ACCOUNTS
of chemical research

The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks

SAMUEL E. BUTCHER*,† AND ANNA MARIE PYLE*,‡,§

†*Department of Biochemistry, University of Wisconsin—Madison, 433 Babcock Drive, Madison, Wisconsin 53706-1544, United States,* ‡*Department of Molecular, Cellular and Developmental Biology and Department of Chemistry, Yale University, New Haven, Connecticut, United States, and* §*Howard Hughes Medical Institute*

RECEIVED ON MARCH 31, 2011

CONSPECTUS

RNA molecules adopt specific three-dimensional structures critical to their function. Many essential metabolic processes, including protein synthesis and RNA splicing, are carried out by RNA molecules with elaborate tertiary structures (e.g. 3QIQ, right). Indeed, the ribosome and self-splicing introns are complex RNA machines. But even the coding regions in messenger RNAs and viral RNAs are flanked by highly structured untranslated regions, which provide regulatory information necessary for gene expression.

RNA tertiary structure is defined as the three-dimensional arrangement of RNA building blocks, which include helical duplexes, triple-stranded structures, and other components that are held together through connections collectively termed RNA tertiary interactions. The structural diversity of these interactions is now a subject of intense investigation, involving the techniques of NMR, X-ray crystallography, chemical genetics, and phylogenetic analysis. At the same time, many investigators are using biophysical techniques to elucidate the driving forces for tertiary structure formation and the mechanisms for its stabilization. RNA tertiary folding is promoted by maximization of base stacking, much like the hydrophobic effect that drives protein folding. RNA folding also requires electrostatic stabilization, both through charge screening and site binding of metals, and it is enhanced by desolvation of the phosphate backbone. In this Account, we provide an overview of the features that specify and stabilize RNA tertiary structure.

A major determinant for overall tertiary RNA architecture is local conformation in secondary-structure junctions, which are regions from which two or more duplexes project. At junctions and other structures, such as pseudoknots and kissing loops, adjacent helices stack on one another, and these coaxial stacks play a major role in dictating the overall architectural form of an RNA molecule. In addition to RNA junction topology, a second determinant for RNA tertiary structure is the formation of sequence-specific interactions. Networks of triple helices, tetraloop–receptor interactions, and other sequence-specific contacts establish the framework for the overall tertiary fold. The third determinant of tertiary structure is the formation of stabilizing stacking and backbone interactions, and many are not sequence specific. For example, ribose zippers allow 2'-hydroxyl groups on different RNA strands to form networks of interdigitated hydrogen bonds, serving to seal strands together and thereby stabilize adjacent substructures. These motifs often require monovalent and divalent cations, which can interact diffusely or through chelation to specific RNA functional groups.

As we learn more about the components of RNA tertiary structure, we will be able to predict the structures of RNA molecules from their sequences, thereby obtaining key information about biological function. Understanding and predicting RNA structure is particularly important given the recent discovery that although most of our genome is transcribed into RNA molecules, few of them have a known function. The prevalence of RNA viruses and pathogens with RNA genomes makes RNA drug discovery an active area of research. Finally, knowledge of RNA structure will facilitate the engineering of supramolecular RNA structures, which can be used as nanomechanical components for new materials. But all of this promise depends on a better understanding of the RNA parts list, and how the pieces fit together.

Published on the Web 09/07/2011 www.pubs.acs.org/accounts
10.1021/ar200098t

1. Introduction

RNA molecules assemble into elaborate tertiary structures, forming globular shapes stabilized by networks of diverse interactions. Tertiary folded RNAs are recognized by proteins, ligands, and other RNA molecules, leading to biochemical events that impact every aspect of cellular metabolism. Understanding the molecular features of RNA tertiary structure is therefore a goal of central importance in biology.

RNA tertiary folding is challenged by many features inherent to the biopolymer: There are only four building blocks, and the polymer is negatively charged, so folding results in large electrostatic potentials that must be neutralized through interactions with ions. Unlike proteins, RNA secondary structural elements are inherently stable, and while this provides a rigid set of assembly subunits, it also can lead to "misfolding" as stable alternative structures can obstruct the assembly pathway.

However, numerous factors facilitate tertiary folding: Magnesium and other cations are abundant in the cell, serving to overcome electrostatic barriers and, in some cases, becoming integral structural components. Osmolytes dehydrate RNA, stimulating collapse.[1] Molecular crowding helps to stabilize folded RNAs.[2] Proteins often stabilize, remodel, and collaborate with the RNA biopolymer.

Here we provide an overview of the features that are important for specifying and stabilizing RNA tertiary structure. As novel structures are solved and new methods developed, we anticipate that the landscape of RNA tertiary interactions and folding strategies will rapidly expand.

2. Stacking and Coaxial Helices

Stacking of the aromatic nucleic acid bases is one of the most important driving forces during formation of RNA structure.[3] In general, an RNA molecule will attempt to maximize base stacking, particularly at helical termini. If two helices are next to each other (e.g., separated by a phosphodiester linkage), their terminal base pairs will stack, and the helices will become colinear, resulting in a "coaxially stacked" substructure (Figure 1).

The earliest glimpses of nucleic acid tertiary structure, in the form of tRNA crystal structures,[4] revealed that coaxial stacking of helices determines the overall molecular shape (Figure 1A). Each of the four helices in the "cloverleaf" tRNA secondary structure chooses a stacking partner, and these pairs of helices stack end-on-end, forming two long helices that ultimately arrange themselves through tertiary interactions.[4] Numerous subsequent studies have shown that the choice of stacking partners among sets of helices is a critical determinant of RNA structural fate (vide infra).

The coaxial stacking of helices at junctions is thermodynamically favorable,[3,5] and the free energy gained is sequence dependent, closely following the trends observed for nearest neighbor interactions in formation of RNA secondary structure.[5] From 0.5 to 3.0 kcal/mol of free energy is released upon formation of a coaxial stack, and these energetic differences can influence choice of coaxial stacking partners at a junction.[5] Tertiary structures such as kissing loops (Figure 2, vide infra) and pseudoknots (Figure 3, vide infra) are composed of coaxially stacked helices and therefore represent RNA structures that are almost completely dictated by the forces driving coaxial stacking.

At RNA junctions, duplexes immediately next to each other tend to coaxially stack, to minimize folding free energy.[6] However, determining which helices are actually

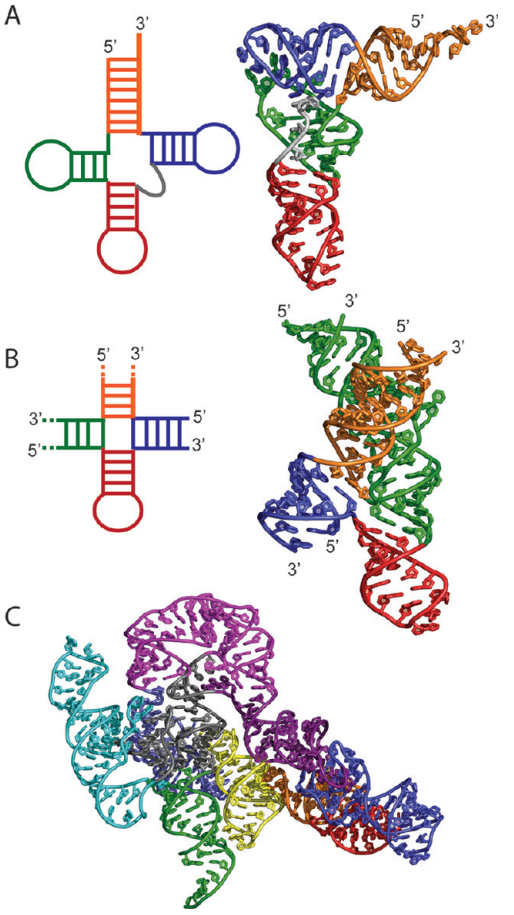

FIGURE 1. Multihelical junctions: (A) yeast phenylalanine tRNA, (left) secondary structure and (right) 3D-structure, PDB 6TNA; (B) the hairpin ribozyme, (left) secondary structure and (right) 3D structure, PDB ID 1M5K; coaxial stacking is observed between the gold and blue helices and between the green and red helices; (C) the RNA subunit of RNase P, PDB ID 3OK7.

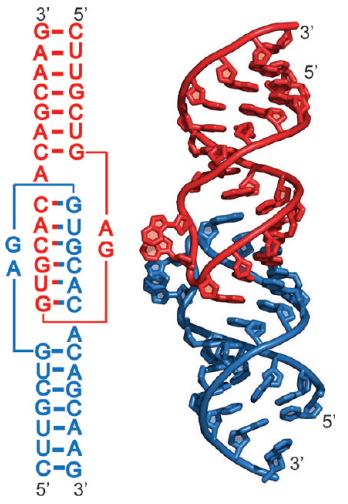

FIGURE 2. The HIV-1 dimerization initiation site kissing loop, PDB ID 1K9W. Hairpin loops are red and blue, respectively.

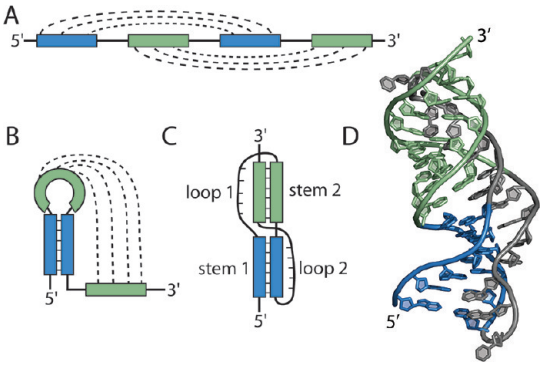

FIGURE 3. Pseudoknot topology and structure: (A) long-range base-pairing interactions; (B) hairpin secondary structure with long-range pseudoknot contacts; (C) coaxial stacking of pseudoknot helices; (D) the telomerase pseudoknot, PDB ID 2K96. Loops 1 and 2 (gray) form a series of base triples with stem 1 (blue) and stem 2 (green).

adjacent can be difficult because "linker" nucleotides are not always flexible. Rather, linker nucleotides often form non-cannonical base pairs that extend the helical terminus, thereby presenting an alternative interface with other helices. For example, A−G, C−C, and G−U pairs are common at helical termini, where they serve as energetically favorable interfaces for coaxial stacks.[7] Although the role of sequence is central, coaxial stacking choice can be strongly dictated by ionic conditions and by the topological constraints of intervening junctions[8,9]

3. Junction Topology and RNA Structure

Junction regions set the stage for specific RNA architectural forms. An emerging view of RNA tertiary structure is that junctions constrain the relative orientation of helices, setting up a limited number of possible conformational states, one of which (the native state) is locked into place through formation of stable tertiary interactions.[8,10,11] The landscape of junction topologies (and their resultant outcome for tertiary structure) are strongly influenced by ionic conditions.[12,13] The simplest examples to consider are two- and three-way junctions.

Computational and structural analysis of two-way junctions has shown that the relative spatial orientations and pitch of the two helices can be described by a set of three polar angles.[8] These angles vary over a narrow range, indicating that the flexibility at junctions is limited, thereby restricting the possible conformational states of a molecule. Three-way junctions in RNA are common, consisting of three helices connected by three "single-stranded" linker regions. Generally, the number of nucleotides in the linker regions are distributed asymmetrically; two of the helices will coaxially stack while the third is positioned at an angle relative to the other two.[8,14] There are three major topological families that differ according to the angle of the third helix.[14] Rules for which helices will form the coaxial stack have been derived from analysis of RNA crystal structures in the PDB[14] and are as follows: (1) If two helices are separated by a 0 nucleotide linker, while the third is separated by a longer linker, the two contiguous helices will form a coaxial stack.[14] (2) If the linker lengths are equal, the Watson−Crick pairs with the most favorable stacking free energy[5] will tend to determine which helices form the coaxial stack.

Four-way junctions are also common among RNA structures, as illustrated by the classic tRNA cloverleaf secondary structure, which folds into an L-shape due to long-range base-pairing (a kissing loop interaction) between the D-loop and TΨC loops (Figure 1A, green and blue helices).[4] The hairpin ribozyme (Figure 1B) is an example of a "perfect" four-way junction,[15] with no linker nucleotides between the four helices. Coaxial stacking and the angles between the helices are governed by tertiary contacts that form between internal loops within two of the helical arms (Figure 1B, gold and green helices).[9] In the absence of these tertiary contacts, the hairpin ribozyme junction fluctuates between different coaxial stacks, which isomerize through an open, unstacked intermediate.[9] Analysis of 62 RNA four-way junctions has elucidated nine families with different configurations.[10]

Multihelical junctions of 5−10 helices have been analyzed and grouped into families.[10] These higher order junctions display coaxial stacking similar to three- and four-way

junctions, from which they are constructed. Illustrative of a complex multihelical RNA is RNase P,[16] with 18 helices (Figure 1C). Interestingly, multihelical junction topologies are often governed by tertiary interactions, as in the so-called A-minor junctions.[17] These topologies are sufficiently understood that they can be used to design supramolecular self-assembling nanostructures.[17]

4. Long-Range Interactions Involving W–C Base Pairs

4.1. Kissing Loops. Long-range base pairings between hairpin stem–loops are known as "kissing" loops. These interactions are mediated by loop nucleotides that interact through complementary Watson–Crick base pairs. Kissing loops have been found in many RNAs, including tRNA, mRNA, and rRNA.[18] Stable loop–loop interactions can occur with as few as two base pairs between the loops.[19] Kissing loop interactions are commonly used by retroviruses to initiate dimerization of genomic RNA.[18] The HIV-1 dimerization initiation site (DIS) is a well-studied example of this interaction (Figure 2).[20] The self-complementary loop forms a stable interaction involving six base pairs that coaxially stack within a continuous helix that contains bulged purines (Figure 2).

4.2. Pseudoknots. Pseudoknots involve long-range Watson–Crick base pairing between the loop nucleotides of a hairpin and a complementary region of RNA. The long-range base pairs form a second helix, which coaxially stacks upon the first (Figure 3). A large number of pseudoknots with diverse structure and function have been identified in many RNAs.[21,22] The pseudoknot found in human telomerase RNA is a particularly interesting example of this type of tertiary structure[23] (Figure 3). The loop regions of this pseudoknot form base triples in the minor groove of stem 1 and major groove of stem 2 (Figure 3D), resulting in an extended triple helix.

5. Minor Groove Triples and A-Minor Interactions

5.1. A Universal Packing Strategy for RNA. Examples of specific minor-groove triple interactions were initially reported in structural analyses of the hammerhead ribozyme[24] and active sites for both the group I[25,26] and group II introns.[27] Extensive analysis of diverse functional RNAs revealed that these triples, in which the third nucleotide is usually (but not exclusively) an adenine, are a recurrent structural motif and a "universal packing mode for RNA".[28] Indeed, these interactions fall into different categories, depending on the extent of van der Waals surface contact and free energy of interaction.[28] Analysis of the large ribosomal subunit crystal structure revealed a preponderance of adenines that formed tertiary contacts with the minor grooves of distal helices.[29] The large subunit rRNAs (totaling over 3000 nucleotides) were observed to form a densely compact tertiary structure, with 186 adenines forming minor groove triple interactions similar to other examples,[24,27,28] and they were named "A-minor" tertiary interactions.[29] Analysis of the rRNA structure revealed that these "A-minor interactions" are even more common than long-range Watson–Crick base pairings. Therefore, coaxial stacking of helices and the A-minor motif are among the most important types of interaction in RNA tertiary structure.[28,29]

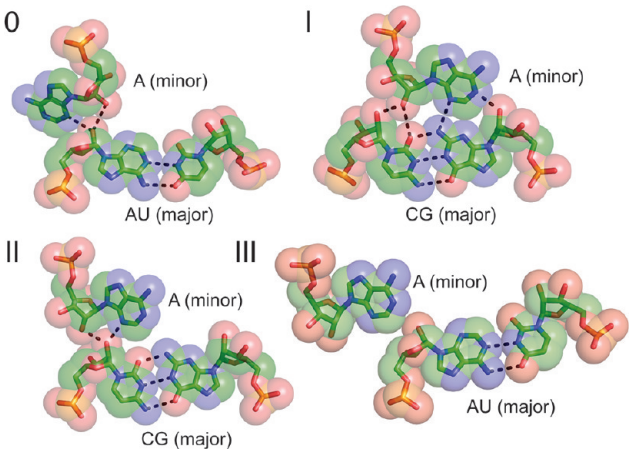

FIGURE 4. Examples of four A-minor interaction subtypes from the large ribosomal subunit, PDB ID 1FFK. Major and minor groove edges are indicated.

There are four variations of the A-minor motif (Figure 4). Each type is defined by the orientation of the 2′ OH group of the adenosine in the minor groove. Types I and II are specific for adenines.[28,29] Types 0 and III can be formed with nucleotides other than adenine, although adenine is still the preferred nucleotide for this type of interaction.[29] The structural basis for the utilization of adenine in these interactions is most evident in the type I form (Figure 4), where it can be seen that the minor groove edge of adenine is complementary in shape to the curve of the helical minor groove.[28] Guanine, on the other hand, contains a bulky amino group on its minor groove edge, which hinders its approach to the minor groove of an A-form duplex. Pyrimidines are too small to span the duplex minor groove, and ketone oxygens project from their minor groove edges. Adenine, on the other hand, has a small, neutral CH group on its minor groove edge (C2), which is ideal for forming the type I and II A-minor interactions (Figure 4). Often, A-minor

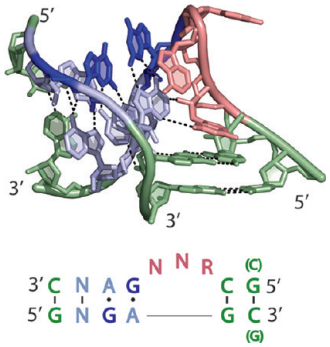

FIGURE 5. (top) The kink turn from the SAM-I riboswitch, PDB ID 2GIS. Guanines that form cross-strand stacks are shown in dark blue. (bottom) The kink-turn consensus sequence.

motifs are found in clusters or "patches" formed by 2−3 consecutive adenines.[29] The A-minor motif is an important building block for larger motifs. For example, the kink turn and tetraloop receptor are structural elements that contain consecutive A-minor interactions (*vide infra*).

6. Kink Turns and Related Motifs

6.1. Structural Features. Kink turns (or K-turns) were identified during analysis of the large ribosomal subunit crystal structure,[30] and they have been found in a diversity of RNAs.[31] The K-turn is a helix−loop−helix motif that bends the RNA helical axis by ∼120°, resulting in a close juxtaposition of helical minor grooves[30] (Figure 5). One helix is canonical, whereas the other is noncanonical with tandem G−A base pairs. The bend is initiated by a three-nucleotide loop that is typically purine-rich (Figure 5). The severe kink in the loop is facilitated by the G−A pairs, which form cross-strand stacking interactions that twist the backbone in a characteristic manner. The adenines of the tandem G−A pairs stack on each other and participate in A-minor interactions across the junction (Figure 5).

6.2. Stability and Plasticity. K-turns are often bound and stabilized by proteins.[31] In isolation, the K-turn is in equilibrium between the kinked form and a more extended conformation.[32] The kinked population is stabilized by high concentrations of metal ions, but it remains in equilibrium with an extended form.[32] Thus it is believed that K-turns do not provide a thermodynamic driving force for RNA tertiary folding, but rather require cooperation from surrounding proteins or RNA tertiary structure in order to stabilize them. Curiously, a helix−loop−helix motif in the *Azoarcus* group I intron has the hallmarks of a K-turn, yet bends in the opposite direction, toward the major groove instead of the minor.[33] This reverse K-turn is stabilized by a tetraloop-receptor interaction, which clamps the canonical helix in an opposite orientation.[33] This observation highlights the dynamic plasticity of the K-turn motif.

6.3. Hook Turns and Other Repetitive Turn Motifs in RNA Structure. The "hook turn" is a recurrent motif that was originally characterized during crystallographic investigations of a loop E motif from the 5S rRNA of a purple sulfur bacterium.[34] This RNA, which undergoes a sudden reverse in direction stabilized by a sheared G−A pair and a reverse-Hoogsteen pair, was found to be recurrent by scanning structures of the ribosome using the motif-finding algorithm Primos.[35] The same algorithm was also used to identify the constituent K-turns and S-turns within the ribosome, revealing that the S-turn occurs in two distinct forms.[35] A computational adaptation of this approach allowed the RNA structural database to be winnowed for new and recurrent elements of RNA structure,[36] leading to the first automated RNA motif discoveries.

7. Tetraloop-Receptor Motifs

Tetraloop-receptor motifs are among the most common types of long-range RNA tertiary interaction.[37] They have been observed in almost every large RNA crystal structure and have even been employed in the fabrication of RNA nanostructures.[38] In every case, this motif involves a terminal hairpin loop that contains a signature sequence, and known examples of interacting tetraloops are the GNRA[37,39] and GANC tetraloops.[40] While other conserved tetraloops are known (such as UNCG and CUYG),[41] these serve as stable caps for hairpin termini and do not typically function as tertiary interaction partners.

There are several types of receptor motifs for tetraloops, and they differ in their level of structural complexity. The most complex type of receptors are highly conserved internal loop motifs, such as the "11-nucleotide motif", which specifically recognizes GAAA tetraloops (Figure 6), and the IC3 motif, which has a more relaxed specificity for GNRA tetraloops.[37,39] The interaction between the GAAA tetraloop and the 11 nucleotide receptor is surprisingly stable.[42] Molecular interactions between GAAA tetraloops and the 11 nucleotide receptor are extensive and include two major components: (1) The second adenine of the GAAA tetraloop stacks on an A−A platform that is formed within the receptor loop. Note that in some instances, the A−A platform can be A−C.[37] The second adenine of the tetraloop also forms supporting hydrogen bonds to the receptor. (2) The remaining bases of the tetraloop engage in a network of hydrogen bonds with G−C base pairs in the adjacent receptor helix

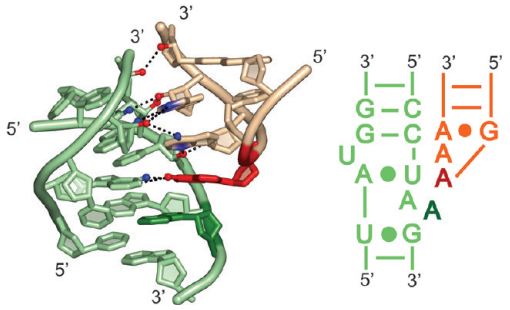

FIGURE 6. The tetraloop receptor. (left) Tetraloop receptor from the *Tetrehymena* group I intron P4–6 domain, PDB ID 1HR2. Receptor is green; the tetraloop is tan. Stacking occurs between an adenine platform (dark green) and the second adenine in the GAAA tetraloop (dark red). (right) Consensus sequence.

(Figure 6).[25] Intriguingly, other types of tetraloop-receptor interactions (below) utilize only one of these two interaction strategies.

A simpler class of receptors are tandem G–C pairs, which represent the first type of tetraloop-receptor interaction to be visualized crystallographically.[24] The interaction first appeared as a set of crystal contacts reported during structural studies of the hammerhead ribozyme.[24] Later dubbed "A-minor motifs" (vide supra), the G–C pairs within the minor groove of these receptors form hydrogen bonds with base and sugar substituents on tetraloop nucleotides 3 and 4.[24] Similar variations on this subfamily were reported within group I introns,[39] and they have subsequently been observed in many RNA crystal structures.

By contrast, the GANC tetraloops interact exclusively through base stacking with their cognate receptors, which are simple extrahelical bulged purines.[40] Occurring only in group IIC introns, these tetraloop-receptor interactions are highly conserved.

8. Interaction through Intercalation: T-Loops and Other Long-Range Motifs

Another way that GNRA-like loops can form long-range tertiary interactions is through intercalation within the tetraloop itself. The basic "U-turn" backbone architecture of GNRA loops (which is shared by tRNA anticodons and other RNAs[4,24]) can tolerate many variations in structural form, including inserted bases[43] and even the absence of a nucleobase.[44,45] In cases of intercalated U-turns (a subtype of "T-loop motifs"[44,45]), the fourth nucleotide of the loop is "missing", and stacked in its place is an intercalated base that is provided from elsewhere in the structure.[44,46] This allows two motifs to slide together, conjoining bases to merge distal regions of structure (Figure 7).

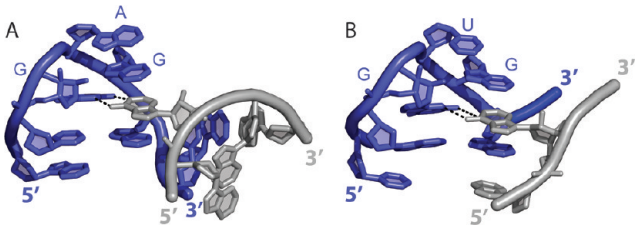

FIGURE 7. The T-loop motif: (A) T-loop motif from the group II intron, PDB ID 3IGI; (B) T-loop motif from the RNase P specificity domain, PDB ID 1NBS. Loop nucleotides (dark blue) are labeled.

9. Triple-Stranded RNA Structures

Base triples occur frequently in RNA tertiary structure. For example, there are 27 base triples in the 50S ribosomal subunit, and 10 triples in the 30S subunit.[47] The number of possible base triples constrained by at least three hydrogen bonds is 840.[47] In the *Tetrahymena* intron, the active site contains a sandwich of four base triples.[48]

Triple-stranded RNA structures contain a Watson–Crick base-paired duplex that hydrogen bonds to a third strand. While some triples occur in the minor groove (A-minor motifs, vide supra), base triples can also occur within the narrow major groove of RNA. Base triples that involve a third strand in the major groove often utilize the "Hoogsteen" face of purines. One example of such an interaction is observed in the telomerase pseudoknot,[49] which has a loop that lies in the helical major groove and forms a series of Hoogsteen base triples (Figure 3). Similarly, a series of stacked major groove triples comprises the conserved "catalytic triplex" within the group II intron active site, where it plays a key role in supporting reaction chemistry.[50] A similar triplex network is hypothesized to exist within the active site of the eukaryotic spliceosome.[46]

Hoogsteen triples are most commonly U–A·U, where the adenine N7 accepts a hydrogen bond from the U imino proton (Figure 8A). The base triple C–G·C^+ is isosteric with U–A·U but requires protonation of the third strand C in order to form a hydrogen bond with the guanine N7.

The SAM-II riboswitch bound to *S*-adenosylmethionine contains a pseudoknot with an extended major-groove triple helix.[51] Another example was recently identified within the Kaposi's sarcoma-associated herpesvirus, which produces a highly abundant noncoding RNA, the polyadenylated nuclear (PAN) RNA. PAN contains an expression and nuclear retention element (ENE) that prevents degradation of its message.[52] The ENE contains an internal loop is made up of uracils, which clamp onto the poly-A tail of the message, thereby forming a triple stranded U–A·U structure[52] (Figure 8).

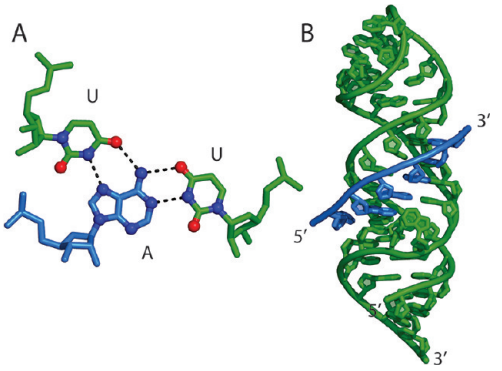

FIGURE 8. (A) Structure of an A−U·A triple base pair and (B) structure of the Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear (PAN) RNA expression and nuclear retention element (ENE), showing a triple-stranded interaction between the ENE RNA (green) and poly-A (blue). From PDB ID 3P22.

10. RNA Quadruplex Structures

Repeated stretches of guanine-rich sequences can form quadruple base pairs that stack on each other to form quadruplexes. Telomeric DNA contains such guanine repeats, and it has recently been discovered that telomeres are transcribed into a large noncoding RNA called TERRA (telomeric repeat-containing RNA).[53] TERRA RNA has been implicated in regulation of telomere length, through telomerase inhibition[54] and chromatin remodeling.[53] The crystal structure of a quadruplex derived from human TERRA has been solved[55] (Figure 9). Guanine quadruplexes are stabilized by potassium ions, which are chelated in the center of the quadruplex by the guanine oxygens (Figure 9). Electrospray mass spectrometry reveals that TERRA quadruplexes form stable multimers.[56] An emerging body of evidence suggests that RNA quadruplex structures play an important role in regulating translation of mRNA and splicing of some pre-mRNAs.[57] Such RNA quadruplexes are recognized by the fragile X mental retardation protein (FMRP), which acts as a repressor of translation.[57]

11. Ribose and the 2′-Hydroxyl Group: Key Components of RNA Tertiary Structure

The 2′-hydroxyl group of the RNA backbone is a stabilizing component in many tertiary interactions, since it can make two hydrogen bonds, acting as both hydrogen-bond donor and acceptor. The prevalence of 2′-OH contacts was evident from crystal structures of tRNA,[4] and their thermodynamic contribution to RNA tertiary structure was revealed through studies of the *Tetrahymena* ribozyme.[58,59] Many subsequent biochemical and structural studies revealed the importance of 2′-hydroxyl groups in ribozymes and other large RNA

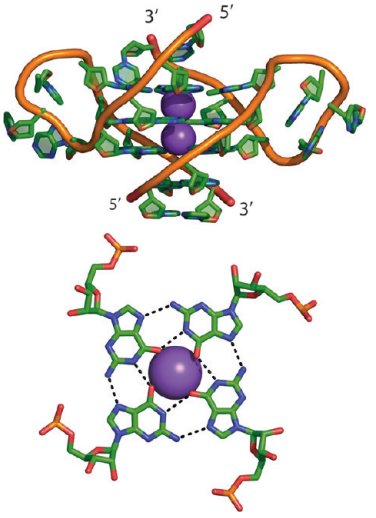

FIGURE 9. Structure of the human telomeric RNA (TERRA) quadruplex, PDB ID 3IBK: (top) side view of the TERRA quadruplex; (bottom) view of a central guanine quartet. Potassium ions are purple.

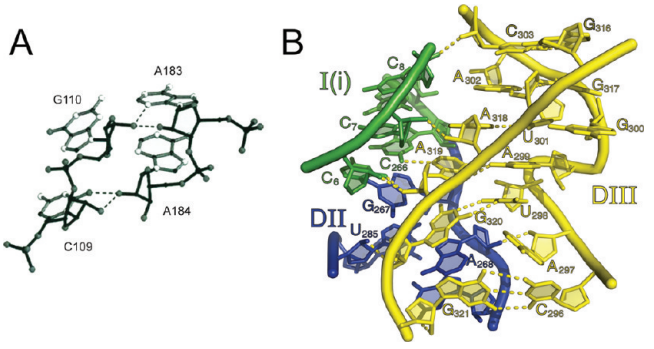

FIGURE 10. Ribose zippers in RNA structure: (A) a ribose zipper from P456;[25] (B) a network of ribose zippers joining domains 1, 2, and 3 in group II intron, PDB ID 3IGI.[46]

molecules, where they contribute to many motifs, including the A-minor interaction (see section 5). Structural studies show that 2′-OH groups usually form networks of interactions, forming arrays with different types of morphologies.

The most common type of 2′-OH array is the ribose zipper motif,[60] which brings two backbone strands into close proximity through interdigitated 2′-OH interactions (Figure 10a), thereby stabilizing neighboring structural features. Ribose zippers were first reported in early studies of ribozyme structure, where they play a key role in supporting core architecture.[25] Inspection of solved structures shows that groups of ribose zippers tend to surround and buttress more sequence-specific types of tertiary interactions, such as tetraloop-receptor interactions, kissing loops, and S-turns (Figure 10b).[46] In this way, ribose zippers provide the "glue" that maintains stability of entire tertiary substructures. A single 2′-hydroxyl group within a tertiary interaction

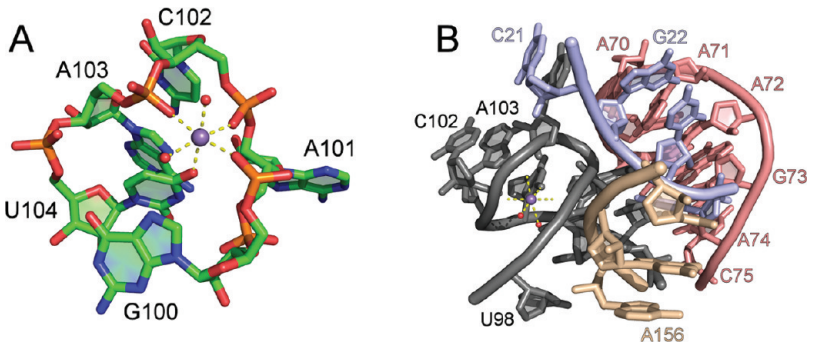

FIGURE 11. A chelated inner-sphere metal ion binding site within the M-box riboswitch RNA: (A) A detailed view of the metal ligation environment. The divalent ion bound to phosphoryl oxygens at nucleotides 100, 102, and 103 and O4 of U104. (B) The structural environment of the metal site shown in panel A. From PDB ID 3PDR.

network may contribute 1−2 kcal/mol of interaction free energy,[43,59] so the stabilizing influence of a ribose zipper is likely to be considerable. While most ribose zipper motifs would seem to be sequence independent (all nucleotides contain a ribose), some types of ribose zipper motifs are found in specific sequence contexts,[40,60] and their formation may be directed by the surrounding structural environment.

12. Tertiary Structural Elements Involving Metal Ions

It is generally believed that physiological concentrations of monovalent ions (~150 mM) can promote secondary structure formation but that magnesium (or molar amounts of monovalent) ion is required for stabilization of RNA tertiary structure. However, there are many different roles for monovalent and divalent ions in RNA structure.[61] For example, while monovalent ions are often engaged in charge screening, they can be chelated and tightly bound within RNA tertiary structures.[62−64] Conversely, divalent cations are not always tightly site-bound, as suggested by crystal structures in which magnesium ion binding sites are evident. Draper and colleagues have shown that many of these magnesium ions are "diffusely bound", exchanging readily and remaining hydrated.[61] These ions play an important role in stabilizing the RNA, but they can often be substituted with exchange-inert complexes such as cobalt hexammine. In some cases, however, site-bound magnesium ions interact through direct coordination with RNA functional groups, becoming dehydrated and integrating themselves into RNA structure.[61] For example, of the 11 crystallographically observed magnesium ions in a 58-mer rRNA fragment that has been studied thermodynamically, only two of these metals are "chelated" or directly site-bound to the RNA.[61] These are located at sites of such negative electrostatic potential that the energetic barriers to phosphate dehydration are likely to be overcome.[61]

Architectural elements that specifically bind chelated metal ions are found within the active sites and folding domains of many ribozymes.[46,65] These have been observed crystallographically and studied biochemically, but no single architectural theme has emerged.

One of the most interesting structural sites for chelated metal ions was identified within the M-box RNA, which is a magnesium riboswitch that sensitively binds and detects divalent cations.[66] This RNA has been studied biochemically and structurally, resulting in the highest resolution structure ever obtained for a large RNA.[66] Given the extraordinary resolution of this structure, it is possible to learn much about the interactions of monovalent and divalent cations within RNAs (Figure 11). As predicted by Draper,[61] only a few of the crystallographically determined metals interact through extensive inner-sphere contacts between the RNA and the ion.[66] However, several are clearly chelated through multiple direct interactions. One of the most interesting sites involves direct coordination by four inner-sphere ligands, including three phosphoryl oxygens and the O4 ketone of uracil (Figure 11a). Like the metal ion binding regions within the group II intron active site[50] or the P4−6 domain of the *Tetrahymena* group I intron,[67] this chelating site is created through a very tight turn in the RNA backbone,[66] resulting in close proximity of multiple backbone strands (sample interphosphate distances are 5.3 and 5.4 Å), which creates a region of extremely negative electrostatic potential (Figure 11b).

13. Noncanonical Base Pairs and Their Role in Tertiary Structure

Tertiary structures are often stabilized by noncanonical (non-Watson−Crick) nucleobase interactions. There are

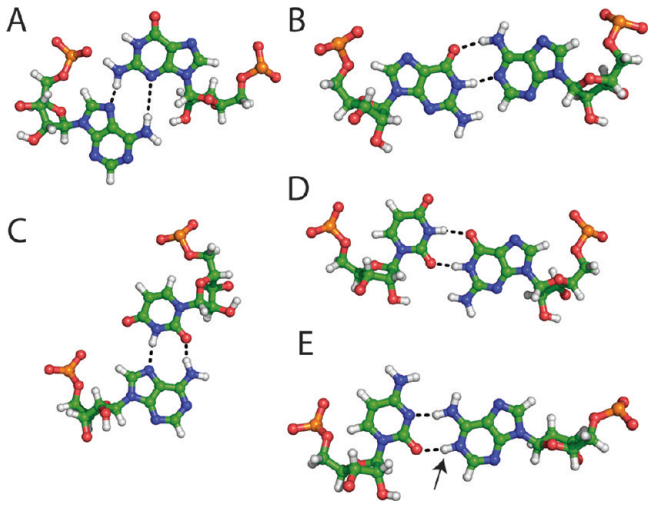

FIGURE 12. The five most common noncanonical base pairs in RNA: (A) sheared GA; (B) GA imino; (C) AU reverse Hoogsteen; (D) GU wobble; (E) AC wobble. The adenine is protonated at N1, indicated with an arrow. Structures A–D are from PDB ID 1GID. Structure E is from PDB ID 1SYZ.

29 possible base pairs that share at least two hydrogen bonds.[68] Of these, there are 12 basic geometric families.[69,70] The relationships of these base pair families have been described in terms of "isostericity matrices", which reflect the degree of structural similarity between noncanonical base pairs.[69] The most commonly occurring noncanonical pairings are the sheared GA, GA imino, AU reverse Hoogsteen, and GU and AC wobble pairs[71] (Figure 12). Note that the AC wobble pair requires protonation of the adenine N1 nitrogen in order to hydrogen bond to the cytosine O2 oxygen (Figure 12E). Although the adenosine mononucleotide has a pK_a of 3.5, in certain structural contexts this pK_a can be elevated by several orders of magnitude toward neutrality.[72,73] The cytosine N3 nitrogen has an intrinsic pK_a of 4.2, and this pK_a can also be perturbed toward neutrality.[74] The presence of these ionizable groups diversifies RNA function and allows ribozymes to perform general acid–base catalysis.[74–76]

14. Conclusion

The fact that we can now describe many recurring motifs in RNA tertiary structure may suggest that we have acquired a complete understanding of RNA's structural repertoire. However, this is certainly not the case. While there are over 69 000 PDB depositions containing proteins, there are only 2021 depositions containing RNA. Yet the macromolecular diversity of RNA in Biology is startling: approximately 80% of the human genome is transcribed into RNA, whereas only 2% of the genome is translated into protein (Encode project). New RNA substructures will be revealed for many years, and we will witness the many ways they can be combined to yield new functions.

We thank Max Bailor, Kevin Keating, and David Draper for helpful discussions. A.M.P. is an Investigator of the Howard Hughes Medical Institute.

BIOGRAPHICAL INFORMATION

Samuel E. Butcher is Professor of Biochemistry at the University of Wisconsin—Madison.

Anna Marie Pyle is Professor of Molecular, Cellular and Developmental Biology and Professor of Chemistry at Yale University.

FOOTNOTES

*E-mail addresses: sebutcher@wisc.edu, anna.pyle@yale.edu.

REFERENCES

1 Lambert, D.; Leipply, D.; Draper, D. E. The osmolyte TMAO stabilizes native RNA tertiary structures in the absence of Mg2+. *J. Mol. Biol.* **2010**, *404*, 138–57.
2 Kilburn, D.; Roh, J. H.; Guo, L.; Briber, R. M.; Woodson, S. A. Molecular crowding stabilizes folded RNA structure by the excluded volume effect. *J. Am. Chem. Soc.* **2010**, *132*, 8690–6.
3 Walter, A. E.; Turner, D. H.; Kim, J.; Lyttle, M. H.; Muller, P.; Mathews, D. H.; Zuker, M. Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9218–22.
4 Quigley, G. J.; Rich, A. Structural domains of transfer RNA molecules. *Science* **1976**, *194*, 796–806.
5 Walter, A. E.; Turner, D. H. Sequence dependence of stability for coaxial stacking of RNA helixes with Watson-Crick base paired interfaces. *Biochemistry* **1994**, *33*, 12715–9.
6 Tyagi, R.; Mathews, D. H. Predicting helical coaxial stacking in RNA multibranch loops. *RNA* **2007**, *13*, 939–51.
7 Kim, J.; Walter, A. E.; Turner, D. H. Thermodynamics of coaxially stacked helixes with GA and CC mismatches. *Biochemistry* **1996**, *35*, 13753–61.
8 Bailor, M. H.; Sun, X.; Al-Hashimi, H. M. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **2010**, *327*, 202–6.
9 Hohng, S.; Wilson, T. J.; Tan, E.; Clegg, R. M.; Lilley, D. M.; Ha, T. Conformational flexibility of four-way junctions in RNA. *J. Mol. Biol.* **2004**, *336*, 69–79.
10 Laing, C.; Schlick, T. Analysis of four-way junctions in RNA structures. *J. Mol. Biol.* **2009**, *390*, 547–59.
11 Leontis, N. B.; Lescoute, A.; Westhof, E. The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **2006**, *16*, 279–87.
12 Lafontaine, D. A.; Norman, D. G.; Lilley, D. M. Structure, folding and activity of the VS ribozyme. *EMBO J.* **2001**, *20*, 1415–24.
13 Al-Hashimi, H. M.; Pitt, S. W.; Majumdar, A.; Xu, W.; Patel, D. J. Mg2+-induced variations in the conformation and dynamics of HIV-1 TAR RNA. *J. Mol. Biol.* **2003**, *329*, 867–73.
14 Lescoute, A.; Westhof, E. Topology of three-way junctions in folded RNAs. *RNA* **2006**, *12*, 83–93.
15 Rupert, P. B.; Ferre-D'Amare, A. R. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **2001**, *410*, 780–6.
16 Reiter, N. J.; Osterman, A.; Torres-Larios, A.; Swinger, K. K.; Pan, T.; Mondragon, A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **2010**, *468*, 784–9.
17 Geary, C.; Chworos, A.; Jaeger, L. Promoting RNA helical stacking via A-minor junctions. *Nucleic Acids Res.* **2011**, *39*, 1066–80.
18 Brunel, C.; Marquet, R.; Romby, P.; Ehresmann, C. RNA loop-loop interactions as dynamic functional motifs. *Biochimie* **2002**, *84*, 925–44.
19 Li, P. T.; Bustamante, C.; Tinoco, I., Jr. Unusual mechanical stability of a minimal RNA kissing complex. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15847–52.
20 Ennifar, E.; Walter, P.; Ehresmann, B.; Ehresmann, C.; Dumas, P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.* **2001**, *8*, 1064–8.
21 Giedroc, D. P.; Cornish, P. V. Frameshifting RNA pseudoknots: Structure and mechanism. *Virus Res.* **2009**, *139*, 193–208.

22 Staple, D. W.; Butcher, S. E. Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* **2005**, *3*, No. e213.

23 Theimer, C. A.; Jady, B. E.; Chim, N.; Richard, P.; Breece, K. E.; Kiss, T.; Feigon, J. Structural and functional characterization of human telomerase RNA processing and cajal body localization signals. *Mol. Cell* **2007**, *27*, 869–81.

24 Pley, H. W.; Flaherty, K. M.; McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **1994**, *372*, 68–74.

25 Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **1996**, *273*, 1678–85.

26 Szewczak, A. A.; Ortoleva-Donnelly, L.; Ryder, S. P.; Moncoeur, E.; Strobel, S. A. A minor groove RNA triple helix within the catalytic core of a group I intron. *Nat. Struct. Biol.* **1998**, *5*, 1037–42.

27 Boudvillain, M.; de Lencastre, A.; Pyle, A. M. A tertiary interaction that links active-site domains to the 5′ splice site of a group II intron. *Nature* **2000**, *406*, 315–8.

28 Doherty, E. A.; Batey, R. T.; Masquida, B.; Doudna, J. A. A universal mode of helix packing in RNA. *Nat. Struct. Biol.* **2001**, *8*, 339–43.

29 Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4899–903.

30 Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. The kink-turn. *EMBO J.* **2001**, *20*, 4214–21.

31 Schroeder, K. T.; McPhee, S. A.; Ouellet, J.; Lilley, D. M. A structural database for k-turn motifs in RNA. *RNA* **2010**, *16*, 1463–8.

32 Goody, T. A.; Melcher, S. E.; Norman, D. G.; Lilley, D. M. The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA* **2004**, *10*, 254–64.

33 Antonioli, A. H.; Cochrane, J. C.; Lipchock, S. V.; Strobel, S. A. Plasticity of the RNA kink turn structural motif. *RNA* **2010**, *16*, 762–8.

34 Szep, S.; Wang, J.; Moore, P. B. The crystal structure of a 26-nucleotide RNA containing a hook-turn. *RNA* **2003**, *9*, 44–51.

35 Duarte, C. M.; Wadley, L. M.; Pyle, A. M. RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucleic Acids Res.* **2003**, *31*, 4755–61.

36 Wadley, L. M.; Pyle, A. M. The identification of novel RNA structural motifs using COMPADRES. *Nucleic Acids Res.* **2004**, *32*, 6650–9.

37 Costa, M.; Michel, F. Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection. *EMBO J.* **1997**, *16*, 3289–302.

38 Geary, C.; Baudrey, S.; Jaeger, L. Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res.* **2008**, *36*, 1138–52.

39 Ikawa, Y.; Naito, D.; Aono, N.; Shiraishi, H.; Inoue, T. A conserved motif in group IC3 introns is a new class of GNRA receptor. *Nucleic Acids Res.* **1999**, *27*, 1859–65.

40 Keating, K. S.; Toor, N.; Pyle, A. M. The GANC tetraloop: A novel motif in the group IIC intron structure. *J. Mol. Biol.* **2008**, *383*, 475–81.

41 Woese, C. R.; Winker, S.; Gutell, R. R. Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8467–71.

42 Vander Meulen, K. A.; Davis, J. H.; Foster, T. R.; Record, M. T.; Butcher, S. E. Thermodynamics and folding pathway of tetraloop receptor-mediated RNA helical packing. *J. Mol. Biol.* **2008**, *384*, 702–17.

43 Abramovitz, D. L.; Pyle, A. M. Remarkable morphological variability of a common RNA folding motif. *J. Mol. Biol.* **1997**, *266*, 493–506.

44 Krasilnikov, A. S.; Mondragon, A. On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA* **2003**, *9*, 640–3.

45 Nagaswamy, U.; Fox, G. E. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA* **2002**, *8*, 1112–9.

46 Keating, K. S.; Toor, N.; Perlman, P. S.; Pyle, A. M. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* **2010**, *16*, 1–9.

47 Walberer, B. J.; Cheng, A. C.; Frankel, A. D. Structural diversity and isomorphism of hydrogen-bonded base interactions in nucleic acids. *J. Mol. Biol.* **2003**, *327*, 767–80.

48 Guo, F.; Gooding, A. R.; Cech, T. R. Structure of the Tetrahymena ribozyme. *Mol. Cell* **2004**, *16*, 351–62.

49 Theimer, C. A.; Blois, C. A.; Feigon, J. Structure of the human telomerase RNA pseudoknot. *Mol. Cell* **2005**, *17*, 671–82.

50 Toor, N.; Keating, K. S.; Taylor, S. D.; Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **2008**, *320*, 77–82.

51 Gilbert, S. D.; Rambo, R. P.; Van Tyne, D.; Batey, R. T. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat. Struct. Mol. Biol.* **2008**, *15*, 177–82.

52 Mitton-Fry, R. M.; DeGregorio, S. J.; Wang, J.; Steitz, T. A.; Steitz, J. A. Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* **2010**, *330*, 1244–7.

53 Luke, B.; Lingner, J. TERRA: Telomeric repeat-containing RNA. *EMBO J.* **2009**, *28*, 2503–10.

54 Redon, S.; Reichenbach, P.; Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* **2010**, *38*, 5797–806.

55 Collie, G. W.; Haider, S. M.; Neidle, S.; Parkinson, G. N. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.* **2010**, *38*, 5569–80.

56 Collie, G. W.; Parkinson, G. N.; Neidle, S.; Rosu, F.; De Pauw, E.; Gabelica, V. Electrospray mass spectrometry of telomeric RNA (TERRA). *J. Am. Chem. Soc.* **2010**, *132*, 9328–34.

57 Melko, M.; Bardoni, B. The role of G-quadruplex in RNA metabolism. *Biochimie* **2010**, *92*, 919–26.

58 Pyle, A. M.; Murphy, F. L.; Cech, T. R. RNA substrate binding site in the catalytic core of the Tetrahymena ribozyme. *Nature* **1992**, *358*, 123–8.

59 Bevilacqua, P. C.; Turner, D. H. Comparison of binding of mixed ribose−deoxyribose analogues of CUCU to a ribozyme. *Biochemistry* **1991**, *30*, 10632–40.

60 Tamura, M.; Holbrook, S. R. Sequence and structural conservation in RNA ribose zippers. *J. Mol. Biol.* **2002**, *320*, 455–74.

61 Draper, D. E. A guide to ions and RNA structure. *RNA* **2004**, *10*, 335–43.

62 Conn, G. L.; Gittis, A. G.; Lattman, E. E.; Misra, V. K.; Draper, D. E. A compact RNA tertiary structure contains a buried backbone-K+ complex. *J. Mol. Biol.* **2002**, *318*, 963–73.

63 Lambert, D.; Leipply, D.; Shiman, R.; Draper, D. E. The influence of monovalent cation size on the stability of RNA tertiary structures. *J. Mol. Biol.* **2009**, *390*, 791–804.

64 Basu, S.; Rambo, R. P.; Strauss-Soukup, J.; Cate, J. H.; Ferre-D'Amare, A. R.; Strobel, S. A.; Doudna, J. A. A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat. Struct. Biol.* **1998**, *5*, 986–92.

65 Stahley, M. R.; Adams, P. L.; Wang, J.; Strobel, S. A. Structural metals in the group I intron: a ribozyme with a multiple metal ion core. *J. Mol. Biol.* **2007**, *372*, 89–102.

66 Ramesh, A.; Wakeman, C. A.; Winkler, W. C. Insights into metalloregulation by M-box riboswitch RNAs via structural analysis. *J. Mol. Biol.* **2011**, *407*, 556–70.

67 Cate, J. H.; Hanna, R. L.; Doudna, J. A. A magnesium ion core at the heart of a ribozyme domain. *Nat. Struct. Biol.* **1997**, *4*, 553–8.

68 Burkard, M. E.; Turner, D. H.; Tinoco, I., Jr. In *The RNA World*, 2nd ed.; Cech, T. R., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1999; pp 675−80.

69 Leontis, N. B.; Stombaugh, J.; Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **2002**, *30*, 3497–531.

70 Leontis, N. B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7*, 499–512.

71 Nagaswamy, U.; Larios-Sanz, M.; Hury, J.; Collins, S.; Zhang, Z.; Zhao, Q.; Fox, G. E. NCIR: A database of non-canonical interactions in known RNA structures. *Nucleic Acids Res.* **2002**, *30*, 395–7.

72 Legault, P.; Farmer, B. T.; Mueller, L.; Pardi, A. Through-bond correlation of adenine protons in a ^{13}C-labeled ribozyme. *J. Am. Chem. Soc.* **1994**, *116*, 2203–4.

73 Huppler, A.; Nikstad, L. J.; Allmann, A. M.; Brow, D. A.; Butcher, S. E. Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure. *Nat. Struct. Biol.* **2002**, *9*, 431–5.

74 Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **2000**, *287*, 1493–7.

75 Wilson, T. J.; Li, N. S.; Lu, J.; Frederiksen, J. K.; Piccirilli, J. A.; Lilley, D. M. Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11751–6.

76 Han, J.; Burke, J. M. Model for general acid−base catalysis by the hammerhead ribozyme. *Biochemistry* **2005**, *44*, 7864–70.