Published in final edited form as:

Wiley Interdiscip Rev RNA. 2014 ; 5(6): 803–822. doi:10.1002/wrna.1247.

Structure and function of pseudoknots involved in gene expression control

Alla Peselis and Alexander Serganov

Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY, USA

Abstract

Natural RNA molecules can have a high degree of structural complexity but even the most complexly-folded RNAs are assembled from simple structural building blocks. Among the simplest RNA elements are double-stranded helices that participate in the formation of different folding topologies and constitute the major fraction of RNA structures. One common folding motif of RNA is a pseudoknot, defined as a bipartite helical structure formed by base-pairing of the apical loop in the stem-loop structure with an outside sequence. Pseudoknots constitute integral parts of the RNA structures essential for various cellular activities. Among many functions of pseudoknotted RNAs is feedback regulation of gene expression, carried out through specific recognition of various molecules. Pseudoknotted RNAs autoregulate ribosomal and phage protein genes in response to downstream encoded proteins, while many metabolic and transport genes are controlled by cellular metabolites interacting with pseudoknotted RNA elements from the riboswitch family. Modulation of some genes also depends on metabolite-induced mRNA cleavage performed by pseudoknotted ribozymes. Several regulatory pseudoknots have been characterized biochemically and structurally in great detail. These studies have demonstrated a plethora of pseudoknot-based folds and have begun uncovering diverse molecular principles of the ligand-dependent gene expression control. The pseudoknot-mediated mechanisms of gene control and many unexpected and interesting features of the regulatory pseudoknots have significantly advanced our understanding of the genetic circuits and laid the foundation for modulation of their outcomes.

RNA molecules play important roles in many cellular processes. To carry out biological functions, some RNAs adopt elaborate three-dimensional structures capable of enzymatic catalysis or serve as binding sites for proteins and small molecules. Unlike protein structures, mostly composed of two distinct secondary structure elements, RNA predominantly folds into double-stranded helices, typically connected sequentially and often joined by a multihelical junction. Alternatively, helices can be arranged based on a different building principle, first recognized by C. Pleij and coworkers in the plant viral RNA¹ and coined a pseudoknot (for historical perspective see [2]).³ This tertiary structural arrangement is defined as a Watson-Crick base pairing that involves a stretch of bases (S2 in Figure 1(a)), located between paired strands (S1), and an outlying partner downstream of the paired strands.⁴ In other words, the simplest H-type pseudoknot, where H stands for hairpin loop,

can be generalized as a hairpin whose loop nucleotides (nts) make standard base pairs with a complementary sequence outside of the loop (Figure 1(b), left). Such pairing favors formation of the tertiary helical stem (S2) and the coaxial stacking of this stem with the helical segment from the hairpin (S1), resulting in an elongated quasi-continuous double helix with one continuous and one discontinuous strand. Two loops cross the deep or major (L1) and shallow or minor (L3) grooves of the helix, respectively (Figure 1(b), right). Overall, pseudoknots adopt knot-shaped three-dimensional conformations but are not topological knots.

In contrast to recurrent structural motifs,⁵ pseudoknots do not contain sequence-specific features and represent a structurally diverse group with loops and helices of different lengths and compositions. Not surprisingly, pseudoknots are found in many RNAs where they are either integrated into complex RNA structures or function as stand-alone elements. Pseudoknots contribute to the formation of ribosomes, ribozymes, telomerase, and participate in various RNA activities including replication, RNA processing, inactivation of toxins, gene expression control, as well as several translation-related activities, such as internal translation initiation, translation stimulation, ribosome rescue, and frameshifting.^{6,7}

Recent research, discussed in several excellent reviews,⁶⁻⁸ revealed most prominent characteristics of viral and eukaryotic pseudoknots involved in frameshifting, internal translation initiation, and other biological phenomena. Pseudoknots participating in the feedback regulation of gene expression in bacteria have received less attention. However, tremendous progress in biochemical and structural studies, particularly of riboswitches, regulatory RNA elements modulating gene expression in response to direct binding of cellular metabolites, has recently led to identification of many pseudoknotted RNAs. In this review, we focus on pseudoknots that directly impact on feedback regulation of protein biosynthesis. We summarize structural and functional data on structurally well-defined regulatory pseudoknots and discuss their features in the context of the ligand binding and gene expression control. We also identify common trends and unique characteristics adopted by the regulatory pseudoknots and compare them with pseudoknots involved in other cellular functions.

TYPES OF PSEUDOKNOTS

The definition of a pseudoknot as a tertiary structural element being formed by a base-pairing of the loop nucleotides with a complementary sequence outside of the hairpin is broad. This definition does not specify the length and structural properties of the outside region, which may involve complex structural elements and be hundreds of nucleotides long. Thus, virtually every canonical tertiary base-pairing can be viewed as a pseudoknot and it is debatable whether long-range interactions should be called pseudoknots or the term should be restricted to simple cases when the complementary sequence of the outside region is located close to the hairpin.

The majority of simple natural pseudoknots are classified as H-type pseudoknots that have two base-paired stem regions (S1 and S2) and two or three single-stranded loops (L1, L2, L3) (Figure 1(b)).⁹⁻¹¹ However, even this simplest folding topology provides significant

structural diversity. In 75% of H-type pseudoknots L2 is missing and an additional 10% have a 1-nt loop,^{9,11} allowing base-paired stems to stack coaxially⁹ (Figure 1(c)). This observation gave rise to the alternative nomenclature of pseudoknot elements, where L2 is assumed missing while groove-spanning L1 and L3 are named L1 and L2. In some pseudoknots, coaxial stacking can be the result of L1 or L3 being absent (Figure 1(d),(e)). Additional structural diversity arises from the conformation of the helix-helix junction, defined by the interhelical angle between the stems, the displacement of the stems relative to each other, and the difference in rotation of the coaxial helices.

Tertiary base-pairing involving a bulged (B), interior (I), and multibranch (M) loops with regions elsewhere as well as two hairpin loops (H-H) (Figure 1(f)) also results in adaptation of a pseudoknot.¹² However, these pseudoknots can still be considered as H-type pseudoknots, despite the presence of substructures embedded within their stems and loops. For this reason, the B, I and M nomenclature is not extensively used. Another variation of a pseudoknot structure, observed in several RNAs and referred to as a double-nested pseudoknot, occurs when the hairpin loop of the first pseudoknot is involved in the formation of an additional pseudoknot, set within or nested in the first pseudoknot.

The classical definition of a pseudoknot as a structure formed by base-pairing between the apical loop of a hairpin and an outside region may be confusing in some cases, for instance, when pseudoknots are formed between bulged loops. To understand whether an RNA structure is a pseudoknot, the RNA molecule can be presented as a circle with base-paired nucleotides connected by chords.¹³ If the chords are not crossing, the structure corresponds to hairpins (Figure 1(g)).¹³ If any chords drawn for two consecutive Watson-Crick base pairs from each helix are crossing, the structure can be defined as a pseudoknot (Figure 1(h)). Such definition of the pseudoknot does not involve description of structural elements and is more formal.

Because of the variety of possible pseudoknot topologies, classification, presentation, and prediction of pseudoknots are challenging tasks. Classification of pseudoknots is required to understand common folding principles and is important for computational predictions of pseudoknots, a problem attracting many researchers. One of the most convenient planar representations and classification of pseudoknots is adapted by the PseudoBase database (<http://pseudobaseplusplus.utep.edu/>).⁹ The vast majority out of over 200 natural pseudoknots can be assigned to one of six geometric types: H, LL, HL_{out}, HL_{in}, HH and HHH, where H depicts a hairpin and L designates a bulge, interior or multiple loop.¹⁴ This classification does not take into account the topology of pseudoknots but provides simple drawings which could be used to delineate connectivity of complex pseudoknotted structures (Figure 1(i), Box 1). Simple pseudoknot drawings can be generated by a web service Pseudoviewer (<http://wilab.inha.ac.kr/pseudoviewer/>).¹⁴ In this review, we have opted for the PseudoBase geometric type representation and the three-loop nomenclatures for the pseudoknot schematics and structures.

PSEUDOKNOTS AS BINDING SITES

Bacterial mRNAs often contain regions responsible for the modulation of gene expression. These regulatory segments can adopt tertiary structures that prevent or facilitate the formation of gene control elements, such as ribosome binding sites (Shine-Dalgarno sequences) or transcription terminators, in response to various environmental cues. Regulation usually involves the interplay between mutually exclusive conformations that cause opposite effects on gene expression. One of these conformations is typically stabilized by binding to proteins or small molecules that are associated with genes under control. Several regulatory systems depend on the mRNA conformations that contain pseudoknot structures; most studied examples are described in this section.

Autorepression of Gene 32 in T-even Bacteriophages

One of the first pseudoknots was described in the mRNA of gene 32 protein (gp32) in T-even bacteriophages.¹⁵ gp32 binds to phage single-stranded DNA (ssDNA) during replication and ensures replisome processing and accuracy (Figure 2(a), top).¹⁶ The protein also specifically binds to the pseudoknot located upstream of the translation initiation start in the gp32 mRNA.^{15,17} When the amount of gp32 exceeds the number of binding sites on ssDNA, the pseudoknot-bound protein serves as a nucleation point for cooperative binding of more protein molecules to ssRNA in the downstream direction (Figure 2(a), bottom). These extra protein molecules shield the Shine-Dalgarno sequence and initiation codon and prevent translation of the gp32 mRNA.

The Nuclear Magnetic Resonance (NMR) structure¹⁸ of the gp32 mRNA pseudoknot from bacteriophage T2 highlights several common features of H-type pseudoknots (Figure 1(i), Table 1). The pseudoknot contains two A-form helical stems (S1 and S2) connected by two loops (L1 and L3) (Figure 2(b),(c)). The helices are collinearly stacked but may bend with an acute angle in the range of $15\pm 15^\circ$. The S2 helix is rotated by approximately 18° with respect to S1 to avoid close contacts between phosphates in the interhelical junction and at the same time to preserve stacking, most significantly at the A15-G16 step. Single-adenosine loop L1 crosses the major groove of S2 but does not appear to form important tertiary interactions. On the contrary, 7 nt loop L3 is packed against the minor groove of S1 thus forming a triplex stabilized by numerous tertiary interactions.¹⁸ The sugar-phosphate backbone makes a characteristic sharp turn after L1 which presumably has strong structural tension and is important for the folding of the pseudoknot.

Unfortunately, molecular details of the pseudoknot-gp32 binding are still lacking and we can only hypothesize about the mode of interactions based on the structure of the gp32 core bound to short ssDNA fragment¹⁹ (Figure 2(d)). The protein structure is composed of two subdomains linked by a 'connecting' region, which together form a deep cleft occupied by ssDNA. Given that L3 of the pseudoknot is required for gp32 binding,¹⁷ this loop might be located in the cleft while pseudoknot helices might be bound to the positively charged surfaces along the same face of the protein. Therefore, the pseudoknot would provide additional interactions with gp32 that cannot be formed with the downstream ssRNA and would fulfill the important regulation role of a nucleation point for cooperative protein binding.

Feedback Regulation of Ribosomal Proteins

In *Escherichia coli* and some other bacteria, ribosome biosynthesis is coordinated by regulation of the ribosomal protein operons. Several primary rRNA-binding ribosomal proteins interact specifically with rRNAs and, if in excess, bind to lower affinity sites overlapped with the Shine-Dalgarno sequence in their own mRNAs. Thus, these proteins repress translation initiation via direct *competition* with 30S ribosomal subunits.

The majority of rRNA and corresponding mRNA sites share similarities sufficient for specific recognition by the same regions in ribosomal proteins. However, the mRNA binding sites for *E. coli* ribosomal proteins S4, S15 and L20 include pseudoknots which are not present in the rRNA binding sites.²⁰⁻²² The S4 mRNA ‘double’ pseudoknot is formed by two base-paired regions between L1 and the downstream region (Figure 3(a)) and has no resemblance to the multi-helical site in the 16S rRNA²². The mRNA binding region for L20 is even more complex, containing a long-range pseudoknot separated by 280 nucleotides.²³ The mRNA site for S15 is the simplest of the three pseudoknots. The S15 mRNA folds into the typical H-type pseudoknot that shares remote similarity with the S15 binding region in the three-way junction of the 16S rRNA.²⁴ The pseudoknot contains 10-bp stem S1, 7-bp stem S2, which is crossed by a single adenosine, and long loop L3, which may form an unstable hairpin (Figure 3(b),(c)). This hairpin allows assignment of the pseudoknot to the HL_{out} type (Figure 1(i), Table 1). The S15-binding site is composed of the evolutionarily conserved C-G/U-G motif in S1, adjacent base pairs of S2, adenine loop L1, and moderately conserved distal base pairs of S2.²⁴ The C-G/U-G motif can be paralleled with the S15 binding determinants in the middle of helix H22 whereas base-pairs in S2 are equivalent to the region of the helix adjacent to the H20-H21-H22 junction in the central domain of 16S rRNA.²⁵ Thus, the S15 binding sites on mRNA and rRNA exhibit molecular mimicry mostly on the structural but not on the sequence level.^{24,26} Interestingly, the S15 pseudoknot is distributed in species related to *E. coli*²⁷ whereas in other bacteria, such as *Thermus thermophilus*²⁸ and *Bacillus stearothermophilus*,²⁹ the S15 regulatory mRNA regions resemble the 16S rRNA junction.

The S4 and S15 pseudoknots contain a Shine-Dalgarno sequence and translation initiation codon in the long loop, and can simultaneously bind to both the ribosomal protein and the 30S subunit while inhibiting formation of the translation initiation complex. Such a repression mechanism, inconsistent with *competition*, was coined *entrapment*.^{30,31} The cryo-electron microscopy (cryo-EM) studies have shown that the pseudoknot mRNA and S15 bind to a ‘stand-by’ site on the 30S subunit so that the initiation codon is positioned closely to the pseudoknot structure³² (Figure 3(d)). As a consequence, the mRNA does not enter the mRNA channel in the ribosome thus keeping the initiator tRNA away from the P-site and preventing recognition of the start codon. S15 stabilizes the pseudoknot and blocks transition from the pre-initiation complex to the productive initiation complex. The pseudoknot appears to be the simplest RNA structure which would provide a helical content for the protein repressor recognition, a long loop for Shine-Dalgarno/anti-Shine-Dalgarno pairing, and repressor-dependent sequestration of the initiator codon.

Gene Expression Control by Riboswitches

Riboswitches are mRNA regions that control genetic circuits in response to direct binding of metabolites present in cells at over-threshold concentrations. Regulation by some riboswitches is reminiscent of gene expression modulation by ribosomal proteins, but because of small sizes, bound metabolites have difficulty physically shielding the Shine-Dalgarno sequence. Therefore, riboswitches typically exploit structural changes in RNA to release or block access to the Shine-Dalgarno sequence through base-pairing, or control transcription elongation through alternative formation of a transcription terminator or antiterminator. The key to riboswitch regulation is stabilization of the evolutionarily conserved metabolite-sensing or 'aptamer' domain of the riboswitch by cognate metabolite binding, typically resulted in base-pairing of the 'switching' RNA region within the domain-closing helix. The formation of the stable metabolite-bound domain facilitates folding of the non-conserved downstream elements that affect mRNA transcription or translation. In the absence of the cognate ligand, the aptamer domain does not adopt a stable structure and the switching region engages in alternative base-pairing with the downstream sequence, causing formation of different regulatory elements with the opposite effect on gene expression. Metabolite-sensing domains adopt a plethora of structures supported, or entirely based on, pseudoknots (Table 1).

Fluoride Riboswitch

Many bacterial and archaeal species mitigate fluoride toxicity by activating expression of genes that encode putative fluoride transporters through fluoride-responsive riboswitches.³³ The fluoride sensor forms the HL_{out}-type pseudoknot composed of short S1, S2, L1, and long L3 that contains an additional hairpin important for the pseudoknot stabilization and ligand binding³⁴ (Figure 4(a),(b)). As in many other pseudoknots, the 5' end of L3 forms triplex interactions with S1 through the 'A-minor motif' characterized by the sugar edge interactions of adenosines A20 and A21 with the minor groove of the helix (Figure 4(c)). In contrast to many pseudoknots, bases of L1 (A6 and U7) are not positioned in the major groove of S2 but are stacked on the terminal bases of L3 (Figure 4(b)). In addition, nucleotides from the 3' end of L3 reinforce S2 by stacking interactions and non-canonical base-pairing at its end, and hydrogen-bonding in its major groove. Not surprisingly, the sharp turn of L1 and the close juxtaposition of L1, L3 and S2 are stabilized by 3 Mg²⁺ cations, which trap the negatively charged fluoride anion in the middle of the cation plane (Figure 4(d)). The ligand binding also strengthens the tertiary interactions in the 3' end of L3 and supports the overall pseudoknot fold that prevents formation of the downstream transcription terminator with the 3' end of S2. Thus, the fluoride riboswitch exploits ligand-induced stabilization of the regions that have strong structural tension and repelling electrostatic interactions.

PreQ₁-I Riboswitch

In many bacteria, biosynthesis of a tRNA-related hypermodified guanine nucleotide queuosine (Q) is controlled by two riboswitch classes responsive to preQ₁ (7-aminomethyl-7-deazaguanine), a precursor of Q.³⁵ The preQ₁-I (class I) riboswitch has the smallest aptamer domain built by a compact H-type pseudoknot that contains three loops

and two short stems (Figure 4(e),(f)).³⁶⁻³⁸ As in other pseudoknots, an adenine-rich L3 forms tertiary triplex interactions with S1. These interactions mostly involve N6 atoms of adenosines and stem bases termed 'A-amino kissing motif' (Figure 4(g)),³⁶ distinguishing them from A-minor motifs (Figure 4(c)). Some of these adenosines, such as A27, are inclined and interact with two consecutive base pairs (Figure 4(g)). The pseudoknot is additionally stabilized by base-pairing of L1 with L3 and interactions of the turn after L1 with a metal cation (not shown). L2 is unusually long and contains 6 nucleotides, either looped out or interacting with the minor groove of S2. Long L2 weakens collinear stacking between S1 and S2 but provides a pairing partner for preQ₁, which binds between the two stems and reinstalls stacking interactions (Figure 4(f),(h)). L2 (C19) and L3 (A32) make specific hydrogen bonds with the ligand and help hold it in the pocket. In summary, preQ₁-mediated regulation relies upon pseudoknot stabilization after the ligand binds to the junction between two stems, circumventing flexibility associated with the elongated loop L2.

PreQ₁-II Riboswitch

In Lactobacillales, preQ₁ modulates expression of genes at the translational level by binding to a much longer class II aptamer (preQ₁-II).³⁹ Similarly to the preQ₁-I riboswitch, the preQ₁-II RNA centers on the pseudoknot fold, which, in contrast, is formed by longer stems, slightly longer L1, much shorter L2, and longer hairpin-containing L3 (Figure 4(i)). Thus, the preQ₁-II pseudoknot conforms to the HL_{out} type. An additional hairpin (black and gray in Figure 4(i),(j)) at the 5' end of preQ₁-II stabilizes S1 by collinearly stacking under it.⁴⁰ Because of a much longer S1, L3 does not follow S1 through the entire length and instead makes minor groove interactions with two regions separated by almost a helical turn. These interactions involve an 'A-amino kissing motif', characterized by hydrogen-bonding of a single adenosine, A52, to residues from consecutive base pairs at the bottom part of S1, and A-minor interactions between L3 and the bottom part of S2. Conserved major groove U•(A-U) triples between L1 and S2 additionally stabilize S2 and prevent ribosome binding to the Shine-Dalgarno sequence, located in the 3' region of S2 and engaged in base-pairing in the folded pseudoknot (Figure 4(k)). As in preQ₁-I, the ligand is strategically positioned to fill a gap between S1 and S2, thereby facilitating collinear stacking of the stems and formation of the pseudoknot.

Remarkably, although the bound ligand is specifically recognized by loops in both preQ₁ riboswitch classes, specific readout of the ligand is completely different (Figure 4(h),(l)). The Watson-Crick edge of the ligand is recognized through Watson-Crick pairing with C19 (L2) in preQ₁-I but through non-canonical pairing with the Watson-Crick face of C30 (L1) in preQ₁-II. The N2-N3-N9 edge of the ligand is bound by two nucleotides from L1 (U9) and L3 (A32) in preQ₁-I but by a single U41 from L2 in preQ₁-II. Unlike in preQ₁-I, a nucleotide from L3 (A70) does not specifically recognize the ligand although participates in building a wall of the ligand-binding pocket in preQ₁-II riboswitch.

The preQ₁ riboswitches are interesting examples of the convergent evolution of RNA structures. The two structures are based on the pseudoknot folds and use similar general means for ligand-dependent structure stabilization. However, riboswitches exhibit great

sequence divergence, demonstrate different ligand recognition patterns, utilize distinct regulatory mechanisms, and exploit dissimilar architectural principles to maintain the structures. These and other differences appear to account for independent origins of the two preQ₁ riboswitch classes and argue against apparent simplification and reduction of preQ₁-II to yield preQ₁-I riboswitches.

SAM-II Riboswitch

Seven structurally distinct classes of bacterial riboswitches respond to the essential cofactor S-adenosylmethionine (SAM) (reviewed in [41]). Among available structures for class I, II, and III SAM riboswitches, the global architecture of only SAM-II riboswitch⁴² conforms to the H-type pseudoknot while two other classes are built on the junctional folds. The SAM-V riboswitch is predicted to form a pseudoknot similar to the SAM-II pseudoknot but with longer loops.⁴³ The SAM-II pseudoknot contains a 6-bp S1 and irregular 11-bp S2, and a 7-nt uracil-rich L1 and 9-nt adenosine-rich L3 loops that forms multiple tertiary contacts in the major groove of S2 and the minor groove of S1, respectively (Figure 4(m),(n)).⁴² The contacts between S1 and L3 involve several skewed stacked adenosines, which use different edges to form either A-amino kissing interactions or ‘inclined A-minor interactions’, characterized by the A-minor hydrogen-bonding of a single adenosine A36 with the C4-G28 base pair and additional interactions with adjacent base pairs (Figure 4(o)). SAM binds the riboswitch in an extended conformation along the major groove of the L1-S2 triplex, forming hydrogen bonds and other interactions with five consecutive base pairs and triples (Figure 4(p)). Electrostatic interactions with the positively charged sulfonium moiety are important for discrimination from SAH, bearing neutrally charged sulfur atom.⁴⁴ The adenine moiety of SAM is positioned in a cavity, created by omission of a pairing partner for U44, where it participates in a triple arrangement with U44 (S2) and U10 (L1). Thus, the base of the riboswitch ligand fills the gap in the long pseudoknot helix thereby mediating uninterrupted stacking interactions with the Shine-Dalgarno sequence located in the 3′ region of S2. In turn, the methionine moiety reinforces the helix by interactions with long loop L1 (U11) and S2 (A19 and A47) to ensure occlusion of the Shine-Dalgarno sequence within the helix, and prevent translation initiation of the downstream gene.

SAH Riboswitch

S-adenosylhomocysteine, a by-product of methylation by SAM, activates expression of SAH-recycling genes through a riboswitch mechanism. The SAH riboswitch structure represents an infrequent LL-type pseudoknot (Figure 1(i)).⁴⁵ In this structure, stems S1 and S2 are not stacked, loop L1 does not exist, long loop L2 contains a hairpin that stacks on S2, long L3 stabilizes S1 through various minor groove interactions and a base pair between C46 (L3) and G30 (L2) that stacks on S1 (Figure 5(a),(b)). A semi-open ligand-binding pocket is almost exclusively built by residues from L2 with small contribution from S1 (C32) and L3 (G47) (Figure 5(c)). The adenosine base of SAH is intercalated into the L2 hairpin close to the L2-S2 junction and is locked in place by base-pairing with G15. In this position, the adenosine moiety stacks on A29 (L2) which in turn stacks on G48 from S2. The methionine moiety of SAH is hydrogen bonded to three nucleotides of L2, including a base-specific contact with G30. These interactions stabilize tertiary base-pairing and stacking between L2 and L3 that propagates into S1. Therefore, ligand binding contributes

to the pairing of the regulatory 3' region indirectly by providing interactions essential for stabilization of both stems of the pseudoknot.

c-di-GMP Riboswitch

Two riboswitch classes modulate expression of various genes in response to second messenger bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP). The c-di-GMP-II riboswitch structure can be viewed as a stem-loop structure that contains a bulged loop (Figure 5(d)). The H-type pseudoknot is then formed by long-range base-pairing of the apical and bulged loops of the RNA (Figure 5(e)).⁴⁶ Thus, the helix that closes the apical loop is stem S1 of the pseudoknot. S1 bends significantly to make loop-loop base-pairing that constitutes stem S2 (Figure 5(e)).⁴⁷ In the bent region, regular helical structure is disrupted and S1 stacks on S2 using a non-canonical base pair. S2 also stacks on the P1 helix that closes the domain structure. As a result of complex folding and multiple non-canonical interactions, both L1 and L3 are positioned in the major groove of S2, where L3 makes tertiary base triples A61•(C68-G40) and G62•(G67-C41) (Figure 5(f)) and stacks on top of the bound c-di-GMP. The terminal nucleotides of S1 and P1 as well as base pairs of S2 create walls and floor of the ligand-binding pocket. Guanine bases of c-di-GMP form base-specific (A69 and G73) and backbone interactions with the P1-S2 stem, and are additionally stabilized by intercalation of adenosine A70 from S2 between the bases of the ligand (Figure 5(g)). Thus, like in the SAM-II riboswitch, the c-di-GMP pseudoknot participates in the formation of the ligand-binding pocket but in contrast to the aforementioned riboswitches, stabilization of the pseudoknot fold contributes to the stabilization of an additional riboswitch element, a regulatory helix P1.

Pseudoknots in the Junctional Riboswitches

Several other metabolite-sensing domains of riboswitches integrate a pseudoknot into the junctional structure. These riboswitches are composed of three or more stem-loops joined by a junction. Pseudoknots typically lock stem-loops through long-range base-pairing critical for reinforcing the RNA structure and, with exception of the tetrahydrofolate (THF) riboswitch, do not bind metabolites.

The THF riboswitch structure is formed by side-by-side placement of two sets of coaxially stacked helices organized by a three-way-junction and a distal pseudoknot.^{48,49} The pseudoknot involves long-range base-pairing between apical and interior loops, reminiscent of the c-di-GMP-II riboswitch.⁴⁷ One ligand molecule sits on the top of the helix formed by the long-range base-pairing and contributes to the formation of the pseudoknot, which, in turn, stabilizes a regulatory helix P1 through stacking interactions.⁴⁹ Another ligand molecule is located in the three-way junction and helps to orient the pseudoknot-forming hairpin towards the interior loop. Since helix P1 closes the entire structure and an additional hairpin precedes the pseudoknot-forming hairpin, the pseudoknot does not fit any of the six classes and would be grouped with the HH-type if P1 is not formed.⁴⁸

Formation of the pseudoknot between apical and interior loops was also observed in the structure of the SAM-I riboswitch⁵⁰ and was predicted in SAM-IV and SAM-I/IV riboswitches.⁵¹ In lysine,^{52,53} guanine,^{54,55} adenine,⁵⁵ deoxyguanosine⁵⁶ and

adenosylcobalamin (AdoCbl)^{57,58} riboswitches, pseudoknots form long-range loop-loop interactions, classified as H-H type (Figure 1(f)) or HHH type (Figure 1(i)). These interactions primarily serve to fasten peripheral stems in the lysine and adenosylcobalamin riboswitches and enforce parallel alignment of two stems in purine riboswitches, thereby strengthening the overall RNA conformation in lysine and purine riboswitches and trapping the ligand in its ligand-binding pocket in the adenosylcobalamin riboswitches.

PSEUDOKNOTS IN RIBOZYMES

The vast majority of naturally occurring catalytic RNAs, or ribozymes, catalyze RNA strand scission and ligation in a variety of RNA processing reactions including RNA splicing, viral replication, tRNA biosynthesis, and mRNA cleavage. In addition, as part of the ribosomal RNA, a ribozyme apparently functions within the ribosome to link amino acids during protein synthesis. Biochemical and structural studies revealed a number of pseudoknots involved in the formation of several types of ribozymes, ranging from small pseudoknot-based self-cleaving ribozymes to large and complex self-splicing introns (Table 1). Some pseudoknotted ribozymes, such as *glmS* ribozyme, have been proven to directly participate in gene expression control.

glmS Ribozyme-Riboswitch

The *glmS* ribozyme-riboswitch is the first known example of a natural ribozyme that has evolved to control gene expression by cleaving itself using a small cellular metabolite.⁵⁹ The ribozyme is located in the 5'-UTR of the *glmS* mRNA encoding an enzyme synthesizing the phosphorylated sugar glucosamine-6-phosphate (GlcN6P). The ribozyme specifically binds GlcN6P when the metabolite is present in cells at over-threshold concentrations. Bound GlcN6P cleaves off the 5'-terminus of the ribozyme leaving a hydroxyl on the 5' end of the 3' portion of the mRNA. RNase J recognizes the non-phosphorylated 5' end and degrades the mRNA thus downregulating production of the enzyme.⁶⁰

The ribozyme adopts a triple-pseudoknotted fold that contains four hairpins P1, P2, P3 and P4-P4.1 (Figure 6(a)).^{61,62} The first pseudoknot (PK1) is formed through base-pairing (P2.2) of the 5' mRNA region with the 3' side of the P2 hairpin loop (Figure 6(a)). The second pseudoknot (PK2) is formed by base-pairing (P2.1) of the P1-P2 linker with the 5' side of the loop in P2. Thus, PK2 is embedded within PK1. The third pseudoknot (PK3) involves base-pairing (P3.1) of the apical loop in P3 with the 3' region of the ribozyme (Table 1).

The ribozyme structure is composed of three coaxial stacks of helices packed side by side (Figure 6(b-g)). P2.1 and the central part of the P1-P3.1 stack contain nested pseudoknots PK1 (Figure 6(b),(e)) and PK2 (Figure 6(c),(f)) that comprise the catalytic core. Pseudoknot PK3 stacks on P2 and branches out into P4-P4.1 stem that buttresses the core via interactions with the minor groove of P2.1 from PK2 (Figure 6(d),(g)). Like fluoride and preQ₁-II riboswitch pseudoknots (Figure 4(b),(j)), PK3 contains a hairpin in L3, attached to the pseudoknot stem via a couple of minor groove triples at the bottom of S1.3 and the packing of a stacked helix into the minor groove of the top part of S1.3 (Figure 6(h)).

The two pseudoknots of the catalytic core are reinforced by multiple stacking interactions between stems and loops so that the core adopts a tight nest-like structure containing a semi-open ligand-binding pocket (Figure 6(i)). Bound GlcN6P lies in a channel lined by the major groove sides of P2.1 (part of L2.1) and P2.2 (S1.1), and loops L2.1 and L3.1 of PK1. Stacked nucleotides A35 and G(+1) ('ligand' stack in Figure 6(i)), brought together by the regions adjacent to P2.1 and P2.2, zip the channel on top and orient the ligand ring almost parallel to the G(+1) base (Figure 6(j)). The ligand makes direct and Mg²⁺-mediated hydrogen bonds with the loops (G65 and A50) and helices (C2, G64, and U51) of PK1 and PK2. The A(−1) nucleotide loops out of the pocket, and twists the RNA backbone into the conformation required for the in-line attack of the scissile phosphate by the 2'-OH. The catalytic amine of GlcN6P orients towards the scissile phosphate (Figure 6(j)) and assists in the general acid-base catalysis. Thus, two pseudoknots of the *glmS* ribozyme create a catalytic core and the cofactor-binding site to position all functional groups in the active site. The third pseudoknot fastens the catalytic core with a peripheral stem.

Other Small Ribozymes

Several different types of ribozymes have been found in pre-mRNA molecules. Examples include ribozymes structurally and biochemically related to the double-pseudoknotted HDV ribozyme,^{63,64} as well as pseudoknotted variants of hammerhead ribozymes.⁶⁵ Although the biological functions of these ribozymes remain unproven, some could be involved in gene regulation based on their genomic contexts, for instance, by producing mRNAs of different length and controlling mRNA degradation.

Allosteric Group I Self-Splicing Intron

Self splicing introns are common components of 'selfish' mobile genetic elements. The introns appear not to carry out any biological functions except self-cleaving from pre-mRNA and other RNA species. Typically, group I introns fold into large complex structures which may possess pseudoknot pairings. One of the group I introns was harnessed by a pathogenic bacterium to regulate expression of a putative virulence gene in tandem with the pseudoknotted c-di-GMP-II riboswitch.⁴⁶ The riboswitch senses elevated concentrations of c-di-GMP and after binding to this second messenger, re-organizes pairing alignments in the catalytic site of the downstream group I intron. This change causes excision of the intron so that two pieces of the disrupted Shine-Dalgarno sequence create a full-length ribosome binding site thus promoting translation of the gene. In the absence of c-di-GMP, the ribozyme cleaves at the alternative site so that mRNA lacks the Shine-Dalgarno sequence and cannot be translated. Thus, two pseudoknotted RNA structures cooperate to modulate alternative mRNA processing.

DIVERSITY AND SIMILARITIES IN REGULATORY PSEUDOKNOTS

The aforementioned examples of pseudoknotted RNAs demonstrate a multitude of structural principles used for formation of pseudoknots and a diversity of means to utilize these structural elements. Practically, all pseudoknots in regulatory RNAs adopt unique folds and do not resemble each other. Simple pseudoknots differ in the length of helical elements, ranging from 4 to 10 bps in S1 and from 3 to 10 bps in S2 (Figure 7). The average stem

lengths of 7 and 6 bps, respectively, are, however, longer than those reported for the collection of pseudoknots with $L2=0$ in the PseudoBase, where S1 peaks at 3 bps and S2 favors 5 or 6 bps.⁶⁶ The length of the loops varies significantly as well. L1 typically ranges from 1 to 6 nts with most species bearing >2 nts, while the most frequent L1 in the database is 1 nt⁶⁶. L2 tends to be 0-1 nts, a common feature with other pseudoknots.⁶⁶ L3, on the other hand, is at least 7 nts long and often bears an additional self-contained hairpin, although L3 peaks at 3 and 6 nts in this pseudoknot collection.⁶⁶

Despite low structural similarity, regulatory pseudoknotted RNAs share a few traits important for the formation of pseudoknots and exertion of genetic control. As was noted earlier for pseudoknots in the PseudoBase collection,⁶⁶ nucleotide composition is biased in L1 and L3. Most L1 loops >1 nt are uracil rich, although an adenine appears to be invariant in the single-nucleotide L1. A single adenosine may be favored in L1 for structural reasons if it participates in specific tertiary interactions with the major groove of S2. Formation of such stable interactions was not documented¹⁸ and any nucleotide can be tolerated in L1.⁶⁷ Nonetheless, substitutions of the adenine by other nucleobases affect stability of the pseudoknot.⁶⁸ L3 often contains stretches of adenines, involved in the formation of triple helices through the tertiary base-pairing with S1. The L3-S1 interactions include canonical A-minor (Figure 4(c)) and related interactions involving adenines, such as inclined A-minor interactions (Figure 4(o)) and an A-amino kissing motif (Figure 4(g)). As illustrated in Figure 4, L3 loops, in most cases, make more extensive tertiary interactions with the shallow groove of S1 than the shorter L1 does with the deep groove of S2. Therefore, S1, located closer to the 5' end, nucleates folding of the pseudoknot.

Formation and stabilization of S2, on the other hand, depend on the presence of external factors such as proteins and metabolites. Not surprisingly, virtually all regulatory pseudoknots sequester regulatory regions, for instance, a Shine-Dalgarno sequence, within S2. In more complex RNAs, such as c-di-GMP-II and THF riboswitches, S2 assists the formation of the regulatory helix P1 through stacking interactions (Figure 5(e)). To stabilize S2, regulatory pseudoknots can directly interact with a ligand, as in the SAM-II and c-di-GMP-II riboswitches, where bound ligands are positioned in the major groove of S2 (Figure 4(j),(n)). In fluoride, preQ₁ and THF riboswitches, pseudoknot formation depends on the collinear stacking between S1 and S2, facilitated by binding of the ligands at or near the helical interfaces (Figure 4(b),(f),(j)). Stacking is enforced by different means: a Mg²⁺-coordinated fluoride anion bridges the stems via extensive contacts with the RNA backbone (Figure 4(d)) while planar preQ₁ ligands intercalate between the stems (Figure 4(h),(l)). It was noticed earlier^{40,42} that preQ₁ and the nucleobase of SAM do not simply ensure uninterrupted stacking; these moieties can be paralleled with nucleobases in the S2-stabilizing base triples observed in the non-regulatory pseudoknots. For instance, the major groove U•(A-U) triple in the pseudoknot from human telomerase RNA (hTR),⁶⁹ critical for the pseudoknot stability, resemble ligand-mediated triples in riboswitch pseudoknots, created by substituting omitted nucleotides with ligand nucleobases.

A noticeable feature in the regulatory pseudoknots is participation of loop nucleotides in the formation of ligand-binding sites. Practically all ligands interact with pseudoknot loops and not only with L1, which spans the groove of S2 and is expected to be closer to bound

ligands, but also with L3, positioned along the groove of S1. In fact, L3 does not simply connect the pseudoknot stems; this long loop could be pre-organized as part of the triple helix with S1 and therefore is often adapted for ligand recognition and formation of ligand-induced tertiary interactions contributing to the pseudoknot stabilization.

Sharp turns of the RNA backbone in loop-stem junctions, alongside with the close position of RNA strands in long loops, bring negatively charged sugar-phosphate moieties close to each other. To compensate the negative charge of the backbone and allow correct RNA folding, the regulatory pseudoknots have another common trait – Mg^{2+} cations bound to the phosphate moieties of RNA. Location of the metal sites differs in the diverse pseudoknot folds but clearly shows a tendency of cations to assist in close positioning of pseudoknot loops and stems.^{34,37,40,47} Mg^{2+} cations appear to contribute to the pseudoknot folding by inducing molecular compaction and restraining flexible elements in the majority of the regulatory pseudoknots *in vitro*, although the cations are not necessarily essential for ligand-binding to the isolated pseudoknotted aptamers.^{36,45,70} Given that pseudoknot folding rate should keep up with a high rate of transcription *in vivo*, contribution of metal cations to the adaptation of conformations competent to ligand binding is likely more significant in live cells.

PSEUDOKNOT FOLDING AND REGULATORY RESPONSE

Biochemical studies have proven that the pseudoknot formation is critical for the ligand-mediated gene expression in simple pseudoknot-based systems. What are the pseudoknot elements actually needed to be pre-folded prior to the ligand binding and how stable the pseudoknot should be, if folded, to bind ligands and exert gene control? The answers totally depend on the regulatory system.

Pseudoknots in the gp32 and ribosomal protein mRNAs do not include regulatory elements and therefore can in principle be entirely folded prior to the protein binding. In fact, pre-folding of the pseudoknot is required for high affinity binding of the ribosomal protein S15 since the protein recognizes the region near the S1-S2 interface and at the tip of S2.⁷¹ On the other hand, the S15 pseudoknot, or at least S2, should be easily unfolded by the ribosome in order to place the initiation codon into the P-site of the ribosome in the absence of S15. Therefore, S2 of the S15 pseudoknot does not need to be very stable since S15 would bind across S2 and reinforce the structure.

Another regulatory system involving a pre-folded pseudoknotted RNA is the *glmS* riboswitch-ribozyme⁷². Down-regulation in this system is triggered by the ligand-induced self-cleavage of the mRNA that requires ligand binding but not alternative pairing. Therefore, the *glmS* RNA adopts a unique stable structure enforced by three pseudoknots for specific ligand recognition and backbone cleavage and not for the interplay of alternative ligand-free and ligand-bound conformations.

Unlike the *glmS* ribozyme and pseudoknots that serves as protein recognition platforms, ligand-bound regulatory pseudoknots in riboswitches typically sequester regions involved in genetic control within S2 of the pseudoknot. Thus, to exert genetic control, a pseudoknot or at least S2 should be either unfolded or adopt an unstable conformation if RNA does not

encounter a cognate ligand. First biochemical tests of riboswitch conformations by in-line probing that report on flexible and essentially non-paired nucleotides suggested that regulatory riboswitch pseudoknots typically form S1 in the absence of cognate ligands.^{33,46} Formation of stable S2 however was not found in all riboswitches. Pre-formation of S2 is possible in some RNAs, for instance, c-di-GMP-II⁴⁶ and fluoride³³ riboswitches, as suggested by a lack of scission in in-line probing of nucleotides which constitute S2. In other RNAs, such as SAM-II⁷³ and SAH,⁷⁴ S2 appears to be partially or transiently folded since some nucleotides in S2 were not cleaved while other nucleotides showed reduced scission (become paired) in the assay upon addition of the ligand. In the THF⁷⁵ and both preQ^{35,39}₁ riboswitches, S2 is not likely to be a prevalent conformation in the ligand-free state since nucleotides of S2 are flexible and show reduced scission after ligand binding.

How can riboswitch ligands recognize disorganized binding pockets to occlude regulatory elements or induce structural re-arrangements for modulation of gene expression? To answer these questions, ligand-dependent folding of relatively small SAM-II and preQ₁ pseudoknots has been characterized in detail by a variety of methods.

Biophysical studies^{76,77} have tracked a hierarchical folding of the SAM-II riboswitch and revealed that the RNA adopts multiple conformations along its folding pathway. In the absence of Mg²⁺ cations, SAM-II RNA adopts an elongated conformation with unstructured S2 and pre-formed stem-loop S1, whose helix makes triple-stranded arrangement with L3 (Figure 8). Physiological concentrations of Mg²⁺ cations make the RNA conformation more compact, probably stabilizing S1-L3 contacts and facilitating partial preorganization of L1 and S2. Thus, the function of Mg²⁺ cations is to restrict the dynamic conformational ensemble of the RNA to conformations competent for ligand binding by adapting transient pseudoknot-like structures which are in quick exchange with partially folded conformations. The pseudoknot becomes more compact and stable upon SAM binding as a result of the formation of a new set of base pairs, absent or sparsely populated in the ligand-free state. Consistent with the selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) analysis,⁴² which has demonstrated that flexibility of S2 becomes significantly reduced upon SAM binding, these new interactions organize S2 and S2-L1 triple arrangements. Thus, SAM binding sequesters the Shine-Dalgarno sequence, located in the 3' strand of S2, within the S2 helix, preventing base-pairing with the 16S rRNA of the 30S ribosomal subunit and repressing translation initiation of the downstream gene. In the absence of SAM, the ribosome can gain access to the non-paired Shine-Dalgarno sequence, turning translation of the gene 'ON'.

Similarly to the SAM-II pseudoknot, an isolated aptamer domain of the 'transcriptional' preQ₁-I riboswitch from *Fusobacterium nucleatum*, in absence of the ligand, adopts multiple conformations in solution and can pre-fold into a pseudoknot-like structure that reaches significant population at the physiological Mg²⁺ concentrations.⁷⁸ In the longer construct encompassing an expression platform and resembling natural mRNA, the RNA appears to be predominantly in the equilibrium between two folds. Both folds contain a preformed S1 stem-loop whereas the 3' portion of the RNA adopts either an anti-terminator (ON state) or a terminator (OFF state) hairpin. Ligand binding traps a transient pseudoknot-like conformation and drives the equilibrium towards the OFF state, causing premature

transcription termination. Aptamer-ligand complex formation is very fast and the riboswitch likely operates through the kinetically controlled pathway, although transcriptional pausing may allow equilibration between the RNA and ligand, i.e. thermodynamically controlled regulation. Interestingly, there is no overlap between the aptamer and terminator structures in *F. nucleatum*⁷⁹ and *B. subtilis*^{36,37} preQ₁-I riboswitches. This “noncommunicative” situation has been resolved by introducing an additional stem-loop, an antiterminator hairpin that overlaps with both, terminator and aptamer structures.

The preQ₁-I riboswitch pseudoknot from the thermophile *Thermoanaerobacter tengcongensis* appears to be more pre-organized than its mesophilic counterpart from *F. nucleatum*. Solution data have revealed an ensemble of compact ligand-free conformations, which produce a stable crystallizable structure closely resembling the ligand-bound conformation.⁸⁰ In the ligand-free structure, the absence of the metabolite causes a nucleotide shift in S2, resulting in unstacking of terminal nucleotides of S2 from the structural core. Such increase in accessibility is illustrative of how the 30S ribosomal subunit can gain access to the Shine-Dalgarno sequence, overlapped with the terminal nucleotides of S2, for translation of the gene.⁸⁰

In contrast to preQ₁-I riboswitches, the preQ₁-II pseudoknot harbors an extra-stem loop structure in L3 upstream of the Shine-Dalgarno sequence. Does this additional element, observed also in the fluoride riboswitch, affect the pseudoknot formation and ligand binding? Single molecule studies show that the preQ₁-II RNA interchanges between pseudoknot-like and stem-loop conformations⁸¹ as was described for the SAM-II riboswitch,⁷⁷ which does not contain an extra stem-loop. However, the pseudoknot conformations are more populated in the preQ₁-II than in the SAM-II riboswitch. Deletion of the extra stem-loop in L3 of the preQ₁-II RNA substantially increases both the on- and off-rates of ligand binding but modestly decreases the lifetime of the ligand-bound and intermediate states, according to the minimal L3-ligand contact⁴⁰ and moderate reduction in binding affinity.⁸¹ It is speculated that the extra stem-loop may influence pseudoknot dynamics in a manner that decreases its propensity to spontaneously fold and increases its responsiveness to ligand binding.⁸¹

The dynamic nature of the ligand-free SAM-II and preQ₁ riboswitches highlights an important question in macromolecule-ligand binding, intensively studied recently with respect to riboswitches.⁸² The solution data suggest that the riboswitch aptamer rapidly samples pre-existing conformations during the ligand-sensing phase, and the cognate ligand appears to capture only ligand-competent conformations present at least in minor population in the dynamic ensemble. Such behavior of the system agrees with postulates of the “conformational selection” or “conformational capture” model of ligand binding. This hypothesis contrasts to the “induced fit” concept which claims that conformational rearrangements between the free and ligand-bound conformations are driven by the nature of the binding interactions. Since many bound ligands are enveloped in riboswitch structures, ligands cannot capture final RNA conformations right away. Therefore, subsequent structural adjustments should depend on the nature of the initial ligand binding interactions in agreement with the induced fit model, even if a transiently bound ligand keeps searching for the ‘closing’ conformation on the local scale.

CONCLUSION

Extensive biochemical and biophysical works have revealed important contributions of pseudoknots in the gene expression control. Pseudoknots are not only restricted to being building blocks in large structures. These small elements intrinsically possess all characteristics required to exert genetic control in response to small and large ligands of diverse chemical nature. Three-dimensional structures of regulatory pseudoknots have uncovered a variety of folds and diverse molecular principles for maintaining the structure, ligand binding, and regulatory activities. Nevertheless, pseudoknots share common features, not on the sequence but on the structural level. These features are mostly important for defining stable pseudoknot topology and can be found in pseudoknots involved in other cellular activities. Despite common traits with other pseudoknots, regulatory pseudoknots in riboswitches possess unique adaptations to exert genetic control. Riboswitch pseudoknots exploit structural tricks to destabilize ligand-free and stabilize ligand-bound conformations. This structural interplay provides a molecular basis for the riboswitch-driven genetic regulation.

Several laboratories have begun to elucidate the mechanisms of the pseudoknot-dependent genetic control at the molecular level. These studies have significantly advanced our understanding of the mechanistic principles involved in the RNA-driven regulation of gene expression. However, the research progress is limited by a few convenient systems and often involves truncated sequence variants that bear little resemblance with natural mRNA species. In addition, the vast majority of the current studies cannot take into account co-transcriptional RNA folding and the chemical environment present in live cells. New methodologies to elucidate regulatory RNA in the natural environment will be a large step forward to understand how genetic networks operate and how we can modulate genetic outcomes. Such experimentation will provide us with the most accurate information for comprehension of complex regulatory circuits and accelerate the development of novel biotechnological tools and therapeutic interventions.

References

1. Rietveld K, Van Poelgeest R, Pleij CW, Van Boom JH, Bosch L. The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. *Nucleic Acids Res.* 1982; 10:1929–1946. [PubMed: 7079175]
2. Dam E, Pleij K, Draper D. Structural and functional aspects of RNA pseudoknots. *Biochemistry.* 1992; 31:11665–11676. [PubMed: 1280160]
3. Studnicka GM, Rahn GM, Cummings IW, Salser WA. Computer method for predicting the secondary structure of single-stranded RNA. *Nucleic Acids Res.* 1978; 5:3365–3387. [PubMed: 100768]
4. Pleij CW, Rietveld K, Bosch L. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.* 1985; 13:1717–1731. [PubMed: 4000943]
5. Lescoute A, Leontis NB, Massire C, Westhof E. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.* 2005; 33:2395–2409. [PubMed: 15860776]
6. Brierley I, Gilbert RJ, Pennell S. RNA pseudoknots and the regulation of protein synthesis. *Biochem Soc Trans.* 2008; 36:684–689. [PubMed: 18631140]
7. Brierley I, Pennell S, Gilbert RJ. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat Rev Microbiol.* 2007; 5:598–610. [PubMed: 17632571]

8. Giedroc DP, Cornish PV. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.* 2009; 139:193–208. [PubMed: 18621088]
9. van Batenburg FH, Gulyaev AP, Pleij CW. PseudoBase: structural information on RNA pseudoknots. *Nucleic Acids Res.* 2001; 29:194–195. [PubMed: 11125088]
10. Westhof E, Jaeger L. RNA pseudoknots. *Curr. Opin. Struct. Biol.* 1992; 2:327–333.
11. Kim N, Shiffeldrim N, Gan HH, Schlick T. Candidates for novel RNA topologies. *J Mol Biol.* 2004; 341:1129–1144. [PubMed: 15321711]
12. Pleij CW. Pseudoknots: a new motif in the RNA game. *Trends Biochem Sci.* 1990; 15:143–147. [PubMed: 1692647]
13. Puglisi JD WJ, Tinoco I Jr. RNA Pseudoknots. *Acc Chem Res.* 1991; 24:152–158.
14. Han K, Byun Y. PSEUDOVIEWER2: Visualization of RNA pseudoknots of any type. *Nucleic Acids Res.* 2003; 31:3432–3440. [PubMed: 12824341]
15. McPheeters DS, Stormo GD, Gold L. Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA. *J Mol Biol.* 1988; 201:517–535. [PubMed: 3262167]
16. Alberts BM, Frey L. T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA. *Nature.* 1970; 227:1313–1318. [PubMed: 5455134]
17. Shamoo Y, Tam A, Konigsberg WH, Williams KR. Translational repression by the bacteriophage T4 gene 32 protein involves specific recognition of an RNA pseudoknot structure. *J Mol Biol.* 1993; 232:89–104. [PubMed: 8331672]
18. Holland JA, Hansen MR, Du Z, Hoffman DW. An examination of coaxial stacking of helical stems in a pseudoknot motif: the gene 32 messenger RNA pseudoknot of bacteriophage T2. *RNA.* 1999; 5:257–271. [PubMed: 10024177]
19. Shamoo Y, Friedman AM, Parsons MR, Konigsberg WH, Steitz TA. Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. *Nature.* 1995; 376:362–366. [PubMed: 7630406]
20. Philippe C, Portier C, Mougél M, Grunberg-Manago M, Ebel JP, Ehresmann B, Ehresmann C. Target site of *Escherichia coli* ribosomal protein S15 on its messenger RNA. Conformation and interaction with the protein. *J Mol Biol.* 1990; 211:415–426. [PubMed: 2407855]
21. Chiaruttini C, Milet M, Springer M. A long-range RNA-RNA interaction forms a pseudoknot required for translational control of the IF3-L35-L20 ribosomal protein operon in *Escherichia coli*. *EMBO J.* 1996; 15:4402–4413. [PubMed: 8861967]
22. Tang CK, Draper DE. Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell.* 1989; 57:531–536. [PubMed: 2470510]
23. Haentjens-Sitri J, Allemand F, Springer M, Chiaruttini C. A competition mechanism regulates the translation of the *Escherichia coli* operon encoding ribosomal proteins L35 and L20. *J Mol Biol.* 2008; 375:612–625. [PubMed: 18037435]
24. Serganov A, Ennifar E, Portier C, Ehresmann B, Ehresmann C. Do mRNA and rRNA binding sites of *E. coli* ribosomal protein S15 share common structural determinants? *J Mol Biol.* 2002; 320:963–978. [PubMed: 12126618]
25. Serganov A, Benard L, Portier C, Ennifar E, Garber M, Ehresmann B, Ehresmann C. Role of conserved nucleotides in building the 16 S rRNA binding site for ribosomal protein S15. *J Mol Biol.* 2001; 305:785–803. [PubMed: 11162092]
26. Mathy N, Pellegrini O, Serganov A, Patel DJ, Ehresmann C, Portier C. Specific recognition of rpsO mRNA and 16S rRNA by *Escherichia coli* ribosomal protein S15 relies on both mimicry and site differentiation. *Mol Microbiol.* 2004; 52:661–675. [PubMed: 15101974]
27. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res.* 2003; 31:439–441. [PubMed: 12520045]
28. Serganov A, Polonskaia A, Ehresmann B, Ehresmann C, Patel DJ. Ribosomal protein S15 represses its own translation via adaptation of an rRNA-like fold within its mRNA. *EMBO J.* 2003; 22:1898–1908. [PubMed: 12682022]
29. Scott LG, Williamson JR. Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with its 5'-translational operator mRNA. *J Mol Biol.* 2001; 314:413–422. [PubMed: 11846555]

30. Philippe C, Eyermann F, Benard L, Portier C, Ehresmann B, Ehresmann C. Ribosomal protein S15 from *Escherichia coli* modulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proc Natl Acad Sci U S A*. 1993; 90:4394–4398. [PubMed: 7685101]
31. Spedding G, Draper DE. Allosteric mechanism for translational repression in the *Escherichia coli* alpha operon. *Proc Natl Acad Sci U S A*. 1993; 90:4399–4403. [PubMed: 7685102]
32. Marzi S, Myasnikov AG, Serganov A, Ehresmann C, Romby P, Yusupov M, Klaholz BP. Structured mRNAs regulate translation initiation by binding to the platform of the ribosome. *Cell*. 2007; 130:1019–1031. [PubMed: 17889647]
33. Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, Breaker RR. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science*. 2012; 335:233–235. [PubMed: 22194412]
34. Ren A, Rajashankar KR, Patel DJ. Fluoride ion encapsulation by Mg^{2+} ions and phosphates in a fluoride riboswitch. *Nature*. 2012; 486:85–89. [PubMed: 22678284]
35. Roth A, Winkler WC, Regulski EE, Lee BW, Lim J, Jona I, Barrick JE, Ritwik A, Kim JN, Welz R, et al. A riboswitch selective for the queuosine precursor preQ₁ contains an unusually small aptamer domain. *Nat Struct Mol Biol*. 2007; 14:308–317. [PubMed: 17384645]
36. Kang M, Peterson R, Feigon J. Structural insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol Cell*. 2009; 33:784–790. [PubMed: 19285444]
37. Klein DJ, Edwards TE, Ferre-D'Amare AR. Cocystal structure of a class I preQ₁ riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat Struct Mol Biol*. 2009; 16:343–344. [PubMed: 19234468]
38. Spitale RC, Torelli AT, Krucinska J, Bandarian V, Wedekind JE. The structural basis for recognition of the PreQ₀ metabolite by an unusually small riboswitch aptamer domain. *J Biol Chem*. 2009; 284:11012–11016. [PubMed: 19261617]
39. Meyer MM, Roth A, Chervin SM, Garcia GA, Breaker RR. Confirmation of a second natural preQ₁ aptamer class in Streptococcaceae bacteria. *RNA*. 2008; 14:685–695. [PubMed: 18305186]
40. Liberman JA, Salim M, Krucinska J, Wedekind JE. Structure of a class II preQ₁ riboswitch reveals ligand recognition by a new fold. *Nat Chem Biol*. 2013; 9:353–355. [PubMed: 23584677]
41. Batey RT. Recognition of S-adenosylmethionine by riboswitches. *Wiley Interdiscip Rev RNA*. 2011; 2:299–311. [PubMed: 21957011]
42. Gilbert SD, Rambo RP, Van Tyne D, Batey RT. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat Struct Mol Biol*. 2008; 15:177–182. [PubMed: 18204466]
43. Poiata E, Meyer MM, Ames TD, Breaker RR. A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA*. 2009; 15:2046–2056. [PubMed: 19776155]
44. Doshi U, Kelley JM, Hamelberg D. Atomic-level insights into metabolite recognition and specificity of the SAM-II riboswitch. *RNA*. 2012; 18:300–307. [PubMed: 22194311]
45. Edwards AL, Reyes FE, Heroux A, Batey RT. Structural basis for recognition of S-adenosylhomocysteine by riboswitches. *RNA*. 2010; 16:2144–2155. [PubMed: 20864509]
46. Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science*. 2010; 329:845–848. [PubMed: 20705859]
47. Smith KD, Shanahan CA, Moore EL, Simon AC, Strobel SA. Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proc Natl Acad Sci U S A*. 2011; 108:7757–7762. [PubMed: 21518891]
48. Huang L, Ishibe-Murakami S, Patel DJ, Serganov A. Long-range pseudoknot interactions dictate the regulatory response in the tetrahydrofolate riboswitch. *Proc Natl Acad Sci U S A*. 2011; 108:14801–14806. [PubMed: 21873197]
49. Trausch JJ, Ceres P, Reyes FE, Batey RT. The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer. *Structure*. 2011; 19:1413–1423. [PubMed: 21906956]
50. Montange RK, Batey RT. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*. 2006; 441:1172–1175. [PubMed: 16810258]

51. 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]
52. Garst AD, Heroux A, Rambo RP, Batey RT. Crystal structure of the lysine riboswitch regulatory mRNA element. *J Biol Chem*. 2008; 283:22347–22351. [PubMed: 18593706]
53. Serganov A, Huang L, Patel DJ. Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature*. 2008; 455:1263–1267. [PubMed: 18784651]
54. Batey RT, Gilbert SD, Montange RK. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*. 2004; 432:411–415. [PubMed: 15549109]
55. Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol*. 2004; 11:1729–1741. [PubMed: 15610857]
56. Pikovskaya O, Polonskaia A, Patel DJ, Serganov A. Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. *Nat Chem Biol*. 2011; 7:748–755. [PubMed: 21841796]
57. Peselis A, Serganov A. Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. *Nat. Struct. Mol. Biol*. 2012; 19:1182–1184. [PubMed: 23064646]
58. Johnson JE Jr, Reyes FE, Polaski JT, Batey RT. B12 cofactors directly stabilize an mRNA regulatory switch. *Nature*. 2012; 492:133–137. [PubMed: 23064232]
59. 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]
60. Collins JA, Irnov I, Baker S, Winkler WC. Mechanism of mRNA destabilization by the glmS ribozyme. *Genes Dev*. 2007; 21:3356–3368. [PubMed: 18079181]
61. Cochrane JC, Lipchock SV, Strobel SA. Structural investigation of the GlmS ribozyme bound to its catalytic cofactor. *Chem Biol*. 2007; 14:97–105. [PubMed: 17196404]
62. Klein DJ, Ferre-D'Amare AR. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*. 2006; 313:1752–1756. [PubMed: 16990543]
63. Salehi-Ashtiani K, Luptak A, Litovchick A, Szostak JW. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*. 2006; 313:1788–1792. [PubMed: 16990549]
64. Webb CH, Riccitelli NJ, Ruminski DJ, Luptak A. Widespread occurrence of self-cleaving ribozymes. *Science*. 2009; 326:953. [PubMed: 19965505]
65. Perreault J, Weinberg Z, Roth A, Popescu O, Chartrand P, Ferbeyre G, Breaker RR. Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. *PLoS Comput Biol*. 2011; 7:e1002031. [PubMed: 21573207]
66. Aalberts DP, Hodas NO. Asymmetry in RNA pseudoknots: observation and theory. *Nucleic Acids Res*. 2005; 33:2210–2214. [PubMed: 15831794]
67. Du Z, Hoffman DW. An NMR and mutational study of the pseudoknot within the gene 32 mRNA of bacteriophage T2: insights into a family of structurally related RNA pseudoknots. *Nucleic Acids Res*. 1997; 25:1130–1135. [PubMed: 9092620]
68. Theimer CA, Wang Y, Hoffman DW, Krisch HM, Giedroc DP. Non-nearest neighbor effects on the thermodynamics of unfolding of a model mRNA pseudoknot. *J Mol Biol*. 1998; 279:545–564. [PubMed: 9641977]
69. Theimer CA, Blois CA, Feigon J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell*. 2005; 17:671–682. [PubMed: 15749017]
70. Zhang Q, Kang M, Peterson RD, Feigon J. Comparison of solution and crystal structures of preQ₁ riboswitch reveals calcium-induced changes in conformation and dynamics. *J Am Chem Soc*. 2011; 133:5190–5193. [PubMed: 21410253]
71. Ehresmann C, Philippe C, Westhof E, Benard L, Portier C, Ehresmann B. A pseudoknot is required for efficient translational initiation and regulation of the *Escherichia coli* rpsO gene coding for ribosomal protein S15. *Biochem Cell Biol*. 1995; 73:1131–1140. [PubMed: 8722030]
72. Soukup GA. Core requirements for glmS ribozyme self-cleavage reveal a putative pseudoknot structure. *Nucleic Acids Res*. 2006; 34:968–975. [PubMed: 16464827]

73. 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]
74. Wang JX, Lee ER, Morales DR, Lim J, Breaker RR. Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Mol Cell.* 2008; 29:691–702. [PubMed: 18374645]
75. 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]
76. Chen B, Zuo X, Wang YX, Dayie TK. Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Res.* 2012; 40:3117–3130. [PubMed: 22139931]
77. Haller A, Rieder U, Aigner M, Blanchard SC, Micura R. Conformational capture of the SAM-II riboswitch. *Nat Chem Biol.* 2011; 7:393–400. [PubMed: 21532598]
78. Santner T, Rieder U, Kreutz C, Micura R. Pseudoknot preorganization of the preQ₁ class I riboswitch. *J Am Chem Soc.* 2012; 134:11928–11931. [PubMed: 22775200]
79. Rieder U, Kreutz C, Micura R. Folding of a transcriptionally acting preQ₁ riboswitch. *Proc Natl Acad Sci U S A.* 2010; 107:10804–10809. [PubMed: 20534493]
80. Jenkins JL, Krucinska J, McCarty RM, Bandarian V, Wedekind JE. Comparison of a preQ₁ riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J Biol Chem.* 2011; 286:24626–24637. [PubMed: 21592962]
81. Souliere MF, Altman RB, Schwarz V, Haller A, Blanchard SC, Micura R. Tuning a riboswitch response through structural extension of a pseudoknot. *Proc Natl Acad Sci U S A.* 2013; 110:E3256–3264. [PubMed: 23940363]
82. Haller A, Souliere MF, Micura R. The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc Chem Res.* 2011; 44:1339–1348. [PubMed: 21678902]

BOX 1 SERIES OF STEPS ON PSEUDOKNOT CLASSIFICATION

To simplify comprehension of complex pseudoknots, PseudoBase database groups pseudoknots into six basic types depending on their planar representation. H-type pseudoknots contain a hairpin whose apical loop makes tertiary base pairs with a single stranded region outside of the loop, forming two collinear stems S1 and S2 connected by three linkers L1, L2 and L3 (Fig. 1(i)). Integration of a secondary structural element, such as a hairpin, into a pseudoknot results in additional classes. The pseudoknot is classified as LL-type if an extra hairpin is inserted into the apical loop of a pseudoknot hairpin, downstream of the tertiary base-pairing. In this pseudoknot, the apical loop of the 5' hairpin can be viewed as a pseudoknot-forming internal loop. If a structural element is inserted into L3, the 5' hairpin pairs with the single-stranded region downstream of the structural element, forming an HL_{out}-type pseudoknot. Conversely, the additional hairpin can be placed in the 3' region of S1, dividing S1 in two helices. This pseudoknot is classified as an HL_{in}-type. The additional hairpin can also be inserted at the 5' end of S1 to split the stem. In such a pseudoknot, a separate structural element is not present downstream of the apical loop of the 3' hairpin, and the presence of two 'uninterrupted' hairpins produces an HH-type pseudoknot. Finally, if two adjacent hairpins make tertiary base-pairing through apical loops, they essentially form a third hairpin and the overall fold is classified an HHH-type pseudoknot.

FIGURE 1.

Pseudoknot schematics. (a) Linear representation of base-pairing (dashed lines) in the H-type RNA pseudoknot. The color code is used throughout all figures. Nucleotides are depicted by small open circles. (b) Two-dimensional representation of the H-type pseudoknot formation. The schematic highlights the base-pairing between an apical loop of a hairpin and an external region resulting in the formation of the pseudocontinuous S1-S2 helix. (c-e) The secondary structure of the H-type pseudoknot as a result of coaxial stacking when loop L2, L1 or L3 is eliminated, respectively.¹⁴ (f) A hypothetical secondary structure presentation depicting various types of pseudoknots formed by different types of loops. (Adapted with permission from [12]. Copyright 1990 Elsevier Science Publishers). (g) Circular representation of two adjacent hairpins. RNA is shown as a semi-circle with nucleotides in small circles and base-pairing depicted by lines. (Adapted with permission from [13]. Copyright 1992 American Chemical Society). (h) Circular representation of a pseudoknot. Note crossing of the chords. (i) Planar representation of the geometric types of pseudoknots that delineates connectivity of complex pseudoknotted structures. (Adapted with permission from [14]. Copyright 2003 Oxford University Press). H depicts a hairpin and L designates a bulge, interior or multiple loop.

FIGURE 2.

Pseudoknot in the phage T4 gp32 mRNA and the mechanism of gp32 autorepression. (a) gp32 (light orange ovals) binds single-stranded DNA as the replication fork advances. If gp32 is overproduced, one molecule of the protein binds to the pseudoknot located in the 5'-UTR of gp32 mRNA. Once the gp32-mRNA interaction has been nucleated, it is extended in a cooperative fashion towards the initiation codon, blocking the Shine-Dalgarno sequence and preventing translation of the mRNA. PK, pseudoknot. SD, Shine-Dalgarno sequence. ORF, open reading frame. Pol, DNA polymerase complex. (b) Secondary structure of the gp32 mRNA pseudoknot. (c) NMR structure of the gp32 mRNA pseudoknot (PDB code 2TPK).¹⁸ (d) Crystal structure of gp32 shown with electrostatic surface potential (PDB code 1GPC).¹⁹

FIGURE 3.

Pseudoknots in the autorepression of ribosomal proteins genes. (a), Secondary structure of the double pseudoknot specifically recognized by the *E. coli* ribosomal protein S4.²² (b) Secondary structure of the pseudoknot in the *E. coli rpsO* mRNA encoding the ribosomal protein S15. Ribosomal protein S15 (oval in wheat color) is positioned along S2 of the pseudoknot according to its location in the S15-mRNA complex. Translation initiation codon (AUG) is in light blue color. (c) An atomic model of the S15-pseudoknot complex derived from the crystal structure of the S15-rRNA complex and molecular modeling,²⁶ supported by biochemical^{24,26} and cryo-EM data (PDB code 2VAZ).³² (d) Schematic representation of the S15-mRNA interactions with the ribosome and the entrapment mechanism of repression.³² If S15 is overproduced in the cell, the protein binds and stabilizes the pseudoknot in its own mRNA. The S15-mRNA complex is then loaded onto the ribosome and makes SD/anti-SD interactions with the 16S rRNA. However, the ribosome cannot melt the pseudoknot structure so that the initiation codon, located in L3 of the pseudoknot in vicinity of S2, cannot reach the ribosome P site and interact with the initiator tRNA. As a consequence, the 3' end of the mRNA rests on the surface of the ribosome rather than inside the mRNA channel of the ribosome. The resulting 'entrapped' complex is stalled in the inactive pre-initiation state. In the absence of S15, the *rpsO* mRNA either unfolds on the ribosome or binds the ribosome in the unfolded conformation so that mRNA enters the mRNA channel in the single-stranded conformation (dashed line). In this

position, the initiation codon can be placed into the P site and interact with the initiator tRNA, thereby initiating translation of the mRNA.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 4.

H-type pseudoknot-based riboswitches. Arrows indicate rotation of zoomed-in structures relative to global views. Secondary structure (a), three-dimensional X-ray structure (b), A-minor groove triples (c), and fluoride binding (d), in the fluoride riboswitch structure (PDB code 4ENC).³⁴ Fluoride anion is shown by a red sphere. Magnesium cations (green spheres) are shown with coordination bonds (green sticks) and coordinated water molecules (light red spheres). Putative hydrogen bonds are depicted by black dashed lines. Secondary structure (e), three-dimensional NMR structure (f), A-amino kissing motif (g), and preQ₁ base-pairing

arrangement from the top view (h) in the preQ₁ riboswitch structure (PDB code 2L1V).³⁶ Secondary structure (i), three-dimensional X-ray structure (j), and preQ₁ recognition from the side (k) and top (l) views in the preQ₁-II riboswitch structure (PDB code 4JF2).⁴⁰ Secondary structure (m), three-dimensional X-ray structure (n), inclined A-minor motif (o), and SAM binding (p) in the SAM-II riboswitch structure (PDB code 2QWY).⁴² Electrostatic interactions between sulfur atom (yellow) and oxygen atoms of RNA are shown by yellow dashed lines.

FIGURE 5.

Infrequent pseudoknot arrangements. Secondary structure (a), three-dimensional X-ray structure (b), and SAH binding (c) in the SAH riboswitch structure (PDB code 3NPN).⁴⁵ Secondary structure (d), three-dimensional X-ray structure (e), major groove triples (f), and c-di-GMP binding (d) in the c-di-GMP riboswitch structure (PDB code 3Q3Z).⁴⁷

FIGURE 6.

Pseudoknots in the *glmS* riboswitch-ribozyme. (a) Secondary structure schematic of the RNA showing formation of hairpins and pseudoknots. Secondary structures, depicted from the crystal structure of the RNA,^{61,62} highlighting pseudoknot PK1 (b), PK2 (c), and PK3 (d). The three-dimensional structure of the RNA (PDB code 2GCV)⁶² highlighting pseudoknots PK1 (e), PK2 (f), and PK3 (g). (h) Zoomed-in view of PK3 showing interactions of L3 with the pseudoknot stem. (i) Formation of the active site by pseudoknots PK1 and PK2. The color code corresponds to PK1 except S1.2 from PK2 depicted in cyan.

(j) Zoomed-in view of the active site, showing recognition of GlcN6P (red) and position of the scissile phosphate (SP, gray sphere).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 7.

Summary of typical features in the regulatory pseudoknots. The pseudoknot schematic is consistent with the average length of the pseudoknot elements and preferred nucleotide composition on L1 and L3 loops. Nucleotides are depicted by circles; nucleotide preference is indicated within large circles. The length range is indicated in the schematic. Red oval shows the location of the ligand binding sites.

FIGURE 8.

A model of the SAM-II riboswitch folding and translation repression. Mg^{2+} cations stabilize the prefolded stem-loop bearing S1 (left panel) by triple interactions with L3 (top middle panel) and facilitate the formation of the transient pseudoknot-like conformations, capable of ligand binding (bottom middle panel). SAM captures such conformers and induces adaptive changes, resulting in the extension of initial pairing in S2 that sequesters the Shine-Dalgarno sequence (right panel). (Adapted with permission from [77]. Copyright 2011 Nature Publishing Group).

TABLE 1

Pseudoknot Classification Based on PseudoBase Classes

Name	Regulation	Type	PDB ID Code	Reference
Adenine	Riboswitch	HHH ^a	1Y26	[55]
Adenosylcobalamin	Riboswitch	HL _{out} ^a	4GXY, 4GMA	[57,58]
c-di-GMP-II	Riboswitch	H-like ^a	3Q3Z	[47]
Deoxyguanosine	Riboswitch	HHH ^a	3SKI	[56]
Fluoride	Riboswitch	HL _{out}	4ENC	[34]
glmS	Ribozyme-Riboswitch		2NZ4, 2GCV	[61,62]
PK1		LL		
PK2		HL _{out}		
PK3		HL _{out}		
gp32	Regulatory Protein Binding Site	H	2TPK, 1GPC	[18,19]
Guanine	Riboswitch	HHH ^a	1U8D, 1Y27	[54,55]
Hammerhead-like	Ribozyme	H, HL _{out} ^a	-	[65]
HDV-like	Ribozyme	LL, H ^a	-	[63,64]
L20	Ribosomal Protein Regulatory Site	HL _{out}	-	[23]
Lysine	Riboswitch	H ^a	3D0U, 3DIL	[52,53]
preQ ₁ -I	Riboswitch	H	2L1V	[36]
preQ ₁ -II	Riboswitch	HL _{out}	4JF2	[40]
S15	Ribosomal Protein Regulatory Site	HL _{out}	2VAZ	[32]
S4	Ribosomal Protein Regulatory Site	Unclassified	-	[22]
SAH	Riboswitch	LL	3NPN	[45]
SAM-I	Riboswitch	HL _{out} ^a	2GIS	[50]
SAM-II	Riboswitch	H	2QWY	[42]
SAM-IV	Riboswitch	HL _{out} ^a	-	[51]
SAM-V	Riboswitch	H	-	[43]
THF	Riboswitch	HH ^a	3SUX, 3SD1	[48,49]

^aClassification of the pseudoknot requires removal of adjacent elements.