Published in final edited form as:

Curr Opin Struct Biol. 2008 February ; 18(1): 120–129. doi:10.1016/j.sbi.2007.12.006.

Towards deciphering the principles underlying a mRNA recognition code

Alexander Serganov and Dinshaw J. Patel

Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY, 10021, USA

Abstract

Messenger RNAs interact with a number of different molecules that determine the fate of each transcript and contribute to the overall pattern of gene expression. These interactions are governed by specific mRNA signals, which in principle could represent a special mRNA recognition ‘code’. Both, small molecules and proteins demonstrate a diversity of mRNA binding modes often dependent on the structural context of the regions surrounding specific target sequences. In this review, we have highlighted recent structural studies that illustrate the diversity of recognition principles used by mRNA binders for timely and specific targeting and processing of the message.

Introduction

Transcription does not simply transfer coding information required for protein biosynthesis from DNA to mRNA. In fact, transcription produces pre-mRNA and mRNA molecules, which carry multiple signals required for processing, modification, transport, translation and degradation of the message. These signals are recognized by mRNA-binding molecules in both sequence-specific and structure-dependent manner, and help define the spatial and temporal constraints for translation of mRNA species. The mRNA recognition signatures, therefore, could be considered a special ‘code’, contributing, along with other layers of gene expression control, to the final pattern of gene expression. This code, however, is unlikely to be universal due to dramatic differences in transcription and mRNA processing amongst evolutionary distant groups, as well as occurrence of species-specific mRNA-recognition systems necessary for adaptation to particular environmental cues [1]. Due to the immense potential and opportunities for manipulation of gene expression at the post-transcriptional level, many structural biology groups have focused their ongoing research efforts towards determination of structures that would uncover the complex network of relationships between mRNA and its partners, thereby contributing towards a comprehensive understanding of the principles underlying a ‘mRNA recognition code’.

Although the last decade of intensive research has provided us with molecular details of many interesting intermolecular interactions involving mRNA, the last two years have been especially informative, with over thirty structures reported of complexes containing mRNA. This review analyzes recently published structural data (spanning 2006–2007) on specific

© 2008 Elsevier Ltd. All rights reserved.

Corresponding authors: serganoa@mskcc.org; pateld@mskcc.org.

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

mRNA recognition events, and complements excellent earlier reviews on protein-RNA [2–5] and metabolite-mRNA [6–8] recognition.

The distinct modes of mRNA recognition

For the purpose of this review, we have considered protein data bank (PDB) entries that describe interactions of mRNA fragments or their mimetics with either small molecules or proteins. We propose dividing all such complexes into three categories (Table 1): (1) structure-specific recognition of folded RNAs (Figure 1a); (2) sequence-specific recognition of single-stranded RNAs (Figure 1b); (3) non-specific recognition of single-stranded RNAs (Figure 1c).

Most complexes belong to the first group and are often characterized by unique structure-specific aspects of mRNA recognition. About half of these complexes are comprised of long (70–150 nt) sensing domains of riboswitch mRNAs bound to their ligands (Figure 1a, left panel). The other half contains protein domain(s) bound to shorter RNAs, which typically adopt stem-loop scaffolds (Figure 1a, middle and right panels). The second group features protein domain(s) that interact with sequence-specificity to short single-stranded mRNA fragments (Figure 1b). The third group, not discussed in the current review, mostly includes proteins and protein assemblies capable of binding various RNA species in a non-sequence specific manner (Figure 1c).

Interactions of small molecules with higher-order mRNA structures

Ribosensors are mRNA sequences that control gene expression in response to various stimuli, such as metabolites (riboswitches), cations (metalloensors), and temperature (thermosensors). Recently, the seminal structure determination of purine-sensing riboswitches [9,10] has been rapidly followed by structures of five more ribosensors: bacterial [11,12] and plant [13] thiamine pyrophosphate (TPP) riboswitches, a S-adenosylmethionine type I (SAM-I) riboswitch [14], a glucosamine-6-phosphate-sensing *glmS* ribozyme [15,16], a thermosensor [17] and a Mg^{2+} ribosensor [18]. Here, we focus on the recent structure of the Mg^{2+} ribosensor, since it was not discussed in an earlier review [6].

Mg^{2+} , the most abundant divalent cation, is critically required for both structure and function of many RNAs, including mRNAs. Therefore, it was surprising to find that Mg^{2+} homeostasis in *Salmonella enterica* and in *Bacillus subtilis* are controlled directly by two different RNA sensors located in the Mg^{2+} transporter MgtA and MgtE mRNAs, respectively [18,19]. Under high Mg^{2+} concentration, both sensors adapt one of two alternative conformations that arrest transcriptional elongation, using an as yet uncharacterized mechanism for the *mgtA* switch, and by a transcription attenuation mechanism for the *mgtE* switch (Fig. 2a–b).

The X-ray structure of the Mg^{2+} -sensing domain (M-box) of *mgtE* switch provides, for the first time, insights into the molecular details of metalloregulation by an RNA sensor [18]. The structure comprises of close packing of the P1-P2-P6 helix against the stem-loop structures P5/L5 and P4-P3/L4 (Fig. 2c–d), that are oriented downward and stapled by extensive tertiary contacts between the helix P2 and J2-1 junction with loops L4 and L5. Four (Mg^{1-4}) of the six Mg^{2+} cations reside in this region, most likely comprising the cations that are key for metal sensing (Fig. 2e): Mg^1 organizes the L5 structure for docking with P2 and L4, Mg^2 bridges L5 and P2, and Mg^3 and Mg^4 stabilize the local conformations of P2 and L4. Similar to Mg^{2+} cations that mediate ligand binding in the TPP riboswitch and the *glmS* ribozyme [12,15], Mg^{2+} cations in the M-box structure predominantly utilize the outer-sphere coordination for interactions with nucleobases. Like

the TPP riboswitch [12], only two inner-sphere contacts with nucleobases are formed, while additional direct contacts are made with non-bridging phosphate oxygens. Since Mg¹ uses four inner-sphere (one nucleobase and three sugar-phosphate backbone) contacts with RNA (Fig. 2e), a feature rarely observed in previous studies, this cation may provide a key contribution to the docking of the J2-1/P2 region with the L4 and L5 loops. These long-distance interactions facilitate the formation of tertiary base contacts and base stacking, which in turn sequester antiterminator nucleotides and, along with Watson-Crick base pairing, contribute to the stabilization of helix P1, leading to the formation of the terminator hairpin and the repression of gene expression (Fig. 2b). By contrast, in purine, TPP and SAM-I riboswitches, formation of the P1 helix is dependent on stabilization of the adjacent junction by bound ligand (Fig. 2a) [9,11–14].

Recognition of three-dimensional mRNA structures by proteins

Similar to riboswitches, some bacterial proteins inhibit translation by interactions with the mRNA region located adjacent to the ribosome-binding site (RBS), thereby preventing ribosome loading onto mRNA. The classical example of such a protein versus ribosome competition mechanism is highlighted for the autoregulation of ribosomal protein synthesis. If produced in excess over their rRNA targets, primary ribosomal proteins, such as L1, interact with their own mRNAs, repressing ribosome binding. This implies a preferential binding to rRNA and overlap of binding sites for rRNA and mRNA on L1, suggesting similarities between both RNA targets.

The structures of ribosomal protein L1 bound to mRNA and 23S rRNA fragments demonstrated that both RNAs indeed have a common structural determinant (upper panel, Fig. 3a) [20–22]. This binding region in mRNA is built by an asymmetrical internal loop that creates a sharp bend between two helices, thereby resembling the kink-turn motif [3]. The RNA core includes the primary recognition determinant, a G-C pair and its neighboring uridine, that is specifically recognized in both complexes by invariant E42, T217, M218 and G219 of domain 1 (lower panel in Figure 3a). To increase binding affinity, rRNA additionally interacts with domain 2 of L1.

Ribosomal protein S15 represents an interesting deviation from the general trend outlined above. In thermophilic bacteria, mRNA and rRNA targets of S15 [23,24] are similar, and the protein represses translation by competition with the ribosome [24]. In *Escherichia coli*, S15 recognizes a pseudoknot structure folded within its own mRNA [25–27], rather than the three-way junction architecture associated with thermophilic bacteria. Although the ribosome can interact with mRNA already bound to S15, it cannot initiate translation in *E. coli* [28]. A long-awaited cryo-electron microscopy study [29] showed that in the stalled ribosome, S15 positions itself along the mRNA pseudoknot, and the S15-mRNA complex is nested on a special platform of the small ribosomal subunit (Figure 3b). This precise positioning allows Shine-Dalgarno-rRNA interactions, but precludes the initiator tRNA from reaching the start codon. Therefore, S15 performs its repressor function by preventing the mRNA pseudoknot from unfolding and entering the ribosome, thereby trapping the ribosome in a translation-incompetent state.

The iron regulatory protein 1 (IRP-1) also has a dual function [30]. The protein either binds iron-responsive elements (IREs) in mRNA to repress translation or degradation, or it binds an iron-sulfur cluster to become a cytosolic aconitase, catalyzing the conversion of citrate to isocitrate. In order to accommodate the stem-loop IRE, domains 3 and 4 are splayed apart in the aconitase bound state [31] and their contacting surfaces are incorporated into two distinct and separated RNA-binding sites (upper panel in Figure 3c). IRE contacts IRP-1 using its lower stem and the terminal loop, which contains a conserved CAGUG motif. The specific

recognition of the loop is accomplished by base-specific bonding of S371, K379 and R260 with a 5'-A15-G16-U17 pseudo-triloop, and is strengthened by van der Waal's contacts with the exposed purines (lower right panel in Figure 3c). The other recognition determinant, the conserved bulged C8, is sandwiched between two arginines within a small pocket and is involved in base-specific hydrogen bonds with the side chain of S681 and backbone of P682, D781, and W782 (lower left panel in Figure 3c). The availability of two separated RNA-binding sites, which recognize the loop and bulged cytosine, greatly increase the selectivity of IRE recognition by IRP-1, thereby resembling the two-point recognition reported previously in tRNA-aminoacyl-tRNA synthetase and some RNA-ribosomal protein complexes [32,33].

Specific recognition of RNA loops by proteins

Loop-specific recognition is utilized by several other proteins of the first group for the readout of certain nucleotide sequences (Table 1). Though some proteins contact the helical RNA segments that close RNA loop regions, the majority of specific interactions are observed with non-paired nucleotides within the loops. These RNA-protein interactions resemble sequence-specific recognition patterns observed in complexes between proteins and single-stranded RNA (discussed below), which often involve small canonical RNA-binding domains, such as zinc-finger domains, the K-homology (KH) domain and the RNA recognition motif (RRM) [2]. Not surprisingly, some proteins described here utilize canonical RNA-binding domains and motifs for RNA binding.

The structure of the translational repressor RsmE bound to the Shine-Dalgarno sequence of *hcnA* mRNA shows how a protein dimer specifically recognizes the consensus sequence 5'-A_UCANGGANG^{U/A} (Figure 3d) [34]. The loop contains six unpaired nucleotides A8-C9-G10-G11-A12-U13, with U13 and the C9-G10-G11 segment bulged out. The protein specifically recognizes the Watson-Crick edges of A8, G10, G11, the Hoogsteen edge of A12, and the major-groove side of C7-G14 and U6-A15 base pairs. In contrast to small canonical RNA-binding domains, the sequence-specific recognition of unpaired nucleotides is mediated primarily by β -strand backbone residues, implying that the protein fold itself is responsible for RNA-binding specificity.

The sequence-specific recognition of the bulged out nucleotides in apical loops is a recurrent theme in complexes of aptamer RNAs with the KH1 domain of NOVA-1 KH1/2 protein (PDB code: 2ANR) and the RNA recognition motif (RRM) domain of human RBMY protein [35]. The NOVA (neuro-oncological ventral antigen) family of proteins is expressed in neurons where it plays a crucial role in the regulation of alternative splicing [36]. The NOVA-1 protein contains three KH domains. An earlier structure [37] has revealed details of the recognition between the KH3 domain and a UCAC tetranucleotide embedded within the hairpin loop of an *in vitro*-selected stem-loop RNA scaffold. However, it has not addressed the question how multiple KH domains can target RNA. The structure of the first two KH domains (KH1/2) bound to tandem UCAN repeats of an *in vitro*-selected stem-loop RNA, attempted to answer this question. These structural efforts revealed that the KH2 domain does not participate in RNA binding and only the KH1 domain interacts with a 5'-UCAG-UCAC-C loop closed by three non-canonical base pairs. This domain primarily binds to the second UCAN repeat in the cleft usually used by KH-domains for ss-DNA and ss-RNA recognition. Despite the Watson-Crick edges of all four nucleotides interacting with the protein, only cytosine and adenine form sequence-specific hydrogen bonds, thereby validating the YCAN sequence consensus found using the SELEX approach [36].

Testes-specific RBMY (RNA-binding motif gene on Y chromosome) protein encoded by the human Y chromosome is important for sperm development. The protein is possibly involved

in pre-mRNA processing and recognizes an *in vitro*-selected RNA hairpin with a 5'-CA^A/UCAA loop and a 5'-GUC-loop-GAY consensus element in the loop-closing part of the stem [35]. In the structure, CAA nucleotides protrude from the CACAA pentaloop and are spread on the β -sheet surface of the RRM, similar to other proteins that utilize the RRM-RNA mode of recognition. All three nucleotides form base-specific contacts with main and side chain atoms; however, only adenines provide base-specific discrimination. Unexpectedly, the protein makes additional contacts with a major groove of the stem using its β -hairpin, thereby demonstrating dual sequence and shape-specific RNA-recognition, a duality that is generally unusual for RRM motifs.

Two proteins, elongation factor SelB and mRNA-binding factor Vts1p, interact with stem-loop structures, whose loop regions, though composed of different sequences, demonstrate conformational similarity to the UNCG tetraloop fold [38]. SelB is essential for incorporation of selenocysteine, the 21st amino acid, into bacterial polypeptides. The factor binds selenocysteine insertion sequence (SECIS) in mRNA with extremely high selectivity, and this binding serves as a signal for delivery of selenocysteyl-tRNA at a UGA stop codon upstream of SECIS hairpin. The high binding specificity is achieved through base-specific interactions of a DNA- and RNA-binding winged-helix (WH) motif with consecutive bulged out guanine and unpaired uridine of the 5'-GGUC-U loop, and interactions with the RNA backbone, which are determined by shape complementarity and electrostatic properties of the protein surface [39,40].

Yeast Vts1p has been implicated in vesicular transport and sporulation; however, its precise role remains unknown. The protein is a homolog of the *Drosophila* protein Smaug, a translational repressor that mediates body patterning during embryogenesis by binding to a mRNA hairpin termed Smaug recognition element (SRE) [41]. The SRE hairpin exhibits consensus sequences 5'-UNGA-N and 5'-GNGC-N which are targeted by α -helical sterile alpha motif (SAM) domain of Vts1p, a domain also implicated in protein-protein and DNA-protein interactions [42]. Three structures of the Vts1p-SAM domain bound to two SRE variants show parallels with SelB-SECIS recognition, such as shape recognition of the loop region and base-specific binding to an unpaired nucleotide, guanosine in this case [42–44]. In contrast to the SelB-SECIS complex, the bulged out nucleotide does not play a significant role in recognition by Vts1p-SAM.

Specific binding of single-stranded mRNA

The majority of the RNA-protein complexes from the second group contain canonical RNA-recognition modules. Nevertheless, the structures of these complexes show interesting details and deviations from typical RNA-recognition modes. These structures illustrate the high complexity of mRNA recognition and significantly expand our knowledge of the code underlying mRNA recognition. Since the RRM domain is the most common RNA-binding motif and is typically used for recognition of specific sequences, it is not surprising that amongst the seven complexes assigned to the second group, only two, RNase Kid and the KH domain of poly(C)-binding protein-2 (PCBP-2), do not contain this motif (Table 1).

The Fox-1 protein regulates tissue-specific alternative splicing by binding to a 5'-GCAUG RNA element. Like the above-mentioned RNA complex of the RBMY protein, the structure of the RRM domain of Fox-1 in complex with U1-G2-C3-A4-U5-G6-U7 demonstrates both canonical and unique modes of RNA recognition (Figure 4a) [45]. The U5-G6-U7 segment is bound in a canonical way by the β -sheet of the protein. These interactions feature typical hydrophobic interactions of H120, F158 and F160 with U5 and G6 nucleotides. However, the binding platform is extended to the β 1 α 1-loop of the RRM motif, where in an unprecedented manner, F126 is caged by the U1-G2-C3 segment. Since aromatic residues at

equivalent positions are predicted in other RRM, this RNA recognition feature may be shared with additional proteins. Binding specificity is provided by a dense network of hydrogen bonds to the bases of the first six nucleotides, while high binding affinity is achieved by numerous electrostatic and hydrophobic interactions. Several intramolecular hydrogen bonds additionally stabilize the RNA conformation, in contrast to its disordered topology in the unbound state.

The structure of another key regulator of alternative splicing, the SRp20 protein, bound to a 5'-CAUC sequence further expands the RNA-binding characteristics of the RRM motif [46]. As anticipated from the consensus sequence 5'-A/UCA/U^A/UC, SRp20-RRM binds RNA in a semi-sequence-specific mode. Although all bases participate in hydrogen bonding with RRM, only the invariant first cytosine is recognized sequence-specifically. In addition to unspecific binding, the AUC segment adopts an unusual RNA topology, possibly preserved for recognition by related proteins.

Two other RRM-containing complexes illustrate another distinctive mode of RRM-RNA binding, namely interactions of RNA with tandem RRMs. In the structure of pre-mRNA splicing factor U2AF⁶⁵ bound to a U₇ strand, the protein utilizes a unique pattern of hydrogen bonds with uracil bases, spread over the surfaces of both RRMs [47]. These hydrogen bonds are frequently formed with protein side chains, which may be rearranged upon RNA binding to accommodate other polypyrimidine sequences. The structure of polyadenylation factor Hrp1 in complex with polyadenylation enhancement element 5'-GUAUAUAUA reveals a sequence-specific mode of recognition of the AUAUAU motif by both main and side chains of RRM2, RRM1 and the linker region. Interestingly, the β1α1-loop of RRM1 contains a tryptophan that is important for RNA binding, and which occupies a position equivalent to F126 in the Fox-1-RNA complex [45]. Related β1α1-loops in Sex-lethal-RNA and HuD-RNA complexes contain a tyrosine at a nearby position in the loop [48,49]. This observation further reinforces the suggestion that an aromatic amino acid outside of the RRM β-sheet can be a strong determinant of RNA recognition.

The structure of the N-terminal part of the multifunctional La protein in complex with 5'-U1-G2-C3-U4-G5-U6-U7-U8-U9 RNA has revealed an unexpected mode of RNA recognition [50]. La protein contains RNA-binding RRM and La domains and interacts with certain pyrimidine-rich mRNAs, as well as with small RNA precursors, which typically bear a UUU_{OH} sequence at their 3' termini. Unexpectedly, recognition of the 3'-terminal U7-U8-U9 segment occurred within a cleft between the La and RRM1 domains, involving contacts with one edge of a canonical β-sheet of the RRM domain and the backside of the winged-helix motif of the La domain (Figure 4b). The majority of the interactions involve conserved aromatic amino acids of the La domain with the U7-U8-U9 segment that adopts a reversed-turn stabilized by stacking of non-adjacent U7 and U9 residues. Hydrogen bonds to the U8 base determine sequence specificity, while interactions of an aspartate carboxylate with hydroxyls of the U9 sugar discriminate against phosphorylated 3'-ends. Since the canonical sites of La and RRM domains are not occupied, the protein could be involved in other RNA interactions.

Conclusions

Our understanding of RNA-recognition principles, which were largely based in previous years on the structures of protein-RNA complexes involved in translation or viral biogenesis, are currently being enhanced by structural information from complexes involving mRNA and other RNA species. Significant progress has been made in our understanding of the mechanism of action of metabolite-sensing mRNAs, based on the crystal structures of several important riboswitches in the metabolite-bound state. These

structures along with the structures of protein repressor-mRNA complexes have stimulated functional studies aimed towards elucidation of the corresponding mechanisms of gene expression control. The ongoing structural studies illustrate the diversity and complexity of mRNA recognition and significantly expand our current understanding of the principles underlying a mRNA recognition code. The increasing sophistication and technical advances in X-ray, NMR and cryo-electron microscopy techniques should undoubtedly lead to the solution of more challenging problems, likely aimed at large multiprotein complexes containing longer mRNA fragments.

Acknowledgments

This work was supported by National Institutes of Health grant GM073618.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest

•• of outstanding interest

1. Serganov A, Patel DJ. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet.* 2007; 8:776–790. [PubMed: 17846637]
2. Auweter SD, Oberstrass FC, Allain FH. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* 2006; 34:4943–4959. [PubMed: 16982642]
3. Batey RT. Structures of regulatory elements in mRNAs. *Curr Opin Struct Biol.* 2006; 16:299–306. [PubMed: 16707260]
4. Lunde BM, Moore C, Varani G. RNA-binding protein modular design for efficient function. *Nat Rev Mol Cell Biol.* 2007; 8:479–490. [PubMed: 17473849]
5. Stefl R, Skrisovska L, Allain F. RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep.* 2005; 6:33–38. [PubMed: 15643449]
6. Edwards TE, Klein DJ, Ferre-D'Amare AR. Riboswitches: small-molecule recognition by gene regulatory RNAs. *Curr Opin Struct Biol.* 2007; 17:273–279. [PubMed: 17574837]
7. Wakeman CA, Winkler WC, Dann CE 3rd. Structural features of metabolite-sensing riboswitches. *Trends Biochem Sci.* 2007; 32:415–424. [PubMed: 17764952]
8. Schwalbe H, Buck J, Furtig B, Noeske J, Wohnert J. Structures of RNA switches: insight into molecular recognition and tertiary structure. *Angew Chem Int Ed Engl.* 2007; 46:1212–1219. [PubMed: 17226886]
9. Batey RT, Gilbert SD, Montange RK. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature.* 2004; 432:411–415. [PubMed: 15549109]
10. Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol.* 2004; 11:1729–1741. [PubMed: 15610857]
11. Edwards TE, Ferre-D'Amare AR. Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure.* 2006; 14:1459–1468. [PubMed: 16962976]
12. Serganov A, Polonskaia A, Phan AT, Breaker RR, Patel DJ. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature.* 2006; 441:1167–1171. [PubMed: 16728979]
13. Thore S, Leibundgut M, Ban N. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science.* 2006; 312:1208–1211. [PubMed: 16675665]
14. Montange RK, Batey RT. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature.* 2006; 441:1172–1175. [PubMed: 16810258]

15. Cochrane JC, Lipchick SV, Strobel SA. Structural investigation of the GlmS ribozyme bound to its catalytic cofactor. *Chem Biol.* 2007; 14:97–105. [PubMed: 17196404]
16. Klein DJ, Ferre-D'Amare AR. Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science.* 2006; 313:1752–1756. [PubMed: 16990543]
17. Chowdhury S, Maris C, Allain FH, Narberhaus F. Molecular basis for temperature sensing by an RNA thermometer. *EMBO J.* 2006; 25:2487–2497. [PubMed: 16710302]
- 18•. Dann CE 3rd, Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC. Structure and mechanism of a metal-sensing regulatory RNA. *Cell.* 2007; 130:878–892. The first structure of the natural magnesium-sensing RNA switch. [PubMed: 17803910]
19. Cromie MJ, Shi Y, Latifi T, Groisman EA. An RNA sensor for intracellular Mg^{2+} . *Cell.* 2006; 125:1–84.
- 20•. Nevskaya N, Tishchenko S, Gabdoulkhakov A, Nikonova E, Nikonov O, Nikulin A, Platonova O, Garber M, Nikonov S, Piendl W. Ribosomal protein L1 recognizes the same specific structural motif in its target sites on the autoregulatory mRNA and 23S rRNA. *Nucleic Acids Res.* 2005; 33:478–485. Along with [21•], this paper describes structural similarities and differences between interactions of L1 protein with mRNA and rRNA. [PubMed: 15659579]
- 21•. Nevskaya N, Tishchenko S, Volchkov S, Kljashtorny V, Nikonova E, Nikonov O, Nikulin A, Kohrer C, Piendl W, Zimmermann R, et al. New insights into the interaction of ribosomal protein L1 with RNA. *J Mol Biol.* 2006; 355:747–759. See annotation to [20•]. [PubMed: 16330048]
22. Nikulin A, Eliseikina I, Tishchenko S, Nevskaya N, Davydova N, Platonova O, Piendl W, Selmer M, Liljas A, Drygin D, et al. Structure of the L1 protuberance in the ribosome. *Nat Struct Biol.* 2003; 10:104–108. [PubMed: 12514741]
23. Scott LG, Williamson JR. Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with its 5'-translational operator mRNA. *J Mol Biol.* 2001; 314:413–422. [PubMed: 11846555]
24. Serganov A, Polonskaia A, Ehresmann B, Ehresmann C, Patel DJ. Ribosomal protein S15 represses its own translation via adaptation of an rRNA-like fold within its mRNA. *EMBO J.* 2003; 22:1898–1908. [PubMed: 12682022]
25. Philippe C, Benard L, Portier C, Westhof E, Ehresmann B, Ehresmann C. Molecular dissection of the pseudoknot governing the translational regulation of *Escherichia coli* ribosomal protein S15. *Nucleic Acids Res.* 1995; 23:18–28. [PubMed: 7532857]
26. Mathy N, Pellegrini O, Serganov A, Patel DJ, Ehresmann C, Portier C. Specific recognition of rpsO mRNA and 16S rRNA by *Escherichia coli* ribosomal protein S15 relies on both mimicry and site differentiation. *Mol Microbiol.* 2004; 52:661–675. [PubMed: 15101974]
27. Serganov A, Ennifar E, Portier C, Ehresmann B, Ehresmann C. Do mRNA and rRNA binding sites of *E. coli* ribosomal protein S15 share common structural determinants? *J Mol Biol.* 2002; 320:963–978. [PubMed: 12126618]
28. Philippe C, Eyermann F, Benard L, Portier C, Ehresmann B, Ehresmann C. Ribosomal protein S15 from *Escherichia coli* modulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proc Natl Acad Sci U S A.* 1993; 90:4394–4398. [PubMed: 7685101]
- 29•. Marzi S, Myasnikov AG, Serganov A, Ehresmann C, Romby P, Yusupov M, Klaholz BP. Structured mRNAs regulate translation initiation by binding to the platform of the ribosome. *Cell.* 2007; 130:1019–1031. Cryo-EM structure of the S15-mRNA complex bound to the ribosome, explaining entrapment mechanism of translational control. [PubMed: 17889647]
- 30•. Walden WE, Selezneva AI, Dupuy J, Volbeda A, Fontecilla-Camps JC, Theil EC, Volz K. Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science.* 2006; 314:1903–1908. This structure uncovers details of the specific IRP-1-IRE recognition and demonstrates large conformational changes in the protein structure upon RNA binding. [PubMed: 17185597]
31. Dupuy J, Volbeda A, Carpentier P, Darnault C, Moulis JM, Fontecilla-Camps JC. Crystal structure of human iron regulatory protein 1 as cytosolic aconitase. *Structure.* 2006; 14:129–139. [PubMed: 16407072]
32. 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; 289:905–920. [PubMed: 10937989]

33. Beuning PJ, Musier-Forsyth K. Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers*. 1999; 52:1–28. [PubMed: 10737860]
34. Schubert M, Lapouge K, Duss O, Oberstrass FC, Jelesarov I, Haas D, Allain FH. Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nat Struct Mol Biol*. 2007; 14:807–813. This study describes unusual binding of the protein dimer with Shine-Dalgarno sequence. [PubMed: 17704818]
35. Skrisovska L, Bourgeois CF, Stefl R, Grellscheid SN, Kister L, Wenter P, Elliott DJ, Stevenin J, Allain FH. The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep*. 2007; 8:372–379. In this structure, sequence-specific RNA-recognition by RRM is complemented by shape-specific binding of β -hairpin to the major groove of RNA. [PubMed: 17318228]
36. Musunuru K, Darnell RB. Determination and augmentation of RNA sequence specificity of the Nova K-homology domains. *Nucleic Acids Res*. 2004; 32:4852–4861. [PubMed: 15367696]
37. Lewis HA, Musunuru K, Jensen KB, Edo C, Chen H, Darnell RB, Burley SK. Sequence-specific RNA binding by a Nova KH domain. implications for paraneoplastic disease and the fragile X syndrome. *Cell*. 2000; 100:323–332. [PubMed: 10676814]
38. 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; 304:35–42. [PubMed: 11071808]
39. Ose T, Soler N, Rasubala L, Kuroki K, Kohda D, Fourmy D, Yoshizawa S, Maenaka K. Structural basis for dynamic interdomain movement and RNA recognition of the selenocysteine-specific elongation factor SelB. *Structure*. 2007; 15:577–586. [PubMed: 17502103]
40. Yoshizawa S, Rasubala L, Ose T, Kohda D, Fourmy D, Maenaka K. Structural basis for mRNA recognition by elongation factor SelB. *Nat Struct Mol Biol*. 2005; 12:198–203. [PubMed: 15665870]
41. Smibert CA, Lie YS, Shillinglaw W, Henzel WJ, Macdonald PM. Smaug, a novel and conserved protein, contributes to repression of nanos mRNA translation in vitro. *RNA*. 1999; 5:1535–1547. [PubMed: 10606265]
42. Aviv T, Lin Z, Ben-Ari G, Smibert CA, Sicheri F. Sequence-specific recognition of RNA hairpins by the SAM domain of Vts1p. *Nat Struct Mol Biol*. 2006; 13:168–176. This paper along with [43–44*], shows SAM domain recognition of the UCG tetraloop-like loop. [PubMed: 16429151]
43. Johnson PE, Donaldson LW. RNA recognition by the Vts1p SAM domain. *Nat Struct Mol Biol*. 2006; 13:177–178. See annotation to [42*]. [PubMed: 16429155]
44. Oberstrass FC, Lee A, Stefl R, Janis M, Chanfreau G, Allain FH. Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. *Nat Struct Mol Biol*. 2006; 13:160–167. See annotation to [42*]. [PubMed: 16429156]
45. Auweter SD, Fasan R, Reymond L, Underwood JG, Black DL, Pitsch S, Allain FH. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J*. 2006; 25:163–173. This study has suggested a novel determinant of RNA recognition adjacent to canonical surface of RRM domain. [PubMed: 16362037]
46. Hargous Y, Hautbergue GM, Tintaru AM, Skrisovska L, Golovanov AP, Stevenin J, Lian LY, Wilson SA, Allain FH. Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J*. 2006; 25:5126–5137. A demonstration of a semi-sequence-specific mode of the RRM-RNA binding. [PubMed: 17036044]
47. Sickmier EA, Frato KE, Shen H, Paranawithana SR, Green MR, Kielkopf CL. Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol Cell*. 2006; 23:49–59. This paper provides structural insights into recognition of semi-conserved polypyrimidine sequences. [PubMed: 16818232]
48. Handa N, Nureki O, Kurimoto K, Kim I, Sakamoto H, Shimura Y, Muto Y, Yokoyama S. Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature*. 1999; 398:579–585. [PubMed: 10217141]
49. Wang X, Hall TM. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat Struct Biol*. 2001; 8:141–145. [PubMed: 11175903]

- 50•. Teplova M, Yuan YR, Phan AT, Malinina L, Ilin S, Teplov A, Patel DJ. Structural basis for recognition and sequestration of UUU_{OH} 3' termini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Mol Cell*. 2006; 21:75–85. This structure has revealed unexpected recognition of UUU_{OH} sequence by La domain. [PubMed: 16387655]
- 51•. Perez-Canadillas JM. Grabbing the message. structural basis of mRNA 3' UTR recognition by Hrp1. *EMBO J*. 2006; 25:3167–3178. This work rationalizes specific binding of AUAUUAU sequence by RRM and further validates importance of the non-canonical RRM determinant for RNA binding. [PubMed: 16794580]
52. Kamphuis MB, Bonvin AM, Monti MC, Lemonnier M, Munoz-Gomez A, van den Heuvel RH, Diaz-Orejas R, Boelens R. Model for RNA binding and the catalytic site of the RNase Kid of the bacterial parD toxin-antitoxin system. *J Mol Biol*. 2006; 357:115–126. [PubMed: 16413033]
53. Du Z, Lee JK, Fenn S, Tjhen R, Stroud RM, James TL. X-ray crystallographic and NMR studies of protein-protein and protein-nucleic acid interactions involving the KH domains from human poly(C)-binding protein-2. *RNA*. 2007; 13:1043–1051. [PubMed: 17526645]
54. Lorentzen E, Dziembowski A, Lindner D, Seraphin B, Conti E. RNA channelling by the archaeal exosome. *EMBO Rep*. 2007; 8:470–476. [PubMed: 17380186]
55. Skordalakes E, Berger JM. Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor. *Cell*. 2006; 127:553–564. [PubMed: 17081977]
56. Frazao C, McVey CE, Amblar M, Barbas A, Vornrhein C, Arraiano CM, Carrondo MA. Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature*. 2006; 443:110–114. [PubMed: 16957732]
57. Andersen CB, Ballut L, Johansen JS, Chamieh H, Nielsen KH, Oliveira CL, Pedersen JS, Seraphin B, Le Hir H, Andersen GR. Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science*. 2006; 313:1968–1972. [PubMed: 16931718]
58. Bono F, Ebert J, Lorentzen E, Conti E. The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell*. 2006; 126:713–725. [PubMed: 16923391]
59. Sengoku T, Nureki O, Nakamura A, Kobayashi S, Yokoyama S. Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell*. 2006; 125:287–300. [PubMed: 16630817]
60. Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X. Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell*. 2006; 124:355–366. [PubMed: 16439209]

Figure 1.

Proposed division of various mRNA-recognition modes based on the structure-and sequence-specificity. Sequence-specific and non-specific interactions are depicted by S and N letters. **(a)**, Examples of structure-based mRNA recognition by small molecules (M) and proteins (group 1 binders). mRNA binders may recognize specific structures (left and center), or particular sequences, presented by structural modules (right). **(b)**, Sequence-specific binding of single-stranded mRNA (group 2 binders). **(c)**, Non-specific binding of single-stranded mRNA (group 3 binders).

Figure 2.

Gene expression control by RNA sensors. **(a)**, Transcriptional attenuation in the TPP riboswitch. In the absence of TPP, mRNA forms an anti-terminator hairpin and transcription proceeds through the open reading frame (ORF). In the presence of TPP, the sensor domain binds the ligand and stabilizes the three-way junction and helix P1, causing the formation of a transcription terminator. **(b)**, The M-box Mg^{2+} sensor functions similar to the TPP riboswitch; however, P1 stabilization is achieved by formation of Mg^{2+} -mediated loop-helix tertiary interactions. **(c)**, Secondary structure diagram based on the tertiary structure of the M-box. Shaded areas depict nucleotides involved in Mg^{2+} coordination. Circles mark non-Watson-Crick base pairs. Antiterminator nucleotides are shown in violet color. **(d)**, Three-dimensional structure of the M-box. Mg^{2+} cations are shown by spheres. **(e)**, Mg1–3 facilitate tertiary contacts via inner- and outer-sphere interactions with labeled nucleotides. Coordination bonds and waters are shown by sticks and small spheres, respectively.

Figure 3.

Recognition of mRNA structures by proteins. Nucleotides essential for recognition are shown by red sticks in top views of panels (a) and (c). Peptide segments contacting RNA are in red in top views of panels (a) and (c) and in (d). (a), Superposition of rRNA (pale cyan ribbon) from the L1-rRNA complex onto the structure of the complex between L1 (green) and mRNA (beige ribbon). Conserved recognition of the G33-C63 pair and U64 by L1 is zoomed out and shown with hydrogen bonds in the panel below. Position of K^+ ion (yellow sphere) is occupied by ammonium group of Lys in the L1-rRNA complex. Hydrogen and coordination bonds are shown in dashed lines. (b), The S15-mRNA complex bound to the small ribosomal subunit. The 5'-terminal stem-loop structure is in light blue color, the pseudoknot is in beige, initiation codon AUG is in magenta, and nucleotides participating in the Shine-Dalgarno-rRNA interactions are in yellow. (c), The IRP1-IRE complex. RNA-protein contacts are shown in detail in the panel below. (d), Interface of RsmE-*hcnA* RNA complex. Amino acids forming sequence-specific hydrogen bonds are shown in sticks.

Figure 4.

Sequence-specific interactions of proteins with single-stranded RNA. **(a)**, Recognition of UGCAUGU element by the RRM of Fox-1 protein. Aromatic amino acids involved in canonical and novel modes of RNA binding are shown by magenta sticks. **(b)**, Recognition of UUU_{OH} sequence by La (in cyan) and RRM (in green) domains of the La protein. Aromatic amino acids underlying the RNA-binding pocket are in magenta color. Amino acids recognizing specific RNA features are shown in sticks with hydrogen bonds depicted by dashed lines.

Table 1Pre-mRNA and mRNA recognition by small molecules and proteins^a

Protein/metabolite	Type of RNA	Function	Technique/Resolution ^b	References, Protein Data Bank codes
Specific binding of folded RNAs				
Metabolite/cation recognition				
Thiamine pyrophosphate (TPP) and analogs	Bacterial riboswitch	Regulation of transcription and translation	2.05–3.5 Å	2GDI [12]; 2HOJ [11]
TPP	Plant riboswitch	Regulation of splicing and processing	2.9 Å	2CKY [13]
S-adenosyl- methionine (SAM)	SAM type I riboswitch	Regulation of transcription	2.9 Å	2GIS [14]
Glucosamin-6- phopshate and analogs	<i>glmS</i> ribozyme	Gene expression control	2.1–2.9 Å	2GCV [16]; 2NZ4 [15]
Mg ²⁺	M-box riboswitch	Regulation of transcription	2.6 Å	2QBZ [18]
Protein recognition				
Ribosomal protein L1	K-turn –like element	Repression of translation	2.1, 2.6 Å	2HW8, 1ZHO [21]
Ribosomal protein S15, ribosome	Pseudoknot and stem-loop	Repression of translation	Cryo-EM	2VAZ [29]
IRP-1	Stem-loop, iron-responsive element	Repression of translation and RNA degradation	2.8 Å	2IPY [30]
RsmE	Stem-loop, consensus ^A / _U CANGGAN G ^U / _A	Repression of translation	NMR	2JPP [34]
KH1/2 domains of NOVA-1	Aptamer, stem-loop, consensus (YCAN) ₂	Alternative splicing	1.94 Å	2ANR [Unpublished]
Human RBMY	Aptamer, stem-loop, consensus C ^A / _U CAA	Pre-mRNA processing?	NMR	2FY1 [35]
SAM domain of Vts1	Stem-loop, Smaug recognition element	Regulation of translation and mRNA stability?	NMR, NMR, 2.8 Å	2B6G [43]; 2ESE [44]; 2F8K [42]
Elongation factor SELB	Stem-loop, SECIS element	Incorporation of selenocysteine	2.31 Å	2UWM [39]
Exonuclease ERI1	Stem-loop	Histone mRNA degradation	3.0 Å	1ZBH [Unpublished]
Specific ssRNA binding				
Fox-1	U-GCAUG-U	Alternative splicing	NMR	2ERR [45]
RRM domain of SRp20	CAUC, consensus ^A / _U C ^A / _U ^A / _U C	mRNA splicing and transport	NMR	2I2Y [46]
Splicing factor U2AF ⁶⁵	U ₇	mRNA splicing	2.5 Å	2G4B [47]
Hrp1	Polyadenylation enhancement element G-(UA) ₄	3'-end modification	NMR	2CJK [51]
La autoantigen	UGCUGU-UUU _{OH}	RNA stability, mRNA binding	1.85 Å	1ZH5 [50]
RNase KID	AUACA	Endonuclease	NMR model	[52]
KH1 domain of poly(C)-binding protein 2	C-rich telomer-like sequence	Diverse functions, mRNA stability	2.6 Å	[53]
Non-specific ssRNA binding				
Archaeal exosome	mRNA mimetic		2.3 Å	2JEA [54]
Rho termination factor	Pyrimidine-rich RNA	Transcription termination	3.5 Å	2HT1[55]
RNase II	Poly(A) RNA	RNA degradation	2.7 Å	2IX1 [56]

Protein/metabolite	Type of RNA	Function	Technique/Resolution ^b	References, Protein Data Bank codes
Exon junction complex	mRNA mimetic	Splicing	2.3 Å, 3.2 Å, 2.2 Å	2HYI [57]; 2J0Q, 2J0S [58]
Vasa	mRNA mimetic	mRNA unwinding	2.2 Å	2DB3 [59]
RNase III	dsRNA	Cleavage of dsRNA	2.05 Å	2EZ6 [60]

^a Ribosomes and RNA polymerases bound to mRNA fragments as well as complexes between proteins and viral RNAs are not considered in the review.

^b Resolution is indicated for X-ray crystallography structures.