that additional riboswitches will be reported as researchers seek to explain other gene-control mysteries.

Thiamine pyrophosphate riboswitches. Each of the riboswitch ligands that have been identified so far functions in fundamental metabolic pathways. TPP, which is a target of one of the most widespread riboswitch classes, is commonly involved as a coenzyme for decarboxylase enzymes. The TPP riboswitch consensus sequence (FIG. 3b), like the coenzyme itself, appears in all three domains of life^{65,66,73}. The riboswitch controls both transcription⁵⁸ and translation^{7,58} using mechanisms that are similar to those described for the coenzyme-B₁₂ riboswitch. Interestingly, the TPP aptamer domain from the *thiC* gene of *E. coli* has ~1,000-fold tighter binding to TPP (apparent K_D of ~100 nM) compared with thiamine, which lacks the pyrophosphate moiety. Therefore, RNAs can form binding pockets that make productive contacts with phosphate moieties. This is surprising, given that the phosphodiester backbone of RNA is exceptionally polyanionic, and so the TPP riboswitch must overcome inherent repulsive effects to accommodate the two additional phosphates of its target molecule.

Flavin mononucleotide riboswitches. Many of the characteristics of the TPP riboswitch are shared with another riboswitch class that selectively binds FMN (FIG. 3c), which is another key coenzyme that is required for the function of many redox-active enzymes. The FMN-specific aptamer has been identified in numerous bacteria by database searching³¹, and is one of the most common riboswitches in prokaryotes along with the coenzyme-B₁₂- and TPP-specific classes (J. E. Barrick and R.R.B., unpublished observations). Similarly, evidence exists for the control of both transcription and translation by FMN riboswitches that are guided by allosteric structural changes^{57,58}. The riboswitch from the *ribDEAHT* operon of *B. subtilis* binds FMN with an apparent K_D of ~10 nM (REF. 57). Again, this is nearly 1,000-fold better than the apparent K_D that was measured for

riboflavin, which differs from FMN only by the absence of a single phosphate residue. This high level of molecular discrimination, which is observed with all the riboswitch classes, is an essential characteristic if RNA is to perform adequately as a genetic switch. In this case, riboflavin is just one of several close analogues of FMN that are present in the cell as a result of biosynthetic processes, or that might be encountered in the environment. If the RNA switch did not reject such compounds, then gene expression might not be accurately coupled to the concentration of the active coenzyme.

Guanine and adenine riboswitches. One of the smallest metabolite-binding domains is present in a purine-specific class of riboswitches. The consensus motif is composed of a three-stem junction that is interspersed with stretches of conserved nucleotides in the central bulge and at the tips of the stem-loops (FIG. 3d). In most instances, this motif binds guanine with an apparent K_D value of ~5 nM (for example, the *xpt-pbuX* variant from *B. subtilis*), whereas adenine is completely rejected by the binding pocket^{59,74}. These observations are consistent with genetic evidence indicating the *xpt-pbuX* operon is controlled by guanine. However, several examples exist, upstream of adenine-metabolism genes, wherein the aptamer carries a C-to-U mutation within its core^{74,75}. Interestingly, these specialized variants of the guanine riboswitch have altered ligand specificity, and now reject guanine and bind adenine with an apparent K_D of ~300 nM.

S-adenosylmethionine riboswitches. Gram-positive prokaryotes such as *B. subtilis* are particularly rich in both diversity and repeat occurrences of riboswitches. All known riboswitch classes are present in *B. subtilis*, whereas more-distantly related organisms such as *E. coli* carry only a subset of the motifs that are depicted in FIG. 3. SAM-dependent riboswitches^{60–62} are the most frequently occurring within Gram-positive organisms. Although SAM is used as a coenzyme for methylase enzymes, in certain bacteria this compound also seems to be an important genetic signal. For example, *B. subtilis* has 11 TRANSCRIPTIONAL UNITS (which encode 26 genes) that are controlled by SAM riboswitches⁴⁹. *B. anthracis* is even more enriched for metabolite-binding RNAs, and has 17 transcriptional units that are controlled by this riboswitch class. The mRNAs in this REGULON encode gene products that are related to sulphur metabolism, including genes that are involved in the biosynthesis of cysteine, methionine and SAM^{49,76}. This arrangement suggests that SAM serves as a master molecule, the concentration of which is monitored by numerous riboswitch variants in each cell. Specifically, a drop in SAM concentration might be indicative of a general lack of sulphur-containing compounds. This is independently detected by each of the mRNAs in the regulon, which results in an increase in the expression of genes that are required for sulphur metabolism.

TRANSCRIPTIONAL UNIT
An RNA transcript, such as mRNA, that is transcribed separately. In the case of operons, one transcriptional unit can encode several proteins.

REGULON
A collection of separate genes, the expression of which is controlled as a unit by a specific signalling compound or factor.

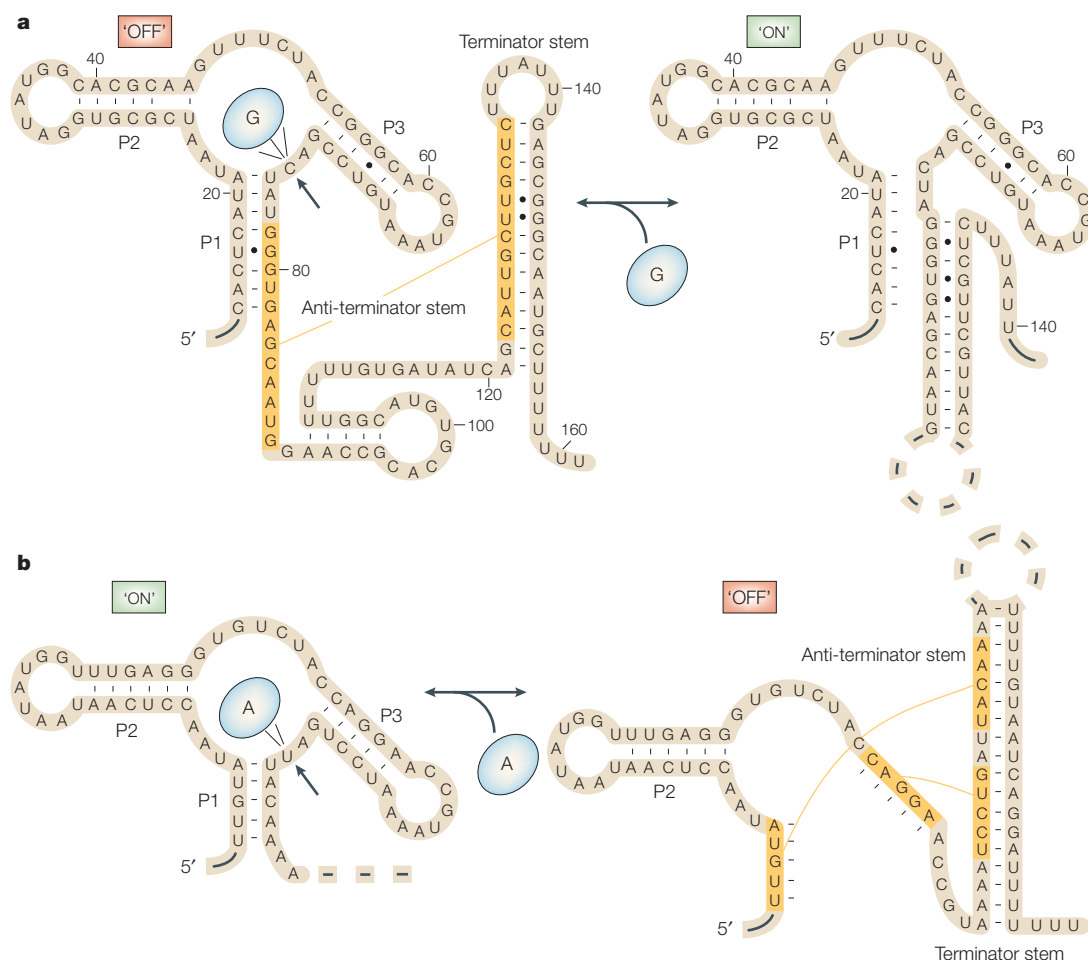


Figure 4 | Differences in ligand specificity and switch direction for purine-specific riboswitches. a | The guanine-specific riboswitch from the *xpt-pbuX* operon of *B. subtilis*⁵⁹. Guanine (G) binding involves base pairing to a cytidine (C) residue of the aptamer and causes formation of an intrinsic terminator stem that turns off gene expression. P1–P3 represent pairing elements that are essential for formation of the aptamer. When G is absent, the anti-terminator stem is formed at the expense of P1. **b** | The adenine-specific riboswitch from the *ydhL* gene of *B. subtilis*⁷⁴. A uridine (U) residue is responsible for base pairing to the ligand. Adenine (A) binding prevents formation of the terminator, thereby causing gene expression to switch on. In this case, the anti-terminator structure is part of the aptamer (stems P1 and P2). In the absence of adenine, some sequences of P1 and P2 now form a part of the intrinsic terminator. Some types of riboswitches probably become kinetically trapped due to the extensive formation of base-paired stems, and are not able to switch freely between 'ON' and 'OFF' states as depicted here. Extensive anti-terminator and terminator base-pairing interactions would be slow to interconvert, and so gene-control decisions would be made by each riboswitch molecule as it commits to a specific folding pathway as established by the presence or absence of bound ligand. Thin black lines represent base pairs, thick black lines represent non-conserved regions and residues that are involved in the formation of terminator or anti-terminator stems are shaded yellow.

Lysine riboswitches. All the compounds that are sensed by the riboswitches described above are fundamental in their biochemical importance and, like coenzyme B₁₂, have substantial chemical similarity to RNA. At least in the case of the purine-specific riboswitches, this RNA-like character is exploited by the aptamer for molecular recognition. Guanine and adenine, even when integrated into larger compounds such as coenzyme B₁₂ and SAM, provide chemical moieties that can be bound through simple Watson–Crick base-pairing^{59,74}. But, if primitive, RNA-world organisms were going to evolve ever-greater metabolic sophistication, molecular recognition by aptamers and by ribozymes would have had to expand to accommodate many

other compounds that have no chemical similarity to nucleic acids.

Perhaps this capability is best manifested by the lysine-dependent class of riboswitches^{59,63,64,77}. Although lysine is much smaller than most other target molecules, the consensus sequence and secondary structure for this aptamer is one of the largest known (FIG. 3f). The demands for accurate molecular recognition by this riboswitch are extreme, as there are many compounds in the cell, including high concentrations of similar amino acids, that must be excluded from triggering changes in gene expression. Again, the lysine aptamer has remarkable selectivity for its target molecule, in this case, despite the absence of nucleotide-like moieties. The

isolated aptamer binds L-lysine with an apparent K_D of 1 μ M, and discriminates against closely related lysine analogues such as ornithine, cadaverine, 5-hydroxylysine, and even the D-isomer of lysine⁶⁴. These observations indicate that RNA aptamers of sufficient complexity can selectively bind even small and structurally unconstrained target molecules. Therefore, we should place no preconceived restrictions on the range of targets that riboswitches could possibly recognize.

Glucosamine-6-phosphate riboswitch. A very different mechanism for riboswitch action controls the *glmS* gene of many Gram-positive organisms. The *glmS* gene encodes an enzyme that produces glucosamine-6-phosphate (GlcN6P). This compound selectively triggers a novel SELF-CLEAVING RIBOZYME that is located in the 5'-UTR of the mRNA^{78,79} (FIG. 3g). The ribozyme is activated 1,000 fold when incubated with GlcN6P, which reduces its half-life for cleavage from over 4 hours to less than 15 seconds. In engineered reporter constructs, mutations that deactivate the ribozyme lead to derepression of gene expression. It is not yet clear whether ribozyme action leads to gene repression by inducing structural alterations in the cleaved UTR, or whether some other processes are involved. Regardless, the integration of ribozymes and aptamers adds new possibilities for gene control by riboswitches, just as molecular engineers have created similar RNAs for biosensing applications^{24–28}.

Diversity of gene-control mechanisms. Like the *glmS* ribozyme, several of the riboswitch classes described above also showcase the diversity of mechanisms that are available for controlling gene expression. For example, the adenine-binding variant associated with the *ydhl* gene of *B. subtilis* serves as an 'ON' switch. In previous examples that were studied in detail, including the *xpt-pbuX* motif (FIG. 4a), riboswitches of all classes turn gene expression off in response to ligand binding. This common observation makes sense when it is considered that most control tasks related to metabolism will require downregulation of biosynthetic genes when an adequate level of the desired compound is produced. In the case of *ydhl*, the function of the gene product seems to be purine transport that expels excess adenine if the need arises⁷⁵. Activation of *ydhl* expression is achieved by the adenine-specific riboswitch by a variation of the transcription-termination mechanism that was described for the coenzyme-B₁₂ riboswitch. In this instance, sequences that comprise the right side of the terminator stem are also required to form part of the aptamer core (FIG. 4b). So, adenine binding precludes formation of the terminator stem, and gene expression proceeds at a high level⁷⁴. RNA can undoubtedly serve either as an ON switch or an OFF switch. However, genetic necessity does not demand a large number of metabolite-induced ON switches, and so these are expected to be far less common.

As with other classes, lysine-dependent riboswitches seem to control gene expression by both transcription-termination^{63,64} and translation-initiation mechanisms⁷⁷.

Although OFF switches have been demonstrated by experiment and seem to be the most common^{63,64}, sequence data indicate that expression of lysine-degradation genes might be activated by ON switches⁷⁷. These main types of gene-control mechanisms are undoubtedly sufficient for most of the regulatory needs in prokaryotes. However, many other mechanisms could be coupled to metabolite binding by RNA, and if these are realized then the level of sophistication of RNA-based gene-control elements would become even closer to that observed for protein factors.

An alternative form of gene regulation seems to be used with TPP-dependent riboswitches of eukaryotes. In both fungi and plants, sequence elements that match the TPP aptamer consensus and bind TPP *in vitro* are affiliated with INTRONS that reside in the 5'- or 3'-UTR of thiamine-metabolism genes⁶⁵. In addition, these elements seem to be essential for splicing events that are necessary for the expression of the gene⁶⁶. Presumably, ligand binding precludes pre-mRNA splicing, which would exemplify a new mechanism whereby high concentrations of a metabolite downregulate expression of biosynthetic or import genes for that compound. Similar mechanisms can be envisioned for lysine riboswitches or any of the other classes as well, although it not yet known whether eukaryotes carry additional riboswitches.

Other RNA switches

Many prokaryotic RNA elements that have been known for years also function, in some respects, as genetic switches, including small non-coding RNAs^{80,81}. Even eukaryotic organisms are known to use RNA elements that interface with proteins to control gene expression or to guide the subcellular location of RNA. For example, recent work on RNA elements that are embedded within cytokine genes indicates that they form complexes with an RNA-dependent protein kinase, which ultimately leads to downregulation of protein synthesis⁸². Similar systems probably control gene expression by a variety of distinct mechanisms in prokaryotes and eukaryotes, and it has been proposed that modern organisms contain many such elements⁸³. But are there distinctions to be made between riboswitches and some of these other RNA elements?

These elements cannot simply be classified by whether a protein factor is involved. For example, thermosensor RNAs^{84,85} and T-box RNAs⁷⁶ (BOX 3) do not necessarily rely on protein factors for function, but also lack some of the characteristics of metabolite-binding riboswitches. Similarly, a metabolite-binding riboswitch could bind its target, and then recruit a protein factor as part of its control mechanism. So, it is difficult in some instances to draw clear-cut distinctions between riboswitches and other gene-control processes that involve RNA structures. However, we think that there is sufficient difference between metabolite-binding riboswitches and RNAs that respond to other factors such that separate classifications are warranted.

SELF-CLEAVING RIBOZYME

Five of the nine known natural ribozymes catalyze self-cleavage using an internal phosphoester transfer reaction.

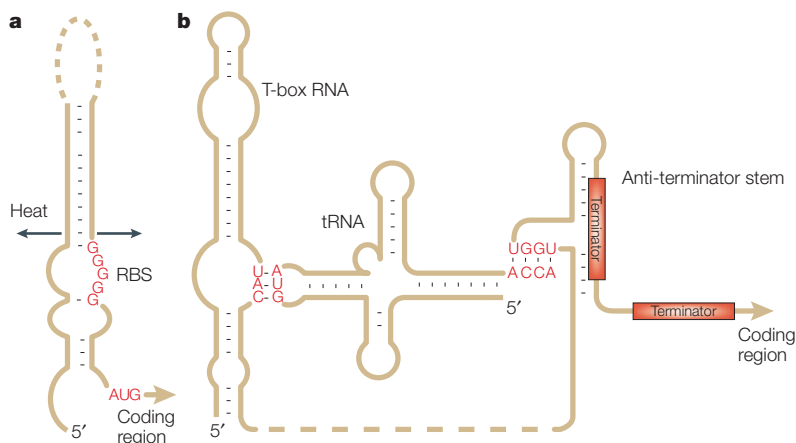
INTRON

A non-coding segment of mRNA that is removed by splicing processes before translation by ribosomes.

ROSE ELEMENT

An RNA sequence in certain bacteria that responds to changes in temperature and controls expression of adjacent heat-shock genes.

Box 3 | Thermosensors and T-boxes



RNA thermosensors and T-box RNAs are gene-control elements that share some characteristics with metabolite-binding riboswitches. The 5'-untranslated region (5'-UTR) of the *prfA* gene from *Listeria monocytogenes* carries a hairpin structure that occludes the ribosome-binding site (RBS; see figure, part a). A shift in temperature from 30°C to 37°C induces a fivefold increase in *prfA* expression⁸⁴, presumably by melting local base pairs and permitting access by the ribosome. Mutational analysis indicates that the stability of the hairpin structure is crucial and, therefore, the RNA alone might serve as a molecular thermometer. This form of control might be widespread, as a similar system seems to function in the ROSE ELEMENT in various rhizobial species⁸⁵ and in the untranslated region of the *E. coli* *rpoH* gene⁹⁶.

T-box RNAs (see figure, part b) are normally found in the 5' UTR of genes that encode AMINOACYL-tRNA SYNTHETASES or related amino-acid-biosynthesis genes of Gram-positive organisms. Each folds into a structure that selectively recognizes a specific tRNA⁷⁶. Recognition is guided largely through the formation of Watson-Crick base pairing between the T-box RNA and the anticodon loop of the target tRNA. Gene expression is activated in the presence of tRNAs that are not aminoacylated, thereby boosting the expression of genes that are needed to maintain an adequate pool of charged tRNAs. The T-box associated with the *tyrS* gene of *B. subtilis* (shown in the figure) functions via a transcription-termination mechanism. The T-box RNA rejects its matched tRNA when the tRNA is already aminoacylated. This permits formation of the terminator stem and subsequently downregulates gene expression (not shown). But, if the tRNA is not aminoacylated (as depicted in the figure), it is bound by the T-box and this permits expression of the adjoining synthetase gene by formation of an anti-terminator structure.

AMINOACYL-tRNA SYNTHETASE

An enzyme that recognizes a specific tRNA and selectively loads each with its cognate amino acid.

LEADER PEPTIDE

A peptide that is encoded upstream of a larger open reading frame, the translation of which is used as a sensor for adequate levels of a particular aminoacylated tRNA.

TRAP

A complex formed by the *trp* RNA-binding attenuation protein from *B. subtilis*. TRAP binds tryptophan and serves as a protein factor for regulating the *trp* operon.

These distinctions can be highlighted best by examining gene-control processes in prokaryotes that have some characteristics of riboswitches. For example, classic translation-mediated transcription-attenuation mechanisms (BOX 4) are controlled by differentially folded secondary structures⁸⁶. Control of the tryptophan operons in *E. coli* and *B. subtilis* are perhaps the most extensively studied examples of this form of gene regulation⁸⁷. In *E. coli*, the speed of ribosome movement while translating a short LEADER PEPTIDE from the nascent mRNA chain for tryptophan biosynthetic genes dictates whether an intrinsic terminator stem^{57,68} will form. The coding region for the leader (*trpL*) is located in what would otherwise be the untranslated region of the *trp* operon mRNA, and instructs the ribosome to incorporate several tryptophan residues. If the terminator stem forms in the 5'-UTR of the nascent mRNA in response

to high concentrations of aminoacylated tRNA^{trp} (rapid translation), then transcription is halted before the main coding regions of the mRNA are transcribed. By contrast, low concentrations of aminoacylated tRNA^{trp} (slow translation, which is indirectly due to low tryptophan concentrations) permits the formation of an anti-terminator stem, which prevents formation of the terminator stem and thereby facilitates transcription of the full mRNA for the *trp* operon.

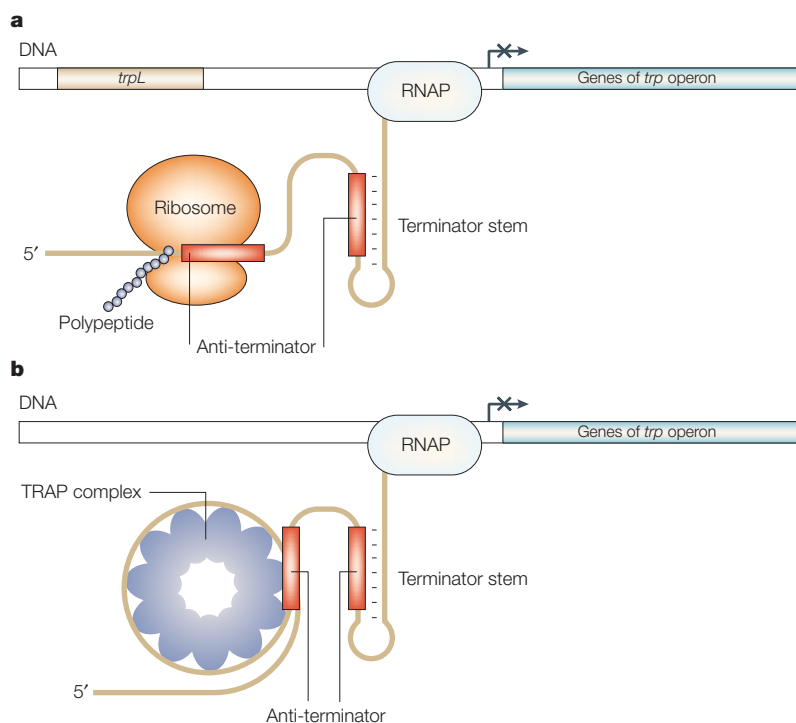
Unlike metabolite-binding riboswitches, the terminator and anti-terminator stems form in response to the action of a protein-containing complex. Specifically, the ribosome serves as the sensor element by its innate ability to recognize aminoacylated tRNA^{trp} when the proper codon is present in the mRNA that is being translated. So, molecular recognition is carried out by the ribosome, and the presence of the tRNA^{trp} target is relayed to RNA polymerase by the selection of one of two folding pathways for the nascent mRNA. Riboswitches are distinct in that they directly serve as the sensor element by recognizing their target metabolite in the absence of proteins. For some riboswitches, this molecular-recognition event is coupled to gene control because it establishes a folding pathway for the 5'-UTR, typically using Watson-Crick base-paired structures like those described above.

In *B. subtilis*, formation of an intrinsic terminator stem occurs if an interaction forms between a tryptophan-binding protein complex (TRAP) and an RNA sequence in the 5'-UTR of the nascent mRNA for the tryptophan operon (BOX 4). The RNA-protein interaction, which is induced by high tryptophan concentrations, prevents the formation of an anti-terminator stem, and the resulting terminator stem causes transcription to abort before the coding regions of the mRNA are made. In this case, TRAP serves as the separate protein factor that senses the target metabolite. Although the 5'-UTR carries the necessary control elements (terminator and anti-terminator stems), its overall RNA structure is made simpler because proteins are involved in recognition of the appropriate target compound and of the mRNA sequence.

Metabolite-binding riboswitches take a dominant role in forming receptor structures that dock their respective metabolites. This is evident on examination of the architectures of riboswitches, which all carry complex folded structures that vary little through evolution because their metabolite ligands are static targets that do not change in chemical composition. Most RNA motifs that are bound by proteins are substantially smaller, and show rapid evolutionary changes in sequence and structure when compared to metabolite-dependent riboswitches⁷⁹. This is due to the fact that the RNA or protein factor in the complex also varies through evolutionary time.

The complexity of these other RNA switches should not be underestimated however. Another structural feature of the nascent *trp* mRNA described above is a third stem-loop, which causes TRANSCRIPTIONAL PAUSING. This pausing, which is overcome once the ribosome begins to translate the leader peptide, seems to be important for

Box 4 | Transcription attenuation and protein-mediated RNA switches



Classical mechanisms of bacterial gene control that involve mRNA structures require the ribosome (transcription attenuation) or protein factors. Transcription attenuation of the *E. coli* *trp* operon makes use of the process of translation to determine the levels of tryptophan in the cell. A sufficient supply of tryptophan allows the cell to produce a level of tryptophan-charged tRNA such that the ribosome rapidly transcribes a tryptophan-rich leader peptide (*trpL*). The ribosome sequesters anti-terminator sequences and, therefore, transcription of the remaining mRNA chain is terminated (as shown in the figure, part a). If tryptophan-charged tRNAs are rare, then the ribosome transcribes the leader more slowly and permits formation of the anti-terminator stem (not shown). The result of this is transcription of the full-length mRNA and expression of the operon, which encodes genes that are needed to boost the tryptophan concentration.

Transcription-termination control in the *B. subtilis* *trp* operon makes use of an 11-subunit protein complex, known as TRAP, wherein each member of the complex binds tryptophan. At high concentrations, tryptophan binds and triggers changes in the protein structure that permit the complex to bind the nascent *trp* mRNA. If bound, sequences of the anti-terminator are sequestered and transcription is terminated (see figure, part b). When tryptophan concentrations are low, the protein complex cannot bind the nascent mRNA, and anti-terminator-stem formation assures that transcription of the complete mRNA proceeds (not shown).

TRANSCRIPTIONAL PAUSING
The temporary stalling of RNA polymerase during transcription that is typically caused by hairpin structures or other sequence elements within the nascent mRNA.

RNase P
A ribonucleoprotein-enzyme complex wherein the RNA component serves as a ribozyme that processes precursor RNAs such as pre-tRNA transcripts.

coordinating transcription and translation timing. In each case, the mRNA structures that are differentially formed are a result of the action of other complexes (either RNA polymerase or the ribosome), which determine the folding pathway of the mRNA. For pausing, anti-termination and termination, the structures formed by the nascent mRNA are usually established by Watson–Crick base pairing. Therefore, the nucleotide sequences of these elements show considerable variability from organism to organism. By contrast, riboswitches rely heavily on critical tertiary contacts that tend to require more-extensive sequence conservation compared with other RNA genetic elements.

Conclusions

The studies that have been conducted so far prove that riboswitches directly bind metabolites in the absence of proteins. The experiments used to assess the mechanism of riboswitch gene control do contain proteins such as RNA polymerase or RNase H. However, these added proteins are usually highly purified, and, therefore, are not likely to lead investigators to a false conclusion by harbouring protein factors that would be essential for a more conventional gene-control mechanism. Neither of these points rule out the possibility that metabolite binding by the RNA recruits or displaces some protein factor that is critical for controlling gene expression inside cells. What is clear from the current data set is that there is no need to invoke the existence of proteins, as each riboswitch seems to have all the components that are necessary to bind its target and influence gene control without the aid of other factors.

The diversity of natural aptamers is in accordance with the finding by molecular engineers that RNA has a robust ability to selectively bind to a variety of organic compounds. Given the versatility of RNA, particularly those molecules that are large in size, it is likely that almost any compound could serve as a target for an RNA aptamer. Therefore, the question regarding the extent of participation by riboswitches in modern genetic control might be one of evolutionary persistence as opposed to capability. How many riboswitches emerged early in evolution and were able to survive in the face of stiff competition from protein factors? Perhaps clues to this mystery can be derived by looking at the results of the long-standing competition between ribozymes and protein enzymes.

It seems that the only ribozymes to survive were those that were difficult to replace by small evolutionary changes (for example, the ribosome, RNase P and self-splicing ribozymes), or which have functions that are particularly well suited to their task (such as self-cleaving ribozymes). Protein factors might be able to provide even more precise molecular recognition capabilities or more versatile control mechanisms than those shown by riboswitches. However, the relative simplicity of mRNAs conducting their own metabolic surveillance and making the necessary changes to gene expression in the absence of protein factors might still be most efficient in many circumstances.

We suspect that many such circumstances exist and that new riboswitches will continue to be discovered as new bioinformatics and biochemical strategies are applied. For example, comparative sequence analysis has been used to identify six new RNA motifs with structural characteristics that are indicative of riboswitch function⁷⁹. However, this approach tends to favour the identification of RNA elements that are widely conserved and can miss RNA elements that are important for gene-control mechanisms in a few closely related species. Other bioinformatics approaches^{88–90} and biochemical strategies, such as the cloning of non-coding RNAs⁹¹, provide complementary routes to the identification of novel RNA motifs.

1. Ptashne, M. & Gann, A. *Genes and Signals* (Cold Spring Harbor Laboratory Press, New York, 2002).
2. Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
3. Dykxhoorn, D. M., Novina, C. D. & Sharp, P. A. Killing the messenger: short RNAs that silence gene expression. *Nature Rev. Mol. Cell Biol.* **4**, 457–467 (2003).
4. McManus, M. T. & Sharp, P. A. Gene silencing in mammals by small interfering RNAs. *Nature Rev. Genetics* **3**, 737–747 (2002).
5. Carrington, J. C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336–338 (2003).
6. Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043–1049 (2002).
- The first demonstrations that mRNAs bind metabolites directly in the absence of proteins are described in this paper and in reference 7.**
7. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).
- This paper and references 32, 57 and 58 report evidence for the existence of two main forms of riboswitch gene control: transcription termination and translation initiation.**
8. Lai, E. C. RNA sensors and riboswitches: self-regulating messages. *Curr. Biol.* **13**, R285–R291 (2003).
9. Müller, S. Another face of RNA: metabolite-induced 'riboswitching' for regulation of gene expression. *Chembiochem* **4**, 817–819 (2003).
10. Winkler, W. C. & Breaker, R. R. Genetic control by metabolite-binding riboswitches. *Chembiochem* **4**, 1024–1032 (2003).
11. Nudler, E. & Mironov, A. S. The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* **29**, 11–17 (2004).
12. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. & Gelfand, M. S. Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.* **20**, 44–50 (2004).
13. Monod, J. & Jacob, F. General conclusions: teleonic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harbor Symp. Quant. Biol.* **26**, 389–401 (1961).
14. Monod, J., Changeux, J.-P. & Jacob, F. Allosteric proteins and cellular control systems. *J. Mol. Biol.* **6**, 306–329 (1963).
15. Kaganov, B. I. *Allosteric enzymes*. (John Wiley & Sons Ltd., New York, 1978).
16. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
17. Gilbert, W. The RNA world. *Nature* **319**, 618 (1986).
18. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).
19. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. & Cech, T. R. Self-splicing RNA: autoexcision and autocatalysis of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31**, 147–157 (1982).
20. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
21. Gold, L., Polisky, B., Uhlenbeck O. & Yarus, M. Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* **64**, 763–797 (1995).
22. Osborne, S. E. & Ellington, A. D. Nucleic acid selection and the challenge of combinatorial chemistry. *Chem. Rev.* **97**, 349–370 (1997).
23. Hermann, T. & Patel, D. J. Adaptive recognition by nucleic acid aptamers. *Science* **287**, 820–825 (2000).
24. Tang, J. & Breaker, R. R. Rational design of allosteric ribozymes. *Chem. Biol.* **4**, 453–459 (1997).
- First demonstration that engineered RNAs can function as allosteric molecular switches and respond to small metabolites.**
25. Soukup, G. A. & Breaker, R. R. Engineering precision RNA molecular switches. *Proc. Natl Acad. Sci. USA* **96**, 3584–3589 (1999).
26. Seetharaman, S., Zivarts, M., Sudarsan, N. & Breaker, R. R. Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nature Biotechnol.* **19**, 336–341 (2001).
27. Breaker, R. R. Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.* **13**, 31–39 (2002).
28. Silverman, S. K. Rube Goldberg goes (ribo)nuclear? Molecular switches and sensors made from RNA. *RNA* **9**, 377–383 (2003).
29. Gold, L., Brown, D., He, Y.-Y., Shtatland, T., Singer, B. S. & Wu, Y. From oligonucleotide shapes to genomic SELEX: novel biological regulatory loops. *Proc. Natl Acad. Sci. USA* **94**, 59–64 (1997).
- Although unpublished speculation that riboswitches might exist had been ongoing for several years, this is an early publication that briefly mentions the possibility.**
30. Gold, L., Singer, B., He, Y.-Y. & Brody, E. SELEX and the evolution of genomes. *Curr. Opin. Genet. Dev.* **7**, 848–851 (1997).
31. Gelfand, M. S., Mironov, A. A., Jomantas, J., Kozlov, Y. I. & Perumov, D. A. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet.* **15**, 439–442 (1999).
- This publication, as well as reference 49, used sequence comparisons to make the first secondary-structure models for genetic-control elements that have since proven to be riboswitches.**
32. Nou, X. & Kadner, R. J. Adenosylcobalamin inhibits ribosome binding to *btuB* RNA. *Proc. Natl Acad. Sci. USA* **97**, 7190–7195 (2000).
33. Stormo, G. D. & Ji, Y. Do mRNAs act as direct sensors of small molecules to control their expression? *Proc. Natl Acad. Sci. USA* **98**, 9465–9467 (2001).
34. Miranda-Rios, J., Navarro, M., & Soberón, M. A conserved RNA structure (*thi* box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc. Natl Acad. Sci. USA* **98**, 9736–9741 (2001).
35. Lundrigan, M. D., Köster, W. & Kadner, R. J. Transcribed sequences of the *Escherichia coli* *btuB* gene control its expression and regulation by vitamin B₁₂. *Proc. Natl Acad. Sci. USA* **88**, 1479–1483 (1991).
36. Ravnum, S. & Andersson, D. I. Vitamin B₁₂ repression of the *btuB* gene in *Salmonella typhimurium* is mediated via a translational control which requires leader and coding sequences. *Mol. Microbiol.* **23**, 35–42 (1997).
37. Richter-Dahlfors, A. A., Ravnum, S. & Andersson, D. I. Vitamin B₁₂ repression of the *cob* operon in *Salmonella typhimurium*: translational control of the *cblA* gene. *Mol. Microbiol.* **13**, 541–553 (1994).
38. Ravnum, S. & Andersson, D. I. An adenosyl-cobalamin (coenzyme-B₁₂)-repressed translational enhancer in the *cob* mRNA of *Salmonella typhimurium*. *Mol. Microbiol.* **39**, 1585–1594 (2001).
39. Soukup, G. A. & Breaker, R. R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**, 1308–1325 (1999).
40. Li, Y. & Breaker, R. R. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. *J. Am. Chem. Soc.* **121**, 5364–5372 (1999).
41. Soukup, G. A., DeRose, E. C., Koizumi, M. & Breaker, R. R. Generating new ligand-binding RNAs by affinity maturation and disintegration of allosteric ribozymes. *RNA* **7**, 524–536 (2001).
42. Nahvi, A., Barrick, J. E. & Breaker, R. R. Coenzyme B₁₂ riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.* **32**, 143–150 (2004).
43. Bond, C. M., Lees, K. A. & Enever, R. P. Photolytic decomposition of three cobalamins. A quantitative study. *J. Pharm. Pharmacol.* **24** (Suppl.), 143P (1972).
44. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. & Gelfand, M. S. Regulation of the vitamin B₁₂ metabolism and transport in bacteria by a conserved RNA structural element. *RNA* **9**, 1084–1097 (2003).
45. Webb, E., Claas, K. & Downs, D. *thiBPQ* encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphate in *Salmonella typhimurium*. *J. Biol. Chem.* **273**, 8946–8950 (1996).
46. Mironov, V. N., Perumov, D. A., Kraev, A. S., Stepanov, A. I. & Skryabin, K. G. Unusual structure in the regulation region of the *Bacillus subtilis* riboflavin biosynthesis operon. *Mol. Biol.* **24**, 256–261 (1990) (in Russian).
47. Krenova, R. A. & Perumov, D. A. Genetic mapping of regulatory mutations of *Bacillus subtilis* riboflavin operon. *Mol. Gen. Genet.* **222**, 467–469 (1990).
48. Kil, Y. V., Mironov, V. N., Gorishin, I. Y., Krenova, R. A. & Perumov, D. A. Riboflavin operon of *Bacillus subtilis*: unusual symmetric arrangement of the regulatory region. *Mol. Gen. Genet.* **233**, 483–486 (1992).
49. Grundy, F. J. & Henkin, T. M. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol. Microbiol.* **30**, 737–749 (1998).
50. Ebbole, D. J. & Zalkin, H. Cloning and characterization of a 12-gene cluster from *Bacillus subtilis* encoding nine enzymes for *de novo* purine nucleotide synthesis. *J. Biol. Chem.* **262**, 8274–8287 (1987).
51. Christiansen, L. C., Schou, S., Nygaard, P. & Saxild, H. H. Xanthine metabolism in *Bacillus subtilis*: characterization of the *xpt-pbuX* operon and evidence for purine- and nitrogen-controlled expression of genes involved in xanthine salvage and catabolism. *J. Bacteriol.* **179**, 2540–2550 (1997).
52. Vold, B., Szulmajster, J. & Carbone, A. Regulation of dihydrodipicolinate synthase and aspartate kinase in *Bacillus subtilis*. *J. Bacteriol.* **121**, 970–974 (1975).
53. Lu, Y., Chen, N. Y. & Paulus, H. Identification of *aecA* mutations in *Bacillus subtilis* as nucleotide substitutions in the untranslated leader region of the aspartokinase II operon. *J. Gen. Microbiol.* **137**, 1135–1141 (1991).
- This paper defines the precise locations of mutations that confer resistance to the toxic effects of a lysine analogue. The resistant phenotype, first described nearly 20 years earlier, is now known to be caused by mutations in a lysine-specific riboswitch.**
54. Kochhar, S. & Paulus, H. Lysine-induced premature transcription termination in the *lysC* operon of *Bacillus subtilis*. *Microbiol.* **142**, 1635–1639 (1996).
55. Patten, J. C. In *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology Vol. 1 (eds Neidhardt, F. C. *et al.*) 528–541 (American Society for Microbiology Press, Washington DC, 1996).
56. Patten, J.-C., Akrim, M. & Méjean, V. The leader sequence of the *Escherichia coli* *lysC* gene is involved in the regulation of *LysC* synthesis. *FEBS Microbiol. Lett.* **169**, 165–170 (1998).
57. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. *Proc. Natl Acad. Sci. USA* **99**, 15908–15913 (2002).
58. Mironov, A. S. *et al.* Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**, 747–756 (2002).
59. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**, 577–586 (2003).
60. McDaniel, B. A. M., Grundy, F. J., Artsimovitch, I. & Henkin, T. M. Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc. Natl Acad. Sci. USA* **100**, 3083–3088 (2003).
61. Epstein, V., Mironov, A. S. & Nudler, E. The riboswitch-mediated control of sulfur metabolism in bacteria. *Proc. Natl Acad. Sci. USA* **100**, 5052–5056 (2003).
62. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Struct. Biol.* **10**, 701–707 (2003).
63. Grundy, F. J., Lehman, S. C. & Henkin, T. M. The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc. Natl Acad. Sci. USA* **100**, 12057–12062 (2003).
64. Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S. & Breaker, R. R. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* **17**, 2688–2697 (2003).
65. Sudarsan, N., Barrick, J. E. & Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* **9**, 644–647 (2003).
- Demonstration that some eukaryotic mRNAs carry metabolite-binding domains that are probably components of riboswitches.**
66. Kubodera, T. *et al.* Thiamine-regulated gene expression of *Aspergillus oryzae* *thiA* requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR. *FEBS Lett.* **555**, 516–520 (2003).
67. Gusarov, I. & Nudler, E. The mechanism of intrinsic transcription termination. *Mol. Cell* **3**, 495–504 (1999).
68. Yarnell, W. S. & Roberts, J. W. (1999) Mechanism of intrinsic transcription termination and antitermination. *Science* **284**, 611–615 (1999).
69. White, H. B. III Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, 101–104 (1976).
70. White, H. B. III in *The Pyridine Nucleotide Coenzymes*. 1–17 (Academic Press, New York, 1982).
71. Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl Acad. Sci. USA* **86**, 7054–7058 (1989).
72. Jeffares, D. C., Poole, A. M. & Penny, D. Relics from the RNA world. *J. Mol. Evol.* **46**, 18–36 (1998).
73. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. Comparative genomics of thiamin biosynthesis in prokaryotes. New genes and regulatory mechanisms. *J. Biol. Chem.* **276**, 5093–5100 (2002).
74. Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Struct. Mol. Biol.* **11**, 29–35 (2004).
- Demonstration of a riboswitch that activates gene expression.**
75. Johansen, L. E., Nygaard, P., Lassen, C., Agerso, Y. & Saxild, H. H. Definition of a second *Bacillus subtilis* *pur* regulon comprising the *pur* and *xpt-pbuX* operons plus *pbuG*, *nupG* (*yxjA*), and *pbuE* (*ydhL*). *J. Bacteriol.* **185**, 5200–5209 (2003).

76. Grundy, F. J. & Henkin, T. M. The T box and S box transcription termination control systems. *Frontiers Biosci.* **8**, D20–D31 (2003).
77. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Res.* **31**, 6748–6757 (2003).
78. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281–286 (2004).
A riboswitch that uses the action of a ribozyme to control gene expression.
79. Barrick, J. E. *et al.* New motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl Acad. Sci. USA* **101**, 6421–6426 (2004).
80. Gottesman, S. *et al.* Small RNA regulators of translation: mechanisms of action and approaches for identifying new small RNAs. *Cold Spring Harbor Symp. Quant. Biol.* **66**, 353–362 (2001).
81. Massé, E., Majdalani, N. & Gottesman, S. Regulatory roles for small RNAs in bacteria. *Curr. Opin. Microbiol.* **6**, 120–124 (2003).
82. Ben-Asouli, Y., Pei-Or, Y., Shir, A. & Kaempfer, R. Human *interferon-γ* mRNA autoregulates its translation through a pseudoknot that activates the interferon-inducible protein kinase PKR. *Cell* **108**, 221–232 (2002).
83. Gold, L., Brodey, E., Hellig, J. & Singer, B. One, two, infinity: genomes filled with aptamers. *Chem. Biol.* **9**, 1259–1264 (2002).
84. Johansson, J. *et al.* An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell* **110**, 551–561 (2002).
85. Chowdhury, S., Ragaz, C., Kreuger, E. & Narberhaus, F. Temperature-controlled structural alterations of an RNA thermometer. *J. Biol. Chem.* **278**, 47915–47921 (2003).
86. Henkin, T. M. & Yanofsky, C. Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays* **24**, 700–707 (2002).
87. Yanofsky, C. Using studies on tryptophan metabolism to answer basic biological questions. *J. Biol. Chem.* **278**, 10859–10878 (2003).
88. Wassarman, K. M., Repola, F., Rosenow, C., Storz, G. & Gottesman, S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.* **15**, 1637–1651 (2001).
89. Lyubetskaya, E. V., Leont'ev, L. A., Gelfand, M. S. & Lyubetsky, V. A. Search for alternative RNA secondary structures regulating expression of bacterial genes. *Mol. Biol.* **37**, 707–715 (2003) (translated from Russian).
90. Klein, R. J. & Eddy, S. R. RSEARCH: Finding homologs of single structured RNA sequences. *BMC Bioinformatics* **4**, 44 (2003).
91. Vogel, J. *et al.* RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.* **31**, 6435–6443 (2003).
The isolation of numerous small non-coding RNAs in bacteria offers the possibility that a rich diversity of RNAs might participate in various cellular functions.
92. Lee, J. F., Hesselberth, J. R., Meyers, L. A. & Ellington, A. D. Aptamer database. *Nucleic Acids Res.* **32**, D95–D100 (2004).
93. Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. *Science* **263**, 1425–1429 (1994).
An early demonstration of the molecular recognition power of RNA aptamers.
94. Kiga, D., Futamura, Y., Sakamoto, K. & Yokoyama, S. An RNA aptamer to the xanthine/guanine base with a distinctive mode of purine recognition. *Nucleic Acids Res.* **26**, 1755–1760 (1998).
95. Burgstaller, P. & Famulok, M. Isolation of RNA aptamers for biological cofactors by *in vitro* selection. *Angew. Chem. Int. Ed. Engl.* **33**, 1084–1087 (1994).
96. Morita, M. T. *et al.* Translational induction of heat shock transcription factor σ^{32} : evidence for a built-in RNA thermosensor. *Genes Dev.* **13**, 655–665 (1999).

Acknowledgements

We thank members of the Breaker laboratory for helpful discussions and comments on the manuscript.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

Entrez: <http://www.ncbi.nlm.nih.gov/entrez>

btuB | *glmS* | *lac* repressor | *prfA* | *pbuX* | *rpoH* | *tyrS* | *ydhL* | *xpt*

Access to this links box is available online.