CHAPTER

THE FUNCTIONS OF GLYCINE-RICH REGIONS IN TDP-43, FUS AND RELATED RNA-BINDING PROTEINS

Boris Rogelj,[1] Katherine S. Godin,[2] Christopher E. Shaw[1] and Jernej Ule*,[2]

[1]*MRC Centre for Neurodegeneration Research, Institute of Psychiatry, King's College London, London, UK*;
[2]*MRC Laboratory of Molecular Biology, Cambridge, UK*
**Corresponding Author: Jernej Ule—Email: jule@mrc-lmb.cam.ac.uk*

Abstract: Glycine-rich regions form intrinsically unstructured domains within RNA-binding proteins. Although they lack a defined structure when alone in solution, these domains can form more defined structural elements when interacting with other proteins or with RNA. TDP-43 and FUS are RNA binding proteins with glycine-rich domains that form abnormal aggregates in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). The vast majority of mutations in familial ALS occur within the glycine-rich domain of TDP-43 as do about a third of FUS mutations. This chapter will review the various functions of some of the best characterised glycine-rich domains in RNA-binding proteins. Furthermore, the chapter will discuss how these findings inform on the possible functions of glycine-rich domains in TDP-43 and FUS.

INTRODUCTION

TAR DNA binding protein (TDP-43, encoded by the *TARDBP* gene) and fused in sarcoma (FUS) are two RNA-binding proteins (RBPs) that have attracted significant interest in recent years due to their involvement in neurodegenerative disorders. Approximately 60% of frontotemporal lobar degeneration (FTLD) and 90% of amyotrophic lateral sclerosis (ALS) cases are characterized pathologically by the TDP-43 inclusions. In addition, mutations in *TARDBP* or *FUS* gene are associated with familial ALS, and more rarely, with frontotemporal dementia (FTD, a syndrome caused by FTLD).[1-5] These familial

RNA Binding Proteins, edited by Lorkovic Zdravko.

cases are also characterized by TDP-43 or FUS inclusions implicating protein aggregation in disease pathogenesis. TDP-43 and FUS are predominantly nuclear proteins but in the pathological state they form inclusion in the cytoplasm of neurons and glia, with a corresponding decrease in nuclear staining.[3,6,7] TDP-43 cleavage products found in the cytoplasmic inclusions have been shown to alter splicing of a TDP-43 regulated exon in minigene studies, indicating that they are capable of interfering with wild-type TDP-43 function.[8] However, it is not yet clear whether inclusions of TDP-43 or FUS have a toxic gain of function or if their toxicity is a result of sequestration of the proteins from the nucleus, disrupting their role in gene expression.

TDP-43 and FUS regulate multiple aspects of gene expression, including transcription, alternative splicing and mRNA stability.[9-38] Both proteins contain glycine-rich regions that are predicted to be highly unstructured. Such glycine-rich regions are common in RBPs, but their function is poorly understood. In addition to TDP-43 and FUS, we will discuss the functions of glycine-rich regions in Sex-lethal (Sxl), heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1), Fragile-X mental retardation protein (FMRP, encoded by *FMR1* gene), small nuclear ribonucleoprotein D1 polypeptide 16kDa (SmD1, also SNRPD1) and U2 small nuclear RNA auxiliary factor 1 (U2AF1, also U2AF35). The glycine-rich regions in these proteins are part of predicted unstructured domains that contain several sub-regions that are enriched in a few additional amino acids. We show that glycine-rich regions have diverse functions in different RBPs, and we speculate how these differences might relate to the additional amino acids that are enriched in the glycine-rich regions.

GENERAL PROPERTIES OF INTRINSICALLY UNSTRUCTURED PROTEINS

Intrinsically unstructured proteins lack a unique structure, either entirely or in part, when alone in solution. Studies of these proteins have revealed some general properties of unstructured regions.[39] Unstructured regions have a strong amino acid compositional bias, rich in hydrophilic and charged residues but often lacking bulky hydrophobic residues. This prevents the formation of a hydrophobic core necessary for a stable three-dimensional fold. Moreover, the large number of side-chain charges present under physiological conditions destabilizes any compact state. Furthermore, unstructured regions are enriched in amino acids that are targets for diverse posttranslational modifications, which can contribute to their conformational variability. These regions often adopt different folds when interacting with different partners. For instance, when unstructured regions are involved in RNA interactions, they can adopt distinct folds upon binding the RNA target. In these cases, it is the RNA tertiary structure that generates scaffolds and a binding pocket for recognition and discrimination amongst minimal elements of protein architecture.[40,41]

Another feature of intrinsically unstructured proteins is that their expression levels are precisely regulated. This includes low stochasticity in transcription and translation, and precise regulation of transcript clearance and proteolytic degradation.[42] In this chapter, we will also discuss several cases of RBPs containing unstructured regions that autoregulate their own expression, most often by regulating alternative splicing of their own transcript. The precise regulation of intrinsically unstructured proteins, together with their conformational variability, allows them to participate in multiple biological processes and to facilitate combinatorial regulation.[39]

SPLICING FUNCTIONS OF GLYCINE-RICH REGIONS

Examination of glycine-rich regions of individual RBPs shows that each contains a small number of additional amino acids that are found in the putative unstructured region. These additional amino acids enriched in the glycine-rich regions contain polar (hydrophilic) residues, such as arginine, asparagine, glutamine, serine and tyrosine. Thus, glycine is the only amino acid enriched in the putative unstructured regions that contains a nonpolar (hydrophobic) side chain. Glycine is the smallest amino acid, with a single hydrogen atom as its side chain, allowing it to fit into tight spaces. These properties might underlie the propensity of glycine-rich regions to promote homo- and heteromeric interactions to create RNP complexes.

The additional amino acids enriched in the glycine-rich regions are likely to influence the interactions of these regions with other proteins or RNA, as well as determine the possibilities for posttranslational modifications. It is only due to our lack of knowledge that these diverse regions are collectively referred to as glycine-rich regions. In this review, we will first discuss the proteins that have enrichment of asparagine, glutamine and serine in glycine-rich regions. As we will discuss below, these proteins contain only one or two functional RNA-binding domains, and the glycine-rich regions often facilitate homo- and heteromeric interactions. This leads to formation of higher-order RNP complexes that allow co-operative RNA binding. We will discuss how this co-operative RNA binding might serve to increase the combinatorial complexity of splicing regulation.

Sxl

Sex-lethal (Sxl) is a Drosophila RBP that contains an unstructured region on the N-terminus, referred to as a glycine-rich N-terminus, and two RRM domains that bind uridine tracts in target RNAs (Fig. 1A). Due to its crucial role in Drosophila sexual differentiation, Sxl was one of the first well-studied splicing regulators. As part of a cascade of genes that are regulated by sex-specific splicing, Sxl binds a uridine tract near the alternative 3′ splice site of transformer (tra) pre-mRNA, thus blocking use of this site to give rise to the female-specific splicing pattern. This is achieved by competing with large subunit of the general splicing factor U2AF (Drosophila gene name is *U2af50,* and the human orthologue is U2AF2, also referred to as U2AF65) for binding to the poly-pyrimidine tract.[43] In addition, Sxl binds male-specific lethal-2 (msl-2) transcript to control its translation and thereby regulate dosage compensation.[44]

Sxl also autoregulates splicing of its own transcript by blocking inclusion of an alternative exon. Splicing autoregulation requires the intact N-terminal region of Sxl, even though this region is not required for translational regulation of msl-2 expression.[43,44] Whereas it has initially been proposed that the N-terminus is not necessary for tra regulation, later studies have proposed the opposite.[44,45] This discrepancy might have resulted from differences in how the groups have performed the experiments, which include different constructs (40-amino-acid deletion of the amino-terminal region in one case and 94-amino-acid deletion in the other). However, it appears likely that the Sxl N-terminal unstructured domain is required to regulate only a subset of its RNA targets.

Several observations point to distinct mechanisms in Sxl autoregulation that might involve the N-terminus. In contrast to the regulation of tra splicing, competition between Sxl and U2AF for binding to the uridine tract upstream of the exon in sxl transcript is

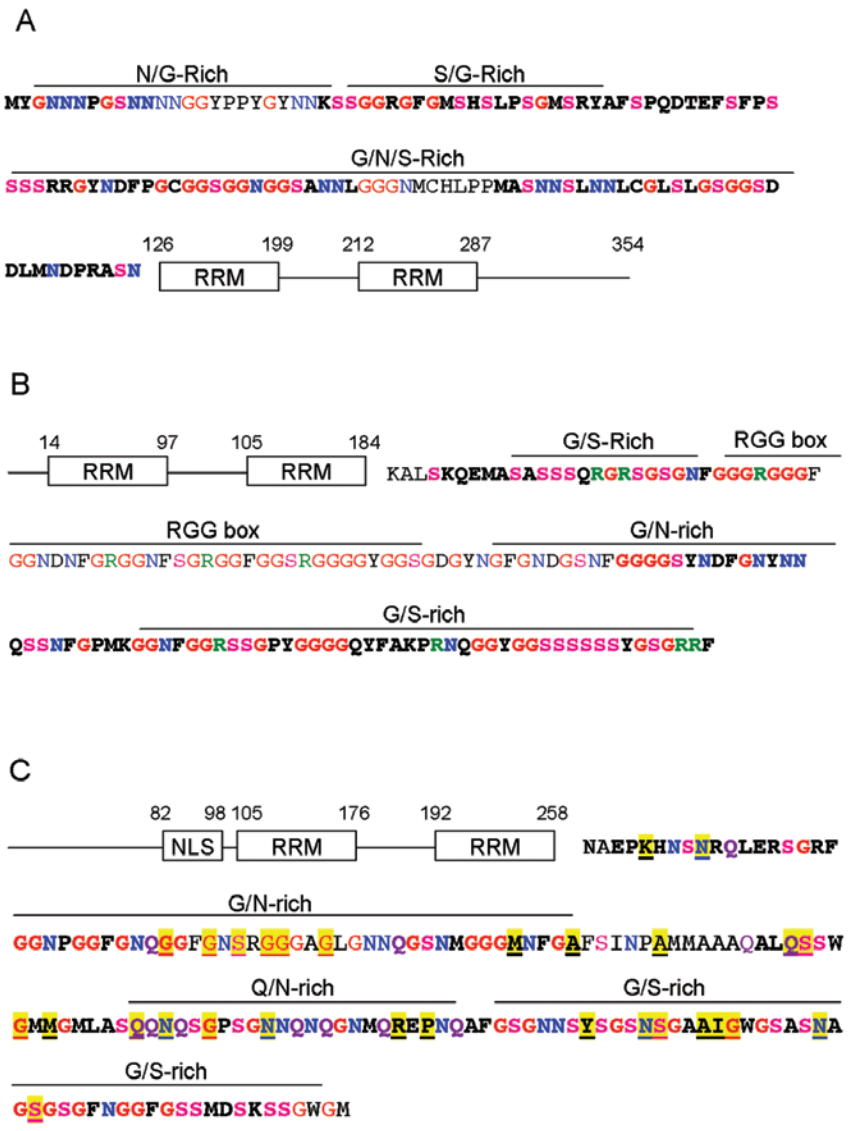

Figure 1. Domain structure and sequence of the glycine/serine/asparagine-rich domains in three RNA-binding proteins. A) Drosophila protein Sex-lethal (Sxl). B) Human protein heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1). C) Human TAR DNA binding protein (TDP-43). The sequence of the protein region predicted to be disordered by Disopred2 software is shown, with highly disordered region in bold. Asparagines are labeled in blue, glutamines in purple, serines in pink, glycines in red and arginines in green. The amino acids in TDP-43 affected by disease-causing mutations are underlined and shaded in yellow. The color version of this image is available at www.landesbioscience.com/curie.

insufficient to explain the autoregulation. This was shown by analysing effects of mutations within the poly(U) stretch upstream of the exon, and by expressing a fusion protein of U2AF and Sxl. Both had effects on splicing of tra, but not sxl transcripts. Instead, it was shown

that Sxl needs to bind multiple uridine tracts upstream and downstream of the alternative exon in its own transcript in order to block its inclusion.[43] It was initially suggested that the N-terminal unstructured domain is implicated in homomeric protein-protein interactions that are required for Sxl multimerization on its own pre-mRNA.[46] However, later studies indicated that protein-protein interactions are mediated mainly by the RBDs.[47,48] Therefore, the role of the unstructured region in Sxl multimerisation remains to be fully resolved.

The Sxl N-terminus interacts with several other RBPs that and could mediate some of its splicing effects.[49] These include U2AF and sans fille (snf), a Drosophila orthologue of small nuclear ribonucleoprotein polypeptide A (SNRPA, also U1A).[50] U2AF plays a crucial role in 3′ splice site recognition, whereas SNRPA is core component of the U1 small nuclear ribonucleoprotein particle (U1 snRNP), which recognises the 5′ splice site. It was proposed that interaction of Sxl with U2AF and snf produces a complex that prevents the formation of a functional spliceosome (Fig. 2A).[43] Whereas the Sxl glycine-rich N-terminus was shown to be involved in interactions with glycine-rich regions in other proteins, such as Hrb87F (hnRNP A/B homologue) and hnRNP L, the interaction of Sxl protein with Snf occurs mainly through the RRM domain.[48,49]

Finally, the mechanism of Sxl autoregulation was also suggested to involve modulation of the second catalytic step of Sxl intron 2 splicing (Fig. 2B).[43] Interestingly, deletion of the whole unstructured region of Sxl (the N-terminal 94 amino acids, Fig. 1A) interfered with this function.[51] Moreover, the N-terminal region was shown to interact with splicing factor 45 (SPF45, also RBM17), a spliceosome component that functions in the second catalytic step of mRNA splicing.[51]

A detailed examination of the Sxl unstructured region shows that it is enriched in glycine (G), asparagine (N) and serine (S) (Fig. 1A). It contains sub-regions that are highly enriched in only one or two of these amino acids. For instance, the first 23 amino acids are enriched in asparagines and glycines, followed by 19 amino acids enriched in serines and glycines. It is likely that each of these sub-regions interacts with different proteins and, therefore, have different functions. However, as even the shortest deletion study of Sxl removed both the asparagine and serine-rich regions, more detailed Sxl deletion analyses would be required to fully understand the function of its unstructured region.[44]

hnRNP A1

hnRNP A1 is one of the most abundant nuclear RBPs. It contains two RRM domains at the N-terminus, which bind with high affinity to RNA sequences containing UAGGGA/U motif. The C-terminus contains a 124-amino-acid region that is predicted to be highly unstructured (Fig. 1B). This region contains domains rich in asparagine and glycine or serine and glycine, which resemble the Sxl N-terminus. In addition, it contains the RGG box, the function of which is discussed in the context of glycine and arginine-rich regions in a later section. Deletion of the C-terminal domain does not affect RNA binding, but it does interfere with autoregulation of hnRNP A1 alternative splicing.[52] Similar to Sxl, autoregulation requires binding to multiple distal RNA sites, and the C-terminal region is responsible for the ability of hnRNP A1 to silence splicing via co-operative RNA binding.[52-56]

Two models were proposed to explain how crosstalk between hnRNP A1 molecules bound at distal RNA sites affects splicing choice. The first proposed model is referred to as ‘RNA looping model’. This model proposes that hnRNP A1 molecules bound at distal

RNA sites can homomultimerise and thereby promote formation of an RNA loop.[52,57] The position of the two high-affinity binding sites determines the type of loop formed, which can allow hnRNP A1 to either silence or enhance exon inclusion.[52,57,58]

The second model is referred to as 'co-operative binding-dependent crosstalk' model. This model suggests that two distal high-affinity sites facilitate co-operative spreading of hnRNP A1 between the two sites, which can lead to displacement of other proteins and unwinding of RNA hairpins.[59] It was shown that initial hnRNP A1 binding to a high-affinity site in the human immunodeficiency virus type 1 (HIV-1) tat exon 3 is followed by co-operative spreading on the RNA. As hnRNP A1 propagates along the exon, it antagonizes binding of a serine/arginine-rich (SR) protein to an exonic splicing enhancer, thereby inhibiting splicing at that exon's alternative 3′ splice site.[56] The C-terminal domain is required for the co-operative protein/protein interactions that underlie the co-operative spreading and was also proposed to be involved in protein/nucleic acid interactions.[53,54] Furthermore, hnRNP A1 was found to unwind an RNA hairpin upon binding, which required the C-terminal region.[59]

It is possible that the distance between the non-adjacent binding sites determines which of the two models explains the co-operative effects of hnRNP A1. If the two non-adjacent sites are positioned close to each other, then hnRNP A1 will be able to act via co-operative spreading that displaces RBPs bound to the intervening RNA region (Fig. 2C). However, if the two RNA binding sites are located a long distance from each other, hnRNP A1 might not be able to spread across the whole intervening RNA, but will therefore instead act via RNA looping.[57] Depending on the type of RNA loop that is promoted, exon inclusion could be either enhanced or silenced (Fig. 2D,E). Interestingly, polypyrimidine tract binding protein (PTB) can act via RNA looping without the need of multimerisation, which might be because it contains four RRM RNA-binding domains that enable its binding to multiple distal sites.[60,61] Therefore, one function of the unstructured domain in hnRNP A1 might be to allow it to form the multimeric complex that can bind to multiple distal sites. Furthermore, interactions between different RNA-binding proteins can also lead to multimeric complexes that bind to multiple distal sites to promote RNA looping, such as for instance interactions between hnRNP A/B and hnRNP F/H or between NOVA and FOX proteins.[58,62] The heteromeric interactions greatly enhance the combinatorial possibilities for regulation of RNA looping, and it is likely that the unstructured domains have an important contribution to these interactions.

TDP-43

TDP-43 has two N-terminal RRM domains, which bind UG repeat sequences with high affinity.[63] In addition, TDP-43 C-terminus is predicted to be unstructured, and has similar amino-acid enrichment as hnRNP A1, containing regions rich in glycine/serine or glycine/asparagine (Fig. 1C). Interestingly, the C-terminal domains of TDP-43 and hnRNP A1 interact with each other, and this interaction is required for TDP-43 to regulate alternative splicing.[64]

Mutations in the gene encoding TDP-43 (*TARDBP*) map to the C-terminal region and are associated with 1–4% of familial and sporadic ALS and have also been found in a few FTD cases (Fig. 1C).[5] In addition, TDP-43 pathology is evident in the majority of sporadic cases that lack any mutations in *TARDBP* gene. Given the similarity between

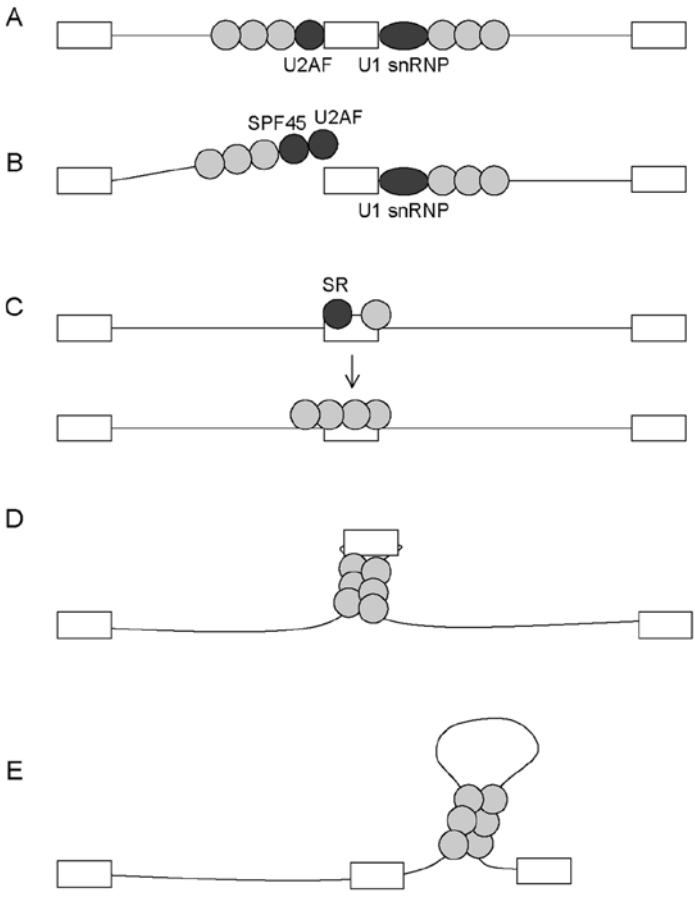

Figure 2. Splicing regulatory mechanisms that involve the glycine-rich regions of RNA-binding proteins. The studies of splicing functions of Sxl, hnRNP A1 and TDP-43 indicate that several splicing functions of these proteins require the glycine-rich unstructured regions. Exons are shown as boxes, introns as the lines connecting these boxes, and proteins as ovals. A) By interacting with Drosophila orthologues of U2AF2 (dark grey) and U1 snRNP (dark grey) at both sides of the exon, Sxl (light grey) was proposed to produce a complex that prevents the formation of a functional spliceosome. B) Sxl (light grey) was proposed to inhibit exon inclusion by interacting with a factor (SPF45, dark grey) that recognizes the 3′ splice site in the second catalytic step of splicing (after cleavage of intron-exon junction). C) hnRNP A1 (light grey) can co-operatively spread on the pre-mRNA, which can act to unwind RNA hairpins and displace other proteins, such as SR proteins (dark grey), from the pre-mRNA. The TDP-43 crosslink enrichment at a distance from its high-affinity UG-rich binding sites indicates that it may have a similar ability for co-operative spreading. D,E) Formation of ribonucleoprotein (RNP) complexes (light grey) via homo- or heteromeric interactions between different RBPs can allow these RNPs to bind at multiple distal sites on the pre-mRNA and thereby change the RNA structure in a manner that can either silence (D) or enhance (E) exon inclusion.

TDP-43 and hnRNP A1 proteins, it is intriguing to speculate why TDP-43, but not hnRNP A1, has the propensity to form the toxic aggregates even in the absence of TDP-43 mutations. A comparison between the putative unstructured regions of the two proteins shows that TDP-43 contains a glutamine/asparagine (Q/N)-rich region, which

is absent in hnRNP A1 (Fig. 1B,C). Interestingly, the Q/N-rich region in TDP-43 was required for its sequestration into detergent-insoluble inclusions formed by polyglutamine proteins.[65,66] Due to this result, it was proposed that the Q/N-rich region might have prion-like properties.[67]

TDP-43 has a propensity to bind RNA regions up to 100 nucleotides away from the UG repeats.[11] This indicated that the RNA binding properties of TDP-43 involve co-operative assembly, similar to hnRNP A1.[53,54,59] Since the C-terminus is required for such co-operative binding in hnRNP A1, we can speculate that the glycine-rich region in TDP-43 might play a similar role. The co-operative RNA binding is not a general property of RBPs, because another protein that also recognizes GU repeats, CUGBP, Elav-like family member 2 (CELF2), did not bind at positions further than 30 nucleotides from the GU repeat. Evidence for co-operative TDP-43 assembly was also seen in the vicinity of alternative exons that are silenced by TDP-43.[11] Such splicing silencing via multiple binding sites resembles the splicing autoregulation of Sxl and hnRNP A1, which requires multimerisation of these RBPs via the glycine-rich regions.[44]

TDP-43 autoregulates the levels of its own transcript by binding to multiple spaced sites on its own 3′UTR, indicating that co-operative assembly on its own RNA acts as a negative feedback loop to maintain the appropriate protein levels.[10,11,68] One aspect of autoregulation involves a splicing regulation of intron within the 3′UTR, which promotes degradation by nonsense-mediated mRNA decay (NMD).[10] A similar mode of autoregulation, where a splicing change leads to NMD, has been documented for PTB and most SR proteins, including autoregulation of an intron in the 3′UTR of SFRS1.[69-73] Unproductive splicing accounts for only part of the SFRS1 autoregulation, which occurs primarily at the translational level.[69] TDP-43 autoregulation via binding to the 3′UTR appears to similarly involve several levels of control, because one study showed that TDP-43 autoregulation is independent of NMD, and instead suggested that exosome-mediated degradation plays the primary role.[68]

Intrinsically unstructured proteins need an especially precise control of their expression levels.[42] Interestingly, some of the disease-causing mutations in TDP-43 were shown to increase the stability of TDP-43 protein, as well as increase its association with FUS protein.[74] Increased stability of TDP-43 could have detrimental effects, because increased expression levels of TDP-43 in yeast, Drosophila and chick are toxic.[75,76] Importantly, overexpression of TDP-43 is toxic only if its RNA binding function is intact.[76] One possible explanation for the requirement of RNA binding might be that increased levels of TDP-43 affect target RNAs in manners that are toxic for the cell, for instance by changing processing or translation of these RNAs. Alternatively, given that TDP-43 appears to be capable of co-operative assembly on RNA, increased expression of TDP-43 might have an increased tendency to multimerise upon RNA binding, which could initiate the formation of toxic aggregates.

PROTEINS WITH GLYCINE AND ARGININE-RICH DOMAINS

Analysis of protein-RNA complexes showed that positively charged amino acid side chains of arginine and lysine contribute to about one-third of the protein-RNA hydrogen-bonds.[77] Therefore, arginine is a likely candidate to modulate interactions

with RNA in domains rich in arginine and glycine, which are common in RNA-binding proteins.[78] Arginine in these domains is a common target for arginine methylation, which can affect protein-RNA interactions, protein localization and protein-protein interactions.[79] hnRNP proteins, such as hnRNP A1, are the primary substrates containing dimethylarginine in the cell nucleus.[80] Dimethylation of hnRNP proteins was shown to regulate their nuclear export.[81] The addition of methyl groups to amino acid side chains increases steric hindrance and removes amino hydrogens that normally participate in hydrogen bonding.[82] Thus, argnine and glycine-rich domains have a potential of interacting with RNA in a manner that would be regulated by arginine methylation.

RGG BOX

The term RGG-box refers to clusters of closely spaced arginine-glycine-glycine tripeptides flanked by aromatic residues (Fig. 3A).[83] RGG boxes in FUS protein were shown to be unstructured when this protein is present free in solution.[19] RGG boxes are present in diverse RBPs, and were reported to have multiple functions. RGG-boxes in hnRNPK, hnRNPD, nucleolin and EWS were implicated in transcriptional regulation.[84-86] RGG boxes in hnRNP U, Ewing sarcoma breakpoint region 1 (EWSR1, also EWS), FUS and fragile X mental retardation 1 (FMR1, also FMRP) were reported to mediate interactions with RNA.[21,87-89] RGG-boxes are also able to mediate protein-protein interactions.[83] For instance, the RGG boxes in FUS interact with SR proteins.[22] RGG boxes in G3BP and nucleolin were suggested to have RNA/DNA helicase activity.[90] Finally, the RGG box in hnRNP A2 was reported to promote nuclear localization in a manner that is regulated by arginine methylation.[91]

The RNA-binding properties of the RGG box have been best characterized in FMRP, where in vitro RNA selection demonstrated that the FMRP RGG box binds intramolecular G-quartets.[88,92,93] G-quartets are nucleic acid structures in which four guanine residues are arranged in a planar conformation stabilized by Hoogsteen-type hydrogen bonds. The planar G-quartets stack on top of each other, giving rise to four-stranded helical structures.[94] (Fig. 3a). The formation of G-quartets is dependent on monovalent cations such as K+ and Na+, which is consistent with the observation that the FMRP RGG box bound its RNA target sites in a potassium-dependent manner.[88,92]

G-quartet formation can involve intermolecular RNA:RNA interactions.[95] Whereas intra-molecular folding of a G-rich strand containing four or more G-tracts is concentration independent, the intermolecular formation of G-quadruplex structure from two G-rich strands containing two G-tracts, or four strands each containing one G-track, is likely to require the aid of protein chaperones.[94] Similarly, nucleolin (which contains the RGG-domain) is required for stabilisation of G-quartet DNA and thereby repress transcription of *c-myc* gene.[96]

Intermolecular G-quartets could represent a flexible motif for mRNA oligomerization and RGG box binding to such G-quartets could stabilise RNA:RNA interactions. RGG boxes are strong targets for arginine methylation, as has been first shown in the case of hnRNP A1.[79,97] Therefore, interactions between G-quartets and RGG boxes could represent an important target for dynamic regulation of ribonucleoprotein complexes via posttranslational modifications of RBPs.

RG DOMAIN

The second type of region enriched in arginine and glycine residues is a repeat of RG dinucleotide, which we will refer to as the RG domain. This domain is predicted to be highly unstructured (Fig. 3B). It is present in many RBPs, including four Sm/Lsm proteins that make up a core heptamer of the U snRNPs (SmB, SmD1, SmD3, and Lsm4). SnRNPs are essential components of the pre-mRNA splicing machinery, and their Sm cores are assembled in the cytoplasm.[98]

Assembly of the Sm cores onto the snRNAs is co-ordinated by survival of motor neuron protein (SMN) and a complex of several additional associated proteins. Interaction of the Tudor domain of SMN with RG domains of SmD1 and SmD3 (also named as SNRPD1 and SNRPD3) is crucial for its function in snRNP assembly.[99] Similar to the RGG box, the RG domains can also be modified by arginine methyltransferases. It was shown that symmetrical dimethylation of SmD1 and SmD3 is required for their high-affinity interaction with the SMN complex.[100] Coilin protein contains a similar RG domain that upon symmetrical dimethylation recruits SMN to Cajal bodies in cell nuclei. Furthermore, RA/RG motifs are also present Piwi proteins. These proteins interact with a specific class of small noncoding RNAs, piwi-interacting RNAs (piRNAs) in germ cells. Symmetrical dimethylation of arginines within these motifs promotes their interaction with a specific set of Tudor domains.[101,102] These examples show that methyltransferases play an important role in regulating assembly of ribonucleoprotein complexes containing RBPs with the arginine- and glycine-rich regions.

R-G DOMAIN

The third type of region enriched in arginine and glycine residues is composed of consecutive glycines flanked by a proximal arginine-rich region. To differentiate from the RG domain discussed earlier, we will refer to this region as R-G domain. An example of such R-G domain is present in the predicted unstructured region of U2AF1 (Fig. 3C). In this protein, the R-G region is located next to the arginine and serine-rich (RS) domain, which is also predicted to be unstructured. Similar R-G domains are also present in several other important SR proteins (i.e., proteins that contain the RS domain), including splicing factor, arginine/serine-rich 1 (SFRS1, formerly ASF/SF2) and small nuclear ribonucleoprotein 70kDa (SNRP70, also U170K), as well as some hnRNP proteins, including heterogeneous nuclear ribonucleoprotein L (hnRNP L).

U2AF1, SFRS1 and SNRP70 are core splicing factors required for the early steps of splicing process. U2AF1 recognises the 3′ splice site, SFRS1 the exonic splicing enhancers and SNRP70 is part of the U1 snRNP that recognises the 5′ splice site. According to the exon definition model, interactions between the RS domains of proteins that recognize different regulatory elements on the exon promote recognition of exon as a unit by the splicing machinery.[103] Recognition of an exon as a unit was proposed to increase the fidelity of splicing.[104] Interestingly, the R-G domain in U2AF1 is more evolutionarily conserved compared to RS domain, as evident in the C. elegans orthologue uaf-2, which contains an extended R-G domain, but lacks the RS domain.[105] However, in contrast to the well-characterised functions of RS domains, little is known about the functions of R-G domains. It would therefore be interesting to explore the potential role of R-G domains in exon definition by promoting interactions between U2AF1, SFRS1 and SNRP70.

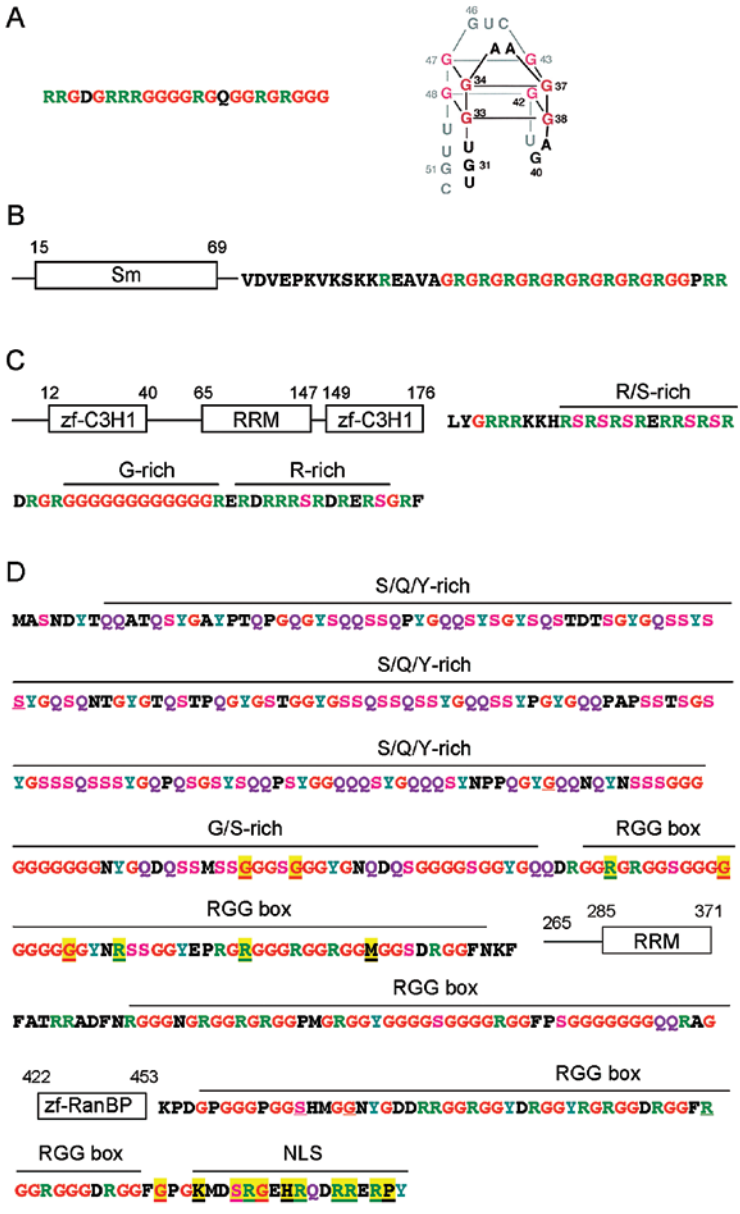

Figure 3. Glycine/arginine-rich domains in RNA-binding proteins. A) Sequence of the RGG box from fragile-X mental retardation protein (FMRP) and the structure of its RNA binding site, the G quartet. (Reprinted from: Darnell JC et al. Cell 2001; 107:489-499;[88] ©2001 with permission from Elsevier.) B) Domain structure and sequence of the small nuclear ribonucleoprotein D1 polypeptide 16kDa (SmD1, also SNRPD1). C) Domain structure and sequence of the U2 small nuclear RNA auxiliary factor 1 (U2AF1, also U2AF35). D) Domain structure and sequence of the human fused in sarcoma (FUS). The sequence of the protein region predicted to be disordered by Disopred2 software is shown, with highly disordered region in bold. Tyrosines are labeled in blue, glutamines in purple, serines in pink, glycines in red and arginines in green. The amino acids in FUS affected by disease-causing mutations are underlined and shaded in yellow.

FUS

FUS, EWS and RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa (TAF15) form a group of highly related FET (FUS, EWS, TAF15) family RBPs.[106] Their N terminus contains a glutamine-, serine-, and tyrosine-rich region that functions as a transcriptional activation domain when fused to a heterologous DNA binding domain.[107,108] This is followed by multiple domains involved in interactions with RNA: an RRM motif flanked by several RGG boxes and a C2C2 zinc finger. With the exception of the RRM and zinc finger domains, the remaining regions of FET proteins are predicted to be highly unstructured (Fig. 3D). FET proteins are involved in several aspects of gene expression, including transcription and splicing regulation and coupling of the transcriptional and splicing machinery.[3,109,110] FETs are targets for tyrosine and serine phosphorylation, as well as arginine methylation of the RGG boxes.[111-113] RGG boxes can self-associate and also play a role in nuclear import.[114] Similar to the association with RG-rich region in SmD1, the RGG box of EWS was also found to interact with SMN.[114]

The zinc finger domain belongs to the RanBP2-type and plays the predominant role in RNA recognition by recognising G-rich sequences.[19] The RNA specificity, if any, of the RGG boxes is not yet known. By analogy with the RGG box of FMRP, it would therefore be important to test if the RGG boxes of FET proteins also associate with G quartet RNA.

Majority of the ALS-associated FUS mutations fall within the PY-type nuclear localization signal present at the C-terminus of FUS.[2-5] They affect the nuclear import of FUS, bringing about cytoplasmic accumulation and in some instances formation of stress granules.[115-118]

Given the high similarity between FUS and EWS and TAF15 proteins, it is intriguing to speculate why FUS, but not other FET family RBPs, form toxic aggregates in neurons when mutated. A comparison between the putative unstructured regions of the three proteins shows that FUS contains a glycine/serine (G/S)-rich region, which is not present in other FET family proteins (Fig. 3D). This region is immediately preceded by an asparagine (Q)-rich region. This is similar to the C-terminus of TDP-43, which contains an N/Q-rich region followed by G/S-rich region (Fig. 1C). Due to this similarity, it was proposed that this domain might have prion-like properties.[67] Interestingly, about a third of the FUS mutations are found in the G/S-rich region, therefore it is of great importance to understand the function of this region of the protein.

CONCLUSION

The RNA-binding proteins discussed in this review indicate that glycine-rich regions can have diverse functions, which mainly depend on the few additional amino-acids that can be present in glycine-rich regions. Analysis of these amino acids indicates that the glycine-rich region in TDP-43 is most similar to proteins such as Sxl and hnRNP A1, whereas the glycine-rich region in FUS is most similar to proteins containing RGG boxes, such as FMR1. Studies of Sxl and hnRNP A1 indicate that their glycine-rich regions are required for multiple mechanisms of splicing regulation (Fig. 2). It is notable that similar to TDP-43, these regions are enriched in serines and asparagines. It remains to be seen if these additional amino acids contribute to the ability of these proteins for co-operative spreading on the target RNAs (Fig. 2C).

It is possible that glycine-rich regions are required for regulation of a specific set of RNA targets. As shown in the cases of Sxl and hnRNP A1, the unstructured region is particularly important for autoregulation, where RBPs need to co-operatively bind multiple sites on their own transcript. So far, however, only a small number of RNA targets of each protein have been evaluated. Therefore, a comprehensive exploration of RNA targets of a specific RBP will be required for better insight into the splicing functions of their glycine-rich regions. A large number of RNAs regulated by TDP-43 were recently identified, therefore it will be important to evaluate the role of its glycine-rich region in the regulation of these RNAs.[10,11]

Another poorly understood aspect of glycine-rich regions is the role of their posttranslational modifications. The best understood is the effect of arginine methylation in R/G-rich domains on protein-protein interactions. However, little is known about the effect of these modifications on protein-RNA interactions. Furthermore, even though G/S-rich regions are common in glycine-rich domains, the effect of serine phosphorylation of these regions has not yet been evaluated. Further experimental and computational studies of glycine-rich regions will be necessary to provide insights into these questions, and thereby contribute to our understanding of the effects of disease-causing mutations in TDP-43 and FUS proteins.

ACKNOWLEDGEMENT

The authors would like to thank Julian Konig and Josh Witten for their instructive comments on this manuscript.

REFERENCES

1. Sreedharan J, Blair IP, Tripathi VB et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science 2008; 319:1668-1672.
2. Vance C, Rogelj B, Hortobagyi T et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science 2009; 323:1208-1211.
3. Lagier-Tourenne C, Polymenidou M, Cleveland DW. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Hum Mol Genet 2010; 19:R46-R64.
4. Kwiatkowski TJ Jr, Bosco DA, Leclerc AL et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science 2009; 323:1205-1208.
5. Mackenzie IR, Rademakers R, Neumann M. TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. Lancet Neurol 2010; 9:995-1007.
6. Mackenzie IR, Munoz DG, Kusaka H et al. Distinct pathological subtypes of FTLD-FUS. Acta Neuropathol 2011; 121:207-218.
7. Neumann M, Sampathu DM, Kwong LK et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 2006; 314:130-133.
8. Igaz LM, Kwong LK, Chen-Plotkin A et al. Expression Of TDP-43 C-terminal fragments in vitro recapitulates pathological features of TDP-43 proteinopathies. J Biol Chem 2009.
9. Kuo PH, Doudeva LG, Wang YT et al. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. Nucleic Acids Res 2009; 37:1799-1808.
10. Polymenidou M, Lagier-Tourenne C, Hutt KR et al. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat Neurosci 2011.
11. Tollervey JR, Curk T, Rogelj B et al. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat Neurosci 2011.
12. Nguyen CD, Mansfield RE, Leung W et al. Characterization of a family of RanBP2-type zinc fingers that can recognize single-stranded RNA. J Mol Biol 2011.

13. Kim SH, Shanware N, Bowler MJ et al. ALS-associated proteins TDP-43 and FUS/TLS function in a common biochemical complex to coregulate HDAC6 mRNA. J Biol Chem 2010.
14. Selamat W, Jamari I, Wang Y et al. TLS interaction with NMDA R1 splice variant in retinal ganglion cell line RGC-5. Neurosci Lett 2009; 450:163-166.
15. Sugiura T, Sakurai K, Nagano Y. Intracellular characterization of DDX39, a novel growth-associated RNA helicase. Exp Cell Res 2007; 313:782-790.
16. Yoshimura A, Fujii R, Watanabe Y et al. Myosin-Va facilitates the accumulation of mRNA/protein complex in dendritic spines. Curr Biol 2006; 16:2345-2351.
17. Fujii R, Takumi T. TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. J Cell Sci 2005; 118:5755-5765.
18. Belly A, Moreau-Gachelin F, Sadoul R et al. Delocalization of the multifunctional RNA splicing factor TLS/FUS in hippocampal neurones: exclusion from the nucleus and accumulation in dendritic granules and spine heads. Neurosci Lett 2005; 379:152-157.
19. Iko Y, Kodama TS, Kasai N et al. Domain architectures and characterization of an RNA-binding protein, TLS. J Biol Chem 2004; 279:44834-44840.
20. Delva L, Gallais I, Guillouf C et al. Multiple functional domains of the oncoproteins Spi-1/PU.1 and TLS are involved in their opposite splicing effects in erythroleukemic cells. Oncogene 2004; 23:4389-4399.
21. Lerga A, Hallier M, Delva L et al. Identification of an RNA binding specificity for the potential splicing factor TLS. J Biol Chem 2001; 276:6807-6816.
22. Yang L, Embree LJ, Tsai S et al. Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing. J Biol Chem 1998; 273:27761-27764.
23. Zinszner H, Sok J, Immanuel D et al. TLS (FUS) binds RNA in vivo and engages in nucleo-cytoplasmic shuttling. J Cell Sci 1997; 110 (Pt 15):1741-1750.
24. Zinszner H, Immanuel D, Yin Y et al. A topogenic role for the oncogenic N-terminus of TLS: nucleolar localization when transcription is inhibited. Oncogene 1997; 14:451-461.
25. McDonald KK, Aulas A, Destroismaisons L et al. TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. Hum Mol Genet 2011.
26. Sephton CF, Cenik C, Kucukural A et al. Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. J Biol Chem 2011; 286:1204-1215.
27. Liu-Yesucevitz L, Bilgutay A, Zhang YJ et al. Tar DNA binding protein-43 (TDP-43) associates with stress granules: analysis of cultured cells and pathological brain tissue. PLoS ONE 2010; 5:e13250.
28. Dreumont N, Hardy S, Behm-Ansmant I et al. Antagonistic factors control the unproductive splicing of SC35 terminal intron. Nucleic Acids Res 2010; 38:1353-1366.
29. Fiesel FC, Voigt A, Weber SS et al. Knockdown of transactive response DNA-binding protein (TDP-43) downregulates histone deacetylase 6. EMBO J 2010; 29:209-221.
30. Volkening K, Leystra-Lantz C, Yang W et al. Tar DNA binding protein of 43 kDa (TDP-43), 14-3-3 proteins and copper/zinc superoxide dismutase (SOD1) interact to modulate NFL mRNA stability. Implications for altered RNA processing in amyotrophic lateral sclerosis (ALS). Brain Res. 2009; 1305:168-182.
31. Strong MJ, Volkening K, Hammond R et al. TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. Mol Cell Neurosci 2007; 35:320-327.
32. Casafont I, Bengoechea R, Tapia O et al. TDP-43 localizes in mRNA transcription and processing sites in mammalian neurons. J Struct Biol 2009; 167:235-241.
33. Bose JK, Wang IF, Hung L et al. TDP-43 overexpression enhances exon 7 inclusion during the survival of motor neuron pre-mRNA splicing. J Biol Chem 2008; 283:28852-28859.
34. Ayala YM, Misteli T, Baralle FE. TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. Proc Natl Acad Sci U S A 2008; 105:3785-3789.
35. Ayala YM, Pagani F, Baralle FE. TDP43 depletion rescues aberrant CFTR exon 9 skipping. FEBS Lett 2006; 580:1339-1344.
36. Ayala YM, Pantano S, D'Ambrogio A et al. Human, Drosophila and C. elegans TDP43: nucleic acid binding properties and splicing regulatory function. J Mol Biol 2005; 348:575-588.
37. Buratti E, Brindisi A, Pagani F et al. Nuclear factor TDP-43 binds to the polymorphic TG repeats in CFTR intron 8 and causes skipping of exon 9: a functional link with disease penetrance. Am J Hum Genet 2004; 74:1322-1325.
38. Buratti E, Dork T, Zuccato E et al. Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. EMBO J 2001; 20:1774-1784.
39. Gsponer J, Babu MM. The rules of disorder or why disorder rules. Prog Biophys Mol Biol 2009; 99:94-103.
40. Patel DJ. Adaptive recognition in RNA complexes with peptides and protein modules. Curr Opin Struct Biol 1999; 9:74-87.
41. Battiste JL, Mao H, Rao NS et al. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. Science 1996; 273:1547-1551.

42. Gsponer J, Futschik ME, Teichmann SA et al. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science 2008; 322:1365-1368.
43. Penalva LO, Sanchez L. RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation. Microbiol Mol Biol Rev 2003; 67:343-359, table.
44. Yanowitz JL, Deshpande G, Calhoun G et al. An N-terminal truncation uncouples the sex-transforming and dosage compensation functions of sex-lethal. Mol Cell Biol 1999; 19:3018-3028.
45. Granadino B, Penalva LO, Green MR et al. Distinct mechanisms of splicing regulation in vivo by the Drosophila protein Sex-lethal. Proc Natl Acad Sci U S A 1997; 94:7343-7348.
46. Wang J, Bell LR. The Sex-lethal amino terminus mediates co-operative interactions in RNA binding and is essential for splicing regulation. Genes Dev 1994; 8:2072-2085.
47. Sakashita E, Sakamoto H. Protein-RNA and protein-protein interactions of the Drosophila sex-lethal mediated by its RNA-binding domains. J Biochem 1996; 120:1028-1033.
48. Samuels M, Deshpande G, Schedl P. Activities of the Sex-lethal protein in RNA binding and protein:protein interactions. Nucleic Acids Res 1998; 26:2625-2637.
49. Wang J, Dong Z, Bell LR. Sex-lethal interactions with protein and RNA. Roles of glycine-rich and RNA binding domains. J Biol Chem 1997; 272:22227-22235.
50. Nagengast AA, Stitzinger SM, Tseng CH et al. Sex-lethal splicing autoregulation in vivo: interactions between SEX-LETHAL, the U1 snRNP and U2AF underlie male exon skipping. Development 2003; 130:463-471.
51. Lallena MJ, Chalmers KJ, Llamazares S et al. Splicing regulation at the second catalytic step by Sex-lethal involves 3′ splice site recognition by SPF45. Cell 2002; 109:285-296.
52. Blanchette M, Chabot B. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. EMBO J 1999; 18:1939-1952.
53. Nadler SG, Merrill BM, Roberts WJ et al. Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteolytic derivative, UP1, with RNA and DNA: evidence for multiple RNA binding domains and salt-dependent binding mode transitions. Biochemistry 1991; 30:2968-2976.
54. Cobianchi F, Karpel RL, Williams KR et al. Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. Large-scale overproduction in Escherichia coli and co-operative binding to single-stranded nucleic acids. J Biol Chem 1988; 263:1063-1071.
55. Mayeda A, Munroe SH, Caceres JF, Krainer AR. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. EMBO J 1994; 13:5483-5495.
56. Zhu J, Mayeda A, Krainer AR. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell 2001; 8:1351-1361.
57. Nasim FU, Hutchison S, Cordeau M et al. High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5′ splice site selection in support of a common looping out and repression mechanism. RNA 2002; 8:1078-1089.
58. Martinez-Contreras R, Fisette JF, Nasim FU et al. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. PLoS Biol 2006; 4:e21.
59. Okunola HL, Krainer AR. Co-operative-binding and splicing-repressive properties of hnRNP A1. Mol Cell Biol 2009; 29:5620-5631.
60. Chan RC, Black DL. Conserved intron elements repress splicing of a neuron-specific c-src exon in vitro. Mol Cell Biol 1995; 15:6377-6385.
61. Amir-Ahmady B, Boutz PL, Markovtsov V et al. Exon repression by polypyrimidine tract binding protein. RNA 2005; 11:699-716.
62. Zhang C, Frias MA, Mele A et al. Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. Science 2010; 329:439-443.
63. Buratti E, Baralle FE. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. J Biol Chem 2001; 276:36337-36343.
64. Buratti E, Brindisi A, Giombi M et al. TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. J Biol Chem 2005; 280:37572-37584.
65. Fuentealba RA, Udan M, Bell S et al. Interaction with polyglutamine aggregates reveals a Q/N-rich domain in TDP-43. J Biol Chem 2010; 285:26304-26314.
66. Elden AC, Kim HJ, Hart MP et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature 2010; 466:1069-1075.
67. Udan M, Baloh RH. Implications of the prion-related Q/N domains in TDP-43 and FUS. Prion 2011; 5.
68. Ayala YM, De CL, vendano-Vazquez SE et al. TDP-43 regulates its mRNA levels through a negative feedback loop. EMBO J 2011; 30:277-288.
69. Sun S, Zhang Z, Sinha R et al. SF2/ASF autoregulation involves multiple layers of posttranscriptional and translational control. Nat Struct Mol Biol 2010; 17:306-312.
70. Ni JZ, Grate L, Donohue JP et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev 2007; 21:708-718.

71. Lareau LF, Inada M, Green RE et al. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 2007; 446:926-929.
72. Wollerton MC, Gooding C, Wagner EJ et al. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell 2004; 13:91-100.
73. Jumaa H, Nielsen PJ. The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. EMBO J 1997; 16:5077-5085.
74. Ling SC, Albuquerque CP, Han JS et al. ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. Proc Natl Acad Sci U S A 2010; 107:13318-13323.
75. Johnson BS, McCaffery JM, Lindquist S et al. A yeast TDP-43 proteinopathy model: Exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. Proc Natl Acad Sci U S A 2008; 105:6439-6444.
76. Voigt A, Herholz D, Fiesel FC et al. TDP-43-mediated neuron loss in vivo requires RNA-binding activity. PLoS ONE 2010; 5:e12247.
77. Bahadur RP, Zacharias M, Janin J. Dissecting protein-RNA recognition sites. Nucleic Acids Res 2008; 36:2705-2716.
78. Godin KS, Varani G. How arginine-rich domains co-ordinate mRNA maturation events. RNA Biol 2007; 4:69-75.
79. Bedford MT. Arginine methylation at a glance. J Cell Sci 2007; 120:4243-4246.
80. Liu Q, Dreyfuss G. In vivo and in vitro arginine methylation of RNA-binding proteins. Mol Cell Biol 1995; 15:2800-2808.
81. Shen EC, Henry MF, Weiss VH et al. Arginine methylation facilitates the nuclear export of hnRNP proteins. Genes Dev 1998; 12:679-691.
82. McBride AE, Silver PA. State of the arg: protein methylation at arginine comes of age. Cell 2001; 106:5-8.
83. Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. Science 1994; 265:615-621.
84. Lee MH, Mori S, Raychaudhuri P. Trans-Activation by the hnRNP K protein involves an increase in RNA synthesis from the reporter genes. J Biol Chem 1996; 271:3420-3427.
85. Dempsey LA, Hanakahi LA, Maizels N. A specific isoform of hnRNP D interacts with DNA in the LR1 heterodimer: canonical RNA binding motifs in a sequence-specific duplex DNA binding protein. J Biol Chem 1998; 273:29224-29229.
86. Alex D, Lee KA. RGG-boxes of the EWS oncoprotein repress a range of transcriptional activation domains. Nucleic Acids Res 2005; 33:1323-1331.
87. Kiledjian M, Dreyfuss G. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. EMBO J 1992; 11:2655-2664.
88. Darnell JC, Jensen KB, Jin P et al. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. Cell 2001; 107:489-499.
89. Ohno T, Ouchida M, Lee L et al. The EWS gene, involved in Ewing family of tumors, malignant melanoma of soft parts and desmoplastic small round cell tumors, codes for an RNA binding protein with novel regulatory domains. Oncogene 1994; 9:3087-3097.
90. Costa M, Ochem A, Staub A et al. Human DNA helicase VIII: a DNA and RNA helicase corresponding to the G3BP protein, an element of the ras transduction pathway. Nucleic Acids Res 1999; 27:817-821.
91. Nichols RC, Wang XW, Tang J et al. The RGG domain in hnRNP A2 affects subcellular localization. Exp Cell Res 2000; 256:522-532.
92. Schaeffer C, Bardoni B, Mandel JL et al. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. EMBO J 2001; 20:4803-4813.
93. Bole M, Menon L, Mihailescu MR. Fragile X mental retardation protein recognition of G quadruplex structure per se is sufficient for high affinity binding to RNA. Mol Biosyst 2008; 4:1212-1219.
94. Lipps HJ, Rhodes D. G-quadruplex structures: in vivo evidence and function. Trends Cell Biol 2009; 19:414-422.
95. Ramos A, Hollingworth D, Pastore A. G-quartet-dependent recognition between the FMRP RGG box and RNA. RNA 2003; 9:1198-1207.
96. Gonzalez V, Guo K, Hurley L et al. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. J Biol Chem 2009; 284:23622-23635.
97. Kim S, Merrill BM, Rajpurohit R et al. Identification of N(G)-methylarginine residues in human heterogeneous RNP protein A1: Phe/Gly-Gly-Gly-Arg-Gly-Gly-Gly/Phe is a preferred recognition motif. Biochemistry 1997; 36:5185-5192.
98. Kambach C, Walke S, Nagai K. Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. Curr Opin Struct Biol 1999; 9:222-230.
99. Paushkin S, Gubitz AK, Massenet S et al. The SMN complex, an assemblyosome of ribonucleoproteins. Curr Opin Cell Biol 2002; 14:305-312.
100. Brahms H, Meheus L, de B et al. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. RNA 2001; 7:1531-1542.

101. Vagin VV, Wohlschlegel J, Qu J et al. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. Genes Dev 2009; 23:1749-1762.
102. Liu K, Chen C, Guo Y et al. Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. Proc Natl Acad Sci U S A 2010; 107:18398-18403.
103. Kohtz JD, Jamison SF, Will CL et al. Protein-protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. Nature 1994; 368:119-124.
104. Robberson BL, Cote GJ, Berget SM. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol 1990; 10:84-94.
105. Zorio DA, Blumenthal T. U2AF35 is encoded by an essential gene clustered in an operon with RRM/cyclophilin in Caenorhabditis elegans. RNA 1999; 5:487-494.
106. Bertolotti A, Lutz Y, Heard DJ et al. hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. EMBO J 1996; 15:5022-5031.
107. May WA, Lessnick SL, Braun BS et al. The Ewing's sarcoma EWS/FLI-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than FLI-1. Mol Cell Biol 1993; 13:7393-7398.
108. Zinszner H, Albalat R, Ron D. A novel effector domain from the RNA-binding protein TLS or EWS is required for oncogenic transformation by CHOP. Genes Dev 1994; 8:2513-2526.
109. Yang L, Chansky HA, Hickstein DD. EWS.Fli-1 fusion protein interacts with hyperphosphorylated RNA polymerase II and interferes with serine-arginine protein-mediated RNA splicing. J Biol Chem 2000; 275:37612-37618.
110. Knoop LL, Baker SJ. The splicing factor U1C represses EWS/FLI-mediated transactivation. J Biol Chem 2000; 275:24865-24871.
111. Felsch JS, Lane WS, Peralta EG. Tyrosine kinase Pyk2 mediates G-protein-coupled receptor regulation of the Ewing sarcoma RNA-binding protein EWS. Curr Biol 1999; 9:485-488.
112. Belyanskaya LL, Gehrig PM, Gehring H. Exposure on cell surface and extensive arginine methylation of ewing sarcoma (EWS) protein. J Biol Chem 2001; 276:18681-18687.
113. Olsen RJ, Hinrichs SH. Phosphorylation of the EWS IQ domain regulates transcriptional activity of the EWS/ATF1 and EWS/FLI1 fusion proteins. Oncogene 2001; 20:1756-1764.
114. Shaw DJ, Morse R, Todd AG et al. Identification of a self-association domain in the Ewing's sarcoma protein: a novel function for arginine-glycine-glycine rich motifs? J Biochem 2010; 147:885-893.
115. Ito D, Seki M, Tsunoda Y et al. Nuclear transport impairment of amyotrophic lateral sclerosis-linked mutations in FUS/TLS. Ann Neurol 2011; 69:152-162.
116. Bosco DA, Lemay N, Ko HK et al. Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. Hum Mol Genet 2010.
117. Gal J, Zhang J, Kwinter DM et al. Nuclear localization sequence of FUS and induction of stress granules by ALS mutants. Neurobiol Aging 2010.
118. Dormann D, Rodde R, Edbauer D et al. ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. EMBO J 2010; 29:2841-2857.