RNA-protein interactions are behind a number of vital processes in the cell. Without the ability of particular proteins to bind RNA, the RNA would no longer be able to carry out its important functions as a component of the ribosome and spliceosome. Other examples of important RNA-protein interactions include binding of tRNA to aminoacyl-tRNA synthetases, a process vital to translation of genetic information into proteins necessary for continued biological function and regulation of post-transcriptional control of gene expression via the binding of RNA to ribonucleoproteins, or RNPs.

Although not as well characterized as the binding between DNA and proteins, RNA-protein binding has been a field that has seen a great deal of growth in recent years. Although it was originally expected that RNA-protein binding motifs might fall neatly into categories the way DNA motifs did, the wide range of secondary and tertiary RNA structures that can be recognized by proteins requires more variety in binding motifs of the proteins, and the rules used to categorize them become correspondingly more complex. At this time all major families of RNA-binding proteins have been structurally characterized and these characterizations have led to a much better understanding of RNA recognition.

### Major Families of RNA-Binding Proteins

**Arginine-Rich Motif**

Positively charged amino acid side chains are important in the binding of many proteins to both DNA and RNA. Certain viral and bacterial proteins are known to contain regions densely populated by Arg and Lys residues. These peptides contribute to recognition of the major groove of the RNA and are commonly known as an arginine-rich motif. This term is slightly misleading because the positions of the arginines are not conserved between proteins and therefore cannot be a true motif. Furthermore, while these arginine-rich sequences within the overall protein are necessary for binding, when removed from the context of the larger protein the fragments tend to become unstructured and are not sufficient for binding.

One example of such a peptide is Rev, which is involved in the regulation of gene expression and is encoded by human immunodeficiency virus (HIV-1). The binding of Rev to the Rev Responsive Element (RRE) of RNA is an essential step in the HIV life cycle and involves the widening of the major groove to facilitate binding. After binding to the RNA hairpin Rev adopts an α-helical conformation. Conformational changes in the peptide after recognizing the major groove are typical of arginine-rich motifs.
Other examples of proteins with arginine-rich motifs in them include Tat, the main binding domain of bovine immunodeficiency virus,\textsuperscript{11} and N protein, from bacteriophage λ.\textsuperscript{12} Both of these proteins also form α-helices which are inserted into the major groove of the RNA, although in the case of N protein the helix is distorted.\textsuperscript{11,12}
There are three major families with αβ motifs that bind to RNA. These families are the RNA recognition motif (RRM or RNP), the K-homology (KH) domain, and the double-stranded RNA binding motif (dsRBM). In addition to these three families, there exist a variety of less common αβ motifs used to bind RNA, such as the Piwi Argonaut and Zwille domain (PAZ).

RNA Recognition Motif

Proteins with the RRM make up the largest family of RNA binding proteins, and the RRM itself is the most completely characterized binding domain. It is composed of 90 amino acids with a sequence of eight conserved residues. The motif itself contains a β1α1β2β3α2β4 fold with the first and third β-strands creating an anti-parallel sheet. The conserved residues, which are mostly aromatic and positively charged, are located in the β1 and β3 strands within the hydrophobic core of the domain. Most RRMs are highly sequence specific and have a high binding affinity.

The RRM directs RNA processing factors to specific RNAs by supplying an RNA binding platform, one of its major functions. The RRM is also involved in post-transcriptional events in eukaryotes. The function of RRM can be highly varied due to the 30+ structures it is capable of adopting by altering loops, positions of secondary structures, and secondary structures themselves. Function can also vary due to its ability to utilize multiple RRM domains, different motif domains, N- and C-terminal extensions, and protein cofactors.
The protein Human U1A contains the RNA recognition motif RRM, which recognizes the RNA hairpin internal loop AUUGCAC sequence and hydrogen bonds to the bases.\textsuperscript{16,17} In the specific case of U1A, the $\beta_2$-$\beta_3$ loop is inserted into the hairpin loop of the RNA. Prior to binding between the protein and the RNA, both the $\beta_2$-$\beta_3$ loop of the protein and the hairpin loop of the RNA are disorganized.\textsuperscript{14,18} The interaction between the protein and RNA thus can be understood as a mutual adaptation of the structures of the molecules with important inter- and intramolecular interactions.\textsuperscript{8}

![Figure 3. Structure of Human U1A protein with bound RNA. $\beta_2$-$\beta_3$ loop is inserted into the hairpin loop of the RNA. (PDBID: 1URN)](image)

**K-Homology Domain**

The K-homology (KH) motif is approximately 70 amino acids in length. Proteins typically contain more than one KH domain, although exceptions exist.\textsuperscript{19} Multiple copies of the KH domain increases specificity and micromolar affinity, which is low when only a single copy is present.\textsuperscript{20} There are several conserved moieties in the domain, including a variable loop that is 3 to 60 residues long, a GXXG loop, and a hydrophobic residue (Ile or Leu).\textsuperscript{21} Also conserved are a $\beta$-sheet, consisting of three $\beta$-strands, and three $\alpha$-helices abutted to the $\beta$-sheet.

Variation in the $\beta$-strands gives rise to two types of KH folds.\textsuperscript{22} The three $\beta$-strands are identified as $\beta_1$, $\beta_2$, and $\beta'$ where $\beta_1$ and $\beta_2$ are parallel and $\beta'$ is antiparallel. Type I folds are usually found in eukaryotes with the three $\beta$-strands aligned in the order $\beta_1$, $\beta'$, $\beta_2$. Type II folds are usually found in prokaryotes, with the strands aligned as $\beta_1$, $\beta_2$, $\beta'$. A second difference between the folds is that type I folds generally have multiple copies of the KH domain, but type II folds usually have only one copy.\textsuperscript{21}
KH can interact with single strands of RNA or DNA. The strand is bound between the α-1 and α-2 helices, and between the GXXG loop and the variable loop.21 The KH domain is found primarily in proteins performing transcriptional and translational regulation by binding single stranded RNA. KH only binds to 4 bases, although multiple KH domains can be found to work in conjunction to bind longer sequences.21

![Figure 4. Crystal structure of Nova-2 protein KH3 domain recognizing single-stranded SELEX RNA sequence 5'-UCAC. (PDBID: 1EC6)](image)

**Double Stranded RNA Binding Motif**

The double stranded RNA binding motif (dsRBM) contains an αββα fold which binds to RNA molecules.23 This motif appears with a variable number of copies of the domain in a protein, with the maximum being 5.24 Studies show that dsRBMs are not sequence specific, but do require A-form helical confirmation of RNA.25

The dsRBM can be divided into two categories, based on the absence or presence of a catalytic domain.24 RNaseIII, Dicer, and Drosha enzymes are involved in dsRNA processing in the RNAi/miRNA pathway and all three contain the dsRBM and a catalytic domain.25 RNA helicase A (RHA), a transcriptional cofactor, also contains dsRBM with a catalytic domain.26 The staufen protein and *Xenopus laevis* RNA-binding protein (Xlrbpa) have the motif without a catalytic domain, and both are involved in RNP localization.27,28

In targeting RNA, the A-form helix geometry is recognized by residues on loops 2 and 4 of the dsRBM domain. Contact occurs between loops 2 and 4 of the αββα fold and 2’OH groups and phosphate groups in the nearby minor and major grooves.25 Another mode of interaction between dsRBM and its targets involves the first helix of a dsRBM protein recognizing helical or non-helical secondary structure elements that possess a minor groove-type surface.29
The dsRBM is unusual in that it binds specifically \textit{in vivo},\textsuperscript{30} but it binds non-specifically \textit{in vitro},\textsuperscript{31,32} and the determination of what dictates its binding specificity is an area of ongoing study.

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{dsRBM.png}
\caption{dsRBM of Rnt1p RNase III complexed with 5' terminal RNA hairpin of snR47 precursor. (PDBID: 1T4L)}
\end{figure}

\textbf{Zinc Fingers}

Zinc fingers are a common structural motif in which conserved cysteines and histidines are coordinated with zinc to form small, independently folded domains.\textsuperscript{33} Each protein typically contains multiple copies of these domains.\textsuperscript{7,33} The classic Cys-Cys-His-His motif (CCHH) makes up about 3 \% of the human genome.\textsuperscript{7} Zinc fingers are found most often in DNA-binding proteins but it has been discovered that some motifs, such as the CCCH- and CCHC-type, can bind to RNA.\textsuperscript{7}

The fold of zinc fingers is composed of two distorted \(\beta\)-hairpins which surround a tryptophan residue and a zinc ion.\textsuperscript{34,35} Zinc fingers bind to double stranded RNA by way of contact with phosphates and hydrophobic stacking.\textsuperscript{33} In single stranded RNA, the aromatic sidechains of the motif insert between dinucleotide bases in an AU-rich region.\textsuperscript{33} Water-mediated hydrogen bonds and hydrogen bonds to protein atoms are involved in binding, though there is no interaction with the phosphate backbone.\textsuperscript{36}
Zinc finger domains can recognize many secondary and tertiary structures of RNA, including internal loops and helical elements closed by hairpin loops. Zinc fingers have a micromolar binding affinity for RNA and are sequence specific. Sequence-specific recognition is accomplished with exposed bases in loops and through interactions mediated by the backbone.

Transcription factor IIIA (TFIIIA) is the main component of the initiation complex needed for the transcription of genes encoding eukaryotic ribosomal 5S RNA. It contains nine zinc fingers and is capable of binding to both DNA and the ribosomal 5S RNA. The central 4 through 6 fingers bind to RNA, with only a few contacts to nucleotide bases.

**Figure 6.** Structure of three-finger TFIIIA-5S RNA complex demonstrating zinc-finger binding. Zinc ions are shown in magenta. (PDBID: 1UN6)

**Multimeric Proteins**

Multimeric proteins form heteromeric or homomeric structures by employing the use of multiple proteins or by repeating the same integral structural motif. In many cases, the details of the actual method of recognition for RNA of these proteins is not well known.
Human Pumilio protein, a translational regulator, utilizes multiple copies of the protein to form its multimeric complex. RNA recognizes and binds to the inner surface of the complex.  

Figure 7. Structure of modular protein Human Pumilio bound with noncognate RNA. (PDBID: 3BSX)

RNA-Targeting Enzymes

Given the wide variety of enzymes that act on RNA, considering all such enzymes as one protein family is absurd, though they all share the function of modifying RNA. Naturally, modes of recognition will vary greatly depending on the enzyme involved. The enzymatic domain of the protein binds to the RNA, often increasing enzymatic activity.

One example of an enzyme recognizing and acting on RNA are aminoacyl-tRNA syntheses (aaRSs), which prevent the mistranslation of specific amino acids. Many aaRSs have been studied, often with focus on key residues. A large amount of variation occurs in the interactions points as well as the nucleotides responsible for specificity. A common feature conserved in all aaRS studied thus far is contact between the anti-codon stem of tRNA and the synthetase. The orientation of the interaction differs depending on the specific aaRS involved. Another motif found is 2-loop, which is in the active site and involves binding ATP, an amino acid, and the acceptor end of tRNA. It requires glycine for flexibility.

RNA Structural Motifs

Like the proteins that bind to it, RNA has structural motifs that have roles in interaction and recognition, though many of these roles have not been thoroughly explored.
In addition to helices, RNA is comprised of various loops which fold characteristically and may also have sequence patterns. The most common motif is a tetraloop, a hairpin loop composed of four conserved residues. These residues fall into specific consensus sequence patterns: GNRA, where N is any base and R is G or A, UNCG, and CUUG. Assumed functions of the tetraloop include initiation folding of RNA molecules, stabilization of helical stems, and the contribution of recognition elements for protein binding. These loops likely play a large role in proteins interactions due to tetraloops always being exposed. Tetraloops are an example of 1-loops, which stay within one helix.

Another common motif is the kink-turn (K-turn), a 15 nucleotide sequence from two different segments of RNA which base pair and create an internal loop and two helices. This is an example a 2-loop, in that two helices are joined. E-loops are similar to kink-turns, but they join asymmetrical helices. Like tetraloops, kink-turns have been found to be significantly involved with bound proteins. There are a wide variety of 2-loops, and they can be double stranded, like kink-turns or E-loops, or single stranded, such as Π-turns and Ω-turns.

References


25. Chang, K; Ramos, A. The double-stranded RNA-binding motif, a versatile macromolecular docking platform. *FEBS J.* 2005, 272, 2109-2117.


40. Beebe, K; Mock, M; Merriman, E; Schimmel, P. Distinct domains of tRNA synthetase recognize the same base pair. Nature. 2008, 451, 90-93.

41. Theobald, A; Springer, M; Grunberg-Manago, M; Ebel, J; Giege, R. Tertiary structure of Escherichia coli tRNA3THR in solution and interaction of the tRNA with the cognate thereonyl-tRNA synthetase. Euro. J. Biochem. 1988, 175, 511-524.


44. Cusack, S.; Yaremchuk, A; Tukalo, M. The crystal structure of the ternary complex of T. thermophilus seryl-tRNA synthetase with tRNASer and a seryl-adenylate analogue reveals a conformational switch in the active site. EMBO J. 1996, 15, 2834–2842.


Contributors

- Therese Gerbich- Truman State University, Kirksville, MO
- Ashley Hoaglin- Truman State University, Kirksville, MO