Published in final edited form as:

Wiley Interdiscip Rev RNA. 2014 January ; 5(1): . doi:10.1002/wrna.1196.

Recognition Modes of RNA Tetraloops And Tetraloop-Like Motifs By RNA Binding Proteins

Roopa Thapar^{1,2,3,*}, Andria P. Denmon³, and Edward P. Nikonowicz³

¹Hauptman-Woodward Medical Research Institute, 700 Ellicott Street, Buffalo, NY 14203, USA

²Department of Structural Biology, SUNY at Buffalo, 700 Ellicott Street, Buffalo, NY 14203, USA

³Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77251-1892, USA

Abstract

RNA hairpins are the most commonly occurring secondary structural elements in RNAs and serve as nucleation sites for RNA folding, RNA-RNA, and RNA-protein interactions. RNA hairpins are frequently capped by tetraloops, and based on sequence similarity, three broad classes of RNA tetraloops have been defined: GNRA, UNCG, and CUYG. Other classes such as the UYUN tetraloop in histone mRNAs, the UGAA in 16S rRNA, the AUUA tetraloop from the MS2 bacteriophage, and the AGNN tetraloop that binds RNase III have also been characterized. The tetraloop structure is compact and is usually characterized by a paired interaction between the first and fourth nucleotides. The two unpaired nucleotides in the loop are usually involved in base-stacking or base-phosphate hydrogen bonding interactions. Several structures of RNA tetraloops, free and complexed to other RNAs or proteins are now available and these studies have increased our understanding of the diverse mechanisms by which this motif is recognized. RNA tetraloops can mediate RNA-RNA contacts via the tetraloop-receptor motif, kissing hairpin loops, A-minor interactions, and pseudoknots. While these RNA-RNA interactions are fairly well-understood, how RNA binding proteins recognize RNA tetraloops and tetraloop-like motifs remains unclear. In this review, we summarize the structures of RNA tetraloop-protein complexes and the general themes that have emerged on sequence and structure-specific recognition of RNA tetraloops. We highlight how proteins achieve molecular recognition of this nucleic acid motif, the structural adaptations observed in the tetraloop to accommodate the protein binding partner, and the role of dynamics in recognition.

Overview

The RNA stem-loop or hairpin is the most common secondary structure motif in RNA¹. It was first described in 1983 by C.R. Woese, H.F. Noller and coworkers to exist in eubacterial 16S-like ribosomal RNA (rRNA)². The loop motif that caps the A-form RNA helix is structurally important to initiate RNA folding³ as it allows the phosphodiester backbone of single-stranded RNA to fold back on itself, thereby facilitating formation of an intramolecular double-stranded RNA helix. Approximately 55% of RNA helices in *E. coli* 16S rRNA and 38% of RNA helices in 23S rRNA are capped by tetraloops^{1, 2, 4}. Several families of RNA tetraloops have been classified based on phylogenetic analysis, the geometry of the backbone, and the interactions between the loop bases. These are the GNRA^{1, 2}, UNCG⁵, CUYG¹, GANC⁶, (A/U)GNN^{7, 8}, and UUUM^{9, 10} tetraloops (where G is guanine, A is

*To whom correspondence should be addressed: Tel: 713-201-7875; rthapar@rice.edu.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing financial interest.

adenine, U is uracil, N is any base, R is a purine, Y is a pyrimidine, and M is either adenine or cytosine). The GNRA and UNCG tetraloops account for >70% of all tetraloops found in rRNA².

RNA tetraloops generally form compact and stable structures. The thermodynamic stability of RNA hairpins depends on the size and sequence of loop nucleotides. In general, tetraloops are more stable compared to smaller or larger loops containing the same stem since they are able to effectively minimize unfavorable base-solvent interactions via base stacking, base-phosphate, and base-ribose hydrogen bonds¹¹. Among all tetraloops, the UNCG, GNRA, and CUUG family of hairpins are the most stable. The stability of these classes is attributed to base pairing between the nucleotides at the 1 and 4 positions, base stacking, and 2' OH-base hydrogen bonds in the loop¹¹.

Hairpin loops participate in tertiary interactions in large RNAs such as ribosomal RNAs¹²⁻¹⁴ and self-splicing RNAs¹⁵ that allow these molecules to fold and function. A well-characterized example is the P4-P6 domain of the Group I intron¹⁵. The crystal structure of this motif shows a network of stabilizing base stacking and 2' OH hydrogen bonding interactions between a GAAA tetraloop and the receptor nucleotides located in a distal helix. The RNA kissing-loop motif is characterized by Watson-Crick (W-C) base pairing between loop nucleotides of two different hairpin loops, as observed in HIV-1 genomic RNA¹⁶. RNA pseudoknots, formed by W-C base pairing of the loop nucleotides with single-stranded RNA adjacent to the loop are diverse and found in ribozymes¹⁷, telomerase RNA¹⁸, and self-splicing introns¹⁹. While these RNA-RNA interactions involving RNA tetraloops are fairly well understood, little is known about the mechanisms by which RNA tetraloops recognize proteins.

RNA tetraloops also function as recognition sites for proteins in ribonucleoprotein complexes. The widely-held view is that solvent exposure of loop nucleotides could play a role in sequence-specific molecular recognition particularly via “induced-fit” mechanisms that are associated with RNA-protein complexes²⁰⁻²². Solution nuclear magnetic resonance (NMR) structures have been reported for the retroviral Rous sarcoma virus nucleocapsid protein bound to a UGCG tetraloop²³, bacteriophage N-peptide-*boxB* GNRA-like tetraloop complexes²⁴⁻²⁶, the Staufen dsRBD-UUCG tetraloop complex²⁷, and the Rnt1p RNase III dsRBD bound to AGNN cognate and non-cognate sequences^{28, 29}. X-ray crystal structures are available for the bacteriophage MS2 coat protein bound to hairpin tetraloops³⁰⁻³³, the restrictocin endonuclease bound to the GNRA sarcin/ricin loop (SRL) of 23S/28S rRNA³⁴, and the SLBP RNA binding domain/ERI-1/histone mRNA stem-loop ternary complex³⁵. These structural studies allow us to gain insight into the versatile modes of hairpin loop recognition.

In this review we examine our current understanding of how specific recognition of RNA tetraloops and RNA tetraloop-like motifs is mediated by RNA binding proteins. Although general reviews on RNA hairpin structure, RNA-protein interactions, and RNA-RNA interactions are available, the molecular mechanism(s) by which RNA tetraloops recognize their protein targets has not been reviewed yet. RNA hairpin structure^{11, 14, 36}, RNA-protein interactions^{37, 38}, and RNA-RNA interactions^{12, 14, 39} have been reviewed and are not discussed in this report. High resolution structures of free and bound RNA tetraloops are now available from several sub-families. The existing structural, thermodynamic, and kinetic data allow us to scrutinize the different modes of tetraloop-protein recognition in detail and emphasize the various themes that have emerged.

SPECIFIC BASE RECOGNITION BY A HYDROPHOBIC POCKET

• The Retroviral RSV Nucleocapsid Protein-UNCG type SL3 ψ -RNA Complex—

The UNCG family of RNA tetraloops form unusually stable hairpins^{40, 41}. Several NMR and crystal structures of UUCG⁴¹⁻⁴⁴ and UACG⁴⁵ containing RNA hairpins exist. The UUCG tetraloop in particular (Fig. 1A) has been characterized by both NMR and X-ray methods^{43, 44} and is essentially a bi-loop. It is distinguished by a bifurcated hydrogen bond between the first U carbonyl and the imino and amino protons of the fourth G (Fig. 1A). The fourth G adopts a *syn* configuration about the glycosidic bond that is required for the stability of this tetraloop. The structure of the UUCG tetraloop is further stabilized by a base stacking interaction between the C and the first U as well as a base-phosphate hydrogen bond between the C amino proton and the U-U bridging phosphate. The backbone turns between the first and second nucleotides. These features are conserved in all characterized UNCG tetraloops.

The Rous Sarcoma Virus (RSV) Nucleocapsid (NC) protein binds the UGCG tetraloop of stem-loop SL-C from the RSV M ψ packaging signal²³ which corresponds to a 160 nucleotide fragment in the RSV 5' untranslated region (5' UTR). The RSV M ψ is sufficient to direct packaging of RNAs into viral particles⁴⁶. An 82 nucleotide sub-fragment of M ψ (termed $\mu\psi$) is the minimal binding site for the RSV NC protein⁴⁷. The predicted structure of $\mu\psi$ consists of three stem-loops, SL-A, SL-B, SL-C. An AUG initiation codon of open reading frame 3 (ORF3) lies between SL-A and SL-B. Isothermal titration calorimetry (ITC) was used to show that the RSV NC protein interacts with $\mu\psi$ RNA with a 1:1 stoichiometry and a K_d of 1.9 nM in 10 mM NaCl. The affinity of the RSV NC protein for the isolated hairpin SL-C is 350 ± 60 nM, but the RSV NC protein shows no appreciable binding towards hairpins SL-A or SL-B. A second binding site for the RSV NC protein is the AUG initiation codon⁴⁸, and both the SL-C hairpin and the AUG site are required for formation of a high affinity ($K_d \sim 1.9$ nM) $\mu\psi$ RNA-NC protein complex⁴⁸. Intriguingly, replacement of the UGCG tetraloop in SL-C by a structurally stable GAGA tetraloop in the context of the intact $\mu\psi$ RNA decreases binding to the RSV NC by 1500-fold, yielding a K_d of 2.8 ± 0.7 μ M⁴⁸. Substitution of the SL-A and SL-B loops by GAGA has no effect on RNA affinity for NC. Replacement of the UGCG tetraloop by a GAGA tetraloop also reduces viral infectivity suggesting that reduced binding of RSV NC to the SL-C GAGA tetraloop affects viral assembly *in vivo*⁴⁸.

The RSV NC protein binds the U²¹⁷GCG tetraloop of SL-C via a CCHC-type zinc knuckle motif (Fig. 1C, 1D). Solution NMR reveals that the specific interaction between the N-terminal zinc knuckle and the U²¹⁷GCG tetraloop²³ is centered on G218 (Fig. 1B, 1C, 1D). Based on NMR and X-ray crystal structures of related UUCG and UACG tetraloops⁴²⁻⁴⁵ and the lack of observed NOEs between G218 and the RNA hairpin in the RSV NC/SL-C complex, it is inferred that G218 is likely to be solvent exposed in the free RNA. When bound to the RSV NC protein, G218 is completely buried in a hydrophobic pocket in the zinc knuckle (Fig. 1C, 1D). The base of G218 forms π - π interactions with the aromatic side-chains of Tyr-22 and Tyr-30 in a hydrophobic pocket defined by the side-chains of Tyr-22, Tyr-30, Leu-20, and Gln-31 (Fig. 1B, 1C, 1D). The G218 base also makes four hydrogen bonds with the main chain of the zinc knuckle (G218 O6/Tyr-22 NH, G218 O6/Gln-31 NH, G218 N1H/Leu-20 CO, G218 NH21/Leu-20 CO). The network of hydrogen bonding and stacking interactions mediated by the guanine base of G218, indicates that this nucleotide plays a key role in specific recognition of SL-C by the hydrophobic pocket of the RSC NC protein. In addition to G218, the base of G220 packs against the side-chains of Tyr-30 and Ala-17, and the ribose of C216 packs against the side-chain of Ala-32 (Fig. 1). A similar mode of nucleotide-protein interaction has been observed in the structure of the HIV-1 NC protein bound to the SL3 ψ -RNA that has an unrelated GGAG tetraloop⁴⁹. The solvent

exposed tetraloop guanine bases at the second and fourth position of the GGAG loop are buried in separate hydrophobic pockets in the two zinc knuckles of the HIV-1 NC and stack against either Phe or Trp side-chains.

Interestingly, the backbone and sugar conformations of the RSV SL-C M_ψ-RNA tetraloop in the complex mirror free UNCG structures (Fig. 1A, B) and neither the U²¹⁷GCG SL-C tetraloop nor the zinc knuckle motif appear to undergo a major conformational change upon binding. Although the free structure of the UGCG tetraloop is not available, it is possible to infer the potential free structure of this tetraloop based on structures of other tetraloops from this family and this structure concords with the NMR data for the UGCG tetraloop in the complex. In the RSV NC/SL-C complex, U217 and G220 are base paired via interactions between the U217 2'-OH and the G220 O6 as well as between the U217 O2 and G220 NH and NH₂, as expected for the free RNA. The U217 O4 and G220 NH and NH₂ also interact via a bifurcated hydrogen bond. The phosphate 5' of G218 forms a hydrogen bond with the C219 NH₂ and the glycosidic bond of G220 is in a *syn* configuration. The U217 base stacks upon the closing C216•G221 base pair of the stem and C219 stacks upon the U217•G220 base pair. These hydrogen bonding and stacking interactions are identical to those observed in structures of the free UNCG RNA tetraloops. Therefore, the structure of the RSV-NC protein bound to the UGCG tetraloop of stem-loop SL-C suggests that the RNA does not undergo a conformational change in the complex and sequence specific recognition of the UNCG loop may occur primarily via the variable second position that is solvent exposed. In contrast, the three conserved nucleotides in the tetraloop are likely important to stabilize the loop structure.

ADAPTIVE BINDING OF A HELIX IN THE MAJOR GROOVE OF A GNRA FOLD

• **The λ and P22 bacteriophage N-peptides in complex with GNRA-like boxB RNAs**—The GNRA-family of RNA tetraloops are widely distributed in rRNAs^{1, 50}, ribozymes⁵¹, phage RNAs²⁴⁻²⁶, the signal recognition particle,^{52, 53} and viral internal ribosomal entry sites⁵⁴. All eight family members of this sub-group have a very similar backbone conformation and are distinguished by a network of hydrogen bonding and stacking interactions^{15, 55-57}. There are several structures available for GNRA tetraloops^{51, 55, 56}. Analogous to the UNCG tetraloop, these sequence motifs form bi-loops where the G and A form a sheared G-A base pair (Fig. 2A). In such a base pair, the G NH₂ forms a hydrogen bond with the phosphoryl oxygen 5' of the A and the G 2'-OH forms a hydrogen bond with the N7 of the purine at the third position. The turn of the phosphate backbone between the first and second nucleotides is similar to that of the U-turn motif⁵⁸ and the last three bases sequentially stack on each other on the 3' side of the stem. The second and third nucleotides can be dynamic and their sugar puckers frequently exchange between the C2'-*endo* and C3'-*endo* conformations⁵⁵. Notably, the functional groups of the last three bases can be an important readout for RNA-RNA and RNA-protein interactions.

It is important to recognize that the various sub-families of tetraloops are structural “motifs” that are defined by their unique backbone geometry and interactions between the loop bases. Kinetic and structural data indicate that although the sequence of the tetraloop imposes structural constraints, these motifs can occur within the context of larger pentaloops and hexaloops⁵⁹. Early kinetic studies on the group II intron showed that insertion of one or two nucleotides into a GAAA tetraloop of the D5 domain (to yield a GAX_nAA type of pentaloop or hexaloop), did not dramatically affect the catalytic activity of the group II intron ribozyme⁵⁹. This suggested that larger loops may form a GAAA-like structure and participate in the tetraloop-receptor motif interaction. The NMR structure of the GAAGA *boxB* pentaloop bound to N-peptide²⁴ later confirmed that the *boxB* pentaloop adopts a GAAA fold that superimposes on the GAAA tetraloop structure solved by NMR^{56, 60} as

well as the crystal structures of GAAA tetraloops from the hammerhead ribozyme⁶¹ and the P4-P6 domain of the group I intron¹⁵ with r.m.s.d's between 1.4–1.6 Å. The structure of a hairpin with a GCAUA pentaloop from the U6 RNA (PDB code 1LC6) has also been shown to adopt a GNRA-like fold that superimposes on a GCAA tetraloop (PDB code 1ZIH) with an r.m.s.d of 2.4 Å⁶². The uridine at the fourth position is extruded to solvent and is not part of the GCAA-like fold (Fig. 2B).

The mode of N-peptide recognition by the *boxB* GNRA-like hairpins present in the nascent phage mRNA transcripts from lambdoid bacteriophages (λ and P22) is an example of molecular recognition mediated by the GNRA-like family of tetraloops (Fig. 2C–2H)^{24–26}. The N-protein plays an important role in regulating transcription elongation and transcription antitermination of phage mRNA transcripts and is crucial for phage development^{63, 64}. The 22 amino acid N-peptides form specific complexes with their respective cis-acting *boxB* hairpins via arginine rich motifs (ARMs). N-peptides bind their respective *boxB* RNAs with high affinity (K_d of ~5 nM)^{65, 66} and have similar affinity for the RNA as the full length N protein. There are three structures available of *boxB*/N-peptide complexes from the λ and P22 bacteriophages^{24–26}. NMR structures have been calculated for two related *boxB* hairpin/N-peptide complexes from λ -phage, a *boxB* hairpin with a GAAGA loop²⁴, and another with a GAAAA loop²⁶. The NMR structures of both complexes are very similar. Although the GAAGA and GAAAA loops are pentaloops, they adopt the GNRA-like hairpin motif using nucleotides 1,2,3, and 5 and are identical to the GAAA tetraloop (Fig. 2A, 2C)^{56, 67}. The fourth nucleotide of the pentaloop is flipped out in both structures. Replacement of the nucleotide at the fourth position or deletion of the fourth nucleotide by any other nucleotide does not affect binding to the phage λ N peptide in EMSA assays^{68, 69}. A K_d value similar to that of the wild-type hairpin/N-peptide interaction (5.2 nM) has been reported for hairpins with substitutions at the fourth position^{24, 65} and a less than 2-fold decrease in binding is observed for a fourth position deletion mutant in EMSA assays^{68, 69}. In contrast, mutation of either the G or the A of the sheared G-A base pair abolishes binding to the N-peptide^{68, 69}, consistent with the GNRA-like fold being essential for N-peptide binding.

In the structure of the GAAGA λ -*boxB* hairpin loop complexed to a 22 amino acid arginine-rich λ N-peptide (Fig. 2C–2E), the peptide forms a bent α -helix that packs against the major groove face on the 5' end of the RNA stem and three nucleotides (G1, A3, and A5) of the loop. A 120° bend between two helical segments (Ala3-Arg10, Arg12-Ala21) is stabilized by ionic interaction of Arg-11 with the loop-closing U•A base pair as well as hydrophobic contacts between the side-chain of Gln-15 and the second and third adenine bases in the loop (Fig. 2E)²⁴. The peptide undergoes a “disorder-to-order” folding transition upon RNA binding to form a surface that is complementary to that of the RNA hairpin. The bent helix conformation is unique to the λ and P22 N-peptide/*boxB* complexes and is not observed in other arginine-rich peptide-RNA complexes such as HIV-1 Rev^{70, 71} or BIV-1 Tat^{72, 73}. Formation of the complex is electrostatically driven and a cluster of five arginines and two lysines form a positively charged α -helical surface that interacts extensively with the negatively charged phosphodiester backbone of the RNA (Fig. 2C, 2E).

Hydrophobic interactions are also important for binding. The aliphatic side-chains of Arg-7 and Arg-8 contact the ribose, as well as base (H5 and H6) protons of the C6 and U7 bases in the stem and the third nucleotide (A10) in the loop. The methyl group of Ala-3 packs against two cytidines in the stem. Most importantly, A9, which corresponds to the variable “N” position in the GNRA motif, stacks with the indole side-chain of Trp-18 (Fig. 2C, 2E). This interaction is important for specific recognition of the λ *boxB* RNA by λ N-peptide since other phage N-peptides that are arginine rich but do not have a Trp at this position are not able to form a specific complex with the λ *boxB* RNA. It is also crucial for

antitermination^{60, 67}. Ultrafast dynamics measurements using fluorescence show that the Trp-18/A9 stacking interaction regulates the conformation of the N-terminal end of the α -helix and is required for proper assembly of peptide λ in the complex with *boxB* RNA⁷⁴. Therefore specific recognition of *boxB* RNA by the λ N peptide is due to shape and charge complementarity of the protein α -helix and the RNA major groove as well as the stacking of Trp-18 with the second adenine base of the GAA(G)A loop.

The NMR structure of the phage P22 N-peptide in complex with P22 *boxB* RNA²⁵ suggests a similar adaptive mode of recognition by the P22 ARM (Fig. 2F-2H). The N-peptide forms a bent α -helix with extensive ionic interactions with the major groove face of the GNRA-like hairpin (Fig. 2G). The P22 *boxB* pentaloop has the sequence GACAA and in the structure of the complex with the P22 N peptide it is the C that is extruded from the GNRA-like fold (Fig. 2F). This cytidine is involved in extensive hydrophobic contacts with the methyl and methylene groups of Arg-10 and Arg-11, the side-chain of Ile-15, and the imidazole ring of His-7 (Fig. 2F).

The broad specificity of the GNRA-like fold for ARMs is largely determined by shape and electrostatic complementarity of the GNRA-like loop for arginine-rich containing peptides. The intermolecular contacts observed between loop nucleotides 1, 2, and 3 with the ARM-containing peptides and extensive ionic interactions with the major groove face of the GNRA-like motif indicates that the GNRA-fold is an important determinant of specificity for ARM-containing peptides. The 5' strand of the RNA stem is involved in extensive ionic and van der Waals contacts with the peptide and these interactions are conserved in all structures. The more restricted specificity of N-peptide-*boxB* recognition within ARM containing proteins arises from differing interactions of the N-peptide with the GNRA-like loop. In the case of the λ phage, the stacking interaction between Trp-18 and A9 is critical. In P22, the bulged cytidine of the loop contacts the aliphatic side-chains of Arg-10, Arg-11 and Ile-15 residues and stacks against the side-chain of His-7. In both the λ and P22 phage structures, the conserved nucleotides of the GNRA-like tetraloop motif are important for the overall fold of the loop but not important for specificity.

In summary, the *boxB*/N-peptide complexes are an example of “induced-fit” or adaptive mode of recognition of an RNA hairpin by a bent α -helix in the major groove of a GNRA fold. The N-peptide is disordered in the absence of the RNA and adapts to the shape of the major groove face of the GNRA-like fold, forming a well-defined helical structure. Although no structure is available for the free GAAGA, GAAAA, and GACAA RNA pentaloops, the structure of the free GCAUA pentaloop from the U6 RNA (Fig. 2B) shows a typical GNRA fold formed by the GCAA residues with the extra uridine nucleotide flipped out and flexible in the context of the pentaloop. While the GAAGA, GAAAA, and GACAA pentaloops have a backbone geometry that is similar to a GNRA-like fold when bound to the N-peptide, the extruded nucleotide in the RNA loop is rigid and ordered, suggesting that motion may be more restricted in the GNRA-like RNA loop when bound to the N-peptide. Therefore, conformational flexibility of both the RNA and the N-peptide may be necessary to exploit surface complementarity in the major groove of the GNRA-like RNA loop.

SHAPE-SPECIFIC RECOGNITION OF THE (A/U)GNN TETRALOOP MINOR GROOVE BY A HELIX

• **(A/U)GNN cognate and non-cognate tetraloops in complex with the endonuclease Rnt1p**—The double-stranded RNA binding domain (dsRBD) is one of the most abundant structural domains found in RNA binding proteins⁷⁵. Canonical dsRBDs consist of 65-70 amino acids with a $\alpha\beta\beta\alpha$ folding topology. dsRBDs are highly specific for dsRNA and do not bind dsDNA or ssRNA. The specificity of dsRBDs towards dsRNA arises from extensive interaction of the protein backbone and side-chains with the 2' OH

groups of the dsRNA in the minor groove⁷⁶. Structural studies of dsRBD-RNA complexes reveal that in general, these domains bind RNA in a structure-specific but sequence-independent manner. Structures of RNA-bound dsRBDs are available for *Xenopus laevis* Xlrpba⁷⁶, *Drosophila* Staufen²⁷, TRBP2⁷⁷, ADAR2⁷⁸, *Aquifex aeolicus* RNase III⁷⁹, *Arabidopsis* HYL1⁷⁷, and Rnt1p^{28, 29, 80}. In all cases, helix α 1, helix α 2, and the β 1– β 2 loop mediate interactions with the consecutive minor, major, and minor grooves on one face of the RNA, respectively.

Rnt1p is an RNase III homolog from budding yeast that has a Dicer-like endoribonuclease domain that cleaves double-stranded RNA and is essential for RNA processing^{81–84}. Rnt1p is unique in that it is the only known RNase III that binds substrates containing a (A/U)GNN tetraloop via its N-terminal dsRBD and cleaves the RNA 12–14 bp away from the tetraloop in a molecular-ruler type of mechanism^{85, 86}. The (A/U)GNN tetraloops are commonly found in yeast small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), and ribosomal RNAs (rRNAs) that are all substrates for Rnt1p⁸⁷. Most Rnt1p substrates have A and G as the first two nucleotides. The only exceptions are U1 snRNA that has a U instead of an A at the first position and the AAGU tetraloop from the snoRNA, snR48⁸⁸. The (A/U)GNN RNA tetraloop-Rnt1p interaction is of high affinity ($K_d \sim 10$ – 12 nM)^{85, 86} and decreases 100-fold upon deletion of the dsRBD indicating that the dsRBD is important for target selection⁸⁶. In addition, replacing the AGGA tetraloop with a GUGA tetraloop decreases the efficiency of cleavage that is directly attributed to a four-fold decrease in the affinity of Rnt1p for the GUGA mutant⁸⁶. There are several structures now available for free Rnt1p dsRBD^{80, 89}, free (A/U)GNN RNA substrates^{8, 90}, and complexes of Rnt1p with cognate²⁸ and non-cognate²⁹ tetraloops. These structures along with information about backbone dynamics from NMR spin-relaxation measurements⁸⁰ have provided detailed insight into the mechanism of Rnt1p dsRBD-tetraloop RNA recognition.

Solution NMR studies⁸ of the free A⁶GAA and AGUU tetraloops (Fig. 3A) reveal a well-ordered fold that is similar to the solution NMR structure of the unmethylated version of the UGAA tetraloop conserved at the 3' end of eukaryotic 18S rRNA⁷, but very different from the UNCG and GNRA family of tetraloops. A characteristic feature of the (A/U)GNN family of tetraloops is the conserved G at the second position that adopts a syn configuration about the glycosidic bond allowing the sugar-phosphate backbone to reverse direction by 180° on the 3' side of the guanosine (Fig. 3A). The sharp turn exposes the non-bridging phosphate oxygen atoms between the G and the N nucleotides while the G base stacks on the preceding adenine and the closing base pair of the stem. This loop conformation creates a binding pocket that is recognized by Rnt1p. The syn configuration of the G and the sharp turn are strong determinants of (A/U)GNN tetraloop structure and Rnt1p dsRBD recognition. In addition, in the free RNA hairpin, the first A and the second G stack over the closing base of the stem at the 5' end, whereas the two adenines in the third and fourth positions stack on the 3' end of the closing base pair of the stem.

Solution NMR and X-ray crystal structures of the free⁸⁹ and RNA-bound²⁸ Rnt1p dsRBD reveal a conserved $\alpha\beta\beta\beta\alpha$ fold with an additional helix α 3 that is unique to Rnt1p (Fig. 3D). In the free dsRBD, helix α 3 is positioned near the C-terminus of α 1, and packs against the α 1– β 1 loop. The solution NMR structure of the Rnt1p dsRBD complexed to a 14 bp hairpin with an AGAA tetraloop from the snR47 snoRNA precursor has been solved²⁸ (Fig. 3C, 3D, 3E). Helix α 1 of the Rnt1p dsRBD is the recognition helix and it binds the bend in the minor groove of the (A/U)GNN tetraloop by contacting the sugar-phosphate backbone, the “NN” bases, and the top of the stem (Fig. 3C). The N-terminus of helix α 1 is extended by three residues in the RNA bound complex and lies near the 5' end of the tetraloop, whereas its C-terminus is near the 3' end. Overall, helix α 1 undergoes a conformational change and adapts to the shape of the RNA tetraloop in the minor groove (Fig 3D, 3E). Helix α 1 bends at

Ser-376 and undergoes an 18° rotation and translation towards the RNA⁸⁰. This conformational change is facilitated by hinge motion of the $\alpha 1$ - $\beta 1$ loop that also changes the orientation of helix $\alpha 3$. In helix $\alpha 1$, the side-chain guanidinium of Arg-372 and hydroxyl groups of Ser-376 hydrogen bond to the 2' OH of the "NN" nucleotides A17 and A18. The aliphatic portion of the Arg-372 side-chain also makes van der Waals contacts with the A17 and A18 bases (Fig. 3C, 3G). In addition, residues Asp-367, Lys-371, and Tyr-375 of helix $\alpha 1$ form contacts with the 2'OH and ribose of three base pairs in the A-form stem below the tetraloop. Further, side-chains from the N-terminus of helix $\alpha 2$ and the $\beta 3$ - $\alpha 2$ loop contact the sugar-phosphate backbone of the major groove of the A-form stem. Intriguingly, only the nonconserved bases are contacted by the Rnt1p dsRBD and there are no contacts made to the conserved (A/U)G bases that point into the major groove and are essential for the general fold of the tetraloop. The RNA tetraloop does not appear to undergo a major conformational change to accommodate helix $\alpha 1$ in this complex. The only significant change in RNA structure is that a small bend in the backbone is observed at the top of the helix and below the tetraloop in response to binding of the Rnt1p dsRBD. A superposition of the lowest energy conformers of the free AGAA tetraloop and the AGAA tetraloop bound to the Rnt1p dsRBD yields an r.m.s.d of 0.99 Å for backbone heavy atoms of the tetraloop and the closing base pair. Similar to the structures of the free RNA hairpins containing AGAA⁸, AGUU⁸, and AGUC⁹¹ tetraloops, the RNA backbone in the Rnt1p complex turns after the second guanine which is in a *syn* configuration, the conserved A and G bases stack on the 5' end and the nonconserved bases stack on the 3' side of the tetraloop. In summary, the structure of the Rnt1p dsRBD bound to the AGAA tetraloop shows that the dsRBD adapts to the shape of the RNA tetraloop in the minor groove and no sequence-specific contacts are made between the protein and the RNA. The RNA undergoes only a small conformational change upon protein binding.

In contrast to the cognate AGAA tetraloop, the structure of the Rnt1p dsRBD/non-cognate AAGU tetraloop complex²⁹ (Fig. 3F-3H) shows a more significant conformational rearrangement in the AAGU RNA tetraloop to accommodate the Rnt1p helix $\alpha 1$. The structures of the free non-cognate (AAGU)⁹⁰ and cognate (AGAA) tetraloops⁸ (Fig. 3A and 3B) are very different with an r.m.s.d of 3.3 Å for backbone heavy atoms. The difference between the tetraloop structures is particularly important at the 3' end of the loop. In the structure of the free non-cognate AAGU RNA, the uracil is flipped out towards the solvent and does not form a base pair with the adenosine at the 5' end of the loop (Fig. 3B). The two adenosines on the 5' side stack continuously with the closing base pair of the stem and the backbone turns between the adenosine and the guanine. Comparison of the structures of the cognate Rnt1p dsRBD-AGAA and non-cognate Rnt1p dsRBD-AAGU complexes shows that the bound states are remarkably similar. A superposition of the lowest energy conformers of the Rnt1p bound AAGU tetraloop²⁹ and the AAGU tetraloop bound to the Rnt1p dsRBD²⁸ complexes yields an r.m.s.d of 0.97 Å for backbone heavy atoms of the tetraloop and the closing base pair (not shown). When bound to the Rnt1p dsRBD, the nonconserved G17 and U18 bases of the AAGU tetraloop undergo a large conformational change to accommodate helix $\alpha 1$ (Fig. 3F, 3B). G17 and U18 point into the minor groove and their backbones interact with the side-chains of Arg-372 and Ser-376 in helix $\alpha 1$ (Fig. 3F, 3H). No interaction is observed between the first two adenosines on the 5' side of the tetraloop, similar to that observed in the cognate AGAA-dsRBD complex. The structural data on the cognate AGAA/dsRBD and non-cognate AAGU/dsRBD complexes is supported by ITC measurements²⁹. The two complexes have the same binding affinity ($K_d = 2.23 \pm 0.27 \mu\text{M}$ for the AAGU/dsRBD vs $2.25 \pm 0.15 \mu\text{M}$ for the AGAA/dsRBD) in 300 mM NaCl, but the AAGU/dsRBD undergoes larger changes in enthalpy ($\Delta H = -1.42 \pm 0.06 \times 10^4 \text{ kcal.mol}^{-1}$ for the AAGU/dsRBD vs $-0.92 \pm 0.02 \times 10^4 \text{ kcal.mol}^{-1}$ for the AGAA/dsRBD) and a significant decrease in entropy ($\Delta S = -21.8 \text{ kcal.mol}^{-1}.\text{K}^{-1}$ for the AAGU/dsRBD vs $-3.94 \text{ kcal.mol}^{-1}.\text{K}^{-1}$ for the AGAA/dsRBD) upon complex formation as

compared to the AGAA/dsRBD complex. Therefore, the structure of the non-cognate AAGU capped hairpin bound to the Rnt1p dsRBD is an example of a cooperative conformational change in both the RNA and the protein components to form a shape-specific complex.

The Rnt1p dsRBD-AGAA tetraloop complex is one of the few RNA-protein complexes where the contribution of motion and dynamics to RNA binding has been investigated⁸⁰. NMR-based backbone dynamics measurements for the free and RNA bound Rnt1p dsRBD suggests that in general the dynamic changes in the protein correlate well with the conformational transitions that the protein undergoes upon binding RNA. In the RNA free state, the Rnt1p dsRBD shows slow motions on the microsecond-millisecond timescale (as measured by the parameter R_{ex}) for helix $\alpha 1$, the $\alpha 1$ - $\beta 1$ loop, and helix $\alpha 3$. When bound to the AGAA tetraloop, the $\alpha 1$ - $\beta 1$ loop becomes more rigid, but residues in helices $\alpha 1$ and $\alpha 3$ have larger R_{ex} values indicating an increase in conformational exchange on the microsecond-millisecond timescale. The increase in slow motions for these regions indicates that the Rnt1p-AGAA tetraloop interface is not rigid and such dynamics at the protein-RNA interface may reflect how the dsRBD can adapt to different (cognate and non-cognate) RNA substrates.

COUPLED FOLDING AND BINDING OF A DISORDERED TETRALOOP

• **The MS2 AUUA Hairpin Loop RNA complex**—The MS2 coat protein binds a hairpin from its RNA genome in a highly sequence and structure-specific manner⁹²⁻⁹⁴. Interaction of the RNA hairpin with the MS2 coat protein is the obligatory first step in translational repression of the viral replicase gene⁹⁵. The free RNA hairpin consists of a seven bp stem with a -7 AUU -4 A loop and an adenosine-bulge (-10 A) (Fig. 4A). In the structure of the free RNA, the bulged -10 A is intercalated between the flanking base pairs of the stem. The -7 AUU -4 A tetraloop is flexible and multiple conformations are observed by NMR⁹⁶. The relaxation measurements suggest that the -6 U, -5 U, and -4 A bases are very dynamic and predisposed to reorganization, whereas -7 A at the start of the loop is stable⁹⁷.

The MS2 coat protein is dimeric at low concentrations (up to ~ 1 μ M) but the dimers self-associate to form phage-like capsids at high concentrations⁹⁸. The MS2 coat protein dimer binds the RNA hairpin with a K_d of 3.3 nM⁹⁹. Some insight into the mechanism of RNA recognition by the MS2 coat protein comes from stopped-flow fluorescence experiments⁹⁹. The kinetics measurements show a biphasic mode of interaction between the MS2 coat protein and the RNA hairpin. The fast phase is diffusion driven and occurs with a rate constant of 2×10^9 $M^{-1} s^{-1}$. The salt dependence of the first phase suggests that it is dominated by electrostatic steering effects. The slower phase appears to correspond to the isomerisation of the protein from a non-native state to an RNA-binding competent state prior to RNA binding.

Several crystal structures of MS2 bound to wild type, mutant RNA hairpins, and RNA aptamers have been solved³¹⁻³³. The crystal structures of the MS2 protein bound to the RNA hairpin reveal that the protein is a symmetric dimer in which the RNA recognition face consists of a five-stranded β -sheet from each monomer⁹⁸ (Fig. 4C). The MS2 protein dimer binds asymmetrically to the RNA hairpin and contacts -4 A and -10 A directly (Fig. 4C, 4D). The -4 A base inserts into a hydrophobic pocket formed by valine and lysine side chains in one monomer and the bulged -10 A base inserts into a hydrophobic pocket in the second monomer. Biochemical studies show that a pyrimidine is required at the -5 position in the loop. The U/C at the -5 position stacks on the side chain of Tyr-85 of one of the MS2 monomers. The adenosine at the -7 position in turn stacks with the -5 U/C above it and the closing base pair of the stem below. The sugars in the -4 , -5 , and -6 positions have C2'-*endo* conformations and the bulged -10 A has a C3'-*endo* conformation. Although the MS2

coat protein does not undergo a large conformational change upon RNA binding, the structure of the RNA tetraloop and bulge in the complex is dramatically different from the NMR solution structures^{96, 97} of the same hairpin loop in the absence of protein. In the free RNA, the sugars of nucleotides in the -4, -5, and -6 positions have *C2'-endo* conformations, as observed in crystal structures of the MS2-RNA complex, however the ⁻⁵U and ⁻⁴A bases are stacked on each other and the ⁻⁷A and ⁻⁶U bases stack on each other. Therefore the structural data suggests that the free RNA loop exchanges between multiple states and either undergoes a large conformational change upon binding MS2 or a minor tetraloop conformer is selected upon binding.

RECOGNITION BY TETRALOOP UNFOLDING

• Recognition of the GNRA Sarcin/Ricin Loop by restrictocin

•The histone mRNA hairpin-SLBP RBD-ERI1 complex: In the examples discussed so far, several intriguing observations have emerged. In many cases, the structure of the free RNA tetraloop remains largely unaltered when bound to the target protein. Second, there are several RNA-protein complexes where non-conserved bases in the tetraloop form RNA-protein contacts important for specificity while the conserved bases appear to be important for preservation of the RNA tetraloop structure. There are, however, two examples that deviate from these observations wherein the RNA tetraloop undergoes a large conformational change and becomes unfolded upon binding the protein.

X-ray crystal structures of the Sarcin/Ricin Loop (SRL) RNA bound to the α -sarcin ribotoxin restrictocin³⁴ show that the highly stable GNRA tetraloop is unfolded in the toxin-bound form (Fig. 5B, 5C). The α -sarcin family of ribonucleases are secreted by fungi¹⁰⁰ and bind the SRL structure in 23S-28S rRNA. Restrictocin specifically cleaves the SRL GAGA tetraloop between the third and fourth loop nucleotides and impairs binding of the elongation factors EF-G and EF-Tu to the ribosome, thereby stopping protein synthesis. The structure of the free SRL hairpin present in the 50S ribosomal subunit consists of a typical GNRA tetraloop (Fig. 5A) and a stem containing a bulged-G motif¹⁰¹ (Fig. 5D). The bulged G-motif is located 12 Å from the cleavage site and is necessary for specific recognition by α -sarcin ribotoxins. To minimize cleavage during crystallization, structures of restrictocin were solved bound to RNA variants containing a 2'-O-methyl guanosine (mG) substitution at the third position in the loop. This substitution is not expected to alter the structure of the GNRA motifs. In one variant a 2'-deoxy adenosine (dA) was also introduced at the fourth position³⁴. In all structures, the restrictocin contacts the major groove face of the SRL RNA via three loops in the protein. Two lysine-rich loops make direct contacts with the guanine nucleotide and the backbone at the bulged-G (Fig. 5D). The structure of restrictocin and the bulged-G motif of the RNA hairpin remain unchanged when compared to the uncomplexed states.

Upon formation of the restrictocin-SRL RNA complex, the GAGA tetraloop shows base flipping, a change in the backbone conformation, and base stacking that is characteristic of an unfolded loop (Fig. 5A, 5B, 5C). Two different unfolded conformations are observed for the tetraloop in the mG and dA SRL RNA substrates, indicating that tetraloop unfolding is not due to the methyl (mG) substitution. In each of these structures, the bases of the third G and fourth A flip out of the GNRA fold. As a result, the sheared G-A base pair characteristic of GNRA motifs is lost. In addition, the bases of the second, third, and fourth nucleotides no longer stack. The backbone also reverses direction between the second and third nucleotides in the restrictocin-SRL complex, in contrast to the free SRL, where the backbone changes direction between the first and second nucleotides. The unfolding of the tetraloop is presumably a result of protein binding. Superposition of the restrictocin-SRL mG complex crystal structure with the structure of the SRL in the 50S ribosomal subunit¹⁰¹ shows that

the toxin is unable to contact both the bulged-G motif and the folded GNRA loop in the 50S ribosomal subunit at the same time due to steric effects. Although it is not clear whether the toxin selects an infrequently populated conformation or induces it, base flipping is required for toxin binding and efficient cleavage. GNRA tetraloops are quite stable and the reported T_m for a C(GAAA)G tetraloop is 65.9°C which corresponds to a ΔG_{fold} of -4.2 kcal/mol at 37°C¹⁰². The thermodynamic stability is due to favourable H-bonding and stacking interactions within the GNRA fold^{11,55}. Therefore the unfolding of the stable loop structure in the restrictocin complex is very unusual.

The second example is the crystal structure of a ternary complex consisting of the histone mRNA hairpin, the unphosphorylated RNA binding and processing domain (RPD) of Stem-Loop Binding Protein (SLBP), and the human exonuclease ERI-1³⁵ (Fig. 6). The mRNAs that encode all five core histone proteins (H1, H2A, H2B, H3, and H4) are unique in that they are the only naturally occurring eukaryotic mRNAs that are not polyadenylated¹⁰³. Instead of a poly(A) tail, the histone pre-mRNA 3' untranslated region (3' UTR) contains a 16 nt conserved stem-loop (Fig. 6D) that is capped by a UYUM tetraloop and is the minimal substrate required for high affinity interaction with SLBP. The solution NMR structure of the U(UUUC)A stem-loop from the UYUM family present in the histone H4-12 gene has been solved by two laboratories in the absence of SLBP^{9, 10} (Fig. 6A). These structures confirm that the six base pair RNA stem adopts a classic A-form helix starting with the highly conserved G-C base pair at the base and extending up the stem to a loop-closing U-A base pair. The 5' and 3' flanking regions at the base of the stem are disordered in solution. There are three unique features of the UUUC tetraloop. First, unlike the UNCG, GNRA, and CUUG classes of tetraloops, there is no evidence for a base pair between the first and fourth nucleotide. The UUUC tetraloop is stabilized only by base stacking interactions. Second, the conformation of the phosphate backbone in the loop is significantly different from that observed in most tetraloops in that it lacks a sharp reverse turn and instead reverses direction over multiple nucleotides. Third, a notable difference between the two NMR structures is the orientation of U12 in the loop which base stacks against other loop uridines in one case⁹ but is solvent exposed towards the minor groove face in the other¹⁰ (Fig. 6A). These observations likely reflect the dynamic nature of the loop and its sensitivity to solution conditions since different salt and pH conditions were used in the two NMR studies.

Two proteins are known to bind this stem-loop: SLBP and an exonuclease, ERI1, that has been implicated in histone mRNA degradation¹⁰⁴. ERI1 is a DEDDh family exonuclease that trims 2-3 nt from the 3' end of the mature histone mRNA stem-loop. It belongs to a family of structure-specific nucleases that bind RNA hairpins or duplexes with a 2-3 nt 3' overhangs and has no sequence-specificity for the RNA¹⁰⁵.

The mode of interaction of SLBP with the histone mRNA stem-loop has been well characterized using biochemical and biophysical tools by a number of laboratories. The contribution of specific RNA nucleotides to SLBP binding has been quantified by mutagenesis using EMSA assays^{10, 106}, nitrocellulose filter binding assays¹⁰⁷, and an elegant high-throughput microfluidics approach¹⁰⁸ that takes into consideration both structural and sequence features of the RNA. These affinity measurements have all been made using the Thr-171 phosphorylated form of baculovirus-expressed SLBP (see below). The results indicate that the second G-C base pair and the sixth U-A base pair in the stem and two uridines in the loop (U12 and U14) are critical for interaction with SLBP. In addition, the flanking adenosines 5' and 3' to the stem-loop contribute to SLBP binding (Fig. 6D). In all studies, the UYUM tetraloop is a strong determinant of SLBP binding. Replacement of the UUUC tetraloop by an AUAC, GUUA, or a GNRA sequence results in a >200 fold decrease in binding affinity towards SLBP¹⁰⁷. Replacement of the loop uridines U12 and U14 by any other nucleotide (including cytidine) strongly reduces binding to

SLBP^{10, 109}. Watson-Crick base pairing of the closing UA base pair is also critical^{10, 107, 109}. Since the functional groups of the tetraloop and closing base pair are oriented towards the major groove, the generally accepted view is that the SLBP RPD would recognize this face of the RNA hairpin, making sequence-specific contacts with loop nucleotides.

NMR and mass spectrometry studies have shown that human and *Drosophila* SLBP are intrinsically disordered phosphoproteins^{110, 111}. Human SLBP is phosphorylated at nine Ser/Thr residues in mammalian cells¹¹² and 23 Ser/Thr sites when isolated from baculovirus¹¹¹. Solution NMR studies show that the SLBP RPD is not stably folded in solution, has three helices, and undergoes a disorder-to-order transition upon RNA binding^{113, 114}. The first 30 amino acids of the SLBP RPD form a helix-turn-helix motif in solution, but do not bind histone mRNA with high affinity¹¹³. A unique feature of SLBP is that phosphorylation of a conserved threonine in a ¹⁷¹TPNK sequence in the SLBP RPD is important for the kinetics of interaction with histone mRNA^{113, 115}. An apparent equilibrium dissociation constant (K_d) between 1.5 nM–9 nM^{10, 107, 113, 115} has been measured for the interaction of Thr-171 phosphorylated human SLBP with the 26-nucleotide stem-loop. Baculovirus expressed phosphorylated full-length SLBP has a slow association rate for the histone mRNA stem-loop ($k_{on} = 1.2 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$) compared to the diffusion-limited rate of $1.0 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$ and a very slow dissociation rate ($k_{off} = 5.7 \times 10^{-4} \text{ min}^{-1}$) yielding a binding constant (K_d) of $47.5 \pm 17 \text{ nM}$ as measured by Surface Plasmon Resonance (SPR)¹¹³. The more basic dephosphorylated full-length SLBP has a 10-fold faster on rate toward the stem-loop, but a 100-fold faster off rate of $7.44 \times 10^{-2} \text{ M}^{-1} \text{ s}^{-1}$, yielding a K_d of $528.6 \pm 73 \text{ nM}$. This corresponds to an ~11-fold decrease in binding affinity for dephosphorylated full-length dSLBP, which is similar to that reported using filter binding measurements¹¹⁵. NMR studies provide evidence for cis-trans proline isomerization about the ¹⁷¹Thr-¹⁷²Pro sequence¹¹³. Mutation of Pro-172 results in loss of RNA binding and embryonic lethality¹¹⁶. The ¹⁷¹Thr-¹⁷²Pro sequence is the binding site for the prolyl isomerase Pin1 *in vitro* and *in vivo*¹¹².

In the crystal structure of the unphosphorylated SLBP RPD-histone mRNA hairpin-ERI1 ternary complex³⁵, the SLBP RPD (residues 125–223) consists of three helices: helix-A (Glu-132–Lys-146), helix-B (Ile-149–Val-158), and helix-C (Arg-180–Phe-196) (Fig. 6E). Consistent with previous reports^{111, 113, 114}, a significant portion (57 residues out of 99 total) of the RPD is disordered, with 37 residues having no visible electron density and 20 residues in loops that are likely flexible given their high B-factor. The site of phosphorylation, Thr-171, is in a flexible loop between helix-B and helix-C and is oriented away from the RNA (Fig. 6E). Helix-C is the recognition helix and it approaches the major groove face of the RNA hairpin from the 5' end, but does not lie in the major groove. Helix-C acts as a molecular ruler and extends from the base of the RNA stem to the top of the tetraloop. A specific base contact is made between the side-chain guanidinium group of Arg-181 and the second guanine base (G7) of the RNA stem, consistent with previous biochemical studies^{107, 109}. Surprisingly, the interaction of SLBP with the tetraloop consists of non-sequence specific π -stacking interactions between Tyr-144 of helix-A and the first uridine (U12) of the loop and His-195 with the non-conserved cytidine, C15 (Fig. 6B and 6F). Substitution of C15 with any other nucleotide has no appreciable effect on SLBP binding^{10, 109}. Tyr-144 also shows a T-shaped stacking interaction with U14 in the crystal structure³⁵. The RNA tetraloop is unfolded in the crystal structure as compared to the two solution NMR structures of the free RNA hairpin (Fig. 6A, 6B, 6C). The first, third, and fourth bases of the tetraloop are flipped out and oriented towards SLBP, while the second uridine (U13) is oriented towards the SAP domain of ERI1 and interacts with Tyr-66, Phe-61, and Lys-111 of ERI1 (Fig. 6F). The ERI1 nuclease approaches the RNA hairpin

from the 3' end with the SAP domain interacting with the tetraloop and non-specific contacts are made with the backbone of the stem and the 3' flanking region (Fig. 6E).

Although the structure is the first example of a ternary complex involving two proteins in contact with an RNA hairpin, there are several inconsistencies with biochemical studies to be resolved. For example, the lack of sequence-specific contacts between SLBP and the tetraloop in the crystal structure is not consistent with biochemical studies of the phosphorylated SLBP-histone mRNA binary complex. It is possible that the orientation of helix-B is different in the context of the phosphorylated SLBP-histone mRNA complex. The solution NMR structure of residues Glu-129-Val-158 indicates the presence of two helices encompassing Glu-132-Ile-142 and Thr-147-Val-158. These helices are longer [Helix-A (Glu-132-Lys-146) and Helix-B (Ile-149-Val-158)], in the RNA-bound crystal structure. In the crystal, Helix-B is almost perpendicular to Helix-A, and lies farther away from the RNA binding interface in the crystal. Helix-B has the highest temperature factors (B-factors) in the entire ternary complex (Fig. 6G), which could be indicative of high-amplitude motions for the peptide backbone atoms of this helix. Helix-B is also involved in crystal contacts with the SAP domain of a neighboring ERI-1 molecule in the lattice (Fig. 6H). Extensive crystal contacts are also observed between the nuclease domains of ERI-1 (Fig. 6H). The effect of these contacts on the conformation of the tetraloop is unclear. Future structural studies by crystallography and NMR on the binary and ternary complexes should help clarify the mode of recognition of the histone mRNA hairpin in detail.

Concluding Remarks

We have described the different modes of RNA tetraloop recognition by proteins as observed in the few complexes that have been structurally characterized in detail. The mechanisms include specific base recognition by a hydrophobic pocket, adaptive binding of a helix in the major groove of a GNRA fold, shape-specific recognition of the (A/U)GNN tetraloop minor groove by a helix, simultaneous folding and binding of an unfolded tetraloop, and recognition by tetraloop unfolding. There are examples of disordered-to-ordered transitions where the peptide is conformationally dynamic and adapts its structure to that of the RNA target, and ordered-to-disordered transitions that involve the RNA tetraloop. In most cases, the structural characteristics of the RNA tetraloop motifs are conserved in the protein-RNA complexes. The five modes of recognition we have described do not exhaust the possibilities, and, no doubt, further studies will reveal additional mechanisms of recognition.

To fully comprehend the modes of RNA tetraloop recognition by proteins we need to understand the contribution of dynamics to recognition. Studies using NMR, time-resolved fluorescence, and molecular dynamics simulations have shown that tetraloops, although stable, populate several minor or unfolded states at equilibrium. How dynamics of the RNA and protein contributes to complex formation has been relatively unexplored. Another area that has not been well studied is the mode of RNA recognition in larger ribonucleoprotein complexes where more than one protein can recognize the tetraloop simultaneously. The SLBP-ERI1-histone mRNA stem-loop is the only example of a ternary complex for which one structure is available, and in that case, the RNA tetraloop adapts to accommodate both proteins in the complex. Therefore, additional structural studies are required on larger assemblies. Finally, a number of proteins are posttranslationally modified by phosphorylation, ubiquitination, acetylation, etc. Several RNAs, especially ribosomal RNAs, are also edited or undergo base modifications. Such modifications are important for proper function, but structural information on modified proteins or RNAs in complexes is currently lacking. We anticipate that these are the areas that will lead to new insights into the diverse modes of RNA tetraloop-protein interactions in the future.

Acknowledgments

Supported by NIH 1RO1-GM076660 and 1RO1-GM076660-04S1 (ARRA) grants (R.T.) faculty start-up funds from HWI to R.T and by NIH 1RO1-GM73969 (E.P.N.). Support for A.P.D. was provided by National Institutes of Health Training Grant 2T32 GM008280. We gratefully thank Dr. Pascale Legault (University of Montreal, Canada) for PDB coordinates of the λ N-peptide- λ boxB RNA complex. We also thank the reviewers for their insightful comments. Our goal was to focus on the most well studied examples and structures that depict a unique mode of molecular recognition. We apologize to individuals whose work we could not include due to space limitations.

References

1. Woese CR, Winker S, Gutell RR. Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc Natl Acad Sci U S A*. 1990; 87(21):8467–71. [PubMed: 2236056]
2. Woese CR, Gutell R, Gupta R, Noller HF. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev*. 1983; 47(4):621–69. [PubMed: 6363901]
3. Tinoco I Jr. Bustamante C. How RNA folds. *J Mol Biol*. 1999; 293(2):271–81. [PubMed: 10550208]
4. Woese CR, Magrum LJ, Gupta R, Siegel RB, Stahl DA, Kop J, Crawford N, Brosius J, Gutell R, Hogan JJ, Noller HF. Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res*. 1980; 8(10):2275–93. [PubMed: 6159576]
5. Tuerk C, Gauss P, Thermes C, Groebe DR, Gayle M, Guild N, Stormo G, d'Aubenton-Carafa Y, Uhlenbeck OC, Tinoco I Jr. et al. CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc Natl Acad Sci U S A*. 1988; 85(5): 1364–8. [PubMed: 2449689]
6. Keating KS, Toor N, Pyle AM. The GANC tetraloop: a novel motif in the group IIC intron structure. *J Mol Biol*. 2008; 383(3):475–81. [PubMed: 18773908]
7. Butcher SE, Dieckmann T, Feigon J. Solution structure of the conserved 16 S-like ribosomal RNA UGAA tetraloop. *J Mol Biol*. 1997; 268(2):348–58. [PubMed: 9159475]
8. Wu H, Yang PK, Butcher SE, Kang S, Chanfreau G, Feigon J. A novel family of RNA tetraloop structure forms the recognition site for *Saccharomyces cerevisiae* RNase III. *EMBO J*. 2001; 20(24):7240–9. [PubMed: 11743000]
9. DeJong ES, Marzluff WF, Nikonowicz EP. NMR structure and dynamics of the RNA-binding site for the histone mRNA stem-loop binding protein. *RNA*. 2002; 8(1):83–96. [PubMed: 11871662]
10. Zanier K, Luyten I, Crombie C, Muller B, Schumperli D, Linge JP, Nilges M, Sattler M. Structure of the histone mRNA hairpin required for cell cycle regulation of histone gene expression. *RNA*. 2002; 8(1):29–46. [PubMed: 11871659]
11. Varani G. Exceptionally stable nucleic acid hairpins. *Annu Rev Biophys Biomol Struct*. 1995; 24:379–404. [PubMed: 7545040]
12. Noller HF. RNA structure: reading the ribosome. *Science*. 2005; 309(5740):1508–14. [PubMed: 16141058]
13. Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci U S A*. 2001; 98(9):4899–903. [PubMed: 11296253]
14. Butcher SE, Pyle AM. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res*. 2011; 44(12):1302–11. [PubMed: 21899297]
15. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*. 1996; 273(5282): 1678–85. [PubMed: 8781224]
16. Ulyanov NB, Mujeeb A, Du Z, Tonelli M, Parslow TG, James TL. NMR structure of the full-length linear dimer of stem-loop-1 RNA in the HIV-1 dimer initiation site. *J Biol Chem*. 2006; 281(23):16168–77. [PubMed: 16603544]
17. Ferre-D'Amare AR, Zhou K, Doudna JA. Crystal structure of a hepatitis delta virus ribozyme. *Nature*. 1998; 395(6702):567–74. [PubMed: 9783582]

18. Theimer CA, Blois CA, Feigon J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell*. 2005; 17(5):671–82. [PubMed: 15749017]
19. Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA. Crystal structure of a self-splicing group I intron with both exons. *Nature*. 2004; 430(6995):45–50. [PubMed: 15175762]
20. Williamson JR. Induced fit in RNA-protein recognition. *Nat Struct Biol*. 2000; 7(10):834–7. [PubMed: 11017187]
21. Leulliot N, Varani G. Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry*. 2001; 40(27):7947–56. [PubMed: 11434763]
22. Frankel AD, Smith CA. Induced folding in RNA-protein recognition: more than a simple molecular handshake. *Cell*. 1998; 92(2):149–51. [PubMed: 9458038]
23. Zhou J, Bean RL, Vogt VM, Summers M. Solution structure of the Rous sarcoma virus nucleocapsid protein: muPsi RNA packaging signal complex. *J Mol Biol*. 2007; 365(2):453–67. [PubMed: 17070546]
24. Legault P, Li J, Mogridge J, Kay LE, Greenblatt J. NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell*. 1998; 93(2):289–99. [PubMed: 9568720]
25. Cai Z, Gorin A, Frederick R, Ye X, Hu W, Majumdar A, Kettani A, Patel DJ. Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. *Nat Struct Biol*. 1998; 5(3):203–12. [PubMed: 9501914]
26. Scharpf M, Sticht H, Schweimer K, Boehm M, Hoffmann S, Rosch P. Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. *Eur J Biochem*. 2000; 267(8):2397–408. [PubMed: 10759866]
27. Ramos A, Grunert S, Adams J, Micklem DR, Proctor MR, Freund S, Bycroft M, St Johnston D, Varani G. RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J*. 2000; 19(5):997–1009. [PubMed: 10698941]
28. Wu H, Henras A, Chanfreau G, Feigon J. Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc Natl Acad Sci U S A*. 2004; 101(22):8307–12. [PubMed: 15150409]
29. Wang Z, Hartman E, Roy K, Chanfreau G, Feigon J. Structure of a yeast RNase III dsRBD complex with a noncanonical RNA substrate provides new insights into binding specificity of dsRBDs. *Structure*. 2011; 19(7):999–1010. [PubMed: 21742266]
30. Valegard K, Murray JB, Stonehouse NJ, van den Worm S, Stockley PG, Liljas L. The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J Mol Biol*. 1997; 270(5):724–38. [PubMed: 9245600]
31. Horn WT, Convery MA, Stonehouse NJ, Adams CJ, Liljas L, Phillips SE, Stockley PG. The crystal structure of a high affinity RNA stem-loop complexed with the bacteriophage MS2 capsid: further challenges in the modeling of ligand-RNA interactions. *RNA*. 2004; 10(11):1776–82. [PubMed: 15496523]
32. Helgstrand C, Grahm E, Moss T, Stonehouse NJ, Tars K, Stockley PG, Liljas L. Investigating the structural basis of purine specificity in the structures of MS2 coat protein RNA translational operator hairpins. *Nucleic Acids Res*. 2002; 30(12):2678–85. [PubMed: 12060685]
33. Grahm E, Moss T, Helgstrand C, Fridborg K, Sundaram M, Tars K, Lago H, Stonehouse NJ, Davis DR, Stockley PG, Liljas L. Structural basis of pyrimidine specificity in the MS2 RNA hairpin-coat-protein complex. *RNA*. 2001; 7(11):1616–27. [PubMed: 11720290]
34. Yang X, Gerczei T, Glover LT, Correll CC. Crystal structures of restrictocin-inhibitor complexes with implications for RNA recognition and base flipping. *Nat Struct Biol*. 2001; 8(11):968–73. [PubMed: 11685244]
35. Tan D, Marzluff WF, Dominski Z, Tong L. Structure of histone mRNA stem-loop, human stem-loop binding protein, and 3'hExo ternary complex. *Science*. 2013; 339(6117):318–21. [PubMed: 23329046]
36. Cheong, C.; Cheong, H-K. eLS. John Wiley & Sons, Ltd.; 2010. RNA Structure:Tetraloops..

37. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol.* 2007; 8(6):479–90. [PubMed: 17473849]
38. Chen Y, Varani G. Protein families and RNA recognition. *FEBS J.* 2005; 272(9):2088–97. [PubMed: 15853794]
39. Cruz JA, Westhof E. The dynamic landscapes of RNA architecture. *Cell.* 2009; 136(4):604–9. [PubMed: 19239882]
40. Molinaro M, Tinoco I Jr. Use of ultra stable UNG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucleic Acids Res.* 1995; 23(15):3056–63. [PubMed: 7544890]
41. Allain FH, Varani G. Structure of the P1 helix from group I self-splicing introns. *J Mol Biol.* 1995; 250(3):333–53. [PubMed: 7608979]
42. Cheong C, Varani G, Tinoco I Jr. Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature.* 1990; 346(6285):680–2. [PubMed: 1696688]
43. Varani G, Cheong C, Tinoco I Jr. Structure of an unusually stable RNA hairpin. *Biochemistry.* 1991; 30(13):3280–9. [PubMed: 1706937]
44. Ennifar E, Nikulin A, Tishchenko S, Serganov A, Nevskaya N, Garber M, Ehresmann B, Ehresmann C, Nikonov S, Dumas P. The crystal structure of UUCG tetraloop. *J Mol Biol.* 2000; 304(1):35–42. [PubMed: 11071808]
45. Comolli LR, Ulyanov NB, Soto AM, Marky LA, James TL, Gmeiner WH. NMR structure of the 3' stem-loop from human U4 snRNA. *Nucleic Acids Res.* 2002; 30(20):4371–9. [PubMed: 12384583]
46. Banks JD, Kealoha BO, Linial ML. An Mpsi-containing heterologous RNA, but not env mRNA, is efficiently packaged into avian retroviral particles. *J Virol.* 1999; 73(11):8926–33. [PubMed: 10515997]
47. Banks JD, Linial ML. Secondary structure analysis of a minimal avian leukosis-sarcoma virus packaging signal. *J Virol.* 2000; 74(1):456–64. [PubMed: 10590135]
48. Zhou J, McAllen JK, Tailor Y, Summers MF. High affinity nucleocapsid protein binding to the muPsi RNA packaging signal of Rous sarcoma virus. *J Mol Biol.* 2005; 349(5):976–88. [PubMed: 15907938]
49. De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science.* 1998; 279(5349):384–8. [PubMed: 9430589]
50. Uhlenbeck OC. Tetraloops and RNA folding. *Nature.* 1990; 346(6285):613–4. [PubMed: 1696683]
51. Wedekind JE, McKay DB. Crystallographic structures of the hammerhead ribozyme: relationship to ribozyme folding and catalysis. *Annu Rev Biophys Biomol Struct.* 1998; 27:475–502. [PubMed: 9646875]
52. Doudna JA, Batey RT. Structural insights into the signal recognition particle. *Annu Rev Biochem.* 2004; 73:539–57. [PubMed: 15189152]
53. Jagath JR, Matassova NB, de Leeuw E, Warnecke JM, Lentzen G, Rodnina MV, Lührink J, Wintermeyer W. Important role of the tetraloop region of 4.5S RNA in SRP binding to its receptor FtsY. *RNA.* 2001; 7(2):293–301. [PubMed: 11233986]
54. Fernandez-Miragall O, Martinez-Salas E. Structural organization of a viral IRES depends on the integrity of the GNRA motif. *RNA.* 2003; 9(11):1333–44. [PubMed: 14561883]
55. Jucker FM, Heus HA, Yip PF, Moors EH, Pardi A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol.* 1996; 264(5):968–80. [PubMed: 9000624]
56. Heus HA, Pardi A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science.* 1991; 253(5016):191–4. [PubMed: 1712983]
57. Pley HW, Flaherty KM, McKay DB. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature.* 1994; 372(6501):111–3. [PubMed: 7526219]
58. Jucker FM, Pardi A. GNRA tetraloops make a U-turn. *RNA.* 1995; 1(2):219–22. [PubMed: 7585251]

59. Abramovitz DL, Pyle AM. Remarkable morphological variability of a common RNA folding motif: the GNRA tetraloop-receptor interaction. *J Mol Biol.* 1997; 266(3):493–506. [PubMed: 9067606]
60. Su L, Radek JT, Labeots LA, Hallenga K, Hermanto P, Chen H, Nakagawa S, Zhao M, Kates S, Weiss MA. An RNA enhancer in a phage transcriptional antitermination complex functions as a structural switch. *Genes Dev.* 1997; 11(17):2214–26. [PubMed: 9303537]
61. Pley HW, Flaherty KM, McKay DB. Three-dimensional structure of a hammerhead ribozyme. *Nature.* 1994; 372(6501):68–74. [PubMed: 7969422]
62. Huppler A, Nikstad LJ, Allmann AM, Brow DA, Butcher SE. Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure. *Nat Struct Biol.* 2002; 9(6):431–5. [PubMed: 11992125]
63. Greenblatt J, Nodwell JR, Mason SW. Transcriptional antitermination. *Nature.* 1993; 364(6436):401–6. [PubMed: 8332211]
64. Santangelo TJ, Artsimovitch I. Termination and antitermination: RNA polymerase runs a stop sign. *Nat Rev Microbiol.* 2011; 9(5):319–29. [PubMed: 21478900]
65. Cillely CD, Williamson JR. Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). *RNA.* 1997; 3(1):57–67. [PubMed: 8990399]
66. Tan R, Frankel AD. Structural variety of arginine-rich RNA-binding peptides. *Proc Natl Acad Sci U S A.* 1995; 92(12):5282–6. [PubMed: 7777498]
67. Su L, Radek JT, Hallenga K, Hermanto P, Chan G, Labeots LA, Weiss MA. RNA recognition by a bent alpha-helix regulates transcriptional antitermination in phage lambda. *Biochemistry.* 1997; 36(42):12722–32. [PubMed: 9335528]
68. Chattopadhyay S, Garcia-Mena J, DeVito J, Wolska K, Das A. Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage lambda. *Proc Natl Acad Sci U S A.* 1995; 92(9):4061–5. [PubMed: 7732031]
69. Mogridge J, Mah TF, Greenblatt J. A protein-RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the lambda N protein. *Genes Dev.* 1995; 9(22):2831–45. [PubMed: 7590257]
70. Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD, Williamson JR. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science.* 1996; 273(5281):1547–51. [PubMed: 8703216]
71. Ye X, Gorin A, Ellington AD, Patel DJ. Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nat Struct Biol.* 1996; 3(12):1026–33. [PubMed: 8946856]
72. Puglisi JD, Chen L, Blanchard S, Frankel AD. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science.* 1995; 270(5239):1200–3. [PubMed: 7502045]
73. Ye X, Kumar RA, Patel DJ. Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem Biol.* 1995; 2(12):827–40. [PubMed: 8807816]
74. Xia T, Becker HC, Wan C, Frankel A, Roberts RW, Zewail AH. The RNA-protein complex: direct probing of the interfacial recognition dynamics and its correlation with biological functions. *Proc Natl Acad Sci U S A.* 2003; 100(14):8119–23. [PubMed: 12815093]
75. Masliah G, Barraud P, Allain FH. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol Life Sci.* 2012
76. Ryter JM, Schultz SC. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J.* 1998; 17(24):7505–13. [PubMed: 9857205]
77. Yang SW, Chen HY, Yang J, Machida S, Chua NH, Yuan YA. Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure.* 2010; 18(5):594–605. [PubMed: 20462493]
78. Stefl R, Xu M, Skrisovska L, Emeson RB, Allain FH. Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure.* 2006; 14(2):345–55. [PubMed: 16472753]
79. Blaszczyk J, Gan J, Tropea JE, Court DL, Waugh DS, Ji X. Nucleolytic assembly of ribonuclease III with double-stranded RNA. *Structure.* 2004; 12(3):457–66. [PubMed: 15016361]

80. Hartman E, Wang Z, Zhang Q, Roy K, Chanfreau G, Feigon J. Intrinsic Dynamics of an Extended Hydrophobic Core in the *S. cerevisiae* RNase III dsRBD Contributes to Recognition of Specific RNA Binding Sites. *J Mol Biol.* 2012
81. Chanfreau G, Legrain P, Jacquier A. Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J Mol Biol.* 1998; 284(4):975–88. [PubMed: 9837720]
82. Chanfreau G, Rotondo G, Legrain P, Jacquier A. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J.* 1998; 17(13):3726–37. [PubMed: 9649442]
83. Elela SA, Igel H, Ares M Jr. RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell.* 1996; 85(1):115–24. [PubMed: 8620530]
84. Kufel J, Dichtl B, Tollervey D. Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. *RNA.* 1999; 5(7):909–17. [PubMed: 10411134]
85. Chanfreau G, Buckle M, Jacquier A. Recognition of a conserved class of RNA tetraloops by *Saccharomyces cerevisiae* RNase III. *Proc Natl Acad Sci U S A.* 2000; 97(7):3142–7. [PubMed: 10716739]
86. Nagel R, Ares M Jr. Substrate recognition by a eukaryotic RNase III: the double-stranded RNA-binding domain of Rnt1p selectively binds RNA containing a 5'-AGNN-3' tetraloop. *RNA.* 2000; 6(8):1142–56. [PubMed: 10943893]
87. Chanfreau G. Conservation of RNase III processing pathways and specificity in hemiascomycetes. *Eukaryot Cell.* 2003; 2(5):901–9. [PubMed: 14555472]
88. Ghazal G, Ge D, Gervais-Bird J, Gagnon J, Abou Elela S. Genome-wide prediction and analysis of yeast RNase III-dependent snoRNA processing signals. *Mol Cell Biol.* 2005; 25(8):2981–94. [PubMed: 15798187]
89. Leulliot N, Quevillon-Cheruel S, Graille M, van Tilbeurgh H, Leeper TC, Godin KS, Edwards TE, Sigurdsson ST, Rozenkrants N, Nagel RJ, Ares M, Varani G. A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNase III. *EMBO J.* 2004; 23(13):2468–77. [PubMed: 15192703]
90. Gaudin C, Ghazal G, Yoshizawa S, Elela SA, Fourmy D. Structure of an AAGU tetraloop and its contribution to substrate selection by yeast RNase III. *J Mol Biol.* 2006; 363(2):322–31. [PubMed: 16979185]
91. Lebars I, Lamontagne B, Yoshizawa S, Aboul-Elela S, Fourmy D. Solution structure of conserved AGNN tetraloops: insights into Rnt1p RNA processing. *EMBO J.* 2001; 20(24):7250–8. [PubMed: 11743001]
92. Valegard K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L. Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature.* 1994; 371(6498):623–6. [PubMed: 7523953]
93. Peabody DS. Translational repression by bacteriophage MS2 coat protein expressed from a plasmid. A system for genetic analysis of a protein-RNA interaction. *J Biol Chem.* 1990; 265(10):5684–9. [PubMed: 2108146]
94. Witherell GW, Gott JM, Uhlenbeck OC. Specific interaction between RNA phage coat proteins and RNA. *Prog Nucleic Acid Res Mol Biol.* 1991; 40:185–220. [PubMed: 2031083]
95. Romaniuk PJ, Lowary P, Wu HN, Stormo G, Uhlenbeck OC. RNA binding site of R17 coat protein. *Biochemistry.* 1987; 26(6):1563–8. [PubMed: 3297131]
96. Borer PN, Lin Y, Wang S, Roggenbuck MW, Gott JM, Uhlenbeck OC, Pelczer I. Proton NMR and structural features of a 24-nucleotide RNA hairpin. *Biochemistry.* 1995; 34(19):6488–503. [PubMed: 7756280]
97. Smith JS, Nikonowicz EP. NMR structure and dynamics of an RNA motif common to the spliceosome branch-point helix and the RNA-binding site for phage GA coat protein. *Biochemistry.* 1998; 37(39):13486–98. [PubMed: 9753434]
98. Ni CZ, Syed R, Kodandapani R, Wickersham J, Peabody DS, Ely KR. Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure.* 1995; 3(3):255–63. [PubMed: 7788292]
99. Lago H, Parrott AM, Moss T, Stonehouse NJ, Stockley PG. Probing the kinetics of formation of the bacteriophage MS2 translational operator complex: identification of a protein conformer unable to bind RNA. *J Mol Biol.* 2001; 305(5):1131–44. [PubMed: 11162119]

100. Jennings JC, Olson BH, Roga V, Junek AJ, Schuurmans DM. Alpha Sarcin, a New Antitumor Agent. II. Fermentation and Antitumor Spectrum. *Appl Microbiol.* 1965; 13:322–6. [PubMed: 14330325]
101. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science.* 2000; 289(5481):905–20. [PubMed: 10937989]
102. Antao VP, Lai SY, Tinoco I Jr. A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res.* 1991; 19(21):5901–5. [PubMed: 1719483]
103. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet.* 2008; 9(11):843–54. [PubMed: 18927579]
104. Hoefig KP, Rath N, Heinz GA, Wolf C, Dameris J, Schepers A, Kremmer E, Ansel KM, Heissmeyer V. Eri1 degrades the stem-loop of oligouridylylated histone mRNAs to induce replication-dependent decay. *Nat Struct Mol Biol.* 2013; 20(1):73–81. [PubMed: 23202588]
105. Kupsco JM, Wu MJ, Marzluff WF, Thapar R, Duronio RJ. Genetic and biochemical characterization of *Drosophila* Snipper: A promiscuous member of the metazoan 3' hExo/ERI-1 family of 3' to 5' exonucleases. *RNA.* 2006; 12(12):2103–17. [PubMed: 17135487]
106. Williams AS, Marzluff WF. The sequence of the stem and flanking sequences at the 3' end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. *Nucleic Acids Res.* 1995; 23(4):654–62. [PubMed: 7899087]
107. Battle DJ, Doudna JA. The stem-loop binding protein forms a highly stable and specific complex with the 3' stem-loop of histone mRNAs. *RNA.* 2001; 7(1):123–32. [PubMed: 11214174]
108. Martin L, Meier M, Lyons SM, Sit RV, Marzluff WF, Quake SR, Chang HY. Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nat Methods.* 2012; 9(12):1192–4. [PubMed: 23142872]
109. Martin L, Meier M, Lyons SM, Sit RV, Marzluff WF, Quake SR, Chang HY. Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nat Methods.* 2012
110. Thapar R, Mueller GA, Marzluff WF. The N-terminal domain of the *Drosophila* histone mRNA binding protein, SLBP, is intrinsically disordered with nascent helical structure. *Biochemistry.* 2004; 43(29):9390–400. [PubMed: 15260482]
111. Bansal N, Zhang M, Bhaskar A, Itotia P, Lee E, Shlyakhtenko LS, Lam TT, Fritz A, Berezney R, Lyubchenko YL, Stafford WF, Thapar R. Assembly of the SLIP1-SLBP Complex on Histone mRNA Requires Heterodimerization and Sequential Binding of SLBP Followed by SLIP1. *Biochemistry.* 2013
112. Krishnan N, Lam TT, Fritz A, Rempinski D, O'Loughlin K, Minderman H, Berezney R, Marzluff WF, Thapar R. The Prolyl Isomerase Pin1 Targets Stem-Loop Binding Protein (SLBP) To Dissociate the SLBP-Histone mRNA Complex Linking Histone mRNA Decay with SLBP Ubiquitination. *Mol Cell Biol.* 2012; 32(21):4306–22. [PubMed: 22907757]
113. Zhang M, Lam TT, Tonelli M, Marzluff WF, Thapar R. Interaction of the histone mRNA hairpin with stem-loop binding protein (SLBP) and regulation of the SLBP-RNA complex by phosphorylation and proline isomerization. *Biochemistry.* 2012; 51(15):3215–31. [PubMed: 22439849]
114. Thapar R, Marzluff WF, Redinbo MR. Electrostatic contribution of serine phosphorylation to the *Drosophila* SLBP--histone mRNA complex. *Biochemistry.* 2004; 43(29):9401–12. [PubMed: 15260483]
115. Borchers CH, Thapar R, Petrotchenko EV, Torres MP, Speir JP, Easterling M, Dominski Z, Marzluff WF. Combined top-down and bottom-up proteomics identifies a phosphorylation site in stem-loop-binding proteins that contributes to high-affinity RNA binding. *Proc Natl Acad Sci U S A.* 2006; 103(9):3094–9. [PubMed: 16492733]
116. Kodama Y, Rothman JH, Sugimoto A, Yamamoto M. The stem-loop binding protein CDL-1 is required for chromosome condensation, progression of cell death and morphogenesis in *Caenorhabditis elegans*. *Development.* 2002; 129(1):187–96. [PubMed: 11782412]

FIGURE 1.

Structure of the N-terminal zinc knuckle of the Rous Sarcoma Virus (RSV) Nucleocapsid (NC) protein bound to the UGCG tetraloop of stem-loop SL-C from the RSV M ψ packaging signal (PDB code 2IH X). **(A)** Summary schematic showing the hydrogen bonding and stacking interactions observed in the free UUCG tetraloop of the P1 helix from group I self-splicing introns (PDB code 1HL X) **(B)** Summary schematic showing the hydrogen bonding and stacking interactions observed in the tetraloop when bound to the zinc knuckle (PDB code 2IH X). The sugar puckers and base configurations are color coded according to the key shown below the schematic. **(C)** The zinc knuckle is shown in purple and the RNA is in green. The side-chains of Tyr22 and Tyr30 stack against the guanine nucleotide G218 of the tetraloop. The configurations of the ribose and bases correspond to the color code shown in the summary schematic (1B). **(D)** Surface representation of the RSV zinc knuckle showing the hydrophobic cleft that accommodates G218.

FIGURE 2.

Structures of λ and P22 bacteriophage N-peptides bound to the GNRA-like boxB RNA tetraloops.

First Row: (Panels A, B) The summary schematic showing the hydrogen bonding and stacking interactions observed in the free GAAA tetraloop (PDB code 1ZIF) and free GCAUA GAAA-like tetraloop (PDB code 1LC6) is shown in panels A and B respectively.

Second Row: (Panels C, D, E) Structure of the λ N-peptide-boxB RNA complex (PDB coordinates are undeponed and were obtained from Dr. Legault (University of Montreal, Canada). The peptide is shown in purple and the RNA is in green. In (C), the summary schematic showing the hydrogen bonding and stacking interactions observed in the tetraloop when bound to the peptide is shown. The specific interaction of the bases with protein side-chains is highlighted. In (D), the shape-specific recognition of the bent helix of the λ N-peptide with the major groove of the RNA is shown. The RNA surface is shown in green. The small purple surface on the RNA shows the position of G11 in the tetraloop. The color codes are similar to that of the schematic shown in (C). In (E), the specific interactions between the protein side-chains and the RNA bases is depicted in cartoon format.

Third Row: (Panels F, G, H) Structure of the P22-ARM peptide-boxB RNA complex (PDB code 1A4T). The peptide is shown in purple and the RNA is in green. The N-peptide forms a bent α -helix with extensive ionic interactions with the major groove face of the GNRA-like hairpin. The summary schematic is in **(F)**, the surface representation in **(G)** and the side-chain to base contacts depicted in cartoon format in **(H)**.

FIGURE 3.

Structure of the Rnt1p dsRBD bound to cognate (PDB code 1T4L) and non-cognate (PDB coe 2LBS) (A/U)GNN tetraloops.

(A) Summary schematic showing the hydrogen bonding and stacking interactions observed in the free AGAA tetraloop (PDB code 1K4A) is shown. In (B), the summary schematic for hydrogen bonding and stacking interactions observed in the free AAGU tetraloop (PDB code 2HNS) is shown. (C) Summary schematic showing the hydrogen bonding and stacking interactions observed in the complex formed by the cognate AGAA tetraloop bound to the Rnt1p dsRBD (PDB code 1T4L). The sugar puckers and base configurations are color coded according to the key shown in Fig. 1C. Protein side-chains that contact the base and the sugars are depicted. (D) Overall structure of the Rnt1p dsRBD bound to the AGAA tetraloop. The protein is shown in purple and the RNA surface is shown in green. The protein contacts the successive minor, major, and minor groove via α -helix1, the β 3- α 2 loop, and the β 1- β 2 loop respectively. (E) A close-up of the interaction of α -helix1 with the RNA

tetraloop is shown. **(F)** Summary schematic showing the hydrogen bonding and stacking interactions observed in the non-cognate AAGU tetraloop bound to the Rnt1p dsRBD complex (PDB code 2LBS). Protein side-chains that contact the base and the sugars are depicted. In panels **(G)** and **(H)**, the specific contacts between the protein side-chains of α -helix1 and the RNA bases of the AGAA and AAGU tetraloops are compared. The AGAA tetraloop is in panel **(G)** and the AAGU tetraloop is in panel **(H)**.

FIGURE 4.

Structure of the MS2 coat protein bound to the AUUA tetraloop from its RNA genome. (A) Summary schematic showing the hydrogen bonding and stacking interactions observed in the AUUA tetraloop in the absence of the MS2 coat protein (PDB code 1D0T). (B) Summary schematic showing the hydrogen bonding and stacking interactions observed in the AUUA tetraloop when bound to the MS2 coat protein (PDB code 1ZDH). The sugar puckers and base configurations are color coded according to the key shown in Fig. 1C. Specific interaction of the bases with protein side-chains is depicted. (C) Overall structure of the MS2 dimer bound to the RNA hairpin. The protein is shown in purple cartoon and the RNA is shown in green. (D) Surface representation of the MS2 coat protein showing hydrophobic clefts that accommodate the ⁻⁴A and ⁻¹⁰A adenosines. The RNA is shown in green and the colors of the ribose sugar puckers correlate with the summary schematic in (B). The C3'-endo configuration is shown in blue and the C2'-endo configuration is shown in pink. (E) Close-up of the MS2-RNA binding interface. The protein side-chains are in gray and the RNA ribose and bases are color coded as shown in the key to Fig. 1B.

FIGURE 5.

Structure of restrictocin bound to the GAGA SRL tetraloop.

(A) Summary schematic showing the hydrogen bonding and stacking interactions observed in the free GAGA tetraloop (PDB code 430D). The sugar puckers and base configurations are color coded according to the key shown in Fig. 1B. (B) Summary schematic showing the hydrogen bonding and stacking interactions observed in the GAGA tetraloop when bound to restrictocin (PDB code 1JBR). Protein side-chains that make specific contacts with the tetraloop are depicted. (C) Superposition of the RNA backbone and side-chains of the free GAGA tetraloop (shown in cyan) and the unfolded GAGA tetraloop when bound to the toxin (shown in green). (D) Overall structure of restrictocin bound to the RNA hairpin. The protein is shown in purple cartoon and the RNA is shown in green. The position of the bulged G motif relative to the tetraloop is highlighted. (E) Close-up of the protein-RNA binding interface. The protein side-chains are in gray and the RNA ribose and bases are color coded as shown in (B).

FIGURE 6.

Structure of the SLBP RNA binding domain, ERI1, histone mRNA hairpin ternary complex (PDB code 4HXH).

(A) Summary schematic showing the hydrogen bonding and stacking interactions observed in the solution NMR structure of the free UUUC tetraloops calculated by NMR. The structure of PDB code 1JWC (Ref 9) is summarized on top and the structure of PDB code 1KKS (Ref 10) is summarized below. (B) Summary schematic showing the hydrogen bonding and stacking interactions observed in the UUUC tetraloop when bound to both SLBP and ERI-1. SLBP protein side-chains that make specific contacts with the tetraloop are depicted in orange, and ERI-1 side-chains that make specific contacts with the tetraloop

are depicted in black. **(C)** Superposition of the RNA backbone and side-chains of the free UUUC tetraloop (Ref 9) (shown in cyan) and the unfolded UUUC tetraloop when bound to both SLBP and ERI-1 (shown in green). There is a loss of base stacking of all uridines in the protein bound complex and a re-orientation of U13. **(D)** The sequence of the histone mRNA hairpin is shown. Nucleotides important for binding T171 phosphorylated SLBP (from Refs 10, 107, and 108) are highlighted in red and orange. G7 and U14 are critical for interaction with phosphorylated SLBP and are shown in red. **(E)** Overall structure of the SLBP RNA binding domain (in orange), the histone mRNA stem-loop (in green) and ERI-1 (in purple) ternary complex (PDB code 4HXXH). Helices in SLBP and the SAP domain are designated. **(F)** Close-up of the histone mRNA tetraloop (in green) bound to the SLBP RNA binding domain (in orange) and ERI-1 (in purple) in the ternary complex. Side-chains in SLBP and the ERI-1 SAP domain that contact the RNA are designated. **(G)** The structure of the ternary complex is color coded to show the variation in the Debye-Waller factor (or B-factor) in the protein and RNA components of the complex. Very high B-factors (shown in red) indicate regions of high mobility or where the electron density is more spread out. Regions with low B-factors are shown in blue. Helix B or SLBP has the highest B-factors in the ternary complex. **(H)** Crystal contacts observed in the structure of the ternary complex are shown. The crystal shows extensive packing interactions between the nuclease domain of ERI-1 from a binary ERI-1/RNA complex (in yellow) that lacks SLBP, and the nuclease domain of a neighbouring ERI-1 molecule (in purple) from the ternary complex. The SLBP RNA binding domain is shown in orange. Helix-B of SLBP contacts the SAP domain of ERI-1 from a neighbouring binary complex. The RNA is shown in green (ternary complex) or cyan (binary complex).