Trends in Biochemical Sciences

CellPress

Review

Emerging Roles of Disordered Sequences in RNA-Binding Proteins

Sara Calabretta[1] and Stéphane Richard[1,*]

RNA-binding proteins (RBPs) maintain RNA metabolism homeostasis in the cell by regulating temporal, spatial, and functional dynamics of RNAs. RBPs achieve RNA binding not only through classical structured RNA-binding domains but also with sequences that are intrinsically disordered and often of low amino acid complexity. RBP–RNA interactions form ribonucleoprotein (RNP) complexes and emerging evidence indicates that RNPs form higher structures or lattices, promoting territories of phase transitions. Herein, we discuss the role of disordered sequences in RBPs, their function in RNPs and protein networks, as well as their regulation by post-translational modifications and how RBP deregulation leads to disease.

RNA-Binding Proteins

Many cellular RNAs exist in association with **RNA-binding proteins (RBPs)** (see Glossary) to form **ribonucleoprotein (RNP) complexes**; defects in the formation or the composition of RNPs lead to diseases [1–3]. Interactions between RBPs and RNA are crucial for maintaining **RNA metabolism** homeostasis at all stages from biogenesis to degradation. Therefore, RBPs are key post-transcriptional gene regulators. It is not surprising that RBPs fulfill versatile roles in the regulation of basic cellular processes [4], such as the regulation of pre-messenger RNA (pre-mRNA) splicing [5], polyadenylation [6], export to the cytoplasm, and translation into protein (Figure 1). Many other roles have been attributed to RBPs including the processing of noncoding RNA, such as microRNAs (miRNAs), circular RNAs (circRNAs), and long noncoding RNAs (lncRNAs), and will not be extensively covered here as these were reviewed recently elsewhere [7–10].

RBPs were historically identified by their ability to associate with RNA using biochemical methods [11–14]. RNA-binding domains were then defined by structure–function studies and structure determination [15–18]. Sequencing and computational analysis of DNA sequences across species led to the identification of >500 human RBPs, each containing at least one RNA-binding domain [19]. Classically, RBPs were categorized based on their RNA-binding domains including the RNA recognition motif (RRM), the K homology (KH) domain, the DEAD motif, the double-stranded RNA-binding motif (DSRM), and the zinc-finger domain [19]. However, new RBP–RNA complexes have emerged through the use of genome-wide RNA target identification via crosslinking and immunoprecipitation (CLIP) technology combined with high-throughput sequencing [20–23]. Bioinformatics focusing on the specific RNA sequences bound by RBPs have revealed many new RBP–RNA interactions [24]. In addition, **interactome** capture coupled with mass spectrometry identified >1300 experimentally confirmed human RBPs [25–28]. A compilation of these datasets has led to the experimentally validated census of ~1500 RBPs [29,30]. Many of these newly identified RBPs do not harbor a canonical RNA-binding domain, but rather contain disordered and low amino acid complexity sequences, such

Trends

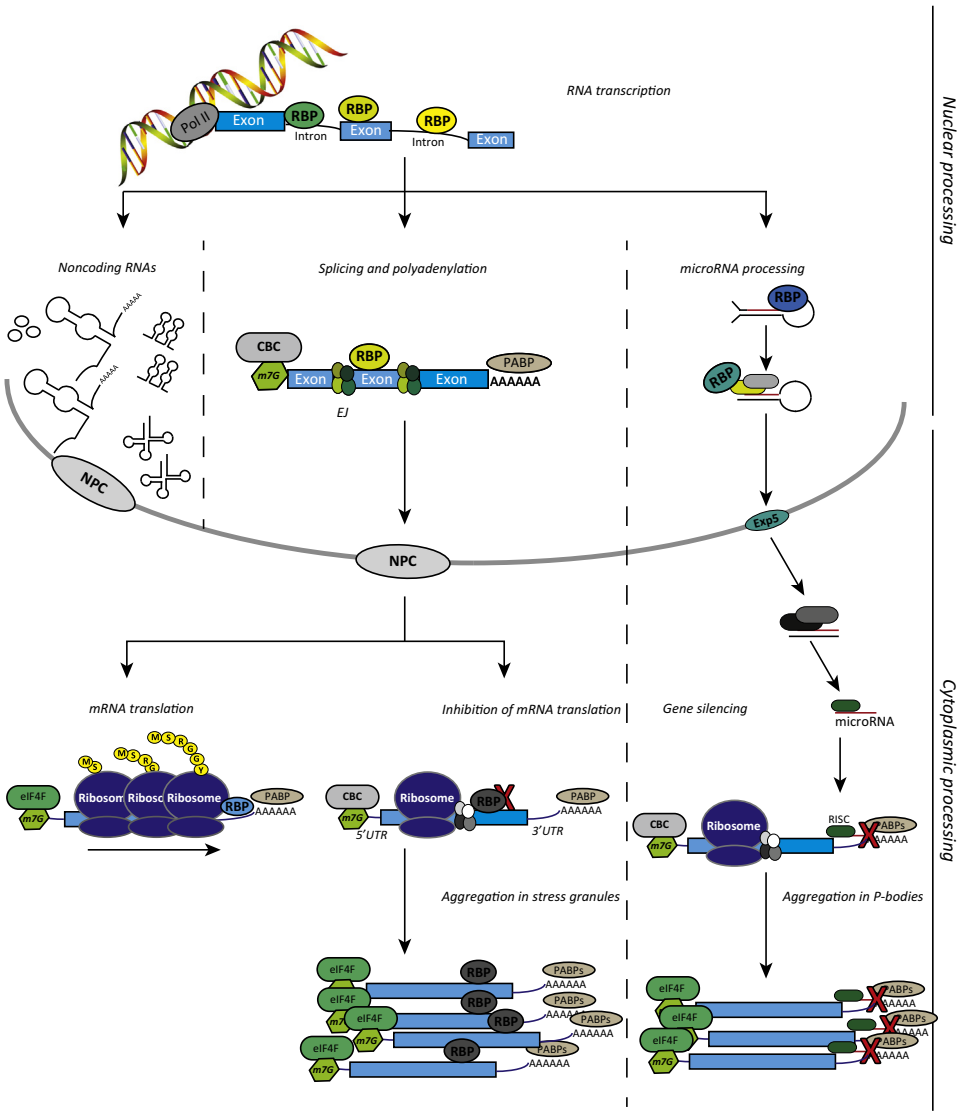
Intrinsically disordered regions (IDRs) are protein sequences that lack a defined and ordered 3D structure and play key roles in RNA-binding proteins (RBPs).

The physiological role of IDRs in RBPs is likely related to the formation of hubs of ribonucleoproteins (RNPs), required for proper RNA metabolism. IDRs undergo disordered-to-ordered transition after binding with interactors.

Mutations of IDRs contained in RBPs have been linked to the onset of diseases. Mutations likely affect the flexibility of intrinsically disordered proteins (IDPs), thereby disrupting the correct balance between disordered and ordered structures.

[1]Terry Fox Molecular Oncology Group and Segal Cancer Center, Bloomfield Center for Research on Aging, Lady Davis Institute for Medical Research and Departments of Oncology and Medicine, McGill University, Montréal, Québec, H3T 1E2, Canada

*Correspondence: stephane.richard@mcgill.ca (S. Richard).



Trends in Biochemical Sciences

Glossary

Hub: top ~20% of the interacting proteins of an interactome.
Interactome: a network of interacting proteins.
Intrinsically disordered proteins (IDPs): proteins that contain one or more IDRs.
Intrinsically disordered regions (IDRs): a sequence in a protein that is not structured by itself, but requires a ligand or binding partner to assume a secondary structure.
Low complexity (LC) sequences: a protein sequence containing limited diversity in amino acid composition and is devoid of hydrophobic residues.
Ribonucleoprotein (RNP) complexes: macromolecules containing proteins and RNAs. Classic examples include the spliceosome, ribosome, and exon junction complex.
RNA-binding proteins (RBPs): proteins that directly interact with RNA. RBPs either interact with RNA with a structured domain and/or with an IDR. The RBP–RNA interaction can occur in a sequence- and/or structure-specific manner; however, some RBPs bind RNA in a sequence- and structure-independent manner.
RNA metabolism: the process that implicates RNAs from their synthesis to their degradation. RNA metabolism defines the processes including pre-mRNA splicing, noncoding RNA regulation, nonsense-mediated decay, as well as RNA export, localization, stability, packaging in RNPs, and mRNA translation.

Figure 1. RNA-Binding Proteins (RBPs) in RNA metabolism. **Nuclear processing:** RNA polymerase II (Pol II) generates RNAs such as the pre-messenger RNA (pre-mRNA) shown. Noncoding RNAs are also transcribed and RBPs are emerging as key players in noncoding RNA metabolism and function. pre-mRNAs are stabilized by the Cap-binding complex (CBC), which binds the 5′cap [5′-5′ triphosphate-linked guanine modified with a 5′ 7-methyl group (m7G)], and by the addition of a poly(A) tail by the poly(A)-polymerase. The poly(A) tail is then recognized by the polyA-binding protein (PABP). Splicing by the spliceosome and RBP splicing factors leads to mature mRNAs. The exon junction (EJ) complex binds the splice junctions and remains associated with the mRNAs until translation or degradation occurs. Additional RBPs bind the mRNA, forming messenger ribonucleoproteins (mRNPs) and contribute to mRNA export into the cytoplasm via the nuclear core complex (NPC). Primary microRNA (miRNA) transcripts are processed to form precursor miRNAs that are then exported to the cytoplasm through Exportin 5 (Exp5) and further processed to obtain the mature miRNAs before its loading into the RNA-induced silencing complex (RISC). The modulation of miRNA processing is also modulated by specific RBPs. **Cytoplasmic processing:** the nonfunctional mRNPs are those that contain a premature termination codon (PTC) or those stored in response to stress. Translation elongation will not proceed and the mRNAs will be either degraded by endonucleases or stored in stress granules. The functional mRNPs recruit productive components of the translation machinery such as the eukaryotic translation initiation factor complex (eIF4F) and ribosomes that initiate protein synthesis. The mRNAs targeted by miRNA are not translated and will be degraded or stored in processing bodies (P-bodies). RBPs are involved in the modulation of all the aforementioned processes.