STUDY MATERIALS

On

RNA BINDING PROTEINS AND RNA MOTIFS

Prepared by-

Padmanav Koushik

M.Sc. in Zoology

(Animal Physiology and Biochemistry)

Certified Nutritionist

GATE qualified 2020

In eukaryotic cells, each mRNA is bound by a dynamic repertoire of RNA binding proteins (RBPs) such that it exists as an mRNA-protein complex (messenger ribonucleoprotein, mRNP, Singh et al. [2015], Rissland [2017]). RBPs contain various structural motifs, such as RNA recognition motif (RRM), dsRNA binding domain, zinc finger and others. They are cytoplasmic and nuclear proteins. However, since most mature RNA is exported from the nucleus relatively quickly, most RBPs in the nucleus exist as complexes of protein and pre-mRNA called heterogeneous ribonucleoprotein particles (hnRNPs). RBPs have crucial roles in various cellular processes such as: cellular function, transport and localization. They especially play a major role in post-transcriptional control of RNAs, such as: splicing, polyadenylation, mRNA stabilization, mRNA localization and translation. The proper pre-mRNA splicing, processing, nuclear export, subcellular localization, and stability and degradation of mRNAs critically depend on these RBP-RNA interactions. Some mRNP components are members of large macromolecular machines, such as the spliceosome or ribosome, that bind mRNA in a coordinated manner to direct processes such as splicing, nuclear export, translation, and mRNA decay. These RBPs typically are deposited on mRNAs according to earlier RNA processing events or through interaction with mRNA landmarks such as the 5'cap, pre-mRNA splice sites, or the poly(A) tail. In addition to these members of common machineries, other RBPs interact with sequence-specific features of individual mRNAs. These proteins often bind mRNAs concurrently with core machineries to regulate specific steps in RNA processing, such as splicing factors binding introns or exons to influence alternative splicing or AU-rich element binding proteins binding 3' UTRs to influence mRNA stability or translation. However, not all RBPs fall into these extreme categories of high sequence specificity or general machineries but instead often operate in the middle ground of a specificity continuum from promiscuous to selective (Mitchell and Parker [2014]). For instance, Pumilio domain-containing proteins bind eight to ten RNA bases with high specificity at one extreme (Zamore et al. [1997]) while DEAD-box helicases have shown little dependence on RNA sequence to rearrange their mRNA substrates (Linder and Jankowsky [2011]). In between, SR (Serine/Arginine-rich) proteins and HNRNPs (hetergeneous nuclear ribonucleoproteins) exhibit discernible sequence preferences but are able to bind a wide range of targets to effect their transcriptome-wide splicing outcomes (Goren et al. [2006], Geuens et al. [2016]).

Functions of RNA binding proteins-

1. RNA processing and modification-

- **Alternative splicing**

Alternative splicing is a mechanism by which different forms of mature mRNAs (messengers RNAs) are generated from the same gene. It is a regulatory mechanism by which variations in the incorporation of the exons into mRNA leads to the production of more than one related protein, thus expanding possible genomic outputs. RBPs function extensively in the regulation of this process. Some binding proteins such as neuronal specific RNA-binding proteins, namely NOVA1, control the alternative splicing of a subset of hnRNA by recognizing and binding to a specific sequence in the RNA (YCAY where Y indicates pyrimidine, U or C). These proteins then recruit splicesomal proteins to this target site. SR proteins are also well known for their role in alternative splicing through the recruitment of snRNPs that form the splicesome, namely U1 snRNP and U2AF snRNP. However, RBPs are also part of the splicesome itself. The splicesome is a complex of snRNA and protein subunits and

acts as the mechanical agent that removes introns and ligates the flanking exons. Other than core splicesome complex, RBPs also bind to the sites of Cis-acting RNA elements that influence exons inclusion or exclusion during splicing. These sites are referred to as exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs) and depending on their location of binding, RBPs work as splicing silencers or enhancers.

- **RNA editing-**

The most extensively studied form of RNA editing involves the ADAR protein. This protein functions through post-transcriptional modification of mRNA transcripts by changing the nucleotide content of the RNA. This is done through the conversion of adenosine to inosine in an enzymatic reaction catalyzed by ADAR. This process effectively changes the RNA sequence from that encoded by the genome and extends the diversity of the gene products. The majority of RNA editing occurs on non-coding regions of RNA; however, some protein-encoding RNA transcripts have been shown to be subject to editing resulting in a difference in their protein's amino acid sequence. An example of this is the glutamate receptor mRNA where glutamine is converted to arginine leading to a change in the functionality of the protein.

- **Polyadenylation**

Polyadenylation is the addition of a "tail" of adenylate residues to an RNA transcript about 20 bases downstream of the AAUAAA sequence within the three prime untranslated region. Polyadenylation of mRNA has a strong effect on its nuclear transport, translation efficiency, and stability. All of these as well as the process of polyadenylation depend on binding of specific RBPs. All eukaryotic mRNAs with few exceptions are processed to receive 3' poly (A) tails of about 200 nucleotides. One of the necessary protein complexes in this process is CPSF. CPSF binds to the 3' tail (AAUAAA) sequence and together with another protein called poly(A)-binding protein, recruits and stimulates the activity of poly(A) polymerase. Poly(A) polymerase is inactive on its own and requires the binding of these other proteins to function properly.

2. mRNA localization-

mRNA localization is critical for regulation of gene expression by allowing spatially regulated protein production. Through mRNA localization proteins are transcribed in their intended target site of the cell. This is especially important during early development when rapid cell cleavages give different cells various combinations of mRNA which can then lead to drastically different cell fates. RBPs are critical in the localization of this mRNA that insures proteins are only transcribed in their intended regions. One of these proteins is ZBP1.

3. Translation-

Translational regulation provides a rapid mechanism to control gene expression. Rather than controlling gene expression at the transcriptional level, mRNA is already transcribed but the recruitment of ribosomes is controlled. This allows rapid generation of proteins when a signal activates translation. ZBP1 in addition to its role in the localization of B-actin mRNA is also involved

in the translational repression of beta-actin mRNA by blocking translation initiation. ZBP1 must be removed from the mRNA to allow the ribosome to properly bind and translation to begin.

RNA motifs

An RNA motif is a discrete sequence or combination of base juxtapositions found in naturally occurring RNAs in unexpectedly high abundance. Because all the motifs examined so far have three-dimensional structures independent of the context in which they are embedded, they are important components of the "kit" of structural elements from which RNAs are constructed.

Types-

1. Helices

Although RNA is composed primarily of Watson–Crick base-paired A-form double helices, other helical forms have been observed. Although RNA double helices are generally thought to not have a sequence-structure relationship (Holbrook et al. 1981), a comprehensive analysis with the large dataset currently available has not been completed. Recently, the structures of a high salt left-handed RNA duplex (Popenda et al. 2004) and a mirror image or L-configuration (Vallazza et al. 2004) Spiegelmer RNA duplex were determined. While the left-handed RNA duplex of repeating (CG) units differed significantly from its Z-DNA counterpart, the Spiegelmer RNA had a very similar structure to that of its right-handed, mirror-image structure even though crystallization conditions were quite different. Structures of RNA quadruplexes have also been determined. These structures include both guanine and adenine tetrads as well as bulged and looped out residues and generally differ from their DNA homologues.

2. Hairpin loops

Hairpin loops link the 3’ and 5’ ends of a double helix. Within the SCOR database classification, structurally characterized hairpin or external loops must close with a Watson–Crick pairing and vary in length from 2 to 14 nucleotides. The most common and well-studied of these are the tetra loops. Of these tetra loops, there are at least four that are characterized by their sequence and conserved structures : the GNRA type, the UNCG type, the ANYA type and the (U/A)GNN type. In ribosomal RNAs about 70% of tetraloops belong to either the GNRAor the UNCG families and are unusually stable thermodynamically compared to other tetraloop sequences (Antao et al. 1991). The well-known GNRA tetraloop is the most frequently observed tetraloop in currently available RNA structures. The GNRA tetraloop is frequently used as part of a tertiary interaction motif in the formation of tetraloop–tetraloop receptor interactions. Other hairpin loop motifs include the T-loop and D-loop motifs of tRNA (Quigley & Rich, 1976), the lonepair triloop, and the sarcin-ricin loop.

3. Internal loops

An internal loop separates double helical RNA into two segments by inclusion of residues that are not Watson–Crick paired in at least one strand of the duplex. Sometimes insertions on only one strand are defined as 'bulge loops', and we include this as a special case of internal loops. Two types of internal loops can be distinguished: symmetric, with the same number of nucleotides inserted on both strands; and asymmetric, with a different number of nucleotides inserted on the opposing strands. Non-canonical base pairing is common in internal loops. A frequently observed motif involves extension of double helical structure through continuous formation of non-Watson–Crick pairs in a symmetric internal loop. This double helical structure is distorted by unwinding, unstacking, and kinking formed by the non-canonical pairs. Fully paired and stacked internal loops of up to eight non-canonical pairs have been structurally observed. Certain asymmetric internal loop motifs have been identified and characterized as resulting in sharp turns important for tertiary structure formation. These include the kink-turn (K-turn), reverse kink-turn (Strobel et al. 2004) and hook-turn.

4. Junction loops/multiloops

Junction loops are formed by the intersection of three or more double helices. These double helices are separated by single-strand sequences of zero or more residues. There are N linker (joining) sequences for N helices in a junction loop, although some of the linker sequences may be of zero length. Although junction loops have not been as systematically or extensively studied as the simpler hairpin and internal loops, some generalizations have been made for the more common three-way and four-way junctions (Lilley, 1998, 2000). Common examples of unction loops are those in tRNA and the hammerhead ribozyme. Coaxial stacking of the helices is a key feature of junction loops as observed in these and many other examples.

5. Binding motifs

A primary function of RNA is to bind ligands, either for structural stabilization, as cofactors, substrates or signals. Ligand binding is critical for ribozyme, riboswitch and splicing functions, as well as in mediating RNA–protein and RNA–RNA intermolecular and tertiary interactions. RNA ligand-binding sites often demonstrate high selectivity and specificity, although there may be more than one motif capable of tightly binding a certain type of ligand.

6. Zinc fingers-

CCHH-type zinc-finger domains are the most common DNA-binding domain within the eukaryotic genome. In order to attain high sequence-specific recognition of DNA, several zinc fingers are utilized in a modular fashion. Zinc fingers exhibit ββα protein fold in which a β-hairpin and a α-helix are joined together via a Zn2+ ion. Furthermore, the interaction between protein side-chains of the α-helix with the DNA bases in the major groove allows for the DNA-sequence-specific recognition. Despite its wide recognition of DNA, there has been recent discoveries that zinc fingers also have the ability to recognize RNA. In addition to CCHH zinc fingers, CCCH zinc fingers were recently discovered to employ sequence-specific recognition of single-stranded RNA through an interaction between intermolecular hydrogen bonds and Watson-Crick edges of the RNA bases. CCHH-type zinc fingers employ two

methods of RNA binding. First, the zinc fingers exert non-specific interaction with the backbone of a double helix whereas the second mode allows zinc fingers to specifically recognize the individual bases that bulge out. Differing from the CCHH-type, the CCCH-type zinc finger displays another mode of RNA binding, in which single-stranded RNA is identified in a sequence-specific manner. Overall, zinc fingers can directly recognize DNA via binding to dsDNA sequence and RNA via binding to ssRNA sequence.

7. RNA-recognition motif (RRM)

The RNA recognition motif, which is the most common RNA-binding motif, is a small protein domain of 75–85 amino acids that forms a four-stranded β-sheet against the two α-helices. This recognition motif exerts its role in numerous cellular functions, especially in mRNA/rRNA processing, splicing, translation regulation, RNA export, and RNA stability. Ten structures of an RRM have been identified through NMR spectroscopy and X-ray crystallography. These structures illustrate the intricacy of protein–RNA recognition of RRM as it entails RNA–RNA and protein–protein interactions in addition to protein–RNA interactions. Despite their complexity, all ten structures have some common features. All RRMs' main protein surfaces' four-stranded β-sheet was found to interact with the RNA, which usually contacts two or three nucleotides in a specific manner. In addition, strong RNA binding affinity and specificity towards variation are achieved through an interaction between the inter-domain linker and the RNA and between RRMs themselves. This plasticity of the RRM explains why RRM is the most abundant domain and why it plays an important role in various biological functions.

8. Double stranded RNA binding motifs-

The double-stranded RNA-binding motif (dsRM, dsRBD), a 70–75 amino-acid domain, plays a critical role in RNA processing, RNA localization, RNA interference, RNA editing, and translational repression. All three structures of the domain solved as of 2005 possess uniting features that explain how dsRMs only bind to dsRNA instead of dsDNA. The dsRMs were found to interact along the RNA duplex via both α-helices and β1-β2 loop. Moreover, all three dsRBM structures make contact with the sugar-phosphate backbone of the major groove and of one minor groove, which is mediated by the β1-β2 loop along with the N-terminus region of the alpha helix 2. This interaction is a unique adaptation for the shape of an RNA double helix as it involves 2'-hydroxyls and phosphate oxygen. Despite the common structural features among dsRBMs, they exhibit distinct chemical frameworks, which permits specificity for a variety for RNA structures including stem-loops, internal loops, bulges or helices containing mismatches.

9. Tertiary interactions

As with secondary structure, the tertiary structure of RNA biomolecules is dominated by a limited number of recurring types of interactions or motifs (Batey et al. 1999). As enumerated in the SCOR database, these are: coaxial helices, kissing hairpin loops, the tetraloop–tetraloop receptor, the A-minor motif/patch, the tRNA D-loop:T-loop interaction, pseudoknots, and ribose zippers.

Table 5. *Some RNA tertiary interaction motifs*

Motif name(s)	Description	Secondary structures	Sequence preference
Ribose Zipper	Formed by hydrogen bonding between consecutive backbone ribose 2′ hydroxyls from two distant regions of the chain, interacting in an anti-parallel manner. Classified as canonical and 6 other types	Double helix: Hairpin or internal loop	Antiparallel 5′-CC-3′(Stem) 3′-AA-5′(Loop) O2′–O2′ and base triples (e.g. A-minor)
A-Minor Motif/ A-patch	A clustering of A-minor interactions, often decreasing in type and order going from the 5′ to the 3′ direction	Internal loop, Hairpin loop	Adenosines
D-Loop:T-Loop	Complex interaction between two conserved hairpins in tRNA, includes interdigitated bases	Hairpin loop: Hairpin loop	Conserved sequences in D-loop and T-loop motifs
Tetraloop:Tetraloop receptor	Conserved in Group I and II introns occuring between a GNRA tetraloop (GNRA fold) and the receptor; an internal loop plus two C-G pairs. It is characterized by a specific hydrogen bond pattern between the first A of the tetraloop and the U·A of the receptor to form an A·U·A triple; between the second A of the tetraloop and the backbone of the receptor C and U; between the third A of the tetraloop and the C:G pair of the receptor	Hairpin loop: Internal loop	5′-CC-UAAG-3′ 3′-GGUA–U-5′
Kissing Hairpin Loop	The kissing hairpin complex is formed by base pairing between single-stranded residues of two hairpin loops with complementary sequences	Hairpin loop: Hairpin loop	Self-complementary often six nucleotides
Coaxial Helices Interhelical stacking	Nucleotide bases from two separate helices stack and align axes to form a pseudo-continuous, coaxial helix. Coaxial helices are highly stabilizing and are dominant in several large RNA structures. Interhelical stacking may occur via a single base or base pair bridge between helices, resulting in continuous helical stacking spanning multiple helices	Double helices often across an internal or junction loop	'Bridging' nucleotide between helices has a preference for adenine
Pseudoknot	When bases pair between nucleotide loops (hairpin or internal) and bases outside the enclosing loop, they form a pseudoknot. This structure often contains coaxial helices. Can be very stable	Hairpin loop: Single strand	Complementary

------00000-------